

Classification of ActiGraph GT9X Accelerometer and Gyroscope Data for Characterizing Youth Physical Activity

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Samuel Robert LaMunion
May 2020

ACKNOWLEDGEMENTS

I would like to extend my gratitude to the members of my committee for their support throughout this process. Thank you all for your patience in teaching me, the time you invested in me, and your guidance to help me see this through and make it as strong of a project as it could be.

To Dr. Crouter – thank you for chairing not one, but two committees for me. Thank you for taking me on six years ago as a Master’s student and for keeping me on to pursue my PhD. I sincerely appreciate your mentorship, your commitment to my success, and your investment in me as a student and person. My experience at Tennessee was greatly shaped and influenced by you and I would not have made it to where I am without you. I am forever grateful for your support of my personal and professional endeavors. I hope I have the chance to continue working with you long after my time at Tennessee has ended.

To Paul – thank you for always having my back and lending a hand on countless occasions. I am proud to call you my lab mate, colleague, and friend and will look back on our time together at Tennessee fondly. This experience truly would not have been what it was without you. It has been an honor to work alongside you as students and I hope that our personal and professional relationship continues long into our careers.

To Ashton – thank you for never giving up on me and always loving me. I know you experienced the worst of this process, but you never wavered in your support of my goals. You kept me focused on our long-term goals and loved me through the good times and the bad. This truly would not have been possible without you and it certainly wouldn’t be worth what it is without you.

To My Family – Mom and Dad, achieving this degree would not have been possible without the many sacrifices you made to support me and your dedication to my education from childhood. Thank you for always loving me and pushing me to be my best. Spencer and Suzanna, thank you both for always believing in me even when I didn’t believe in myself. You may not know it but you both pushed me to be my best.

To Ahmed Samaha and Dr. Brian Parr – you each played an integral part in my pursuing graduate school in the first place and you both have helped me from the start. I sincerely appreciate the investment you both made in me during my time in Aiken and for your continued investment in me now. This wouldn’t have been possible without you.

To Ashleigh Moyer and The University of Tennessee Center for Student Engagement Staff – thank you for taking a chance on me and getting me to Tennessee in the first place. My time with CSE as a Master’s student has shaped my graduate experience more than you can ever know. You all became my Tennessee family and made me love Tennessee. Thank you for everything!

To my friends (near and far) – thank you for helping me find balance in all of this. I might not have maintained my sanity without you. Cheers!

To the undergrads- this project would not have fully come together without the many hours of labor you all contributed over the years. It may not have seemed like much at the time, but you helped make this happen and you helped make it a little more enjoyable at the same time. Thanks for all the laughs and for agreeing to work with us in the first place. Shout out to A. Blythe, Brice, Jesse, Weldon, Marisa, Josie, Polina, and Willie. You guys rock!

ABSTRACT

PURPOSE: The purpose of this study is two-fold: 1) to determine if using gyroscope sensor data in place of accelerometer sensor data improves the classification of PA in youth compared to using only accelerometer sensor data, and 2) to determine if using a combined sensor approach improves classification of PA in youth compared to using either sensor independently.

METHODS: These aims were evaluated two ways: 1) a within-sample cross-validation using semi-structured simulated free-living activities from 99 youth participants ages 6-18 years old, and 2) an out-of-sample validation using unstructured free-living activities from 42 youth participants ages 6-18 years old. PA data were collected using a GT9X device worn on the hip, left wrist, and right wrist. Participant's PA behaviours were directly observed for both phases of the study to create criterion activity labels for each second of data. Four machine learning model types were developed (decision tree, random forest, logistic regression (lasso), and neural network) to classify 1) 14 individual activity classes, and 2) six activity group classes. Each model was trained using 1) accelerometer only data, 2) gyroscope only data, and 3) accelerometer + gyroscope data. **RESULTS:** For the within-sample cross-validation the combined sensor models achieved average classification accuracies of 61.5% for the individual activity classes and 82.0% for the activity groups, which outperformed the individual sensor models by an average of 10.4% for the individual activity classes and 5.3% for the activity groups, across all attachment sites. For the out-of-sample validation, the accelerometer only and combined sensor models classification accuracy decreased by 4.0%, 76.7% to 72.7% and 82.0% to 78.0%, respectively while the gyroscope only models maintained a classification accuracy of 76.7%. Adding transitions as a seventh activity group resulted in decreases in classification accuracy of 11.1% (within-sample) and 12.1% (out-of-sample) when compared to models

excluding transitions. **CONCLUSION:** These studies conclude that using accelerometer and gyroscope sensor data together improves activity classification compared to using data from either single sensor alone when evaluated using semi-structured simulated free-living activity data and when using unstructured free-living activity data.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION.....	1
SECTION 1: BACKGROUND	2
SECTION 2: STATEMENT OF THE PROBLEM.....	6
SECTION 3: STATEMENT OF THE PURPOSE.....	7
SECTION 4: RESEARCH QUESTIONS.....	7
REFERENCES.....	8
CHAPTER 2: REVIEW OF LITERATURE	13
SECTION 1: OVERVIEW	14
SECTION 2: INTRODUCTION TO PHYSICAL ACTIVITY ASSESSMENT	14
SECTION 3: PHYSICAL ACTIVITY ASSESSMENT USING INERTIAL SENSORS	16
<i>3.1 Data Acquisition</i>	<i>17</i>
<i>3.2 Pre-Processing.....</i>	<i>24</i>
<i>3.3 Segmentation.....</i>	<i>28</i>
<i>3.4 Features</i>	<i>28</i>
SECTION 4: MODELING WEARABLE DEVICE DATA	33
<i>4.1 Machine Learning Overview.....</i>	<i>34</i>
<i>4.2 Machine Learning Model Types</i>	<i>37</i>
SECTION 5: CHARACTERIZING PHYSICAL ACTIVITY USING WEARABLE DEVICE AND INERTIAL SENSOR DATA.....	44
<i>5.1 Classifier Development and Within-Sample Cross-Validation Performance.....</i>	<i>44</i>
<i>5.2 Classifier Performance in Application and Out-of-Sample Validation Performance.....</i>	<i>58</i>
SECTION 6: ALTERNATIVE APPROACHES TO ADDRESS MODEL PERFORMANCE ISSUES.....	67
<i>6.1 Multiple Sensor Locations of the Same Sensor Type</i>	<i>67</i>
<i>6.2 Multiple Sensor Types in the Same Location</i>	<i>68</i>
<i>6.3 The Case for the Gyroscope.....</i>	<i>69</i>
SECTION 7: SUMMARY.....	78
REFERENCES.....	79
CHAPTER 3: STUDY 1 METHODS.....	96
SECTION 1: PARTICIPANTS.....	97
SECTION 2: PROCEDURES.....	98
SECTION 3: EQUIPMENT	100
<i>3.1 Wearable Device.....</i>	<i>99</i>
<i>3.2 Direct Observation.....</i>	<i>99</i>
SECTION 4: DATASET ASSEMBLY.....	101
REFERENCES.....	103

CHAPTER 4: DEVELOPMENT AND WITHIN-SAMPLE VALIDATION.....	105
ABSTRACT.....	106
SECTION 1: INTRODUCTION	108
SECTION 2: METHODS.....	110
2.1 <i>Participants.....</i>	<i>110</i>
2.2 <i>Procedures</i>	<i>111</i>
2.3 <i>Data Acquisition</i>	<i>113</i>
2.4 <i>Wearable Device Data.....</i>	<i>115</i>
2.5 <i>Statistical Analysis</i>	<i>119</i>
SECTIONS 3: RESULTS	121
3.1 <i>Hip Individual Activity Class Results</i>	<i>121</i>
3.2 <i>Hip Activity Group Results</i>	<i>125</i>
3.3 <i>Left Wrist Individual Activity Class Results.....</i>	<i>128</i>
3.4 <i>Left Wrist Activity Group Results</i>	<i>132</i>
3.5 <i>Right Wrist Individual Activity Class Results</i>	<i>135</i>
3.6 <i>Right Wrist Activity Group Results</i>	<i>139</i>
SECTION 4: DISCUSSION	143
4.1 <i>Summary</i>	<i>143</i>
4.2 <i>Performance Overview</i>	<i>144</i>
4.3 <i>Comparison to Previous Research.....</i>	<i>148</i>
4.4 <i>Left and Right Wrist Justification</i>	<i>149</i>
4.5 <i>Model Recommendation.....</i>	<i>151</i>
4.6 <i>Strengths</i>	<i>152</i>
4.7 <i>Limitations</i>	<i>153</i>
4.8 <i>Next Steps and Practical Considerations</i>	<i>154</i>
4.9 <i>Conclusion</i>	<i>156</i>
<i>Disclosure</i>	<i>157</i>
REFERENCES.....	158
CHAPTER 5: STUDY 2 METHODS.....	166
SECTION 1: PARTICIPANTS.....	167
SECTION 2: PROCEDURES.....	167
SECTION 3: EQUIPMENT	168
3.1 <i>Wearable Device.....</i>	<i>169</i>
3.2 <i>Direct Observation.....</i>	<i>169</i>
SECTION 4: DATASET ASSEMBLY.....	172
CHAPTER 6: OUT-OF-SAMPLE VALIDATION	173
ABSTRACT.....	174
SECTION 1: INTRODUCTION	176
SECTION 2: METHODS.....	179

2.1 Participants	179
2.2 Procedures	180
2.3 Data Acquisition	182
2.4 Wearable Device Data	184
2.5 Statistical Analysis	185
SECTIONS 3: RESULTS	188
3.1 Out-of-Sample Validation of LaMunion et al. (Chapter 4) Six-class Activity Group Models	188
3.2 Within-Sample Cross-Validation of Random Forest Models with Transitions	189
3.3 Out-of-Sample Validation of Random Forest Models with Transitions	194
SECTION 4: DISCUSSION	195
4.1 Summary	195
4.2 Performance Overview	195
4.3 Contribution and Recommendation	198
4.4 Strengths	201
4.5 Limitations	202
4.6 Next Steps and Practical Considerations	203
4.7 Conclusion	204
Disclosure	204
REFERENCES	205
APPENDICES	209
APPENDIX A: FIGURES	210
APPENDIX B: TABLES	217
APPENDIX C: STUDY 1 RECRUITMENT FLYER	316
APPENDIX D: STUDY 1 INFORMED CONSENT	318
APPENDIX E: HEALTH HISTORY QUESTIONNAIRE	322
APPENDIX F: STUDY 1 INFORMED ASSENT	325
APPENDIX G: STUDY 2 RECRUITMENT FLYER	328
APPENDIX H: STUDY 2 INFORMED CONSENT	330
APPENDIX I: STUDY 2 INFORMED ASSENT	334
VITA	337

LIST OF FIGURES

Figure 1. Project overview flow chart.....	211
Figure 2. Study 1 participant enrollment, description of data loss, and participant allocation for analyses.....	212
Figure 3. Overall classification accuracy results from 1) decision tree (RPART), 2) logistic regression (LASSO), 3) random forest (RF), and 4) neural network (NNET) models for hip A) individual activity classes, and B) activity groups, left wrist C) individual activity classes, and D) activity groups, and right wrist E) individual activity classes, and F) activity groups.....	213
Figure 4. Out-of-sample validation model classification accuracy of LaMunion et al. (Chapter 4) six-class activity group classifiers 1) decision tree (RPART), 2) logistic regression (LASSO), 3) random forest (RF), and 4) neural network (NNET) models for A) hip, B) left wrist, C) right wrist data.	214
Figure 5. Overall classification accuracy results for seven-class random forest classifiers including transitions for a A) within-sample cross-validation using semi-structured simulated free-living activity data and B) out-of-sample validation using unstructured free-living data.	215
Figure 6. Overall classification accuracy results from within-sample cross-validation and out-of-sample validation of random forest models developed to classify six activity classes excluding transitions and seven activity classes including transitions using A) hip, B) left wrist, and C) right wrist data.	216

LIST OF TABLES

Table 1. Individual activity class labels, activity group labels, with the original activity labels and operational definitions.	218
Table 2. Study 1 dataset breakdown by number of data points in each individual activity class and activity group.	219
Table 3. Physical characteristics of participants for development and within-sample validation.....	220
Table 4. Confusion Matrix Performance Summary for Individual Activity Class for the Hip classifiers by Feature Set.	221
Table 5. Performance Summary for Individual Activity Classes for the Hip Classifiers by Feature Set.	221
Table 6. Confusion Matrix Performance Summary by Activity Group for the Hip classifiers by Feature Set.	222
Table 7. Performance Summary by Activity Group for the Hip Classifiers by Feature Set.	222
Table 8. Confusion Matrix Performance Summary for Individual Activity Class for the Left Wrist classifiers by Feature Set.	223
Table 9. Performance Summary for Individual Activity Classes for the Left Wrist Classifiers by Feature Set.	223
Table 10. Confusion Matrix Performance Summary by Activity Group for the Left Wrist classifiers by Feature Set.	224
Table 11. Performance Summary by Activity Group for the Left Wrist Classifiers by Feature Set.	224
Table 12. Confusion Matrix Performance Summary for Individual Activity Class for the Right Wrist classifiers by Feature Set.	225
Table 13. Performance Summary for Individual Activity Classes for the Right Wrist Classifiers by Feature Set.	225
Table 14. Confusion Matrix Performance Summary by Activity Group for the Right Wrist classifiers by Feature Set.	226
Table 15. Performance Summary by Activity Group for the Right Wrist Classifiers by Feature Set.	226
Table 16. Individual Class Confusion Matrix Results for the Hip Accelerometer Only Feature Set Decision Tree Model.	227
Table 17. Individual Class Performance Summaries for the Hip Accelerometer Only Feature Set Decision Tree Model.	227
Table 18. Individual Class Confusion Matrix Results for the Hip Gyroscope Only Feature Set Decision Tree Model.	228
Table 19. Individual Class Performance Summaries for the Hip Gyroscope Only Feature Set Decision Tree Model.	228
Table 20. Individual Class Confusion Matrix Results for the Hip Combined Sensor Feature Set Decision Tree Model.	229
Table 21. Individual Class Performance Summaries for the Hip Combined Sensor Feature Set Decision Tree Model.	229
Table 22. Individual Class Confusion Matrix Results for the Hip Accelerometer Only Feature Set Lasso Model.	230
Table 23. Individual Class Performance Summaries for the Hip Accelerometer Only Feature Set Lasso Model.	230
Table 24. Individual Class Confusion Matrix Results for the Hip Gyroscope Only Feature Set Lasso Model.	231

Table 25. Individual Class Performance Summaries for the Hip Gyroscope Only Feature Set Lasso Model.	231
Table 26. Individual Class Confusion Matrix Results for the Hip Combined Sensor Feature Set Lasso Model.	232
Table 27. Individual Class Performance Summaries for the Hip Combined Sensor Feature Set Lasso Model.	232
Table 28. Individual Class Confusion Matrix Results for the Hip Accelerometer Only Feature Set Random Forest Model.	233
Table 29. Individual Class Performance Summaries for the Hip Accelerometer Only Feature Set Random Forest Model.	233
Table 30. Individual Class Confusion Matrix Results for the Hip Gyroscope Only Feature Set Random Forest Model.	234
Table 31. Individual Class Performance Summaries for the Hip Gyroscope Only Feature Set Random Forest Model.	234
Table 32. Individual Class Confusion Matrix Results for the Hip Combined Sensor Feature Set Random Forest Model.	235
Table 33. Individual Class Performance Summaries for the Hip Combined Sensor Feature Set Random Forest Model.	235
Table 34. Individual Class Confusion Matrix Results for the Hip Accelerometer Only Feature Set Neural Network Model.	236
Table 35. Individual Class Performance Summaries for the Hip Accelerometer Only Feature Set Neural Network Model.	236
Table 36. Individual Class Confusion Matrix Results for the Hip Gyroscope Only Feature Set Neural Network Model.	237
Table 37. Individual Class Performance Summaries for the Hip Gyroscope Only Feature Set Neural Network Model.	237
Table 38. Individual Class Confusion Matrix Results for the Hip Combined Sensor Feature Set Neural Network Model.	238
Table 39. Individual Class Performance Summaries for the Hip Combined Sensor Feature Set Neural Network Model.	238
Table 40. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Decision Tree Model.	239
Table 41. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Decision Tree Model.	239
Table 42. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Decision Tree Model.	240
Table 43. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Decision Tree Model.	240
Table 44. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Decision Tree Model.	241
Table 45. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Decision Tree Model.	241
Table 46. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Lasso Model.	242

Table 47. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Lasso Model.	242
Table 48. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Lasso Model.	243
Table 49. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Lasso Model.	243
Table 50. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Lasso Model.	244
Table 51. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Lasso Model.	244
Table 52. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Random Forest Model.	245
Table 53. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Random Forest Model.	245
Table 54. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Random Forest Model.	246
Table 55. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Random Forest Model.	246
Table 56. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Random Forest Model.	247
Table 57. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Random Forest Model.	247
Table 58. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Neural Network Model.	248
Table 59. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Neural Network Model.	248
Table 60. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Neural Network Model.	249
Table 61. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Neural Network Model.	249
Table 62. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Neural Network Model.	250
Table 63. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Neural Network Model.	250
Table 64. Individual Class Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Decision Tree Model.	251
Table 65. Individual Class Performance Summaries for the Left Wrist Accelerometer Only Feature Set Decision Tree Model.	251
Table 66. Individual Class Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Decision Tree Model.	252
Table 67. Individual Class Performance Summaries for the Left Wrist Gyroscope Only Feature Set Decision Tree Model.	252
Table 68. Individual Class Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Decision Tree Model.	253

Table 69. Individual Class Performance Summaries for the Left Wrist Combined Sensor Feature Set Decision Tree Model.....	253
Table 70. Individual Class Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Lasso Model.	254
Table 71. Individual Class Performance Summaries for the Left Wrist Accelerometer Only Feature Set Lasso Model.	254
Table 72. Individual Class Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Lasso Model.	255
Table 73. Individual Class Performance Summaries for the Left Wrist Gyroscope Only Feature Set Lasso Model.	255
Table 74. Individual Class Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Lasso Model.	256
Table 75. Individual Class Performance Summaries for the Left Wrist Combined Sensor Feature Set Lasso Model.	256
Table 76. Individual Class Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Random Forest Model.....	257
Table 77. Individual Class Performance Summaries for the Left Wrist Accelerometer Only Feature Set Random Forest Model.....	257
Table 78. Individual Class Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Random Forest Model.....	258
Table 79. Individual Class Performance Summaries for the Left Wrist Gyroscope Only Feature Set Random Forest Model.....	258
Table 80. Individual Class Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Random Forest Model.....	259
Table 81. Individual Class Performance Summaries for the Left Wrist Combined Sensor Feature Set Random Forest Model.....	259
Table 82. Individual Class Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Neural Network Model.....	260
Table 83. Individual Class Performance Summaries for the Left Wrist Accelerometer Only Feature Set Neural Network Model.....	260
Table 84. Individual Class Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Neural Network Model.....	261
Table 85. Individual Class Performance Summaries for the Left Wrist Gyroscope Only Feature Set Neural Network Model.....	261
Table 86. Individual Class Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Neural Network Model.....	262
Table 87. Individual Class Performance Summaries for the Left Wrist Combined Sensor Feature Set Neural Network Model.....	262
Table 88. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Decision Tree Model.....	263
Table 89. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Decision Tree Model.....	263
Table 90. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Decision Tree Model.....	264

Table 91. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Decision Tree Model.....	264
Table 92. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Decision Tree Model.....	265
Table 93. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Decision Tree Model.....	265
Table 94. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Lasso Model.....	266
Table 95. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Lasso Model.....	266
Table 96. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Lasso Model.....	267
Table 97. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Lasso Model.....	267
Table 98. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Lasso Model.....	268
Table 99. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Lasso Model.....	268
Table 100. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Random Forest Model.....	269
Table 101. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Random Forest Model.....	269
Table 102. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Random Forest Model.....	270
Table 103. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Random Forest Model.....	270
Table 104. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Random Forest Model.....	271
Table 105. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Random Forest Model.....	271
Table 106. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Neural Network Model.....	272
Table 107. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Neural Network Model.....	272
Table 108. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Neural Network Model.....	273
Table 109. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Neural Network Model.....	273
Table 110. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Neural Network Model.....	274
Table 111. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Neural Network Model.....	274
Table 112. Individual Class Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Decision Tree Model.....	275

Table 113. Individual Class Performance Summaries for the Right Wrist Accelerometer Only Feature Set Decision Tree Model.....	275
Table 114. Individual Class Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Decision Tree Model.....	276
Table 115. Individual Class Performance Summaries for the Right Wrist Gyroscope Only Feature Set Decision Tree Model.....	276
Table 116. Individual Class Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Decision Tree Model.....	277
Table 117. Individual Class Performance Summaries for the Right Wrist Combined Sensor Feature Set Decision Tree Model.....	277
Table 118. Individual Class Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Lasso Model.....	278
Table 119. Individual Class Performance Summaries for the Right Wrist Accelerometer Only Feature Set Lasso Model.....	278
Table 120. Individual Class Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Lasso Model.....	279
Table 121. Individual Class Performance Summaries for the Right Wrist Gyroscope Only Feature Set Lasso Model.....	279
Table 122. Individual Class Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Lasso Model.....	280
Table 123. Individual Class Performance Summaries for the Right Wrist Combined Sensor Feature Set Lasso Model.....	280
Table 124. Individual Class Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Random Forest Model.....	281
Table 125. Individual Class Performance Summaries for the Right Wrist Accelerometer Only Feature Set Random Forest Model.....	281
Table 126. Individual Class Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Random Forest Model.....	282
Table 127. Individual Class Performance Summaries for the Right Wrist Gyroscope Only Feature Set Random Forest Model.....	282
Table 128. Individual Class Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Random Forest Model.....	283
Table 129. Individual Class Performance Summaries for the Right Wrist Combined Sensor Feature Set Random Forest Model.....	283
Table 130. Individual Class Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Neural Network Model.....	284
Table 131. Individual Class Performance Summaries for the Right Wrist Accelerometer Only Feature Set Neural Network Model.....	284
Table 132. Individual Class Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Neural Network Model.....	285
Table 133. Individual Class Performance Summaries for the Right Wrist Gyroscope Only Feature Set Neural Network Model.....	285
Table 134. Individual Class Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Neural Network Model.....	286

Table 135. Individual Class Performance Summaries for the Right Wrist Combined Sensor Feature Set Neural Network Model.....	286
Table 136. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Decision Tree Model.....	287
Table 137. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Decision Tree Model.....	287
Table 138. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Decision Tree Model.....	288
Table 139. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Decision Tree Model.....	288
Table 140. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Decision Tree Model.....	289
Table 141. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Decision Tree Model.....	289
Table 142. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Lasso Model.....	290
Table 143. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Lasso Model.....	290
Table 144. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Lasso Model.....	291
Table 145. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Lasso Model.....	291
Table 146. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Lasso Model.....	292
Table 147. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Lasso Model.....	292
Table 148. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Random Forest Model.....	293
Table 149. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Random Forest Model.....	293
Table 150. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Random Forest Model.....	294
Table 151. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Random Forest Model.....	294
Table 152. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Random Forest Model.....	295
Table 153. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Random Forest Model.....	295
Table 154. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Neural Network Model.....	296
Table 155. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Neural Network Model.....	296
Table 156. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Neural Network Model.....	297

Table 157. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Neural Network Model.....	297
Table 158. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Neural Network Model.....	298
Table 159. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Neural Network Model.....	298
Table 160. List of Study 2 activities and their operational definitions.....	299
Table 161. Study 2 dataset breakdown by number of data points in each activity class and group corresponding to the recoded activity group labels.....	302
Table 162. Physical characteristics of Study 2 participants.....	303
Table 163. Confusion Matrix Performance Summary by Activity Group for the Hip Random Forest Model by Feature Set.....	304
Table 164. Performance Summary by Activity Group for the Hip Random Forest Model by Feature Set.....	304
Table 165. Confusion Matrix Performance Summary by Activity Group for the Left Wrist Random Forest Model by Feature Set.....	305
Table 166. Performance Summary by Activity Group for the Left Wrist Random Forest Model by Feature Set.....	305
Table 167. Confusion Matrix Performance Summary by Activity Group for the Right Wrist Random Forest Model by Feature Set.....	306
Table 168. Performance Summary by Activity Group for the Right Wrist Random Forest Model by Feature Set.....	306
Table 169. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Random Forest Model.....	307
Table 170. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Random Forest Model.....	307
Table 171. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Random Forest Model.....	308
Table 172. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Random Forest Model.....	308
Table 173. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Random Forest Model.....	309
Table 174. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Random Forest Model.....	309
Table 175. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Random Forest Model.....	310
Table 176. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Random Forest Model.....	310
Table 177. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Random Forest Model.....	311
Table 178. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Random Forest Model.....	311
Table 179. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Random Forest Model.....	312

Table 180. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Random Forest Model.....	312
Table 181. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Random Forest Model.....	313
Table 182. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Random Forest Model.....	313
Table 183. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Random Forest Model.....	314
Table 184. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Random Forest Model.....	314
Table 185. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Random Forest Model.....	315
Table 186. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Random Forest Model.....	315

Chapter 1: Introduction

Section 1: Background

Accurately capturing unstructured free-living physical activity (PA) behaviour is critical to robustly assess, document, and evaluate the relationship between PA and health outcomes. However, it can be challenging to adequately assess PA in free-living environments. One popular method for objectively assessing PA is through the use of wearable devices such as accelerometer-based PA monitors. These devices are strategically placed on the body and record movement continuously. These movement data can then be used to characterize different components of PA behaviour, most commonly energy expenditure (EE) and PA intensity. Historically, acceleration data represented as activity counts, a proprietary metric derived from the raw acceleration signal, has been used to model the aforementioned components of PA. Activity counts are generated through a process of filtering the raw acceleration signal to attenuate noise in the sensor signal and then integrating the acceleration data in some fixed-length window of time (i.e. epoch) to represent the volume of movement that occurred in the specified window.¹ Activity counts are a useful metric for describing PA or quantifying PA based on total movement accumulation but lack resolution to distinguish intricacies of individual activities with similar movement frequencies and patterns.² With technological advances including on-board device memory and improved battery life coupled with increased storage and improved processing capabilities and computing power, investigators have shifted to using higher resolution raw acceleration data to assess free-living PA.² It is believed that working with raw acceleration data provides greater flexibility for analyzing the data and with the higher resolution it may be possible to distinguish characteristics of specific activities (e.g. walking versus computer work) or activity groups (e.g. locomotion versus household chores) more robustly and more consistently.² Additionally, working with raw acceleration data improves the

compatibility between data from different device manufacturers which improves comparability between studies.^{3,4}

There have been numerous approaches used to classify PA using wearable device data expressed as both counts and raw acceleration data.⁵⁻⁹ To date, most approaches have focused on classifying activities grouped by activity domain. To do so, it is common to calculate relevant statistical features (i.e. transformations of the raw sensor data).¹⁰⁻¹³ These features are then selected or extracted from the dataset using objective criteria or by using a statistical model that chooses appropriate features as a step in the modeling process.¹⁰⁻¹³ While true that many features can be calculated from acceleration data, more features are not always necessary or useful for improving PA classification performance since features may contain redundant information.¹⁴ To reduce redundant features and eliminate features with low predictive power for the desired classification task, feature selection is typically performed leaving only relevant non-redundant features in the final dataset for classification.¹⁴ Features are typically evaluated in two domains 1) time domain and 2) frequency domain. Time domain features are engineered by processing data in specified windows of time (e.g. mean, median, variance, standard deviation, minimum, maximum, range)¹⁰⁻¹³ and frequency domain features are calculated by extracting signal information (e.g. Fourier transform, wavelet transformation, or calculation of dominant frequency, spectral energy).¹⁰⁻¹³

Using raw acceleration data, a variety of analytical methods have been used to classify wearable sensor PA data ranging from logistic regression to machine learning models including neural networks, support vector machines, and decision tree-based approaches, all of which range in complexity and performance for the desired classification tasks.⁵⁻⁹ Additionally, it has been established that there is no single statistical modeling approach that is superior for all PA

classification tasks.^{15,16} With this in mind, it is standard practice to evaluate a series of classification techniques to determine the optimal methods for the specific classification task.^{15,16} In some cases, it has been demonstrated that by combining machine learning approaches superior classification performance can be achieved compared to using a single independent classification approach.^{15,16} This combination of machine learning methods is known as ensembling and typically involves combining a set of inherently weaker classifiers to achieve a single more robust final classification by pooling the individual classifier results (e.g. random forest models).^{15,16}

While the aforementioned efforts have yielded generally high classification accuracy at the group level, typically >85%, overall improvements in classifier performance have plateaued, regardless of the methodological approach and analytical technique used. The implications of this plateau may not be readily apparent when evaluating classifier performance from within-sample cross-validations, a typical first step in the development process. However, when the classifiers are applied to out-of-sample data in free-living situations, classifier performance decreases as models suffer from poor generalizability to new data.^{3,17-20} For example, in the field of PA assessment, nearly all PA classification research that has been done to date has used only accelerometer sensor data and associated features as the source for classifier development. A recent review by Farrahi et al.⁶ shows that on average, a PA classification accuracy of approximately 85.0% (hip) to 90.0% (wrist) is achieved in both adult and youth samples. At first glance, this level of accuracy seems acceptable, however, when the classifiers developed in these studies are independently validated using unstructured free-living data, performance decreases sharply, and on average, peaks at approximately 70.0-75.0%.⁶ Additionally, there is evidence to suggest that the individual level variability in classifier performance is greater in out-of-sample

validations than in within-sample cross-validations.²¹ The nuances and intricacies of unstructured free-living PA behaviour are complex and challenging to model using information from a single sensor type. This plateau in predictive performance is potentially related to there being insufficient relevant information in the data to adequately distinguish between independent activity classes or activity groups when using only accelerometer sensor data as the source for classifier development. One potential answer to this issue is to use additional sensor data, such as gyroscope data, to complement the acceleration data and more completely characterize PA.

It has been previously hypothesized that additional sensor types such as a gyroscope sensor might provide complementary information regarding human movement when used simultaneously with an accelerometer sensor.²²⁻²⁸ In adults it has been well documented that using accelerometer and gyroscope sensor data together to characterize PA results in improved classifier performance by approximately 10.0% compared to using data from either single sensor individually.²²⁻²⁷ Most of these data have come from studies in other disciplines such as engineering and computer science where smartphones are commonly used to assess PA.²²⁻²⁷ These studies, while informative, generally have small sample sizes and a small number of activities, limiting the generalizability and applicability of the study results. In 2014, ActiGraph (ActiGraph Corporation, Pensacola, FL), a leading research-grade activity monitor manufacturer, released the GT9X Link (GT9X).²⁹ The GT9X contains the same primary accelerometer as its predecessor, the wGT3X-BT and also contains an inertial measurement unit (IMU). The IMU houses a secondary triaxial accelerometer, a triaxial gyroscope, a triaxial magnetometer, and a thermometer to monitor the sensor temperature. Hibbing et al. 2018³⁰ first introduced the GT9X gyroscope into the literature by demonstrating its utility as a predictor of EE using a two-regression approach. When used in concert with the primary accelerometer data, the gyroscope

contributed a 5.0% improvement in EE estimates compared to using only the accelerometer sensor data. One potentially larger takeaway from this study was the contribution of the gyroscope sensor data for classification of sedentary behaviour (SB). The gyroscope sensor data was a perfect classifier of SB across the hip, both wrists, and both ankles compared to 63.7% to 99.4% for the accelerometer sensor data. Additionally, Marcotte et al.³¹ showed that the gyroscope sensor in the GT9X can be useful for detecting the number and degree of turning motions, a common movement that occurs during many activities of daily living such as household chores as well as during intermittent sporting activities. Capturing additional information about intermittent tasks that involve pivoting and turning may be important for assessing the broad PA behaviours that occur in free-living. It has been shown previously that the number, frequency, and degree of turns contribute to EE.^{32,33} When considered across the day, this may be an important step to improving PA assessment using wearable device and inertial sensor data.

Section 2: Statement of the Problem

To date, most PA classification using wearable devices has been conducted using only accelerometer sensor data. Presently, PA classification has reached a plateau in performance in both youth and adult studies. Regardless of the statistical features and analytical technique used, similar classification performance is achieved in independent model validation, specifically with unstructured free-living PA datasets. One potential reason for this plateau in classification performance is the insufficient availability of relevant information in the acceleration sensor data to adequately discriminate individual activities and activity groups. Potential solutions to improve classification performance include: 1) using gyroscope sensor data, instead of

accelerometer sensor data or 2) using a combination of accelerometer and gyroscope sensor data for classifying PA.

Section 3: Statement of the Purpose

The purpose of this study is two-fold: 1) to determine if using gyroscope sensor data in place of accelerometer sensor data improves the classification of PA in youth compared to using only accelerometer sensor data, and 2) to determine if using a combined sensor approach improves classification of PA in youth compared to using either sensor independently.

Section 4: Research Questions

***Question 1:** Does using only gyroscope sensor data as an independent classifier improve the classification of physical activity type in youth compared to using only accelerometer sensor data?*

Hypothesis 1: It is hypothesized that using only gyroscope sensor data and relevant features to classify physical activity type in youth will yield comparable classification performance to using only accelerometer sensor data.

***Question 2:** Does using accelerometer and gyroscope sensor data together improve the classification of physical activity type in youth compared to using only accelerometer or only gyroscope sensor data?*

Hypothesis 2: It is hypothesized that using accelerometer and gyroscope sensor data to classify physical activity type in youth will result in improved classification accuracy of individual activity classes and activity groups compared to using accelerometer or gyroscope sensor data alone.

References

1. Chen KY, Bassett DR. The Technology of Accelerometry-Based Activity Monitors: Current and Future. *Med Sci Sports Exerc.* 2005;37(Supplement):S490-S500.
2. Clark CCT, Nobre GC, Fernandes JFT, et al. Physical activity characterization: does one site fit all? *Physiol Meas.* 2018;39(9).
3. Mannini A, Rosenberger M, Haskell WL, Sabatini AM, Intille SS. Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle. *Med Sci Sports Exerc.* 2017;49(4):801-812.
4. Rowlands AV, Mirkes EM, Yates T, et al. Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent? *Med Sci Sports Exerc.* 2018;50(2):257-265.
5. Clark CC, Barnes CM, Stratton G, McNarry MA, Mackintosh KA, Summers HD. A Review of Emerging Analytical Techniques for Objective Physical Activity Measurement in Humans. *Sports Med.* 2017;47(3):439-447.
6. Farrahi V, Niemela M, Kangas M, Korpelainen R, Jamsa T. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait Posture.* 2019;68:285-299.
7. Twomey N, Diethe T, Fafoutis X, et al. A Comprehensive Study of Activity Recognition Using Accelerometers. *Informatics.* 2018;5(2).
8. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors--a review of classification techniques. *Physiol Meas.* 2009;30(4):R1-33.
9. de Almeida Mendes M, da Silva ICM, Ramires VV, Reichert FF, Martins RC, Tomasi E. Calibration of raw accelerometer data to measure physical activity: A systematic review. *Gait Posture.* 2018;61:98-110.

10. Figo D, Diniz PC, Ferreira DR, Cardoso JMP. Preprocessing techniques for context recognition from accelerometer data. *Pers Ubiquit Comput.* 2010;14(7):645-662.
11. Jain A.K. DRPW, Mao J. Statistical Pattern Recognition: A Review. *IEEE T Pattern Anal.* 2000;22(1).
12. Kate RJ, Swartz AM, Welch WA, Strath SJ. Comparative evaluation of features and techniques for identifying activity type and estimating energy cost from accelerometer data. *Physiol Meas.* 2016;37(3):360-379.
13. Preece SJ, Goulermas JY, Kenney LP, Howard D. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE T Bio-Med Eng.* 2009;56(3):871-879.
14. Capela NA, Lemaire ED, Baddour N. Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients. *PLoS ONE.* 2015;10(4):e0124414.
15. Chowdhury AK, Tjondronegoro D, Chandran V, Trost SG. Ensemble Methods for Classification of Physical Activities from Wrist Accelerometry. *Med Sci Sports Exerc.* 2017.
16. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and Wrist Accelerometer Algorithms for Free-Living Behavior Classification. *Med Sci Sports Exerc.* 2016;48(5):933-940.
17. Bastian T, Maire A, Dugas J, et al. Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough. *J Appl Physiol.* 2015;118(6):716-722.

18. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *J Appl Physiol*. 2011;111(6):1804-1812.
19. Gyllenstein IC, Bonomi AG. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. *IEEE Trans Biomed Eng*. 2011;58(9):2656-2663.
20. Montoye AHK, Westgate BS, Fonley MR, Pfeiffer KA. Cross-validation and out-of-sample testing of physical activity intensity predictions with a wrist-worn accelerometer. *J Appl Physiol*. 2018;124(5):1284-1293.
21. Mannini A, Intille SS. Classifier Personalization for Activity Recognition Using Wrist Accelerometers. *IEEE J Biomed Health Inform*. 2019;23(4):1585-1594.
22. Ferrari A, Micucci D, Mobilio M, Napoletano P. Human Activities Recognition Using Accelerometer and Gyroscope. Paper presented at: European Conference on Ambient Intelligence 2019.
23. Jain A, Kanhangad V. Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sens J*. 2017;18(3):1169-1177.
24. Wang A, Chen G, Yang J, Zhao S, Chang C-Y. A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sens J*. 2016;16(11):4566-4578.
25. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. *J Med Internet Res*. 2012;14(5):e130.
26. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ. Fusion of smartphone motion sensors for physical activity recognition. *Sensors (Basel)*. 2014;14(6):10146-10176.

27. Shoaib M, Scholten H, Havinga PJ. Towards physical activity recognition using smartphone sensors. Paper presented at: IEEE 10th International Conference on Ubiquitous Intelligence and Computing and IEEE 10th International Conference on Autonomic and Trusted Computing 2013.
28. Parkka JA, K; van Gils, M; Manttari, A; Nieminen, H. Estimating Intensity of Physical Activity: A Comparison of Wearable Accelerometer and Gyro Sensors and 3 Sensor Locations. *Proceedings of the 29th Annual International Conference of the IEEE*. 2007.
29. ActiGraph Corporation. ActiGraph GT9X User Manual. 2014.
30. Hibbing PR LS, Kaplan AS, Crouter SE. Estimating Energy Expenditure with ActiGraph GT9X Inertial Measurement Unit. *Med Sci Sports Exerc*. 2018;50(5):1093-1102.
31. Marcotte RT, Bassett DR, Weinhandl JT, Crouter SE. Application of the ActiGraph GT9X IMU for the assessment of turning during walking and running. *Biomed Phys Eng Express*. 2018;4(6).
32. Hatamoto Y, Yamada Y, Fujii T, Higaki Y, Kiyonaga A, Tanaka H. A novel method for calculating the energy cost of turning during running. *Open Access J Sports Med*. 2013;4:117-122.
33. Hatamoto Y, Yamada Y, Sagayama H, Higaki Y, Kiyonaga A, Tanaka H. The relationship between running velocity and the energy cost of turning during running. *PLoS ONE*. 2014;9(1):e81850.

Chapter 2: Review of Literature

Section 1: Overview

This literature review will cover material related to wearable devices and inertial sensors used for characterizing physical activity (PA). While the relevant literature is broad in scope and spans multiple disciplines, this review will focus primarily on how PA has been assessed using wearable devices and inertial sensors, including identifying current shortcomings in the use of wearable device data for PA characterization. This is intended to correspond to the research questions outlined in Chapter 1 and support the novelty and relevance of the studies accompanying this review. Much of the available data in the literature using wearable devices and inertial sensors to characterize PA comes from studies using adult populations. There is a subset of the literature focused on youth, but a few components are missing from that area. Among those is the use of data from multiple sensor types to characterize PA. The general framework this review is structured around is known as the Activity Recognition Chain (ARC)¹, a five-step process for working with wearable device and inertial sensor data and obtaining a desired outcome. In the case of this review, that outcome will be primarily centered around activity recognition and classification. The ARC is comprised of five general steps: 1) data acquisition, 2) pre-processing, 3) segmentation, 4) feature extraction, and 5) classification.¹

Section 2: Introduction to Physical Activity Assessment

PA is defined as any movement of the body that requires skeletal muscle contraction and increases the expenditure of energy above that which is required at rest.² While seemingly simple in definition, PA is a complex and multi-dimensional behaviour comprised of related but independent constructs including physical activity, energy expenditure, and sedentary behaviour.³ The complex nature of PA can make it challenging to objectively assess, which has

long been of interests to researchers, scientific investigators, and medical and health professionals because of the well-established relationship between PA and positive health outcomes associated with regular PA participation.⁴⁻⁷ Similarly, the consequences of chronic physical inactivity include increased risk for developing a host of chronic disease conditions linked to decreased quality of life, morbidity, and pre-mature mortality.⁸⁻¹⁰ Some of the most prevalent chronic disease conditions in the United States (U.S.) include obesity, hypertension, Type 2 diabetes mellitus, coronary heart disease, and cancer.¹¹ The current recommendation states that U.S. adults obtain 150 to 300 minutes of moderate-to-vigorous- aerobic PA per week and perform muscle-strengthening exercises on at least three days per week, yet many individuals fail to meet this guideline for maintaining or improving their health.¹²

A multitude of methods exist to assess PA and are sub-divided into two primary categories 1) subjective and 2) objective. Subjective methods of PA assessment include self-report surveys, recall questionnaires, PA diaries and logs, and phone or in-person interviews. Objective methods include device-based assessments using wearable devices and direct observation. Subjective PA assessment methods have been the primary means for assessing PA at the population level in the U.S. for decades, largely because of the ability to obtain a large sample of responses and a relatively low cost to administer on a large scale. For example, the majority of surveillance system PA data is collected using self-report survey data.¹³ Only one surveillance system, the National Health and Nutrition Examination Survey (NHANES), has an objective device-based component and it was only introduced in the early 2000s.¹⁴ However, while useful for gaining broad insight into general PA behaviour at the population level, self-report methods of PA assessment lack temporal resolution and are subject to both recall and social desirability biases which in turn results in low precision estimates of PA¹⁵. Conversely,

objective device-based assessments of PA eliminate biases, provide greater temporal resolution, and provide the capability to continuously monitor movement which makes wearable devices popular among PA assessment researchers.

Section 3: Physical Activity Assessment Using Inertial Sensors

An important distinction when discussing the tools used to assess PA is between the physical sensors and wearable devices. The sensors make a physical measurement while the wearable devices acquire the measured data and convert it into a useable output that can be read and interpreted by an individual user or an instrument.¹⁶ The inertial sensors all have similar physical properties, defined by tolerances, sensitivities, dynamic ranges, precision, and accuracy.¹⁶ To achieve valid and reliable measurements with these devices, regular calibration checks are required for maintaining consistent and reliable device output.¹⁶ All of these sensor parameters are relevant when using device data to model various aspects of PA and should be evaluated accordingly when selecting a device for a task.¹⁶

In the field of PA assessment, wearable devices containing inertial sensors have become a popular tool among PA researchers for the ability to non-invasively and continuously monitor movement. The output from these wearable devices is frequently used to relate to different aspects of PA: 1) energy expenditure (EE), 2) activity intensity, and 3) activity type; with the rationale being that higher frequency of movement stems from higher intensity activity and results in greater EE. Additionally, individual activities or domain-specific groups of activities with similar movement patterns result in characteristic patterns in the sensor signal. The device data can then be categorized based on the unique distinguishing characteristics in the sensor signal specific to the individual activity or activity group. While this review is primarily focused on the classification aspect of PA characterization, the motivation to improve the quality of data

and methods used for classification is linked to the EE and intensity components of PA assessment. If activities can be classified accurately then it could become relatively straightforward to estimate EE in a way that is easy to implement and interpret.^{17,18} This rationale ties the classification component of PA assessment to the EE estimation component in that one may benefit from the other but neither component is completely dependent on the other.

3.1 Data Acquisition

The first step in the ARC is data acquisition, which in the context of this review will refer to human movement data collected using inertial sensors contained within wearable devices and smartphones.¹ The most common types of sensors used for PA assessment are accelerometers. Additional sensor types include gyroscopes, and inertial measurement units (IMUs – devices containing an accelerometer, a gyroscope, and sometimes a magnetometer), although these have been less commonly used for PA assessment. Accelerometers measure linear movement (acceleration and deceleration) and gyroscopes measure rotational movement (angular velocity). Magnetometers are most commonly used to measure the strength of the Earth's magnetic, from which position and heading (i.e. cardinal direction) can be determined. However, they are susceptible to environmental interference which can make the signal noisy and difficult to for PA assessment. Accelerometers and gyroscope have been used individually but more emerging research is exploring the utility of using the data streams together for the added benefit of more information about the movement being assessed.¹⁹⁻²¹ The following section provides an overview of the different varieties of these three sensor types and how they have generally been used in PA research.

3.1.1 Accelerometers

Historically, in the field of PA assessment, inertial sensor data has come from accelerometer sensors or from an accelerometer-based wearable device where the primary outcome data are transformations of accelerometer sensor data. Accelerometers measure both static (i.e. acceleration due to gravity) and dynamic (i.e. acceleration due to movement) linear accelerations.²² For PA assessment research, ActiGraph devices have been the most widely used research-grade monitors.²³ ActiGraph has manufactured several generations of research-grade accelerometer-based PA monitors and these devices have traditionally been attached at the hip to be near the center of mass. However, in more recent years the wrists have become a popular alternative or complement to hip-worn devices. Other attachments sites include the thigh and ankle which are less practical but still occasionally used for research purposes. In other disciplines where human activity recognition research is common, such as engineering and computer science, smartphones containing multiple sensor types are frequently used. One of the primary sensors in smartphones is an accelerometer. There are two primary types of accelerometers 1) piezoelectric and 2) microelectromechanical systems (MEMS) accelerometers and each have played a substantial role in the field of PA assessment.

i. Piezoelectric Accelerometers

Early accelerometer sensors were costly to manufacture and generally used a piezoelectric bimorph cantilever beam with a seismic mass mounted on the end.^{22,24} When the sensor experiences a change in acceleration, the force on the seismic mass causes the bimorph cantilever beam with the piezoelectric element to bend and deform.^{22,24} This deformation results in a charge developing on the sensor generating a voltage signal that is proportional to the magnitude of

acceleration the sensor undergoes.²² The piezoelectric bimorph cantilever beam configuration is frequently referred to as uniaxial since the cantilever beam is most sensitive to deformation due to dynamic acceleration in the plane in which it is oriented.²² Uniaxial piezoelectric accelerometers have the capability to measure omnidirectional accelerations but there is little to no way to differentiate the plane from which a dynamic acceleration occurs or to distinguish the relative contribution of accelerations in more than one plane to the total acceleration the sensor undergoes.²² This is a result of the overall sensitivity of the accelerometer sensor which is mainly a function of the physical properties of the sensor components.²² Uniaxial sensors could be configured orthogonally to measure differentiable dynamic accelerations in multiple planes.²² There are also newer integrated circuit piezoelectric sensors that are available as multiaxial sensors, most commonly in a triaxial configuration.²²

There are, however, some limitations to using piezoelectric accelerometers, principle among these limitations is the inability to measure static acceleration reliably.^{22,24} This is due to the sensor only providing alternating current coupled output; meaning the sensor can only reliably measure time-varying dynamic accelerations.²⁴ Another concern is charge leakage over time.²² For example, a time-varying dynamic acceleration will cause a deformation in the bimorph cantilever beam and a proportional electrical charge to form.^{22,24} Over time the charge will start to dissipate even when held under a static acceleration of the same magnitude that formed the original charge.²² The rate of dissipation is dependent on the physical properties of the piezoelectric component of the

accelerometer sensor and can vary between sensors.²² An additional limitation of piezoelectric sensors is susceptibility to drift over time due to mechanical and environmental causes which requires regular sensor calibration.^{22,24}

ii. MEMS Accelerometers

Newer sensor technology such as MEMS are used in nearly every iteration of current generations of research-grade accelerometer-based devices and in smartphones.²⁴ The accelerometers contained in smartphones aid in determining the orientation of the phone screen from portrait to landscape views as well as function to quantify how many steps you take, much like they do in purposefully worn accelerometer-based activity monitors. These sensors function based on a capacitive response where the sensor capacitance changes proportionally to the magnitude of the acceleration the sensor undergoes which results in a stable signal.²⁴ MEMS sensors produce direct current (DC) coupled output, which allows for the sensor to detect and capture both time-varying dynamic accelerations and static accelerations due to gravity.²⁴ This capability allows for a richer acceleration signal that can be used to detect positional information as well as information about movement.²⁴ MEMS sensors are semiconductors and require power to operate which affects battery life and requires regular charging.²⁴ Additionally, MEMS sensors are more robust to calibration concerns due to the stable transducers used and calibration that occurs during manufacturing to maintain the sensor stability.²⁴

3.1.2 Gyroscopes

While gyroscopes have been used in smartphone research in computer science and engineering, they have seen virtually no use in the field of PA assessment. When they have been, it has been to analyze gait patterns²⁵⁻²⁸, to detect falls and fall risk in elderly populations²⁹⁻³¹, to detect changes in postural position and orientation³², and to quantify turning motions³³, including the number of turns and the degree of the turn being made, during walking and running.³⁴ Like accelerometers, gyroscopes are in most smartphones and aid in device orientation and image stabilization as the sensor is highly responsive to the turning and rotation movements made by users. There are three primary varieties of gyroscopes 1) mechanical gyroscopes, 2) optical gyroscopes, and 3) MEMS gyroscopes.

i. Mechanical Gyroscopes

Gyroscopes operate on the principle that an isolated spinning mass will maintain its angular position with respect to an inertial frame of reference.³⁵ When the mass undergoes a constant external torque at a constant angular speed, the axis of rotation undergoes a precession motion at a constant angular speed in a direction that is normal to the direction from which the constant external torque is applied.³⁶ When external forces act on the center of the rotating mass, the angular position of the axis of rotation is unaffected.³⁵ Most mechanical gyroscopes consist of a spinning mass mounted on two fixed gimbals. When the mass is rotating on its axis as intended, the mass should remain parallel to itself and be resistive to changes in its orientation. Mechanical gyroscopes exist in two primary forms: 1) mechanical displacement and 2) mechanical rate gyroscopes.

ii. Optical Gyroscopes

Optical gyroscopes operate by detecting differences in time it takes two counter-propagating beams traveling in opposite directions to propagate through an optical path.³⁵ Rotation induces changes in the length of the optical path resulting in phase differences between the two counter-propagating beams.³⁵ Optical gyroscopes operate under the Sagnac principle³⁷, which states that two counter-propagating optical beams being propagated through a ring-shaped optical path change their phase if the ring-shaped optical path is rotating.³⁵ This makes it possible to then relate the magnitude in the change of the phase to the relative angular speed of the optical ring.^{35,37} Optical gyroscopes primarily exist as 1) ring laser and 2) fiber optic gyroscopes.

iii. MEMS Gyroscopes

MEMS gyroscopes generally operate using a vibrating mechanical element to detect angular velocity.³⁵ Coriolis acceleration is observed in a rotating frame of reference and is proportional to the angular velocity.³⁵ The Coriolis acceleration causes energy to be transferred between two modes of vibration and the vibrating element in the sensor detects this transfer of energy.³⁵ Unlike mechanical gyroscopes, there are no rotating components which makes for easier miniaturization and manufacturing of the sensor which has also helped reduced the overall cost of the sensors.³⁵ MEMS gyroscope are what are contained in most IMUs, smartphones, and other multi-sensor wearable devices.

3.1.3 Inertial Measurement Units

IMUs are used in many applications outside of PA assessment for navigation tasks to establish position and for system stability. The advantage of using an IMU over individual accelerometer and gyroscope sensors is that with IMUs, the sensors are contained in a single unit so all sensors can collect data simultaneously using a single device at a single attachment site.³⁸ This lends itself well to being able to combine the sensor signals using a sensor fusion technique, which is advantageous for some applications, including some aspects of PA assessment.³⁸ ActiGraph released its latest iteration of accelerometer-based PA monitors in 2014, the GT9X Link (GT9X). The GT9X contains the same primary triaxial accelerometer as the previous generation wGT3X-BT but now also includes an IMU, which contains a secondary triaxial accelerometer, a triaxial gyroscope, a triaxial magnetometer, and a thermometer to measure the sensor temperature inside the unit housing. The introduction of the GT9X put IMU use squarely into play for active research in the field of PA assessment, but to date, little research has been conducted using the GT9X IMU sensor data.

3.1.4 Inertial Sensor Output

Accelerometer sensors record linear accelerations and decelerations in gravitational units (g 's) and gyroscope sensors record angular velocity in $\text{radians} \cdot \text{sec}^{-1}$ or $\text{degrees} \cdot \text{sec}^{-1}$. Both sensor signals are recorded as continuous time series data. The rate at which data is acquired is called the sampling rate or sampling frequency and for most human movement applications ranges from 30-100 Hertz (Hz) in increments of 10 Hz. In other disciplines, the sampling frequency can be much higher if greater temporal resolution is required. The critical consideration when using inertial sensors to assess PA is to use a sampling frequency great enough to adequately capture the movements being made to satisfy the Nyquist-Shannon^{39,40} theorem criteria that the sampling

frequency must be at least two times the greatest frequency that needs to be captured. The upper limit for most vigorous, high-frequency human movements is between 8-10 Hz so a minimum sampling frequency of 20-30 Hz should be sufficient to capture the full range of human movements.^{22,24} These estimates are for trunk-based movements near the center of mass where hip-mounted wearable devices are typically worn, however, it should be noted that high-frequency movements of the limbs may peak at closer to 25 Hz and may require a greater minimum sampling frequency to sufficiently capture the full range of human movements when devices are worn on the wrists or ankles.²²

There is added benefit when working with the raw sensor signals. There is a lot of valuable information about the movement being assessed that is contained within the information-rich signals. The raw signals, can however, be noisy, and otherwise difficult to deal with as the higher the sampling frequency, the more individual data points there are to work with. While the raw sensor signal is not always used during the modeling process, it is common to engineer derivations of the raw sensor signal (i.e. statistical features) to represent various temporal and spectral components of the signal more completely before the data is reduced to a more manageable window size, most commonly 1-second windows.⁴¹

3.2 Pre-Processing

The second step in the ARC is pre-processing, where the acquired data are cleaned to reduce artifact and noise by smoothing or filtering. The filtering and smoothing process may require calibration of the sensor to correct the collected data or -resampling of the sensor signal into a lower or higher resolution for further processing. The data filtering step is one of the most prominent components for inertial sensors as the raw sensor signal generally requires filtering before use and to make the sensor data more manageable to work with.^{22,24} Generally, a low-

pass, high-pass, or band-pass (a band of high- and low-pass filter frequencies) filter is applied to the raw sensor data to 1) improve the linearity of the signal (i.e. the relationship between measured versus actual values) by attenuating data outside of the physiological range of possibility, and 2) reduce artifact in the signal due to electrical noise and sensor drift due to temperature or the sensor being out of calibration from aging or damage due to impact.^{16,22,24}

Most wearable device manufacturers typically apply an unpublished proprietary low-pass or anti-aliasing filter to make sure the data satisfies the Nyquist-Shannon^{39,40} theorem criteria, which, if not sufficiently met, results in a phenomenon known as aliasing. Aliasing is the corruption of the sampled signal data and distortion of the represented original signal, which compromises the data considerably.²⁴ The frequency limits of this filter vary by device manufacturer but are generally broad and in many cases require a secondary filter to achieve a cleaner signal representative of the human movement being measured.²⁴ In many cases, the secondary filter is a band-pass filter and may be applied to attenuate signal noise and artifact outside of the desired frequency range.²⁴ For example, ActiGraph's default band-pass filter for current devices is 0.25 – 2.5 Hz.⁴² This, however, is only applied to the data when converting the raw acceleration signal to ActiGraph's activity count metric.

Historically, in the field of PA assessment, raw *g* acceleration data has not been the unit most commonly used. Instead, the field has used an integrated unit of the filtered acceleration signal known as activity counts, a proprietary metric used to quantify movement in a given window of time (i.e. epoch). The process to generate counts requires a series of steps. First, the bidirectional analog voltage signal is filtered and amplified and then converted to a digital series of values representing the original acceleration data.²² The amplification step is dependent on the sensor hardware, the magnitude of the analog voltage signal representing the original

acceleration data, and the analog to digital (A/D) conversion factor (e.g. 8-bit conversion).²² Next, the bidirectional digital acceleration values must undergo full-wave rectification (i.e. taking the absolute value of the signal) to make all of the values positive.²² Then the rectified signal is band-pass filtered and can be integrated (i.e. summed) by taking the area under the curve in a specified window of time.²² The result is an activity count value per unit of time (i.e. epoch) to represent the body movements that occurred during that time window with higher count values corresponding to greater movement frequencies. There are, however, some limitations to using counts to characterize PA. One primary limitation is the loss of resolution that occurs due to signal integration.²² The ability to segment data of different intensities or movement frequencies is diminished because of this lack of resolution since some periods of high-frequency movement and low-frequency movement occur sequentially and may fall in the same epoch if a longer epoch is selected.²² Conversely, if a short epoch length is selected the resulting data may not be practically useful for characterizing PA since nuances in movement frequencies in such small windows have little to no practical physiological meaning.²² Historically, count data has been used in 1-minute epochs. This results in counts being used more descriptively to characterize PA data by quantifying overall PA and time spent in different PA intensities.^{43,44} This makes the choice of epoch length critical to how the data are presented, and results are interpreted and must be carefully considered.²²

In recent years, as technology has evolved and improved, more researchers have opted to use the raw acceleration data instead of counts. One benefit to using raw sensor data to characterize PA includes greater compatibility and comparability.^{45,46} Using the raw sensor data should result in greater compatibility across device brands as the data collection and data modeling processes can become more standardized and the resulting sensor data should be near

equivalent for devices of different brands worn at the same attachment site during the same activities.^{45,46} Similarly, using the raw signal instead of proprietary, manufacturer-specific units (i.e. counts) should result in greater comparability between studies using different device brands as the data will all be in non-proprietary units.^{45,46}

Once a stable, clean, and manageable sensor signal is attained, additional pre-processing steps may involve normalization and unit conversion.^{1,47} In the case of IMUs, an additional pre-processing step may include signal fusion.^{1,47} This involves combining the signal from two or more types of sensors, most frequently the accelerometer and gyroscope signals from an IMU. One reason to fuse the signals is to achieve a single, more information-rich data stream which can be useful in some applications. Another reason is to achieve greater signal stability to reduce the measurement error in each individual signal.^{1,47} Alternatively, sensor fusion is of interest in some cases. This involves pre-processing the individual sensor signals individually so the information unique to each signal is retained and then finally the data from the two sensors are synchronized and merged, in other words, fusing the data from the two sensors into a single data source. The use of both accelerometer and gyroscope sensor data in assessing PA is the best example of sensor fusion since the two signals remain independent of one another, the data are just leveraged from both sensor types to classify PA. Lastly, a final step in pre-processing is the synchronization and aggregation of multiple data streams into a single dataset. For example, in the field of PA assessment, this is often the synchronization and merging of the time series sensor data with criterion activity labels from direct observation and criterion EE data from indirect calorimetry.

3.3 Segmentation

The third step in the ARC involves segmentation of the pre-processed data into sections believed to contain meaningful information about the target outcome of interest.^{1,41} The three primary ways to segment time series data are 1) activity-defined windows, 2) event-defined windows, and 3) sliding windows (i.e. windowing).^{41,48} Of these three approaches, windowing is the most commonly used as it is useful for static PA behaviours (e.g. posture changes), dynamic and periodic PA behaviours (e.g. walking and running), and intermittent PA behaviours (e.g. household chores, sports).⁴⁹ Windowing segments the data into fixed window lengths that typically do not overlap. Window sizes vary from less than 1-s to greater than 1-min but are most commonly 1-s or smaller.⁴¹ Activity-defined windows are based on the detection of changes in activity which then partitions the data stream by the number of activities detected. Lastly, event-defined windows require the detection of specific events in the data to further subset the data in partitions. For example, when walking, the heel strike of the foot is a distinguishable event that can be detected in the acceleration signal. This specific event can be used to partition the data stream by each step that occurs.

3.4 Features

The fourth step in the ARC pertains to feature extraction.¹ Features are statistical transformations and representations of the raw sensor data. In order to model wearable device data most efficiently and effectively, these statistical transformations of the raw data are calculated to represent different aspects of the device output and increase the variance in the dataset. There are a large number of features that can be computed and extracted from the raw sensor data, but they generally fall in the following categories: 1) time domain, and 2) frequency domain. Figo et al.⁴⁷ covers in specific detail the motivation and mathematical computation for a

vast array of both time and frequency domain features. Preece et al.⁵⁰ provides an overview of how features play a role in classification and demonstrate the importance of choosing the right features for a task. Liu et al.⁵¹ is a frequently referenced paper in the field of PA assessment as they provide a fairly comprehensive list of possible features to extract from accelerometer sensor data. These studies indicate that the choice of features impact the outcomes of a study and should be chosen carefully. Another point of emphasis is that different features will result in performance differences between classifiers and should be considered accordingly when development classifiers for a specific outcome.

i. Time Domain Features

Time domain features quantify and represent aspects of the device data over a specified unit of time, hence the time domain designation.⁴⁷ Additionally, time domain features may be the most frequently used and are likely the most identifiable to those new to machine learning. Examples of common time domain features include the following statistical metrics that generally describe different aspects of the sensor signal with respect to time: mean, median, variance, standard deviation, coefficient of variation, minimum, maximum, range, correlation, integration (i.e. area under the curve), and zero-crossings (i.e. the number of time a bi-directional sensor signal crosses zero and changes signs).⁴⁷

ii. Frequency Domain Features

Frequency domain features are particularly useful for quantifying and recognizing repeated patterns in a sensor signal as pattern repetition often corresponds to similar movements being performed and may even include a rhythmic or cyclic pattern to indicate an exact activity such as walking or

running.⁴⁷ One of the most commonly used techniques to transform the raw sensor signal to the frequency domain is to deconstruct the signal into its constituent parts by computing a Fourier transform of the signal. Both the discrete and fast Fourier transforms (FFT) are frequently used to extract valuable signal characteristics such as the DC component (i.e. the first spectral coefficient in the frequency spectrum).⁴⁷ Additionally, other frequency domain features are obtained through a full spectrum analysis that provides information about the energy of the sensor signal and how it is distributed across a range of frequencies. From this information the dominant frequency can be obtained (i.e. the frequency band in which the majority of the signal energy is found) as well as peak power of the dominant frequency.⁴⁷ The frequency domain also includes wavelet transformation, which is frequently used to extract time-frequency feature information from the signal.⁴⁷

3.4.1 Feature Extraction

Discussion of features is synonymous with machine learning for both regression-based and classification tasks.^{48,52-63} As evidenced in the above summary, there is a multitude of features⁴⁷ that can be extracted from the raw sensor signals but there is no established practice for what features are most relevant, have the greatest predictive power, and are the most discriminative for a given task. Thus, the feature extraction process is considered a critical step when modeling wearable device data. There is a balance to consider when extracting features from the original data source, however, and that is somewhat contingent on the number of data samples available. As the number of features approaches and begins to exceed the number of data samples, the “curse of dimensionality”⁶⁴ begins to come into question.⁶⁵ This refers to the

high dimensionality of the dataset created by the large number of features extracted.⁶⁵ The problem this creates is that the feature space for discriminating between data points is minimized and the predictive power of the available information with the large number of features is diminished. Thus, it is common to perform either dimensionality reduction or feature selection procedures to only include the most relevant and strongest predictors in the final modeling step. This reduces redundant and potentially covarying features in the dataset which shrinks the dimensionality of the data and will additionally improve the computational cost and efficiency when modeling.

3.4.2 Dimensionality Reduction

To reduce the dimensionality of the dataset, a procedure such as principal component analysis can be used to map the original features into a lower dimension feature subspace.⁶⁶ The goal here is to not only reduce the dimensionality of the data but also to maximize the variance in the dataset.⁴⁸ For example, the first principal component will have a direction that corresponds to the direction of the greatest variability in the data.⁴⁸ Reducing the dimensionality of the dataset results in lower computational cost and greater computational efficiency.

3.4.3 Feature Selection

Another more common method to reduce the feature subspace is to perform a feature selection procedure to choose the most useful subset of features out of a given feature space for the task at hand. There are both forward selection and backward elimination feature selection methods. Forward selection starts with a single variable or small subset of variables and then progressively incorporates more variables sequentially making the subset get progressively larger until an optimal subset has been achieved. Backward elimination starts with the entire feature set and progressively eliminates the least useful features leaving only a subset of the most useful

features for the outcome of interest. There are three primary methods of feature selection: 1) wrapper, 2) filter, and 3) embedded.⁵⁴

i. Wrapper

Wrapper methods are specific to the outcome of interest and irrespective of the model type. Essentially, the model is treated as a black box that blindly assesses the predictive power of subsets of features for the outcome of interest.⁵⁴ Wrapper methods are most commonly used with a “greedy” search approach of either forward selection or backward elimination to test all possible combinations of features to achieve the best subset for the model type being used for the problem.⁵⁴ Some examples of general wrapper methods include recursive feature elimination, sequential feature selection, and genetic algorithms.

ii. Filter

Filter methods are considered to be generalizable because the features selected are irrespective of the model type and outcome of interest. The reason for this is that subsets of features are selected by filter methods as a pre-processing step.⁵⁴ Because filter methods are frequently used as a pre-processing step and to reduce dimensionality, it is not uncommon to then apply a wrapper method for the outcome of interest with the specific model being used for the task at hand.⁵⁴ Examples of common filter methods include correlations, information gain, and variance threshold.

iii. Embedded

Embedded methods are feature selection procedures built-in to or embedded within the model.⁵⁴ For example, regularization of logistic regression models and tree-based approaches (i.e. classification and regression trees - CART⁶⁷) have built-in methods for selecting subsets of features or ranking features based on relative importance for the outcome of interest.⁵⁴ In cases where dimensionality and computational cost are not a major concern, it is common to use embedded methods and allow the model to select feature subsets objectively based on the task at hand.

Section 4: Modeling Wearable Device Data

The final step of the ARC is modeling the sensor data for an outcome of interest.¹ As previously mentioned, wearable device data are frequently used to model different PA outcomes, including estimating EE, tabulating time spent in different PA intensities, and classify PA type. Modeling these outcomes falls into two general categories for: 1) regression-based tasks for continuous outcomes, and 2) classification-based tasks for categorical outcomes. EE is a continuous outcome so modeling wearable device and inertial sensor data for EE estimation are all regression-based tasks. One example of this would be regressing the continuous time-series data from the device, either counts or the raw sensor signal, against a criterion measure (i.e. indirect calorimetry or direct observation). The resulting regression equation can then be applied to new data to estimate how much EE is expended during an activity or bout of activity. To date, this has been the bulk of the wearable device research conducted in the field of PA assessment. There are other methods used to estimate EE, including machine learning models, however, at

the core, the tasks these algorithms are built to solve are regression-based because the outcome of interest is continuous.

The second task is classification. This applies to a categorical outcome such as a finite number of activity classes represented as either a binomial or multinomial outcome. In this case, the time-series data from the wearable device, either counts or the raw acceleration signal, are used to differentiate between different activities and bin the data based on what the probability a data point or set of data points belongs to a certain activity class based on the criterion labelled input variables. There is considerable interest in classifying human movement data using wearable device data.^{1,52,68-71} As previously mentioned, the PA classification aspect of PA assessment using wearable devices is the component of interest for the proposed studies accompanying this review.

4.1 Machine Learning Overview

Machine learning functions at a basic level as simple pattern recognition. It is the automated process of recognizing, describing, grouping, predicting, and classifying data and has a wide variety of applications and is used in many different disciplines of study.⁶⁰ For example, in the case of wearable device data, this might be the recognition of patterns in the sensor signal, grouping based on the magnitude of the count values, or describing the variations in the time-series data from a wearable device. This information can be used to train a machine (i.e. statistical model) to learn information from the inputted data and then make predictions or classifications from what it has learned about the data.^{59,60,72} Machine learning and pattern recognition are broadly distinguished as either 1) supervised learning, or 2) unsupervised learning tasks.

i. Supervised Learning

Supervised learning involves using labelled data to train models to predict or classify an outcome of interest.³⁸ In the context of this paper, this involves taking criterion labelled input data (i.e. merged accelerometer data and direct observation activity labels, so the model learns what movement data are associated with different types of activities) and training a classifier to be able to make a classification on new unknown data without the activity labels. To accomplish this, the data are frequently split into training and holdout (i.e. testing) datasets where the training dataset is used to develop the model and then the model is tested on the holdout data it has not yet seen to gauge how it performs and might generalize. In cases where there is not a lot of available data, a model is trained using leave-one-participant-out cross-validation (LOOCV), leave-one-group-out cross-validation (LOGOCV), or k-fold cross-validation.^{59,73,74} In the case of LOOCV and LOGOCV, each participant or group of participants are held out of the development process and a model is developed using the data from other participants or groups and then tested on the single holdout participant or group of participants. Alternatively, in the case of k-fold cross-validation, the dataset is partitioned into 'k' partitions and a model is developed using all but one partition and then tested on the held out partition and the process is repeated until all partitions have been used as holdout data. The performance on each holdout partition is then aggregated to give an indicator of overall model performance.

ii. Unsupervised Learning

In many cases, there is no criterion labelled data available to train a model.

These instances are where unsupervised learning is necessary. Since no information is available to guide the training process, oftentimes clustering algorithms, such as k-means clustering, are used to determine how data are grouped objectively and then potential insights can be gained from the clusters of data.³⁸ A second use for unsupervised learning is for reducing dimensionality. High dimensional data occurs when there are more input features than there are training data points (i.e. “the curse of the dimensionality”)⁶⁵. This is a process that examines the interrelatedness of the input features and reduces the total number of features by eliminating redundant features from the dataset. This results in a sparser dataset which is advantageous in terms of both computational cost and efficiency.

4.1.1 Bias-Variance Tradeoff

When building a supervised learning model, it is important to consider not only how well the model does with the training data but also how well it will do with holdout data and then eventually entirely new out-of-sample data. This consideration is frequently referred to as the bias-variance tradeoff.^{57,59,72,75} Bias is the error (difference) between the average prediction the model makes and the correct class that the model is attempting to predict. High bias results in oversimplified models that do not fit the training or testing data very well. Variance is the error that results from the trained model being sensitive to small variations and nuances in the training data. High variance results in a model that fits the training data well but does not fit the testing data well.

Ideally, a balance can be struck that results in a good fitting model that performs well on the training data but also generalizes and performs well when tested with the holdout data. More often, the trained model either under- or over-fits the training data. Underfitting the training data results in a model with high bias and therefore does not perform well with the training or testing data. Conversely, overfitting the training data results in high variance. In most cases when fitting a model, there is a concern of overfitting.⁷⁶ To reduce the possibility of the final model overfitting the training data, individual model hyperparameters are tuned and optimized and precautions are taken, such as random re-sampling and cross-validation, to reduce the likelihood of overfitting.^{59,72,77} The model excellently classifies the training data as the model has learned all of the intricacies and nuances of the training data structure.

4.2 Machine Learning Model Types

4.2.1 Bayesian Networks

Bayesian networks use Bayes theorem to describe the probability of an event based on known conditions related to the event. This particular type of statistical inference is called Bayesian inference which is used to develop probabilistic graphical models to compute probabilities of an event or series of events.

i. Naïve Bayes

Naïve Bayes is based on Bayes probability theory and is considered a base level classifier and has been used extensively for classifications tasks in various disciplines.^{55,78} While naïve Bayes is considered to be a simpler approach than many other machine learning techniques, it is generally competitive when compared to other more sophisticated approaches for the same task. The probability of an outcome of interest, given the input features, is computed using

the training data provided as a conditional probability distribution.⁷⁹ Naïve Bayes assumes the input features are conditionally independent of one another given the outcome of interest.⁷⁹ This is done to simplify the probability calculation. Naïve Bayes has no parameters to optimize making it a computationally inexpensive classifier. However, this comes at a cost, as naïve Bayes performance typically deteriorates as the size of the dataset grows.^{55,78}

ii. Hidden Markov Models

Hidden Markov models are a specific type of Bayesian network based on the Markov process, which assumes that future events are independent of past events with respect to present events.⁸⁰ Hidden Markov models allow for the computation of the joint probability of “hidden” states given the information in available observations.⁸⁰ This representation of the probability distributions over a sequence of observations makes Hidden Markov models useful for modeling time series data.⁸⁰

4.2.2 Logistic Regression

Logistic regression is a widely known and used statistical classification technique. Hosmer and Lemeshow⁸¹ showed that the relationship between an outcome of interest and the input features can be determined by the ratio of odds from the training data provided.^{55,78} The odds ratios are then used to predict the probability of what class a given data point belongs to.^{55,78} By itself, logistic regression is a fairly standard base-level classifier, used similarly to Naïve Bayes as a starting point, as it is computationally inexpensive in most cases and has many applications due to the output being interpreted as probabilities. However, logistic regression

does frequently require a large sample size to make stable classifications and it can be susceptible to multicollinearity.^{55,78}

i. Regularization (Shrinkage) of Logistic Regression

In cases where there are a large number of input features, logistic regression can become more computationally expensive and the resulting coefficients can become more difficult to interpret. Logistic regression models with large sets of input features are also the ones most susceptible to multicollinearity. One way to address this issue, as well as the concern of overfitting, is to use a shrinkage parameter which shrinks the value of coefficients for non-relevant or unimportant features.⁸² Regularizing logistic regression results in lower variance which means the model is less prone to overfitting and should generalize to new data more effectively than standard logistic regression. The shrinkage parameter, λ , serves as a penalty term, that penalizes non-relevant variables by shrinking their coefficients, so those variables do not influence the model as heavily. Regularization comes in two primary forms, ridge regression and least absolute shrinkage selector operator (Lasso).

a. Ridge

In the case of Ridge regression, a penalty equal to the squared magnitude of the coefficients multiplied by the value of λ is added to the loss function.⁵⁹ This penalizes the variables the model overemphasizes by shrinking the coefficients towards zero but not to zero.⁵⁹ The larger the value of λ , the closer the coefficients of overemphasized variables are driven to zero.⁵⁹ While this is useful in cases where there are not many features and plenty of data points,

Ridge does not eliminate any non-relevant variables from the model which does not ultimately reduce the model variance.⁵⁹

b. Lasso

In the case of Lasso, a penalty equal to the absolute value of the magnitude of the coefficients multiplied by the value of lambda is added to coefficients the model overemphasizes.⁵⁹ Unlike Ridge, this drives the coefficients of overemphasized variables exactly to zero, thereby removing them from the model entirely.⁵⁹ This results in a simpler, sparser model that is essentially a subset of the original input features. This is useful as a step in determining the most relevant features for a classification task as well as serving as an automated subset selection process when logistic regression is the model type of interest.

4.2.3 K-Nearest Neighbors

K-nearest neighbors is considered an instance-based machine learning approach that works by comparing a data point to the surrounding data points (i.e. neighbors) to determine what class the data point should be classified as based on its proximity to other data points.^{50,55,60,78} The intuition is that data points from the same class will be generally clustered or grouped together. The 'k' value is the parameter that is optimized and is essentially the number of neighbor data points the instance being classified is compared to in order to decide which class it should belong to. In the case of classification, this can be viewed as a majority voting procedure.

4.2.4 Support Vector Machines

Support vector machine models are complex but often result in generally high performance for a variety of classification tasks. Support vector machines function by mapping training data points in a multi-dimensional feature space to allow for greater separation of data points than would be possible in a single dimension feature space.^{55,78} Hyperplanes are then created to separate the data points in the multidimensional feature space to maximize the minimum distance between data points that are in different classes.^{55,78} The data points are classified based on the hyperplane in which they reside. Support vector machines routinely perform well in cases where there are a large number of input features and are useful for modeling non-linear trends in data by using different kernel functions.^{55,78}

4.2.5 Artificial Neural Networks

Artificial neural networks were developed to function similarly to a biological neurological pathway. Statistically, this works by taking a series of input data and feeding the data into intermediate nodes, often called the hidden layer(s).^{55,78} The connections between input nodes and the nodes in the hidden layer are assigned a weight parameter that factors into the transformation of the input data in each hidden layer node.^{55,78} Each connection between nodes has a weight value assigned and this weight parameter is what is generally modified and adjusted during the learning process based on the initial weight values.^{55,78} The weighted input data are then processed through an activation function that propagates a classification or prediction on to the next hidden layer node(s) or as the final output.^{55,78} The number of nodes and the number of layers of nodes can be varied based on each individual problem. Artificial neural networks are commonly used for a variety of machine learning tasks but typically perform better with large volumes of data which is not always feasible or practical. Additionally, artificial neural networks

are generally computationally expensive in terms of both resources and time required to train the model.

In recent years, a new spin on artificial neural networks has emerged and is frequently referred to as deep learning. This machine learning technique is considered to be at the forefront of artificial intelligence. Deep learning models are at their core neural networks. The deep learning aspect of the name refers to the model depth or the number of nodes and layers in the model which is typically much greater than a standard neural network. The main benefit of deep learning is that instead of training the neural network for a given task using a training dataset, the deep learning model learns on its own and continues to evolve as new data become available without the required training step.

4.2.6 Decision Trees

Decision tree models are rule-based classifiers that create splits in the data based on perceived variable relevance.^{55,78} To create these splits, information gain, entropy, or Gini criteria are frequently used to determine when a split should occur and create a new branch and new leaves of the tree.^{55,67,78} Decision trees are logic-based and not math-based and can thus often be expressed visually making them more approachable in terms of explanation and interpretation. Decision trees are fairly computationally inexpensive in terms of training time, but they can be prone to overfitting if not optimized and pruned appropriately.

4.2.7 Ensemble Models

It is common practice to train several machine learning models to see which may be best suited for a given problem.^{50,55} One way to accomplish this would be to use a series of models, often called an ensemble, and use the predictions or classifications from each model to inform

the final decision.⁷⁷ This can be done using multiple weak classifiers and using a majority voting or weighted voting procedure to establish what the final outcome should be.⁷⁷ This, in theory, should achieve higher performance than using multiple weak classifiers individually.⁷⁷ While this ensemble process can be done manually using a variety of classifier types, the following three tree-based models are frequently used as standalone ensemble approaches.⁷⁷

i. Bootstrap Aggregation (Bagging)

Bagging or bagged decision tree models function by growing multiple trees using subsets of the data and all available predictors. The decisions of each of the individual trees are then averaged and the final decision is based on the most commonly occurring class.

ii. Boosting

Boosted decision tree models function by growing multiple trees in sequence with each tree using a subset of the data. This process typically starts out by growing simpler trees and the subsequent trees grown in the sequence learn from the errors made in the previous trees thereby “boosting” the performance of trees further along in the sequence.

iii. Random Forest

Random forest models function by taking the bagging procedure one step further by growing multiple trees but instead of each tree using all predictors to make decisions about the subset of data, each tree using a random subset of the available predictors as well. The decisions of each of the many trees are pooled using a weight or majority voting procedure to make stronger final classifications about the outcome of interest. Random forest models are popular for many

machine learning tasks as they generally perform well, are computationally affordable in most instances, and are robust to noise and less prone to overfitting than standard decision trees.

Section 5: Characterizing Physical Activity using Wearable Device and Inertial Sensor Data

Now that the five steps of the ARC have been covered to provide context for the proposed studies accompanying this review, this section of the document will focus on how the wearable device data has been used to characterize PA over the last two decades when wearable devices began to be implemented in studies assessing PA behaviour. While not limited to these, there have been a number of studies conducted in both youth and adult samples that use machine learning techniques to model count and raw acceleration data for classifying PA type. The most common machine learning algorithm classifiers used to date include: hidden Markov models⁸³, artificial neural networks^{55,84-100}, K-nearest neighbors^{55,88,91}, support vector machines^{55,88,90-92,101-107}, decision trees^{55,90-92,107,108}, logistic regression^{17,55,90,102,107}, naïve Bayes^{55,88,90,107}, and random forest^{55,88,91,92,104,105,109-111} models.

5.1 Classifier Development and Within-Sample Cross-Validation Performance

The current structure for developing a classifier or predictive model using wearable device or inertial sensor data requires some structure to adequately capture the information needed to develop the algorithm. In many studies, referred to as calibration studies, participants are instructed to complete a battery of structured or semi-structured activities for a specified length of time. While participants are performing the activities, they are directly observed, or video recorded to make a time-stamped record of when activities start and stop and what

behaviours and postures occur during the activity window. These time stamped direct observation logs are later merged with the time series device data, so a record exists where for each epoch, a criterion label exists to correspond with the sensor data. In doing so, supervised classifiers and predictive models can be developed, and model performance can be assessed directly since the data labels are known quantities. Below are summaries of some key studies in the field of PA assessment in both youth and adults using both counts and raw acceleration data. These studies use machine learning to classify PA and the performance results reported are all within-sample cross-validation summaries.

Bao and Intille⁴⁹ collected data on 20 apparently healthy adult participants (13 males, 7 females, mean age 21.8 yrs) who each wore five biaxial accelerometers (ADXL210E; sampling frequency 76.25 Hz). Devices were positioned on the upper arm, wrist, hip, ankle, and thigh. Twenty different activities were prescribed but performed under free-living conditions without direct observation and instead the criterion labels came from users annotating the start and stop time of each activity. Activities included walking, sitting, standing, watching tv, running, stretching, scrubbing, laundry, brushing teeth, riding an elevator, waling carrying a load, working on a computer, eating/drinking, reading, cycling, resistance training, vacuuming, lying down, stair climbing, and riding an escalator. Data were collected on two separate days, once where all activities were performed in an obstacle course style format where the activities were goals for participants to complete as if they were engaging in that activity under free-living conditions. On the second visit, a random subset of the 20 activities were performed in a specific order but participants still noted the start and end time of each activity period. These structured data served as the validation dataset. Time and frequency domain features were extracted from 6.7-s windows with a 50% overlap. Naïve Bayes, instance-based learning, decision table, decision tree

classifiers were developed for classifying PA type. Classifiers were evaluated using user-specific data and using a LOOCV procedure where classifiers were trained using the structured data and tested using the obstacle course data. For the user-specific training classification accuracy ranged from 34.9% to 71.6% and 46.8% to 84.3% for the LOOCV procedure using data from all five attachment sites. The decision tree classifier achieved the highest classification accuracy, 77.3% for the user-specific procedure and 73.0% for the LOOCV procedure. To determine the relative contribution of each attachment site, the difference in overall classification was examined for each attachment site individually. Classification accuracy decreased between 29.5% (thigh) and 63.7% (arm) depending on the attachment site. The authors did note that if data from two attachment sites was retained, in this case either thigh and wrist or hip and wrist, classification accuracy only decreased by approximately 4.1% and remained above 70.0% overall.

Pober et al.⁸³ collected data on six apparently healthy adult participants (four males, two females, mean age 24.9 yrs). Participants wore an ActiGraph 7164 uniaxial accelerometer-based activity monitor on the right hip while performing a series of four semi-structured activities for seven minutes each under laboratory conditions including 1) treadmill walking, 2) treadmill walking uphill, 3) vacuuming, and 4) seated computer work. A supervised quadratic discriminant analysis model using 15-s count values and an unsupervised hidden Markov model using second-by-second data were trained and cross-validated using a LOOCV procedure. Within-sample classification accuracies of 70.9% and 80.8% were achieved using the quadratic discriminant analysis and hidden Markov models, respectively.

Crouter et al.¹¹² collected data on 48 apparently healthy adults (24 females, 24 males, mean age 35 yrs) for the purpose of developing a novel approach to estimate EE using a two-regression approach instead of the standard single regression method in an effort to better model

the data from locomotion activities and intermittent lifestyle activities. These data were later used by Staudenmayer et al.⁸⁶ to classify PA type using an artificial neural network. Participants completed at least one of three activity routines comprised of six activities each lasting 10 minutes in duration while wearing an ActiGraph 7164 accelerometer-based activity monitor on the hip (sampling frequency 30 Hz; 1-s epochs). Activities included: 1) lying, 2) standing, 3) computer work, 4) filing paperwork, 5) stair climbing at a self-selected pace, 6) cycling at a self-selected work rate, 7) overground walking (~3 MPH), 8) overground walking (~4 MPH), 9) basketball (1v1), 10) racquetball (1v1), 11) overground running (~5 MPH), 12) overground running (~7 MPH), 13) vacuuming, 14) sweeping/mopping, 15) washing dishes, 16) mowing the lawn (push mower), 17) raking grass/leaves, and 18) washing windows. Activities were collapsed into four categories: 1) low level of activity, 2) household/other, 3) locomotion, and 4) vigorous sport. The following time domain features were extracted from the count data in one-minute windows: the 10th, 25th, 50th, 75th, and 90th percentiles of the vertical axis counts, and the lag-1 autocorrelation. Using a within-sample LOOCV procedure to estimate performance, a classification accuracy of 88.8% was achieved.

Bonomi¹¹³ collected data on activities of daily living in 52 health adult participants (29 males and 23 females; mean age: 29.2 yrs) under laboratory conditions. Participants wore a triaxial accelerometer (Tracmor; 20 Hz) on the waist and performed the following series of PA behaviours: 1) lying, 2) standing, 3) washing dishes, 4) sweeping, 5) treadmill walking, 6) over-ground walking (indoors and outdoors), 7) walking - ascending and descending stairs, 8) stationary cycling, 9) over-ground cycling, 10) treadmill running, and 11) over-ground running. Speeds were not reported for the walking and running activities. The activities were collapsed into six activity groups: 1) lying, 2) sitting/standing, 3) dynamic/transitions, 4) walking, 5)

cycling, and 6) running. The following features were extracted from 6.4-s windows and normalized: mean, standard deviation, kurtosis, skewness, range, cross-axis correlation, accelerometer angle, spectral energy, spectral entropy, peak frequencies, and cross-spectral densities. A decision tree, a neural network, and a support vector machine were developed to classify the six-class activity groupings. Additionally, a majority voting procedure was created to aggregate the decisions of the three individual classifiers. LOOCV was used to estimate model performance on the training data where classification accuracy ranged from 91.4% to 95.1% across the six activity classes.¹¹⁴

De Vries et al.⁸⁵ collected data on 49 apparently healthy adult participants (21 males, 28 females, mean age 38 yrs). Participants wore an ActiGraph GT1M accelerometer-based activity monitor on the right hip and right ankle and performed a series of structured activities under laboratory conditions including sitting, standing, stair climbing, walking, and cycling for a total of approximately 45 minutes. The following time domain features were extracted in one-minute windows: 10th, 25th, 75th, and 90th percentiles, absolute deviation, coefficient of variability, and lag-one autocorrelation. Artificial neural networks were developed to classify the five activity classes using 1) only hip data, 2) only ankle data, and 3) hip and ankle data combined. Classification accuracies across the five classes averaged 80.4%, 77.7%, and 83.0% for the hip, ankle, and combined models.

In a separate study, De Vries et al.⁸⁵ collected data on 58 apparently health youth participants (31 males, 27 females, mean age 11 yrs). Participants each wore an ActiGraph GT1M uniaxial accelerometer-based activity monitor and an ActiGraph GT3X triaxial accelerometer-based activity monitor on the right hip and right ankle. The following series of semi-structured self-paced activities were performed by each participant under laboratory

conditions in 20-minute period: sitting, standing, walking, running, skipping rope, soccer, and cycling. The following time domain features were extracted from one-minute windows: 10th, 25th, 50th, and 90th percentiles, lag-one autocorrelation, absolute deviation, and coefficient of variability. Four artificial neural networks were developed, one for each attachment site and each monitor. Classifier performance was estimated using a within-sample LOOCV procedure. The uniaxial classifiers average classification accuracy was 72.0% (hip) and 57.0% (ankle) compared to 77.0% (hip) and 68% (ankle) for the triaxial classifiers.

Ruch et al.¹¹⁵ collected data on 41 healthy youth participants ages 9-12 years old. Participants wore a uniaxial ActiGraph GT1M device on the hip and wrist for seven days. During the week of measurement, participants were video recorded for an average of 7.3 hours in bouts of 1-3 hours. Video observations were recorded during 1) class instruction at school, 2) physical education participation, 3) unstructured free-living activity in the home, 4) time spent outdoors, and 5) during structured leisure-time activity. During analysis of the video recordings, activities were grouped homogenously into nine principal classes including: 1) stationary activities (e.g. sitting, lying, standing), 2) walking, 3) running, 4) jumping, 5) floor exercise (e.g. gymnastics, tumbling), 6) biking, 7) horseback riding, and 8) crawling, and 9) scooter, however, the “scooter” class was eliminated due to insufficient data existing in the dataset. The full dataset was randomly partitioned into training (n = 21) and testing (n = 20) subsets. Three model types were developed and trained on the 1-s data from the 21 training participants including k-nearest neighbor (KNN), custom decision tree, and a normal density discriminant function model. The three models were used together to classify holdout data from the test dataset using a majority vote based on the decision from each of the three individual classifiers. If no majority vote could be achieved due to individual classifier disagreement, the final classification was labeled as “not

assigned”. Overall, a maximum classification accuracy of 67% was achieved by the KNN and majority vote developed using both hip and wrist data together. The maximum classification accuracy using only hip data was 59% (KNN). The relatively poor performance of the classifiers in this study may be largely attributable to the dataset being comprised of unstructured free-living activity. A primary contribution of this study is the majority vote component. This is essentially an ensemble approach and can easily be modified and adapted for use in other studies and may help improve overall classification accuracy and performance when translating from structured lab-based activities to unstructured free-living activities.

Zhao et al.¹⁰² collected data on 69 healthy children ages 3-5 years old who completed a structured protocol over a 7-hour period while being directly observed. Participants wore a triaxial ActiGraph GT3X+ (30 Hz) on the right hip. The multinomial logistic regression and support vector machine models were trained to classify 60-s count data as one of the following 11 principal classes: 1) sleep [duration (45-120) min], 2) watching television [seated; 20 min], 3) coloring [sitting at a desk; 10 min], 4) video gaming [seated; 10 min], 5) puzzles [seated; 10 min], 6) toys [15 min], 7) ball toss [15 min], 8) active video gaming [standing; 10 min], 9) dance [video instruction; 15 min], 10) aerobics [video instruction; 15 min], 11) running in place (video instruction; 6 min). K-means clustering was used to identify natural clustering of the minute-by-minute data and then subsequently recategorize the data into six principal activity categories based on the intensity of the activity. Ten-fold cross-validation was used to train the support vector machine and multinomial logistic regression classifiers. Both models were trained with and without sleep included. In general, the support vector machine models had lower cross-validation error than the multinomial logistic regression models using the same features. Of all combinations of variables used to develop the different models, the support vector machine

model without sleep data using individual axis count values, vector magnitude, steps, 1- and 2-min lag and lead time, and position (i.e. inclinometer) as a continuous variable resulted in the lowest overall cross-validation error (24.7%) and highest overall classification accuracy (82.0%). A primary contribution of this study was the use of k-means clustering to objectively group activities based on the features of the data instead of grouping them by domain or other similar grouping strategy that is more frequently used to group activities together for classification.

Trost et al.⁸⁷ collected data on 100 apparently healthy youth participants ages 5 and 15 years old (mean age 11.0 yrs). Participants wore an ActiGraph GT1M uniaxial accelerometer-based activity monitor (sampling frequency 30 Hz) on the right hip while performing 12 structured activities that were collapsed into five activity categories as follows: 1) sedentary [lying, computer work], 2) walking [slow and brisk overground walking, treadmill walking, 3) overground running, 4) light intensity household activities and games [sweeping, laundry, catch], and moderate-to-vigorous intensity household activities and games [basketball and aerobic dance]. Activities were performed across two laboratory visits in two groups of six with each activity lasting approximately 5 minutes. One second epochs were selected from the middle 2.5 to 4.5 minutes of each activity where the following time domain features were extracted: 10th, 25th, 50th, 75th, and 90th percentiles, and lag-one autocorrelation. These features were extracted in 10-s, 15-s, 20-s, 30-s, and 60-s windows using the second-by-second data from the 2 minutes of available data of each activity. Average overall classification accuracy ranged from 80.2% to 86.8% across the five activity groups depending on the window size chosen with accuracy increasing as window length increased from 10-s to 60-s.

Mannini et al.¹⁰³ collected data on 33 apparently healthy adult participants who wore a triaxial accelerometer-based device (Wocket – sampling frequency 90 Hz) on the wrist and

ankle. Each participant performed 26 individual activities of daily living ranging from sedentary to vigorous intensity with each activity lasting three to five minutes in duration. The original activities were collapsed into four broad activity groups: 1) ambulation [overground walking carrying a load, ascending and descending stairs, treadmill walking (2, 3, and 4 MPH), treadmill walking (3 MPH at 0%, 6%, and 9% incline), overground walking], 2) cycling [overground (level surface, uphill, downhill), stationary (50 W, 70 RPM)], 3) sedentary [computer, reading/writing, sorting/organizing, lying, sitting, standing, video gaming (handheld and Nintendo Wii)], and 4) other [painting, sweeping]. A support vector machine was developed to classify the raw accelerometer sensor data as one of the four activity categories. The following time and frequency domain features were extracted from 12.8-s non-overlapping windows: mean and standard deviation of vector magnitude, total power, first and second dominant frequencies, ratio between the dominant frequency and the preceding and succeeding windows, and two wavelet-based features. Classifier performance ranged from 84.7% for the wrist to 95.0% for the ankle in a within-sample cross-validation using LOOCV.

Trost et al.¹⁷ collected data in 52 apparently youth participants ages 7 to 18 years old (mean age 13.7 yrs). Participants wore an ActiGraph GT3X+ on the hip and non-dominant wrist (sampling frequency 30 Hz) and completed the following 12 standardized activity trials over two laboratory visits with each activity lasting 5 minutes: 1) lying down, 2) handwriting, 3) laundry, 4) catch, 5) slow over-ground walking, 6) aerobic dance, 7) computer gaming, 8) sweeping, 9) brisk over-ground walking, 10) basketball, 11) over-ground running, and 12) brisk treadmill walking. Two minutes of data were extracted from the middle of each activity and the raw acceleration signal from those two minutes was used to develop regularized lasso logistic regression models for the hip and non-dominant wrist. The following time and frequency domain

features were extracted in non-overlapping 10-s windows and included percentiles, mean, lag one autocorrelation, standard deviation, and coefficient of variation. The within-sample LOOCV procedure resulted in maximum overall classification accuracies of 91.0% for the hip model and 88.4% for the non-dominant wrist model.

Ellis et al.¹¹¹ collected data on 40 apparently healthy adult participants who performed a randomized series of eight activity trials under laboratory conditions for approximately one-minute each. Activities included: slow walking, brisk walking, jogging, stair climbing, laundry, window cleaning, dusting, sweeping, and washing dishes. Activities were collapsed into four categories: 1) walking, 2) running, 3) household, and 4) stairs. Participants wore an ActiGraph GT3X+ (sampling frequency 30 Hz) on the left hip, right hip, and non-dominant wrist. The time and frequency domain features were extracted from the vector magnitude or individual axis data in one-minute non-overlapping windows: mean, standard deviation, coefficient of variation, minimum, maximum, percentiles, lag-one autocorrelation, third and fourth moments, skewness, kurtosis, dominant frequency, power at the dominant frequency, total signal energy, entropy, fast Fourier transform coefficients, cross-axis correlation, mean pitch, mean roll, mean yaw, standard deviation of pitch, roll and yaw, and principal direction of motion. A random forest classifier was developed to predict activity type using within-sample LOOCV. The average classification accuracy was 92.3% for the hip classifier and 87.5% for the wrist classifier across the four classes. When the eight individual activity classes were evaluated, average classification accuracies 70.2% and 80.2% were achieved using the hip and wrist classifiers, respectively.

Montoye et al.¹⁰⁰ collected data on 44 apparently healthy adult participants (22 male, 22 female, 18-44 yrs). Participants wore a GENEActiv triaxial accelerometer-based activity monitors on each wrist (sampling frequency 20 Hz) and an ActiGraph GT3X+ triaxial

accelerometer-based activity monitors on the right thigh and right hip (sampling frequency 40 Hz, reintegrated and downsampled to 20 Hz). Participants performed 14 different activities for three to 10 minutes each over a 90-minute period. Activities were performed at a self-selected pace in an order determined by the participant to simulate free-living conditions. The 14 activities were collapsed into the following 10 activity categories: 1) sedentary [lying, reading, computer], 2) standing, 3) lifestyle [laundry, sweeping], 4) slow walking overground, 5) brisk walking overground, 6) jogging overground, 7) stationary cycling, 8) stair climbing, 9) exercise [body weight squats and bicep curls], and 10) device non-wear. Four different feature sets were compiled from a combination of the following time domain features in 5-s windows: mean, variance, covariance, minimum, maximum, mean axis orientation, variance of axis orientation, 10th, 25th, 50th, 75th, and 90th percentiles of the acceleration signal for each axis. Artificial neural networks were developed using data from each of the four attachment sites and evaluated using a within-sample LOOCV procedure. For the individual activity classes, classification accuracies averaged 66.2% (range: 36.5% - 92.5%) for the hip, 71.4% (38.4%-93.1%) for the thigh, 80.9% (66.9%-95.3%) for the left wrist, and 81.1% (59.9%-96.1%) for the right wrist. The lowest classification accuracy was for reading and the highest classification accuracy was for jogging for all four attachment sites. For the activity groups, classification accuracies averaged 72.5% (56.9% - 92.5%) for the hip, 84.0% (69.6%-93.1%) for the thigh, 86.6% (69.9%-95.3%) for the left wrist, and 86.7% (69.4%-96.1%) for the right wrist. The lowest classification accuracy achieved was for the standing activity class for the hip and thigh and the brisk walking activity class for the wrists. The highest classification accuracy achieved was for the jogging activity class for all four attachment sites. This study illustrates that higher PA classification accuracy can be achieved using data from wrist worn devices as compared to hip and thigh worn devices.

Pavey et al.¹¹⁰ collected data on 21 apparently healthy adult participants who each completed seven structured activities for three minutes each under laboratory conditions in a single session lasting approximately 45 minutes: 1) lying, 2) sitting still, 3) standing still, 4) active sitting, 5) active standing, 6) walking, and 7) running. Participants additionally completed a 24-hr free-living assessment. Participants wore a GENEActiv (sampling frequency 30 Hz) accelerometer-based activity monitor on their non-dominant wrist. A random forest classifier was trained to classify four activity classes: 1) sedentary, 2) stationary (sitting or standing), 3) walking, and 4) running. Data from the middle two minutes of each activity were extracted and time and frequency domain features were extracted in 10-s non-overlapping windows from the list outlined by Liu et al.⁵¹: mean, standard deviation, percentiles, mean absolute deviation, lag one autocorrelation, power, dominant frequency, dominant frequency magnitude, and entropy. The aim of the study was to cross-validate the random forest classifier using the laboratory data and then perform a free-living validation of the classifier using the 24-hr data and ActivPAL data as the criterion measure for comparison of stepping versus non-stepping time. The overall classification accuracy from within-sample LOOCV was 92.7% across all four activity classes and ranged from 80.1% (sedentary) to 95.7% (stationary) for the individual activity classes. For the free-living validation, the walking and running groups were collapsed into a stepping category and the sedentary and stationary activity groups were collapsed into a non-stepping category. The classification accuracy of the random forest classifier for identifying stepping and non-stepping time was 93.7% and kappa was 0.47 using the ActivPAL as the criterion measure of stepping and non-stepping time.

Trost et al.¹⁰⁵ collected data on eleven pre-schooled-aged children ages three to six years old (mean age 4.0 yrs). Participants wore an ActiGraph GT3X+ (sampling frequency 100 Hz) on

the right hip and non-dominant wrist and completed 12 semi-structured activities lasting approximately 5 minutes each over the course of two laboratory visits (6 activities per visit). Activities included: 1) watching television, 2) being read to while seated, 3) making a collage on the wall while standing, 4) walking, 5) playing an “active” game against a member of the research team, 6) completing an obstacle course, 7) playing a computer game while seated, 8) playing with toys while seated, 9) hunting treasure, 10) tidying up toys, 11) riding a bicycle, and 12) running. The following time and frequency domain features were extracted during the middle two-four minutes of each activity trial: mean, standard deviation, minimum, maximum, interquartile range, percentiles, coefficient of variation, signal sum, signal power, peak-to-peak amplitude, median crossings, dominant frequency, magnitude of dominant frequency, and signal entropy. The within-sample LOOCV performance resulted in maximum overall classification accuracies of 80.2%, 78.1%, and 81.8% for the random forest models using hip, wrist, and hip + wrist data combined, respectively. Similarly, the support vector machine models achieved 81.3%, 80.4%, and 85.2% maximum classification accuracy for the hip, wrist, and hip + wrist models combined, respectively.

Chowdhury et al.⁹¹ collected data on 17 apparently healthy youth participants ages 12-17 years old. Participants wore an ActiGraph GT3X+ (30 Hz) on the non-dominant wrist. Participants completed a series of 12 structured activity trials in a laboratory environment that included: 1) lying, 2) laundry, 3) writing, 4) catch, 5-6) over-ground walking (slow and brisk), 7) aerobic dance, 8) computer gaming, 9) sweeping, 10) basketball, 11) over-ground jogging, and 12) treadmill walking. Activities were performed in groups of six across two laboratory visits with each activity lasting approximately 5 minutes in duration. For the analyses, the original 12 activities were collapsed into seven unique activity categories including lying, sitting, standing,

walking, running, basketball, and dance. Data were normalized and the following time and frequency domain features were calculated from 10-s non-overlapping windows: mean, standard deviation, minimum, maximum, variance, median, skewness, kurtosis, percentiles, spectral energy, dominant frequency, the magnitude of dominant frequency, zero-crossings, and cross-axis correlations. The authors used a series of machine learning models to classify PA type that included random forest, bagged decision tree, boosted decision tree, K-nearest neighbors, support vector machine, artificial neural network, decision tree, and three custom fusion ensemble methods, weighted majority voting, naïve Bayes, and behavioral knowledge space. Model performance was reported as F1-scores, a value representing the harmonic mean of precision and recall. Overall, the highest performance was achieved using the custom weighted majority voting ensemble approach which had an F1 of 86.0%. The random forest and support vector machine models had similar performance with F1 scores of 85.0% and 84.0%, respectively. All models achieved F1 scores >80.0% except for the decision tree model (76.0%) indicating all model types generally achieved reasonable performance similar to the best performing model.

As evidenced in the above study summaries, there are a few consistent themes in the current state of the PA literature focused on machine learning techniques for classifying PA type. Most studies use the same statistical features, similar activities and activity groups, and classifier types for classifying PA. In these studies, generally high classification accuracy is achieved during a within-sample cross-validation averaging >85.0%. This is largely due to the more structured nature of most activity protocols and the use of within-sample cross-validation procedures to estimate performance and how the classifiers might generalize to new data. Evidence suggests that the type of within-sample cross-validation procedure used can impact the

estimated performance and in some circumstances can artificially inflate model performance by as much as 10.0%.¹¹⁶ With this in mind, it is necessary to consider all circumstances when evaluating classifier performance, including how the classifier is validated. The end goal with these classifiers is to apply them to new data collected in true free-living conditions. The next section examines the literature in youth and adults looking at out-of-sample and free-living validations of classifiers developed using structured laboratory or semi-structured simulated free-living data to provide an overview of how well classifiers perform in application.

5.2 Classifier Performance in Application and Out-of-Sample Validation Performance

As mentioned above, a critical component to the classifier development process is how well the classifier performs in application. This section outlines some key studies in the field of PA assessment in youth and adults using both counts and raw acceleration data. These studies use machine learning techniques to classify PA and the performance results reported are all out-of-sample validations or independent free-living evaluations of classifiers developed using structured or semi-structured activity data collected in laboratory conditions.

Gyllensten and Bonomi¹¹⁴, examined the reproducibility of the Bonomi¹¹³ laboratory-trained classifiers, a need they identified as crucial to improving the generalizability of models in application. The study was conducted in a sample of 20 healthy adult participants (10 males and 10 females, mean age 30 yrs). Additional supervised laboratory data were collected using the Tracmor (sampling frequency 20 Hz) to perform an out-of-sample validation of the original classifiers (support vector machine, neural network, decision tree, and majority voting). Unsupervised free-living data were also collected using the Intelligent Device for Energy Expenditure and Physical Activity (IDEEA; 32 Hz). The IDEEA is a monitoring system comprised of five accelerometers strategically placed at different attachment sites to capture

more total body movements. The free-living criterion labels were created using a combination of the IDEEA data and participant activity diaries. The out-of-sample validation performed using the supervised laboratory data resulted in average classification accuracies ranging from 92.9% to 95.9% across all four classifiers, very similar to the caliber of performance achieved for the within-sample validation using LOOCV. The free-living validation using the unsupervised data resulted in average classification accuracies ranging from 72.2% to 75.7% across the four classifiers. These results illustrate a couple of important points. One, when classifiers developed and trained using structured activity data are evaluated using out-of-sample data, similar performance can be achieved. However, this is largely dependent on how the new out-of-sample data were collected. If the activities and sample population were generally the same, performance may be maintained. Conversely, if the nature of the activities is different (structured versus unstructured), classifier performance may decrease considerably (19.0% decrease from out-of-sample structured to out-of-sample unstructured). This is contingent on the activities performed and the source of the criterion data labels for evaluation.

Freedson et al.⁸⁹ collected data on 277 apparently healthy adult participants (dataset 1). Each participant completed 11 activities from a list of 23 activities. Each activity was performed for 7 minutes each and broadly sub-divided into treadmill activities and sports/activities of daily living. The treadmill activities consisted of walking at three different speeds at 0% and 3% grade. The activities of daily living consisted of stair climbing, moving a box, and then a combination of two of the following 13 activities: 1) sweeping, 2) mopping, 3) gardening, 4) trimming hedges, 5) mowing, 6) raking, 7) dusting, 8) laundry, 9) vacuuming, 10) washing dishes, 11) painting, 12) tennis (1v1), and 13) basketball. Additional data were collected on 65 apparently healthy adult participants who each performed one of three activity routines, each

routine comprised of six activities, where each activity was performed for 10 minutes each (dataset 2). The activities included: laundry, ironing, light cleaning, aerobics, car driving, frisbee golf, trimming hedges, gardening, using a wheelbarrow, carrying a box while walking, tennis (1v1), and self-paced overground walking and running. For both datasets, participants wore an ActiGraph GT1M uniaxial accelerometer-based activity monitor. The following features were extracted from the second-by-second data: 10th, 25th, 50th, 75th, and 90th percentiles and lag-one autocorrelation. Participant data from dataset 1 ($n = 277$) were used to develop and train a neural network for classifying PA type. The neural network classifier was then validated using the data from dataset 2 ($n = 65$). The individual activities from dataset 1 were collapsed into three categories: household, locomotive, and sports categories. Activities from dataset 2 were collapsed into the same categories and sedentary activities were excluded from this analysis. Within-sample cross-validation using dataset 1 yielded an overall classification accuracy of 97.3%. Out-of-sample validation of the model developed on dataset 1 and applied to dataset 2 yielded an overall classification accuracy of 80.9%, a change of 16.4% from within-sample to out-of-sample.

Bastian et al.¹¹⁷ conducted a study in 59 apparently healthy adult participants. Data were collected using a MotionLog accelerometer-based activity monitor (sampling frequency 100 Hz) worn on the right hip. Accelerometer data were down sampled to 25 Hz and the following time and frequency domain features were extracted from the individual axes and vector magnitude: mean, total signal energy, and dominant frequency. Participants performed 23 structured activities under laboratory conditions ranging in intensity from sedentary to vigorous in bouts of 5 minutes. The 23 individual activities were collapsed into eight activity groups as follows: 1) lying [lying in different positions on back (lying still, moving limbs)], 2) slouching [slouching in

a chair], 3) sitting [sitting still, sitting typing, sitting throwing a ball, sitting drinking], 4) standing [standing still, standing drinking, standing throwing a ball], 5) small utilitarian movements [pacing, stalling, sweeping, cleaning house], 6) walking [treadmill walking at different speeds, stair climbing], 7) running [treadmill running], and 8) cycling [stationary]. A Bayesian network was developed and overall classification accuracy was calculated from a within-sample LOOCV procedure. Overall classification for the eight activity classes was 80.9% and ranged from 52.7% (sitting) to 95.4% (walking). A second set of 20 participants were recruited to perform simulated free-living activities outside of the lab. The same eight activity classes were used but more naturalistic activities were added to each group: 1) lying, 2) slouching, 3) sitting [sitting during transit], 4) standing [taking an elevator, standing during transit], 5) small utilitarian movements [shopping], 6) walking [overground walking at different speeds, uphill, downhill, level, stair climbing], 7) running [overground running], and 8) cycling [overground, flat, uphill, downhill]. When the original model was applied to the semi-free-living data, precision averaged 44.7% and ranged from 9.0% for the small utilitarian movement group to 84.7% for walking. The original classifier was redeveloped using all of the laboratory data and half of the semi-free-living data and tested on the remaining semi-free-living data. In this scenario, precision averaged 66.0% and ranged from 9.4% for the small utilitarian movements group to 90.9% for cycling. These results indicate that applying a model developed using strictly structured laboratory data does not generalize well to new data. However, this study does illustrate that this can be somewhat mitigated by using structured and semi-structured data during the development process.

Kerr et al.¹⁰⁹ collected data in three different samples of adult participants: 1) two female research assistants, 2) 40 cyclists (males and females), and 3) 36 overweight females. Activities were either prescribed ahead of time (research assistants) or offline coded by video observation

using video collected using the SenseCam, a wearable body-camera (cyclists and overweight females). All data were collected under true free-living conditions using a hip-worn ActiGraph GT3X+ (sampling frequency 30 Hz). The following 43 time and frequency domain features were extracted from the individual sensor signals or the vector magnitude in one-minute windows: mean, standard deviation, median, coefficient of variation, entropy, principle direction of motion, autoregression coefficients, fast Fourier transform coefficients, total signal power, dominant frequency, and angular features (pitch, roll, yaw). Random forest classifiers were developed to classify six activity categories: 1) cycling, 2) riding in a vehicle, 3) sitting, 4) standing, 5) standing and stepping, and 6) walking and running. The random forest classifier for each dataset was within-sample cross-validated using LOOCV. Subsequently, an out-of-sample validation was performed using data from the other two datasets. For the within-sample cross-validation, average classification ranged from 84.0% (cyclists) to 93.0% (research assistants). For the out-of-sample validations, classification accuracy ranged from 78.0% (overweight female model applied to the cyclist's data) to 90.0% (research assistant model applied to the cyclists data). This study shows that, in general, classifier performance decreases when applied to out-of-sample data, even if the original classifier was developed using unstructured free-living activity data as classification accuracy decreased on average 6.0% when a classifier was applied to different free-living data.

Mannini et al.¹⁰⁶ replicated previously outlined methods used in adults in a sample of youth participants. The novelty of the study was the mixed use of adult and youth data together to re-train the support vector machine classifier and perform a within-sample cross-validation. Additionally, this study included out-of-sample validations of the original adult support vector machine classifier using the youth data, and an out-of-sample of the youth support vector

machine on the original adult data. Twenty youth participants ages 11-15 years old participated in the study and again wore triaxial accelerometer-based device (Wocket - 90 Hz) attached to the dominant wrist and ankle. Each participant completed 33 supervised activity trials in a structured laboratory environment with each activity lasting three to five minutes in duration. The individual activities were grouped into four primary categories as follows to match the groupings used in the original adult study: 1) Ambulation [over-ground walking, three bouts of treadmill walking (2-3 MPH, 3-4 MPH, and 4-5 MPH)], 2) Cycling [over-ground cycling, stationary cycling (50 W, 70 RPM)], 3) Other [basketball, soccer, catch, cleaning/tidying up], 4) Sedentary [computer, reading/writing, sorting/organizing, lying, sitting, standing, video gaming (handheld and Nintendo Wii)]. The following time and frequency domain features were extracted from 12.8-s non-overlapping windows from the signal magnitude (i.e. square root of the sum of the square of the three individual accelerometer axes): mean, standard deviation, maximum, minimum, first and second dominant frequencies, power of the first and second dominant frequencies, power ratio (i.e. ratio of power at dominant frequency to total power), range, and fragmentation. The fragmentation features were a new addition from the original study in adults and a comparison of which feature set had greater predictive power was completed in several phases of the analyses.

For the primary purpose of the Mannini et al.¹⁰⁶ study, they tested a support vector machine, developed using adult data on new youth data, and the maximum overall classification accuracies achieved were 85.9% for the dominant wrist and 89.7% for the ankle model. For the wrist, performance was similar to that in adults but for the ankle, performance was about 5.0% less but still considered high (~90.0%). In a secondary aim, focused specifically on the features used and model optimization, maximum overall classification accuracies of 91.0% (wrist) and

92.4% (ankle) were achieved in youth and 87.0% (wrist) and 94.8% (ankle) in adults. The out-of-sample validation of the youth classifier using adult data and the adult classifier using youth data was a key component of the study. When training the classifier using the original adult data and testing the classifier using the youth data, classification accuracy ranged from 58.8% to 69.8% for the wrist and 79.3% to 84.7% for the ankle depending on which feature set was used. Similarly, when training the classifier using youth data and testing the classifier using adult data, classification accuracy ranged from 71.7% to 71.8% for the wrist and 84.9% to 87.4% for the ankle depending on which feature set was used. This indicates that for both youth and adult support vector machine classifiers, performance decreases when applied to out-of-sample data as compared to the within-sample cross-validation performance using LOOCV. Lastly, the youth and adult data were pooled into a single dataset and the classifier was trained and cross-validated using LOOCV. Classification accuracy using the combined dataset was 88.5% for the wrist and 91.6% for the ankle, indicating that generally high classification accuracy can be achieved when pooling activity data across multiple sample populations. This is important as classifiers are typically developed using a specific sample population and generally only considered valid for the sample population used in development. This may indicate that for classifier generalizability, using mixed datasets may be advantageous.

Montoye et al.⁹² collected data in two different samples of apparently healthy adult participants: 39 participants in dataset 1 (19 males, 20 females, mean age 22.1 yrs) and 24 participants in dataset 2 (12 males, 12 females, mean age 46.3 yrs). Both datasets were collected using GENEActiv triaxial accelerometer-based activity monitors (sampling frequency 20 Hz dataset 1 and 60 Hz dataset 2) worn on the left wrist. For dataset 1, participants performed 13 activities for between three and 10 minutes each over a 90-minute period. For dataset 2,

participants performed between 12 and 21 activities for two to 15 minutes each over an 80-minute period. Activities were performed in a random order at a self-selected pace to simulate free-living conditions. Individual activities were collapsed into the following three activity categories as follows: 1) sedentary [lying, reading, computer use, watching television, writing, playing cards], 2) household [standing, sweeping, laundry, dusting, making the bed, vacuuming], and 3) ambulatory/exercise [overground slow and brisk walking, overground jogging, stationary cycling, stair climbing, bodyweight squats, bicep curls, treadmill walking, treadmill jogging]. Six different feature sets were compiled from a combination of the following time and frequency domain features using the triaxial acceleration data in 30-s windows: mean, standard deviation, minimum, maximum, 10th, 25th, 50th, 75th, and 90th percentiles, principal frequency, peak amplitude, and integral and percentage of total integral. Random forest, neural networks, decision trees, and support vector machine classifiers were developed and evaluated using a within-sample cross-validation and an out-of-sample validation of each classifier for each dataset [i.e. 1). (train dataset 1 - test dataset 1), 2) (train dataset 2 - test dataset 2), 3) (train dataset 1 - test dataset 2), 4) (train dataset 2 - test dataset 1)]. In a within-sample LOOCV procedure, classification accuracy ranged from 87.8% to 92.8% for dataset 1 and 76.7% to 80.2% for dataset 2. For the out-of-sample validation where each classifier was tested using the other dataset, 75.5% to 77.3% for dataset 1 and 74.0% to 79.1% for dataset 2. This results in a decrease in classifier performance ranging from 2.1% to 14.4% on average depending on the features used and the classifier type.

Ahmadi et al.¹¹⁸ conducted an out-of-sample validation of the random forest and support vector machine classifiers Trost et al.¹⁰⁵ previously developed in pre-school-aged youth. Thirty-one children (22 males, 9 females, mean age 4 yrs) were directly observed for 20-minute periods

of free play while wearing an ActiGraph GT3X+ triaxial accelerometer-based activity monitor (sampling frequency 100 Hz) on the right hip and non-dominant wrist. Activities were grouped in five categories: sedentary, light activities and games, moderate-to-vigorous activities and games, walking, and running. The same features as the original Trost et al.¹⁰⁵ study were extracted in 15-s non-overlapping windows. Classification accuracies ranged from 66.4% to 69.4% for the hip classifiers and 59.1% to 59.3% for the wrist classifiers. In comparison to the original classifier performance estimated from a within-sample LOOCV procedure, the out-of-sample performance was between 11.0% to 15.0% and 19.0% to 21.0% lower for the hip and wrist classifiers, respectively.

As evidenced by the studies summarized in this section, it is clear that classifier performance decreases in application. This is, when a classifier is developed using structured or semi-structured data collected under laboratory conditions using only an accelerometer sensor at a single attachment site, classifier performance will deteriorate when the classifier is evaluated using out-of-sample or free-living data. The above study summaries indicate that this pattern exists in both youth and adult studies and for multiple classifier types (i.e. random forest, neural networks), for data collected at multiple different attachment sites, and for classifiers developed using counts and raw acceleration. Ultimately, the classifiers being developed are intended to be used in application with unlabelled data collected in free-living conditions to evaluate the types of activities individuals engage in. Thus, it is critical to have robust yet generalizable classifiers that maintain high performance when used in application. The results of the studies summarized in this section show a few primary issues with the current classifier development structure. There appears to be a definitive statistical ceiling in place for model generalizability in application. Using only accelerometer sensor data results in large decreases in classifier performance when

classifiers are used in application. The studies available in the literature generally use a small number of activities or activity groups that are categorized based on gross movements and posture or activity domain. A few studies develop classifiers using the individual activity classes, but this is not common practice. A few studies examine using mixed methods for collecting data for classifier development which has had mild success but still results in a decrease in classifier performance in application. Tong et al.¹¹⁹ recently raised the question of how well classifiers developed using data from a single inertial type should be expected to perform in application. The authors note that, by and large, since the Bao and Intille⁴⁹ study in 2004, the field of PA assessment has not made progress equivalent to what might be expected based on the technological advances and improvements in statistical modeling since then. The next section covers some possibilities for how this statistical ceiling could potentially be overcome. The focus in the following section is on using data from multiple sensors or multiple sensor types to improve the characterization of PA using wearable device and inertial sensor data.

Section 6: Alternative Approaches to Address Model Performance Issues

6.1 Multiple Sensor Locations of the Same Sensor Type

One potential avenue to improve PA characterization is to mount multiple sensors of the same type to various attachment sites on the body (e.g. accelerometers mounted on the hip, wrist, ankle) similarly to what Bao and Intille⁴⁹ proposed, instead of individual monitors independent of one another, these sensors will be synchronized together in a single datastream. In the field of PA assessment, this was first attempted using the Intelligent Device for Energy Expenditure and Activity (IDEEA). The IDEEA uses a system of accelerometers strategically placed at different locations on the body.¹²⁰⁻¹²² The motivation for using multiple attachment sites is to better capture total body and limb-specific movements in an effort to improve EE

estimates and better detect different activities. In theory, this should improve PA assessment compared to using single accelerometer sensor in a single wear location. Zhang et al.¹²² reported IDEEA accuracies for EE estimations of $98.9 \pm 6.0\%$ compared to portable indirect calorimetry during a laboratory-based activity protocol and $95.2 \pm 2.3\%$ compared to room calorimetry in an adult sample. Similarly, Whybrow et al.¹²¹ reported that the IDEEA significantly overestimated EE by $0.5 \text{ megajoules} \cdot \text{day}^{-1}$ ($119.5 \text{ kilocalories} \cdot \text{day}^{-1}$) compared to room calorimetry and $2.0 \text{ megajoules} \cdot \text{day}^{-1}$ ($478.0 \text{ kilocalories} \cdot \text{day}^{-1}$) compared to doubly labelled water (DLW). Jiang et al.¹²⁰ reported on the activity identification component of the IDEEA and determined that for posture-based detection, the IDEEA correctly identified 91.6% of all sitting, standing, lying, and reclining activities corresponding to 92.4% of all monitored time being assigned a correct activity label. The system was more sensitive and effective at detecting sitting and standing activities compared to reclining and lying. While the intuition behind the approach should result in an improvement in PA assessment, in practice, the practicality of the IDEEA system approach is not feasible for free-living use due to the increased user burden over a single sensor wear location.

6.2 Multiple Sensor Types in the Same Location

Another potential way to overcome the hypothesized statistical ceiling effect from using only accelerometer sensor data is to use data from additional sensor types, such as a gyroscope, a thermometer, or heart rate sensor, to complement the accelerometer data using a single sensor location. As technology has advanced, there are single unit devices containing multiple sensor types. It is intuitive that data from other sensor types, captured simultaneously, should complement the accelerometer data and additional information about the activity would be captured compared to using an accelerometer alone. One example of this type of device is the

Sensewear Armband (SWA), a device that houses multiple sensor types (bi-axial accelerometer sensor, heat flux sensor, skin thermometer, galvanic skin response sensor, and a near-body temperature sensor) in a single unit worn on the upper arm. Incorporating multiple sensor types in a single device avoids the increased user burden imposed by the IDEEA approach and allows for the collection of multiple input types simultaneously to allow for better estimations of EE to be made. Santos-Lozano et al.¹²³ showed that the SWA significantly overestimated EE compared to indirect calorimetry in all intensity categories [sedentary (rest), light ($3 \text{ km} \cdot \text{hr}^{-1}$), moderate ($5 \text{ km} \cdot \text{hr}^{-1}$), and vigorous ($7 \text{ km} \cdot \text{hr}^{-1}$)] except very vigorous ($9 \text{ km} \cdot \text{hr}^{-1}$) where the EE estimate was still higher than indirect calorimetry but non-significant.¹²³ Lopez et al.¹²⁴ conducted a validation of the SWA and a hip-worn ActiGraph GT3X+ using indirect calorimetry as a criterion measure of EE in healthy children, adolescents, and adults. Participants completed a 40-minute structured activity protocol. The mean absolute percent errors for the SWA EE estimates were 24%, 22%, and 20% for children, adolescents, and adults, respectively.¹²⁴ Comparatively, mean absolute percent errors for the GT3X+ using the Santos-Lozano algorithm¹²⁵ were 28%, 22%, and 28% for children, adolescents, and adults, respectively.¹²⁴ The SWA was primarily intended for research use but never widely adopted because it was never fully transparent if or how the additional sensor types in the device were used to estimate EE. It should be noted that the SWA is no longer commercially available for purchase.

6.3 The Case for the Gyroscope

The SWA used multiple sensor types but only contained a single inertial sensor, a biaxial accelerometer making it somewhat similar to single unit accelerometer-based PA monitors. In order to more completely capture movement, a second type of inertial sensor may be more useful than thermal sensors. One such approach would be to use an accelerometer and a gyroscope.

The rationale behind this is that accelerometers measure static and dynamic linear acceleration and would not adequately capture many activities that involve turning while gyroscopes measure angular velocity or rotation and should capture movements involving turning. Several studies have documented that turning increases EE based on the degree and frequency of a turn indicating more emphasis should be placed on capturing turning movements when assessing PA with wearable devices.¹²⁶⁻¹²⁸ There is additional evidence from other disciplines supporting the use of accelerometer and gyroscope data together to improve PA classification as opposed to using data from either single sensor individually. Below are summaries of the relevant literature from computer science and engineering that show to utility of using data from both sensors simultaneously.

Other disciplines, such as engineering and computer science, have used IMUs and smartphones containing both gyroscope and accelerometer sensors to capture PA behaviour but this has not been incorporated into specific PA study protocols aiming to capture human movement using body-worn devices as is the case in the field of PA assessment. While studies from other disciplines have helped inform some of the essential processes for human recognition using inertial sensor and wearable device data and helped establish benchmark datasets, they are generally limited in the scope of behaviours available for classification, the sample size is typically small, and the sample population is generally limited to adults.

Wu et al.¹²⁹ collected accelerometer and gyroscope data (30 Hz) using iPod Touch devices in a sample of 16 apparently healthy adult participants. Thirteen activities were performed by each participant including variations of sitting, standing, walking, jogging, and stair climbing in different contexts (i.e. over ground vs treadmill; indoor vs outdoor) but were collapsed into four primary categories: 1) sitting, 2) jogging, 3) walking, and 4) stair walking.

Time and frequency domain features were extracted in 2-s windows with a 50% overlap from both sensor signals including the mean and variance of the individual sensor axes, a vector magnitude of acceleration, and a Fast Fourier Transform coefficient magnitude of each individual accelerometer axis. The following classifier types were developed using the sensor data: decision tree, multilayer perceptron, naïve Bayes, logistic regression, and k-nearest neighbor. Additionally, two ensemble classifiers, bagged and boosted k-nearest neighbor, were included but no additional performance gains were achieved compared to using the standard k-nearest neighbor classifier. In general, the standard k-nearest neighbor classifier achieved the highest classification accuracy when using data from either the accelerometer or gyroscope independently. When data from the two sensors were combined, within-sample cross-validation (10-fold) classification accuracy improved by, on average, 6.5% and ranged from 3.1% to 13.4% based on the activity. Every activity class benefitted from including gyroscope data with the accelerometer data with the exception of sitting, because classification accuracy was already 100.0% using only accelerometer data.

Anjum et al.¹³⁰ developed a smartphone application to recognize activities using the smartphone sensor data. The dataset included the following seven activities: 1) walking, 2) running, 3-4) ascending and descending stairs, 5) car driving, 6) cycling, and 7) inactivity (unspecified what this activity class contained, believed to be sitting still). The authors developed a series of machine learning classifiers to classify the seven activity classes including naïve Bayes, decision tree, k-nearest neighbor, and support vector machine classifiers. The activity data were collected using a Samsung Galaxy Y smartphone and it was placed in the following positions for each activity: pants pocket, hand, purse, and shirt pocket. The authors computed time domain features for the accelerometer and gyroscope sensor data. However, after assessing

feature contribution, all gyroscope features were eliminated during the development process resulting in only accelerometer sensor features being used to develop the final models. The resulting classifiers all achieved true positive rates ranging from 73.8% to 95.2% where the decision tree was the best performing classifier and ultimately the model implemented in the Android smartphone application for further evaluation.

Shoaib et al.¹³¹ collected data on four healthy adult male participants using Samsung Galaxy S2 smartphones containing an accelerometer, a gyroscope, and a magnetometer (sampling frequency 50 Hz). Participants performed the following six activities: 1) walking, 2) running, 3) sitting, 4) standing, 5) walking upstairs, and 6) walking downstairs for between 3 and 5 minutes for each activity. Phones were positioned in the right pants pocket, on the waist affixed to a belt, the right arm, and the right wrist. The following time domain features were extracted from 2-s windows with a 50% overlap from the individual sensor axes and the vector magnitude of each sensor to derive an orientation independent value for each sensor since the devices were worn at different attachment sites. While the study included all three sensor types, the primary focus remained on the accelerometer, gyroscope, and a combined approach using the accelerometer and gyroscope data together. The magnetometer performed poorly when compared to the accelerometer and gyroscope individually and was therefore not pursued further in the study. The following machine learning classifiers were used to classify the sensor data: naïve Bayes, support vector machine, neural network, logistic regression, k-nearest neighbor, and decision trees. The results are all presented in figures and no raw values are reported in the text but the results. However, the figures indicate that a combined approach using accelerometer and gyroscope data together results in higher true positive rates by approximately 5.0-10.0% compared to using either single sensor alone. Additionally, of the two single sensor approaches,

the accelerometer only approach was better than the gyroscope only approach at all attachment sites with the exception of the pocket.

In a follow-up study, Shoaib et al.¹³² collected data on ten healthy adult male participants using Samsung Galaxy S2 smartphones containing an accelerometer, a gyroscope, and a magnetometer (sampling frequency 50 Hz). As Shoaib et al.¹³¹ determined previously, the contribution of the magnetometer was minimal and was therefore not a primary focus of the study analyses. Participants performed the following seven activities: 1) walking, 2) running, 3) standing, 4) jobbing, 5) cycling, 6-7) ascending and descending stairs for 3 to 4-minutes per activity. The Samsung smartphones were attached in five positions: 1-2) right and left front pants pockets, 3) on the waist affixed to a belt, 4) right upper arm, and 5) right wrist. The following time and frequency domain features were extracted from the individual axes and vector magnitude of each sensor in 2-s windows with a 50% overlap to create four feature sets: Feature set 1) mean and standard deviation, feature set 2) median, zero crossings, and root mean square, feature set 3) variance, zero crossings, and root mean square, and feature set 4) coefficients sum, and spectral energy. Nine different machine learning classifiers were developed including: Bayesian networks, naïve Bayes, support vector machine, logistic regression, k-nearest neighbor, two rule-based classifiers, a decision tree, and a random forest. The results of the study indicate that the classifier type, the features used, and the attachment site all play a role in which sensor data may be most relevant for activity classification. There were some cases where the gyroscope was best and others where the accelerometer was best based on the combination of circumstances listed above. For example, for the ascending and descending stairs activities, the gyroscope was better than the accelerometer by approximately 10.0% for the right pocket position using feature set 2. However, at the belt position, the accelerometer was approximately 12.0% better than the

gyroscope when using feature set 1. Additionally, the authors noted that when combining accelerometer and gyroscope data together, the average improvement in classification accuracy across attachment sites, feature sets, and classifier types was approximately 10.0% for descending stairs and 7.0% for ascending stairs.

Wang et al.¹³³ were among the first to compare PA classification performance using accelerometer and gyroscope sensor data individually and together. Triaxial accelerometer and gyroscope data from a Samsung Galaxy SII smartphone (50 Hz) were used to develop a naïve Bayes and k-nearest neighbor classifier. Features were extracted in 2.56-s windows with a 50.0% overlap resulting in 128 sensor readings in each window. There were 348 total accelerometer features extracted, 164 time domain and 184 frequency domain. For the gyroscope, there were 211 total features extracted, 106 time domain and 105 frequency domain. This result in 561 total features across both sensors. Four different feature selection methods were used for each classifier type: 1) principal component analysis, 2) wrapper, 3) fast correlation-based filter, and 4) a novel hybrid filter-wrapper selection method. The two classifiers were evaluated using classification accuracy, precision, and F1 score from a 5-fold within-sample cross-validation using accelerometer data only, gyroscope data only, and accelerometer and gyroscope data together to see which sensor was better for classifying PA across six basic activity classes: 1) walking, 2) sitting, 3) standing, 4) lying, 5) ascending stairs, and 6) descending stairs. The results of the study indicate that using accelerometer data alone is superior to using gyroscope data alone for PA classification by approximately 25.0% for classification accuracy, precision and F1. Combining the accelerometer and gyroscope data results in slight improvements (2.0-3.0%) in classification accuracy, precision, and F1 score compared to using only accelerometer data by itself and large improvements (~25.0%) compared to using gyroscope data by itself.

Jain and Kanhangad¹³⁴ explored the utility of time and frequency domain feature sets individually and combined to see which combination of features yields higher classification accuracy. Two previously outlined datasets were used for the analyses, both collected in adults, using smartphones.^{129,135} The results of the study show that combining time and frequency domain features results in improved classification accuracy by approximately 3.0 to 5.0% compared to using only time or frequency domain features by themselves. While not explicitly examined and quantified in the results, the authors also presented classification accuracies for time, frequency and time + frequency domain features using only accelerometer sensor data and then using accelerometer and gyroscope sensor data together. This comparison showed that for both time and frequency domain features individually and time and frequency domain features together, combining the accelerometer and gyroscope sensor data yielded higher classification accuracy by, on average, 5.0% to 10.0% compared to using only accelerometer data by itself.

Ferrari et al.¹³⁶ showed across three public-use benchmark datasets (UCI-HAR, MobiAct, and MotionSense) comprised of adult smartphone data that using accelerometer and gyroscope data together yields an approximately 10.0% increase in classification accuracy, on average, compared to using data from either single sensor by itself. The UCI HAR¹³⁷ dataset was obtained using Samsung Galaxy SII smartphones worn on the waist of 30 adult participants who each performed six activities: 1) walking, 2) ascending stairs, 3) descending stairs, 4) sitting, 5) standing, and 6) lying. The MobiAct^{138,139} dataset was obtained using Samsung Galaxy S3 smartphones worn in the front pocket of 66 adult participants who each performed 15 activities: 1) standing, 2) walking, 3) jogging, 4) jumping, 5-6) ascending and descending stairs , 7) standing, 8-9) sit-to-stand and stand-to-sit transitions, 10) lying, 11-12) getting in and out of a car, and 13-15) simulated falling. The MotionSense¹⁴⁰ dataset was obtained using iPhone 6

smartphones worn in the front pocket collected from 24 adults who each performed six activities: 1-2) ascending and descending stairs, 3) walking, 4) jogging, 5) sitting, and 6) standing. Sixty-three total statistical features were extracted from the triaxial accelerometer and gyroscope sensor signals and were categorized into four feature sets: 1) the raw sensor axes, 2) the magnitude of the raw sensor axes, 3) 21 features of the magnitude of the individual axes, and 4) 21 features of the individual raw axes. The authors evaluated a k-nearest neighbor and a support vector machine classifier for classifying the activity classes in each dataset. Classifier performance (classification accuracy) was compared to a novel deep learning classifier developed in the study called Residual Network (ResNet) for each dataset. Across the three datasets, the combined sensor approach was 2.2% better than the accelerometer only and 8.5% better than the gyroscope only approaches for the k-nearest neighbor and support vector machine classifiers. The ResNet performance was superior in every case except for the combined sensor approach for the MotionAct dataset where the k-nearest neighbor was approximately 2.0% better. Performance was high (>90.0%) for the ResNet classifier for each dataset whether using accelerometer only, gyroscope only, or a combination of the sensor data. The combined sensor approach was better than either single sensor approach for ResNet, but due to already high performance with the single sensor approaches, the difference was marginal at only 2.0% for both sensor types.

To date, IMUs have been used in a limited capacity for the assessment of PA outside of using smartphones, which has mostly been done in the computer science and engineering disciplines. Hibbing et al.¹⁴¹ used the ActiGraph GT9X, a triaxial accelerometer-based device, that includes an IMU containing a secondary triaxial accelerometer, a triaxial gyroscope, a magnetometer, and a thermometer to measure the temperature inside the sensor housing. Thirty

healthy adults completed 10 semi-structured activities for 7 minutes each.¹⁴¹ Data were collected using GT9X devices worn on the hip, left and right wrist, and left and right ankle.¹⁴¹ To determine the usefulness of the gyroscope and magnetometer data for estimating EE compared to and in addition to the accelerometer data, Hibbing et al.¹⁴¹ used a two-regression approach to developed models using 1) accelerometer data only, 2) accelerometer data + gyroscope data, and 3) accelerometer data + gyroscope data + magnetometer data. Regardless of the attachment site, the model including accelerometer and gyroscope data improved EE estimates by approximately 5% compared to using accelerometer data only.¹⁴¹ As an initial step in the two-regression development process, sedentary behaviour (SB) needs to be identified and ideally filtered out leaving only active activity data for the two regression equations to be applied to. A SB threshold was developed using 1) accelerometer data only and 2) gyroscope data only.¹⁴¹ The gyroscope only SB classifier perfectly-identified SB for all five attachment site models compared to an average accuracy of 91% for the five attachment site models using only accelerometer data.¹⁴¹ The discriminatory performance of the classifier developed using gyroscope sensor data indicated that gyroscope sensor data may be better used for activity classification rather than estimation of EE.

Marcotte et al.¹²⁸ showed that the gyroscope in the ActiGraph GT9X IMU can be useful for detecting the number and degree of turning motions, a common movement that occurs during many activities of daily living such as household chores as well as during intermittent sporting activities. These two studies (Hibbing et al.¹⁴¹ and Marcotte et al.¹²⁸) first introduced the GT9X gyroscope into the PA assessment literature and have demonstrated the utility of the sensor data for two important tasks, EE estimation and turning quantification and identification. However,

there are presently no studies using the GT9X gyroscope in youth, and no studies in any sample population using GT9X accelerometer and gyroscope data for PA classification.

Section 7: Summary

The general trend of these studies was that most classifiers were developed using hip data collected during structured laboratory and semi-structured simulated free-living activities. Additionally, most studies reported within-sample cross-validation results as the indicator of classifier performance which generally resulted in high accuracy (>85%). However, when applied to true free-living data in an out-of-sample independent validation, performance decreases for all classifier types by as much as 20.0%. This decrease in classifier performances indicates the presence of a statistical ceiling in application and an issue with classifier generalizability. Other disciplines have demonstrated that using data from both accelerometer and gyroscope sensors simultaneously yields improvements in PA classification when compared to using data from either sensor alone. However, this improvement in classification has yet to be examined when classifiers are used in application. Thus, there is a current gap in the literature pertaining to using accelerometer and gyroscope data together in youth, using the GT9X gyroscope for PA classification, and examining classifier performance using accelerometer and gyroscope data in out-of-sample scenarios in application.

References

1. Bulling A, Blanke U, Schiele B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*. 2014;46(3):1-33.
2. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep*. 1985;100(2):126-131.
3. Pettee GKK MJ, Woolsey AT. Framework for Physical Activity as a Complex and Multidimensional Behavior. *J Phys Act Health*. 2012;9:S11-S18.
4. Kokkinos P. Physical activity, health benefits, and mortality risk. *ISRN Cardiol*. 2012;2012:718789.
5. Reiner M, Niermann C, Jekauc D, Woll A. Long-term health benefits of physical activity--a systematic review of longitudinal studies. *BMC Public Health*. 2013;13:813.
6. U.S. Department of Health and Human Services. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. 2018.
7. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. *CMAJ*. 2006;174(6):801-809.
8. Gonzalez K, Fuentes J, Marquez JL. Physical Inactivity, Sedentary Behavior and Chronic Diseases. *Korean J Fam Med*. 2017;38(3):111-115.
9. Knight JA. Physical inactivity: associated diseases and disorders. *Ann Clin Lab Sci*. 2012;42(3):320-337.
10. Kokkinos P, Sheriff H, Kheirbek R. Physical inactivity and mortality risk. *Cardiol Res Pract*. 2011;924-945.
11. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scand J Med Sci Sports*. 2015;25 Suppl 3:1-72.

12. U.S. Department of Health and Human Services. *Physical Activity Guidelines for Americans 2nd Edition*. Washington, D.C. 2018.
13. LaMunion SR, Fitzhugh EC, Crouter SE. Challenges and Opportunities Related to the Objective Assessment of Physical Activity Within U.S. Health Surveys. *Ann Epidemiol*. 2020.
14. U.S. Department of Health and Human Services. About the National Health and Nutrition Examination Survey. 2018; https://www.cdc.gov/nchs/nhanes/about_nhanes.htm. Accessed August 15, 2018.
15. Butte NF, Ekelund U, Westerterp KR. Assessing physical activity using wearable monitors: measures of physical activity. *Med Sci Sports Exerc*. 2012;44(1 Suppl 1):S5-12.
16. Chen KY, Janz KF, Zhu W, Brychta RJ. Redefining the roles of sensors in objective physical activity monitoring. *Med Sci Sports Exerc*. 2012;44(1 Suppl 1):S13-23.
17. Trost SG, Zheng Y, Wong WK. Machine learning for activity recognition: hip versus wrist data. *Physiol Meas*. 2014;35(11):2183-2189.
18. Ruch N, Joss F, Jimmy G, Melzer K, Hanggi J, Mader U. Neural network versus activity-specific prediction equations for energy expenditure estimation in children. *J Appl Physiol*. 2013;115(9):1229-1236.
19. Incel OD. Analysis of movement, orientation and rotation-based sensing for phone placement recognition. *Sensors (Basel)*. 2015;15(10):25474-25506.
20. Kailas A. Basic human motion tracking using a pair of gyro+ accelerometer MEMS devices. Paper presented at: 2012 IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom) 2012.

21. Zeng H, Zhao Y. Sensing movement: Microsensors for body motion measurement. *Sensors (Basel)*. 2011;11(1):638-660.
22. Chen KY, Bassett DR. The Technology of Accelerometry-Based Activity Monitors: Current and Future. *Med Sci Sports Exerc*. 2005;37(Supplement):S490-S500.
23. Bassett DR, Troiano RP, McClain JJ, Wolff DL. Accelerometer-based physical activity: total volume per day and standardized measures. *Med Sci Sports Exerc*. 2015;47(4):833-838.
24. ActiGraph Corporation. Accelerometer Technologies, Specifications, and Limitations White Paper. 2013.
25. Aminian K, Najafi B, Büla C, Leyvraz P-F, Robert P. Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes. *J Biomech*. 2002;35(5):689-699.
26. Sabatini AM, Martelloni C, Scapellato S, Cavallo F. Assessment of walking features from foot inertial sensing. *IEEE T Bio-Med Eng*. 2005;52(3):486-494.
27. Tong K, Granat MH. A practical gait analysis system using gyroscopes. *Med Eng Phys*. 1999;21(2):87-94.
28. Tsuruoka Y, Ochi F, Tsuruoka M. Bio-feedback system analysis in walking using three gyro sensors. Paper presented at: 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028) 1999.
29. Bourke AK, Lyons GM. A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. *Med Eng Phys*. 2008;30(1):84-90.
30. Li Q, Stankovic JA, Hanson MA, Barth AT, Lach J, Zhou G. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. Paper presented at:

2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks 2009.

31. Rakhman AZ, Nugroho LE. Fall detection system using accelerometer and gyroscope based on smartphone. Paper presented at: The 1st International Conference on Information Technology, Computer, and Electrical Engineering 2014.
32. Alonge F, Cucco E, D'Ippolito F, Pulizzotto A. The use of accelerometers and gyroscopes to estimate hip and knee angles on gait analysis. *Sensors (Basel)*. 2014;14(5):8430-8446.
33. Marcotte RT, Bassett DR, Weinhandl JT, Crouter SE. Application of the ActiGraph GT9X IMU for the assessment of turning during walking and running. *Biomed Phys Eng Express*. 2018;4(6).
34. Wong WY, Wong MS, Lo KH. Clinical applications of sensors for human posture and movement analysis: a review. *Prosthet Orthot Int*. 2007;31(1):62-75.
35. Passaro VMN, Cuccovillo A, Vaiani L, Carlo M, Campanella CE. Gyroscope Technology and Applications: A Review in the Industrial Perspective. *Sensors (Basel)*. 2017;17(10).
36. Halliday D, Resnick R, Walker J. *Fundamentals of physics*. John Wiley & Sons; 2013.
37. Sagnac G. L'éther lumineux démontré par l'effet du vent relatif d'éther dans un interféromètre en rotation uniforme. *CR Acad Sci*. 1913;157:708-710.
38. Kok M, Hol JD, Schön TB. Using inertial sensors for position and orientation estimation. *Foundations and Trends in Signal Processing*: 2017;11(1-2):1-153.

39. Shannon CE. Communication in the Presence of Noise. *Proc Inst Radio Eng.* 1949;37(1):10-21.
40. Nyquist H. Certain Topics in Telegraph Transmission Theory. *J Am Inst Elect Eng.* 1928;47(2):617-644.
41. Banos O, Galvez JM, Damas M, Pomares H, Rojas I. Window size impact in human activity recognition. *Sensors (Basel).* 2014;14(4):6474-6499.
42. John D, Miller R, Kozey-Keadle S, Caldwell G, Freedson P. Biomechanical examination of the 'plateau phenomenon' in ActiGraph vertical activity counts. *Physiol Meas.* 2012;33(2):219-230.
43. Clark CC, Barnes CM, Stratton G, McNarry MA, Mackintosh KA, Summers HD. A Review of Emerging Analytical Techniques for Objective Physical Activity Measurement in Humans. *Sports Med.* 2017;47(3):439-447.
44. Clark CCT, Nobre GC, Fernandes JFT, et al. Physical activity characterization: does one site fit all? *Physiol Meas.* 2018;39(9).
45. Rowlands AV, Mirkes EM, Yates T, et al. Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent? *Med Sci Sports Exerc.* 2018;50(2):257-265.
46. Rowlands AV, Yates T, Davies M, Khunti K, Edwardson CL. Raw Accelerometer Data Analysis with GGIR R-package: Does Accelerometer Brand Matter? *Med Sci Sports Exerc.* 2016;48(10):1935-1941.
47. Figo D, Diniz PC, Ferreira DR, Cardoso JMP. Preprocessing techniques for context recognition from accelerometer data. *Pers Ubiquit Comput.* 2010;14(7):645-662.

48. Preece SJ, Goulermas JY, Kenney LP, Howard D. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE T Biomed Eng.* 2009;56(3):871-879.
49. Bao L, Intille SS. Activity recognition from user-annotated acceleration data. *IEEE Pervas Comput, Conference Proceedings.* 2004;3001:1-17.
50. Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors--a review of classification techniques. *Physiol Meas.* 2009;30(4):R1-33.
51. Liu S, Gao RX, John D, Staudenmayer JW, Freedson PS. Multisensor data fusion for physical activity assessment. *IEEE Trans Biomed Eng.* 2012;59(3):687-696.
52. Allahbakhshi H, Hinrichs T, Huang H, Weibel R. The Key Factors in Physical Activity Type Detection Using Real-Life Data: A Systematic Review. *Front Physiol.* 2019;10:75.
53. Gupta P, Dallas T. Feature selection and activity recognition system using a single triaxial accelerometer. *IEEE Trans Biomed Eng.* 2014;61(6):1780-1786.
54. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res.* 2003;3(Mar):1157-1182.
55. Kate RJ, Swartz AM, Welch WA, Strath SJ. Comparative evaluation of features and techniques for identifying activity type and estimating energy cost from accelerometer data. *Physiol Meas.* 2016;37(3):360-379.
56. Rosati S, Balestra G, Knaflitz M. Comparison of Different Sets of Features for Human Activity Recognition by Wearable Sensors. *Sensors (Basel).* 2018;18(12).
57. Munson MA, Caruana R. *On Feature Selection, Bias-Variance, and Bagging.* 2009; Berlin, Heidelberg.

58. Bishop CM. *Pattern Recognition and Machine Learning*. Springer (India) Private Limited; 2013.
59. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer New York; 2009.
60. Jain A.K, DRPW, Mao J. Statistical Pattern Recognition: A Review. *IEEE T Pattern Anal.* 2000;22(1).
61. Li F, Shirahama K, Nisar MA, Koping L, Grzegorzec M. Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors. *Sensors (Basel)*. 2018;18(2).
62. Liu S, Gao RX, Freedson PS. Computational methods for estimating energy expenditure in human physical activities. *Med Sci Sports Exerc.* 2012;44(11):2138-2146.
63. Banos O, Damas M, Pomares H, Prieto A, Rojas I. Daily living activity recognition based on statistical feature quality group selection. *Expert Sys Appl.* 2012;39(9):8013-8021.
64. Keogh E, Mueen A. Curse of Dimensionality. In: Sammut C, Webb GI, eds. *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US; 2017:314-315.
65. Bellman R, Corporation R, Collection KMR. *Dynamic Programming*. Princeton University Press; 1957.
66. Duda RO, Hart PE, Stork DG. *Pattern Classification (2nd Edition)*. Wiley-Interscience; 2000.
67. Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*. Taylor & Francis; 1984.

68. Twomey N, Diethe T, Fafoutis X, et al. A Comprehensive Study of Activity Recognition Using Accelerometers. *Informatics*. 2018;5(2).
69. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y. Physical Human Activity Recognition Using Wearable Sensors. *Sensors (Basel)*. 2015;15(12):31314-31338.
70. Saez Y, Baldominos A, Isasi P. A Comparison Study of Classifier Algorithms for Cross-Person Physical Activity Recognition. *Sensors (Basel)*. 2016;17(1).
71. Vrigkas M, Nikou C, Kakadiaris IA. A Review of Human Activity Recognition Methods. *Front Robot AI*. 2015;2.
72. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: with Applications in R*. New York: Springer; 2013.
73. Refaeilzadeh P, Tang L, Liu H. Cross-Validation. In: Liu L, ÖZsu MT, eds. *Encyclopedia of Database Systems*. Boston, MA: Springer US; 2009:532-538.
74. Larson SC. The shrinkage of the coefficient of multiple correlation. *J Educ Psychol*. 1931;22(1):45.
75. Rajnarayan D, Wolpert D. Bias-Variance Trade-offs: Novel Applications. In: Sammut C, Webb GI, eds. *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US; 2014:1-15.
76. Webb GI. Overfitting. In: Sammut C, Webb GI, eds. *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US; 2010:744-744.
77. Dietterich TG. Ensemble Methods in Machine Learning. Proceedings of the First International Workshop on Multiple Classifier Systems; 2000.

78. Singh A, Thakur N, Sharma A. A Review of Supervised Machine Learning Algorithms. International Conference on Computing for Sustainable Global Development (INDIACom); 2016.
79. Lewis DD. Naive (Bayes) at Forty: the independence assumption in information retrieval. Proceedings of the 10th European Conference on Machine Learning; 1998; Chemnitz, Germany.
80. Ghahramani Z. An introduction to hidden Markov models and Bayesian networks. *Hidden Markov models: applications in computer vision*: World Scientific; 2001:9-41.
81. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Wiley; 2013.
82. Zhang X. Regularization. In: Sammut C, Webb GI, eds. *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US; 2010:845-849.
83. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. *Med Sci Sports Exerc*. 2006;38(9):1626-1634.
84. De Vries SI, Engels M, Garre FG. Identification of children's activity type with accelerometer-based neural networks. *Med Sci Sports Exerc*. 2011;43(10):1994-1999.
85. De Vries SI, Garre FG, Engbers LH, Hildebrandt VH, Van Buuren S. Evaluation of neural networks to identify types of activity using accelerometers. *Med Sci Sports Exerc*. 2011;43(1):101-107.
86. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol*. 2009;107(4):1300-1307.

87. Trost SG, Wong WK, Pfeiffer KA, Zheng Y. Artificial neural networks to predict activity type and energy expenditure in youth. *Med Sci Sports Exerc.* 2012;44(9):1801-1809.
88. Mu Y, Lo HZ, Ding W, Amaral K, Crouter SE. Bipart: Learning Block Structure for Activity Detection. *IEEE Trans Knowl Data Eng.* 2014;26(10):2397-2409.
89. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *J Appl Physiol.* 2011;111(6):1804-1812.
90. Zhang S, Rowlands AV, Murray P, Hurst TL. Physical activity classification using the GENEa wrist-worn accelerometer. *Med Sci Sports Exerc.* 2012;44(4):742-748.
91. Chowdhury AK, Tjondronegoro D, Chandran V, Trost SG. Ensemble Methods for Classification of Physical Activities from Wrist Accelerometry. *Med Sci Sports Exerc.* 2017.
92. Montoye AHK, Westgate BS, Fonley MR, Pfeiffer KA. Cross-validation and out-of-sample testing of physical activity intensity predictions with a wrist-worn accelerometer. *J Appl Physiol.* 2018;124(5):1284-1293.
93. Montoye AH, Mudd LM, Biswas S, Pfeiffer KA. Energy Expenditure Prediction Using Raw Accelerometer Data in Simulated Free Living. *Med Sci Sports Exerc.* 2015;47(8):1735-1746.
94. Montoye AH, Begum M, Henning Z, Pfeiffer KA. Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data. *Physiol Meas.* 2017;38(2):343-357.

95. Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Evaluation of the activPAL accelerometer for physical activity and energy expenditure estimation in a semi-structured setting. *J Sci Med Sport*. 2017.
96. Montoye AH, Conger SA, Connolly CP, et al. Validation of accelerometer-based energy expenditure prediction models in structured and simulated free-living settings. *Meas Phys Educ Exerc Sci*. 2017;21(4):223-234.
97. Montoye AH, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Wrist-independent energy expenditure prediction models from raw accelerometer data. *Physiol Meas*. 2016;37(10):1770-1784.
98. Montoye AH, Dong B, Biswas S, Pfeiffer KA. Validation of a wireless accelerometer network for energy expenditure measurement. *J Sports Sci*. 2016;34(21):2130-2139.
99. Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Validation and Comparison of Accelerometers Worn on the Hip, Thigh, and Wrists for Measuring Physical Activity and Sedentary Behavior. *AIMS Public Health*. 2016;3(2):298-312.
100. Montoye AH, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Comparison of activity type classification accuracy from accelerometers worn on the hip, wrists, and thigh in young, apparently healthy adults. *Meas Phys Educ Exerc Sci*. 2016;20(3):173-183.
101. Strath SJ, Kate RJ, Keenan KG, Welch WA, Swartz AM. Ngram time series model to predict activity type and energy cost from wrist, hip and ankle accelerometers: implications of age. *Physiol Meas*. 2015;36(11):2335-2351.
102. Zhao W, Adolph AL, Puyau MR, Vohra FA, Butte NF, Zakeri IF. Support vector machines classifiers of physical activities in preschoolers. *Physiol Rep*. 2013;1(1):e00006.

103. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. *Med Sci Sports Exerc.* 2013;45(11):2193-2203.
104. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *J Appl Physiol.* 2015;119(4):396-403.
105. Trost SG, Cliff DP, Ahmadi MN, Tuc NV, Hagenbuchner M. Sensor-enabled Activity Class Recognition in Preschoolers: Hip versus Wrist Data. *Med Sci Sports Exerc.* 2018;50(3):634-641.
106. Mannini A, Rosenberger M, Haskell WL, Sabatini AM, Intille SS. Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle. *Med Sci Sports Exerc.* 2017;49(4):801-812.
107. Zhang S, Murray P, Zillmer R, Eston RG, Catt M, Rowlands AV. Activity classification using the GENE: optimum sampling frequency and number of axes. *Med Sci Sports Exerc.* 2012;44(11):2228-2234.
108. Bonomi AG, Goris AH, Yin B, Westerterp KR. Detection of type, duration, and intensity of physical activity using an accelerometer. *Med Sci Sports Exerc.* 2009;41(9):1770-1777.
109. Kerr J, Patterson RE, Ellis K, et al. Objective Assessment of Physical Activity: Classifiers for Public Health. *Med Sci Sports Exerc.* 2016;48(5):951-957.
110. Pavey TG, Gilson ND, Gomersall SR, Clark B, Trost SG. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *J Sci Med Sport.* 2017;20(1):75-80.

111. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas*. 2014;35(11):2191-2203.
112. Crouter SE, Clowers KG, Bassett DR, Jr. A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol*. 2006;100(4):1324-1331.
113. Bonomi AG. Physical activity recognition using a wearable accelerometer. *Sensing Emotions*: Springer; 2010:41-51.
114. Gyllenstein IC, Bonomi AG. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. *IEEE Trans Biomed Eng*. 2011;58(9):2656-2663.
115. Ruch N, Rumo M, Mader U. Recognition of activities in children by two uniaxial accelerometers in free-living conditions. *Eur J Appl Physiol*. 2011;111(8):1917-1927.
116. Dehghani A, Glatard T, Shihab E. Subject cross validation in human activity recognition. arXiv preprint arXiv:1904.02666 [cs.LG]. 2019.
117. Bastian T, Maire A, Dugas J, et al. Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough. *J Appl Physiol*. 2015;118(6):716-722.
118. Ahmadi MN, Brookes D, Chowdhury A, Pavey T, Trost SG. Free-living Evaluation of Laboratory-based Activity Classifiers in Preschoolers. *Med Sci Sports Exerc*. 2019.
119. Tong C, Tailor SA, Lane ND. Are Accelerometers for Activity Recognition a Dead-end? arXiv preprint arXiv:2001.08111 [cs.CV]. 2020.
120. Jiang Y, Larson JL. IDEEA activity monitor: validity of activity recognition for lying, reclining, sitting and standing. *Front Med*. 2013;7(1):126-131.

121. Whybrow S, Ritz P, Horgan GW, Stubbs RJ. An evaluation of the IDEEA activity monitor for estimating energy expenditure. *Br J Nutr.* 2013;109(1):173-183.
122. Zhang K, Pi-Sunyer FX, Boozer CN. Improving Energy Expenditure Estimation for Physical Activity. *Med Sci Sports Exerc.* 2004:883-889.
123. Santos-Lozano A, Hernandez-Vicente A, Perez-Isaac R, et al. Is the SenseWear Armband accurate enough to quantify and estimate energy expenditure in healthy adults? *Ann Transl Med.* 2017;5(5):97.
124. Lopez GA, Brond JC, Andersen LB, Dencker M, Arvidsson D. Validation of SenseWear Armband in children, adolescents, and adults. *Scand J Med Sci Sports.* 2017.
125. Santos-Lozano A, Santin-Medeiros F, Cardon G, et al. Actigraph GT3X: validation and determination of physical activity intensity cut points. *Int J Sports Med.* 2013;34(11):975-982.
126. Hatamoto Y, Yamada Y, Fujii T, Higaki Y, Kiyonaga A, Tanaka H. A novel method for calculating the energy cost of turning during running. *Open Access J Sports Med.* 2013;4:117-122.
127. Hatamoto Y, Yamada Y, Sagayama H, Higaki Y, Kiyonaga A, Tanaka H. The relationship between running velocity and the energy cost of turning during running. *PLoS ONE.* 2014;9(1):e81850.
128. Marcotte RT, Bassett DR, Weinhandl JT, Crouter SE. Application of the ActiGraph GT9X IMU for the assessment of turning during walking and running. *Biomed Phys Eng Express.* 2018;4(6).

129. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. *J Med Internet Res*. 2012;14(5):e130.
130. Anjum A, Ilyas MU. Activity recognition using smartphone sensors. Paper presented at: IEEE 10th Consumer Communications and Networking Conference (CCNC) 2013.
131. Shoaib M, Scholten H, Havinga PJ. Towards physical activity recognition using smartphone sensors. Paper presented at: IEEE 10th International Conference on Ubiquitous Intelligence and Computing and IEEE 10th International Conference on Autonomic and Trusted Computing 2013.
132. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ. Fusion of smartphone motion sensors for physical activity recognition. *Sensors (Basel)*. 2014;14(6):10146-10176.
133. Wang A, Chen G, Yang J, Zhao S, Chang C-Y. A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sens J*. 2016;16(11):4566-4578.
134. Jain A, Kanhangad V. Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sens J*. 2017;18(3):1169-1177.
135. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL. A public domain dataset for human activity recognition using smartphones. Paper presented at: 21st European Symposium on Artificial Neural Networks (ESANN) 2013.
136. Ferrari A, Micucci D, Mobilio M, Napoletano P. Human Activities Recognition Using Accelerometer and Gyroscope. Paper presented at: European Conference on Ambient Intelligence 2019.

137. Reyes-Ortiz JL, Anguita D, Ghio A, Parra X. Human activity recognition using smartphones data set. *UCI Machine Learning Repository; University of California, Irvine, School of Information and Computer Sciences: Irvine, CA, USA*. 2012.
138. Chatzaki C, Pediaditis M, Vavoulas G, Tsiknakis M. Human daily activity and fall recognition using a smartphone's acceleration sensor. Paper presented at: International Conference on Information and Communication Technologies for Ageing Well and e-Health 2016.
139. Vavoulas G, Chatzaki C, Malliotakis T, Pediaditis M, Tsiknakis M. The MobiAct Dataset: Recognition of Activities of Daily Living using Smartphones. Paper presented at: International Conference on Information and Communication Technologies for Ageing Well and e-Health 2016.
140. Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H. Mobile sensor data anonymization. Paper presented at: Proceedings of the International Conference on Internet of Things Design and Implementation 2019.
141. Hibbing PR LS, Kaplan AS, Crouter SE. Estimating Energy Expenditure with ActiGraph GT9X Inertial Measurement Unit. *Med Sci Sports Exerc*. 2018;50(5):1093-1102.

Chapter 3: Study 1 Methods

The data presented in this chapter are part of a larger study aiming to identify periods of transition between bouts of activity in structured laboratory and unstructured free-living activities. This chapter is focused on presenting the methods for collecting the structured laboratory activity data and includes only procedural information and how the dataset was assembled.

Section 1: Participants

Data presented in this study were collected as part of a larger study and the protocol and participants have been reported on previously^{1,2}. A sample of 100 youth participants between the ages of 6 and 18 years old were recruited from the Knoxville, TN area by word of mouth and flyers (*Appendix C*). Written informed consent (*Appendix D*) and a health history questionnaire (*Appendix E*) were obtained from a parent or legal guardian of each participant prior to enrollment. Each participant provided written informed assent (*Appendix F*) before study participation was initiated. The recruitment strategy aimed to achieve an approximately even distribution of biological sex across the age range in four target age groupings: 1) 6-9 years old, 2) 10-12 years old, 3) 13-15 years old, and 4) 16-18 years old. Participants were excluded if any contraindications to exercise were reported by the parent/legal guardian on the health history questionnaire including: any musculo-skeletal injuries within the past six months that would prevent safe participation in regular physical activity, any metabolic condition or prescription medication that may affect metabolism, and any cardio-respiratory conditions, including asthma, that may limit regular physical activity participation. All study methods and procedures were approved by The University of Tennessee Knoxville Institutional Review Board.

Section 2: Procedures

As part of the larger study protocol, participants were asked to complete 30 minutes of supine rest as a measure of resting metabolic rate (RMR) using a Cosmed K4b² portable indirect calorimeter (Cosmed s.r.l., Rome, Italy). The primary focus of the present study is on physical activity (PA) classification and does not include an energy expenditure (EE) component, however, the wearable device data from the 30 minutes of supine rest during RMR was included as “lying” activity data for the subsequent analyses.

All research visits were conducted at The University of Tennessee Knoxville Applied Physiology Laboratory in the Health, Physical Education, and Recreation building. Data was collected across two laboratory visits on two separate days. The first study visit included completion of the informed consent and assent process and participant screening before final enrollment. Once enrolled, participants had anthropometric measurements taken including: standing and sitting height (cm) using a wall-mounted or SECA 213 portable stadiometer and body mass (kg) and body fat percentage measured using a Tanita BC-418 bioelectrical impedance analyzer. Anthropometric measurements were taken with participants in light weight athletic clothing, with their shoes and socks removed.

To start the first study visit, the 30-min RMR assessment was made after completing the anthropometric measurements. Following the RMR assessment, participants completed a battery of eight different structured activities from a list of 16 total activities (*Appendix B – Table 1*). The remaining eight activities were performed during the second laboratory visit so all participants completing both research visits performed all 16 activities. Prior to the first visit, eight activities were pseudo-randomly selected by a member of the research team based on laboratory space availability. Participants were instructed to perform each of the eight selected

activities two non-consecutive times during the visit, once for approximately 60-90 seconds and once for 4-5 minutes. Participants selected the order in which they wanted to perform the activities and the bout length was randomized by a member of the research team. Minimal limitations were placed on the possible activity combinations but due to scheduling and space constraints, some activities were selected based on available space at the time or postponed until the second day of testing if the required space was not available.

All participants wore an ActiGraph GT9X (GT9X) activity monitor on the left and right wrists, and the right hip. The hip monitor was worn on an elastic belt over the right iliac crest along the anterior axillary line and secured using a manufacturer specified clip. The wrist-worn devices were worn in manufacturer specific watch bands that the devices clipped into. The watch bands were positioned proximal to the styloid processes with the devices oriented where the LCD display could be read naturally by the participant like a watch screen would be. Participants were directly observed throughout each research visit using the Noldus Pocket Observer application for Android tablets (Noldus International Technology, Wageningen, Netherlands). Direct observation was focused on coding activity behaviours, periods of transition, and posture. Participants were also video recorded during each study visit but had the option to opt-out if they were not comfortable with the video component of the observation. The video recordings served as a backup in case the live coded observation records needed to be reviewed and revised for any reasons including adding missing behaviours and adjusting timestamps if an error was made during the live coding process. During the live coded observations, if the observer noticed an error was made regarding an early or late timestamp or a mis-coded activity period, a comment was made to alert the research staff during post-processing to go back and review that time point in the video to make a correction as needed.

Section 3: Equipment

3.1 Wearable Device

ActiGraph GT9X: The GT9X is a small (3.5 x 3.5 x 1 cm), light weight (14 grams), water resistant (1 meter, 30 minutes) research-grade activity monitor with a programmable display window that can be worn on the hip, wrist, or ankle. The GT9X contains a primary triaxial accelerometer with a dynamic range of ± 8 gravitational units (g's) and can be initialized to collect data at sampling frequencies of 30 - 100 Hz in increments of 10 Hz. The GT9X also contains an inertial measurement unit (IMU) that samples at a default 100 Hz. The IMU houses a secondary triaxial accelerometer with a dynamic range of ± 16 g's, a triaxial gyroscope with a dynamic range of ± 2000 degrees per second, a magnetometer with a dynamic range of ± 4800 micro-Tesla, and thermometer that measures the temperature inside the device housing. For the present study, GT9X devices were initialized to collect data at 90 Hz (primary accelerometer) and with the full IMU enabled (100 Hz) and Idle Sleep Mode disabled.

3.2 Direct Observation

Noldus Observer XT: The Noldus Observer XT (version 12.5) is a computer software program designed for direct observation. All research visits were live coded in real-time using the Noldus Pocket Observer application (version 3.2) for Android tablets on a Samsung Galaxy Tab 4 tablet (Samsung, Seoul, South Korea). As a secondary record, research visits were also video recorded using a Canon Vixia HFR700 camcorder.

To use the Pocket Observer application, a coding scheme must first be created in the Noldus XT software program. A two-class focal sampling coding scheme was used to first code the activity behaviour being performed and then the posture of the participant during the activity

behaviour. A 19 item list of activity behaviours was created: the 16 activities listed in *Appendix B - Table 1* plus an activity label for the start time of the observation period and activity routine (“Start Time”), a label for transitions between activities (“Transition”), and an unknown activity label to use when a participant performed an unlisted activity such as taking a water or bathroom break (“Unknown Activity”). After an activity behaviour was coded, the second layer of the coding scheme included modifier labels for the posture of the participant during an activity including lying, sitting, standing, or stepping. This coding scheme allowed for non-overlapping activity labels to be assigned in real-time by selecting the appropriate activity label button and subsequently the corresponding posture in the Pocket Observer mobile application. This created a continuous record of activities with time stamps that indicated the duration of activities marked by the beginning (e.g. when an activity was marked) and end of an activity label (e.g. when a new activity or transition was marked).

Live-coded observations collected using the Pocket Observer application were imported to a Windows computer into the Noldus Observer XT software program. The criterion labelled direct observation data were aligned with the GT9X monitor data using the Windows system clock as the anchor point for alignment. A comment containing the Windows system clock time was inserted in the Pocket Observer observation notes at the start of the observation period. Additionally, a video of the Windows system clock used to initialize the GT9X devices was taken at the start and end of each data collection period.

Section 4: Dataset Assembly

Criterion-labelled activity data from the direct observation files were synchronized and merged with the second-by-second GT9X files so each second of monitor data had a corresponding behaviour label with the primary activity behaviour as well as posture. This was

done for each of the three attachment sites, hip, left wrist, and right wrist making three unique datasets. The full dataset for each attachment site was randomly split into training and testing datasets by participant ID to achieve similar distribution of demographic characteristics. There were 74 participants allocated to the training dataset and 25 participants held out as a testing dataset. The same participant IDs were used for each attachment site, so the training and testing datasets are comprised of the same participants for each attachment site dataset. For the purposes of the subsequent analyses, the original 16 activity labels were subdivided into two target classification outcomes (*Appendix B – Table 1*) 1) 14 individual activity classes and 2) 7 activity groups. For the purposes of the analyses, the “unknown activity”, “transition”, and “Start Time” data were cleaned from the dataset leaving only seconds with one of the 16 original activity behaviour labels. The total number of available data points for each attachment site, including the distribution across individual activity classes and activity groups is available in *Appendix B – Table 2*.

An overview of the analysis plan is shown in *Appendix A - Figure 1*. Information about further data processing and statistical analyses are available in Chapter 4.

References

1. LaMunion SR, Blythe AL, Hibbing PR, Kaplan AS, Clendenin BJ, Crouter SE. Use of Consumer Monitors for Estimating Energy Expenditure in Youth. *Appl Physiol Nutr Metab.* 2020;45(2):161-168.
2. Hibbing PR, Bassett DR, Coe DP, LaMunion SR, Crouter SE. Youth Metabolic Equivalents Differ Depending on Operational Definitions. *Med Sci Sports Exerc.* 2020. Online Ahead of Print.

Chapter 4: Development and Within-Sample Validation

Abstract

Purpose: The purpose of this study was to develop and cross-validate four machine learning classifiers for classifying two target outcomes: 1) fourteen individual activity classes and 2) six activity groups using three feature sets from ActiGraph GT9X (GT9X) 1) primary accelerometer sensor data only, 2) gyroscope sensor data only, and 3) accelerometer and gyroscope sensor data combined in youth. **Methods:** A sample of 100 youth aged 6-18 years old performed sixteen different semi-structured simulated free-living activities ranging from sedentary to vigorous intensity split across two laboratory visits on two separate days. Participants wore a GT9X on the right hip, left wrist, and right wrist and were directly observed for the duration of each laboratory visit to create timestamped criterion labels of the activities performed. GT9X primary accelerometer (90 Hz) and gyroscope (100 Hz) data were reduced to 1-s averages. Seventeen total features were calculated, 15 from the sensor data, and two biological variables [continuous numerical age (y) and leg length (cm)]. Decision trees, lasso logistic regression, random forest, and neural networks classifiers were developed for each feature set, attachment site, and target outcome. The sample was split into a training (n = 75) dataset for classifier development and a testing (n = 25) dataset for cross-validation. Overall model classification accuracy was used for model comparison. **Results:** For the individual activity class models, the combined sensor feature set (average classification accuracy: 61.5%; classification accuracy range across attachment sites: 49.2-67.5%) outperformed the accelerometer only feature set (53.9; 41.0-61.4%) by an average of 7.7% and the gyroscope only feature set (48.3%; 35.2-59.0%) by an average of 13.2%, across all three attachment sites. For the activity group classifiers, the combined sensor feature set (82.0%; 77.7-85.1%) outperformed the accelerometer only feature set (76.7%; 71.8-80.9%) by an average of 5.3% and the gyroscope only feature set (76.6%; 72.3-

80.0%) by an average of 5.4%, across all three attachment sites. **Conclusion:** The results of this study indicate that using the GT9X primary accelerometer and gyroscope data together improves PA classification for both individual activities and activity groups resulting in an average improvement in classification accuracy of 7.9%.

Section 1: Introduction

Objectively assessing physical activity (PA) and sedentary behaviour (SB) are a primary interest for researchers to better understand their relationship with health outcomes. One popular method to assess PA and SB is to use wearable devices, which can be used in both structured and free-living environments. To date, most wearable devices used to assess PA have been comprised of a single accelerometer sensor. The raw accelerometer data can be expressed as the raw sensor signal in g's or as proprietary activity counts. Historically, most methods have used a regression-based approach with count data primarily for the purpose of relating accelerometer data to energy expenditure (EE) or activity intensity. However, more recently, there has been increasing interest in working with raw accelerometer data as well as using the raw sensor data to classify PA type. Modeling raw accelerometer data can include threshold-based approaches but more frequently includes machine learning and pattern recognition approaches. Machine learning leverages the rich information in the raw sensor signal to better characterize PA behaviours. In general, these machine learning approaches are an improvement over threshold-based approaches for modeling accelerometer data but have not been as widely accepted and adopted by end-users because they are not as easy to implement or interpret in most cases.

To date, nearly all PA classification research has used only accelerometer sensor data and associated features as the primary source for classifier development. A review by Farrahi et al.¹ shows that within-sample cross-validation classification accuracy, on average, ranges from 85.0% to 90.0% for hip and wrist classifiers in adults and youth. At the group level, this level of accuracy is generally considered acceptable. However, it has recently been hypothesized that using data from only an accelerometer sensor may not be enough to adequately classify PA behaviours.² One potential avenue to improve the classification of PA is to leverage additional

data to make the classification. Previous research in other disciplines have explored using both accelerometer and gyroscope sensor data together to classify PA. There is evidence to suggest that in adults, using smartphone data, using accelerometer and gyroscope data together yields improved classification accuracy by approximately 10.0%, compared to using data from either sensor individually. Additionally, using smartphone data, it has previously been shown that the contribution of each individual sensor type can vary by the device attachment site, the type of activity being performed, and the type of classifier used to model the sensor data.³ This indicates that for some cases, one sensor type will be superior to the other sensor type.³ However, one caveat to this research is that previous efforts have been focused specifically in adult populations, typically have small sample sizes, and only evaluate a small number of activities, generally consisting of gross body movements and some postures (e.g. walking, stair climbing, sitting, standing).³⁻⁸ However, there is currently no support for these findings in youth.

In 2014, ActiGraph (ActiGraph Corporation, Pensacola, FL), an industry leading manufacturer of research-grade wearable devices, released the GT9X Link (GT9X) device.⁹ The GT9X utilizes the same primary triaxial accelerometer as its widely used predecessor (wGT3X-BT) but also includes an inertial measurement unit (IMU) housing a secondary triaxial accelerometer, a triaxial gyroscope, a triaxial magnetometer, and a temperature sensor to monitor the temperature of the IMU sensors. Hibbing et al. 2018¹⁰ first introduced the GT9X gyroscope into the literature by demonstrating its utility as a predictor of EE using a two-regression approach. Paired with the primary accelerometer data, the gyroscope sensor data improved EE estimates by ~5% compared to using the primary accelerometer alone. However, a larger takeaway from the study was that the gyroscope was a perfect classifier of SB for the hip, wrists, and ankles attachment sites (Area Under the Curve (AUC) = 100%), compared to the

accelerometer which yielded an AUC between 63.7% and 99.4%. The GT9X provides the perfect opportunity to collect both accelerometer and gyroscope data simultaneously, in a device familiar to the research community focused on PA assessment using wearable devices.

Thus, the purpose of this study is two-fold: 1) to determine if using only GT9X gyroscope sensor data improves classification of semi-structured youth PA compared to using only GT9X accelerometer sensor data, and 2) to determine if using a combined sensor approach with GT9X accelerometer and gyroscope sensor data improves classification of semi-structured youth PA compared to using data from either single sensor individually.

Section 2: Methods

2.1 Participants

Data presented in this study were collected as part of a larger study and the protocol and participants have been reported on previously^{11,12}. A sample of 100 youth participants between the ages of 6 and 18 years old were recruited from the Knoxville, TN area by word of mouth and flyers. Written informed consent and a health history questionnaire were obtained from a parent or legal guardian of each participant and written informed assent was provided by each participant prior to enrollment in the study. The recruitment strategy aimed to achieve an approximately even distribution of biological sex across the age range in four target age groupings: 1) 6-9 years old, 2) 10-12 years old, 3) 13-15 years old, and 4) 16-18 years old. Participants were excluded if any contraindications to exercise were reported by the parent/legal guardian on the health history questionnaire including: any musculo-skeletal injuries within the past six months that would prevent safe participation in PA, any metabolic condition or prescription medication that may affect metabolism, and any cardio-respiratory conditions,

including asthma, that may limit regular PA participation. All study methods and procedures were approved by The University of Tennessee Knoxville Institutional Review Board.

2.2 Procedures

As part of the larger study protocol, participants were asked to complete 30 minutes of supine rest as a measure of resting metabolic rate (RMR) using a Cosmed K4b² portable indirect calorimeter (Cosmed s.r.l., Rome, Italy). The primary focus of the present study is on physical activity (PA) classification and does not include an energy expenditure (EE) component, however, the wearable device data from the 30 minutes of supine rest during RMR was included as “lying” activity data for the subsequent analyses.

All research visits were conducted at The University of Tennessee Knoxville Applied Physiology Laboratory in the Health, Physical Education, and Recreation building. Data was collected across two laboratory visits on two separate days. The first study visit included completion of the informed consent and assent process and participant screening before final enrollment. Once enrolled, participants had anthropometric measurements taken including: standing and sitting height (cm) using a wall-mounted or SECA 213 portable stadiometer and body mass (kg) and body fat percentage measured using a Tanita BC-418 bioelectrical impedance analyzer. Anthropometric measurements were taken with participants in light weight athletic clothing, with their shoes and socks removed.

To start the first study visit, the 30-min RMR assessment was made after completing the anthropometric measurements. Following the RMR assessment, participants completed a battery of eight different structured activities from a list of 16 total activities (*Appendix B – Table 1*). The remaining eight activities were performed during the second laboratory visit so all participants completing both research visits performed all 16 activities. Prior to the first visit,

eight activities were pseudo-randomly selected by a member of the research team based on laboratory space availability. Participants were instructed to perform each of the eight selected activities two non-consecutive times during the visit, once for approximately 60-90 seconds and once for 4-5 minutes. Participants selected the order in which they wanted to perform the activities and the bout length was randomized by a member of the research team. Minimal limitations were placed on the possible activity combinations but due to scheduling and space constraints, some activities were selected based on available space at the time or postponed until the second day of testing if the required space was not available.

All participants wore an ActiGraph GT9X (GT9X) activity monitor on the left and right wrists, and the right hip. The hip monitor was worn on an elastic belt over the right iliac crest along the anterior axillary line and secured using a manufacturer specified clip. The wrist-worn devices were worn in manufacturer specific watch bands that the devices clipped into. The watch bands were positioned proximal to the styloid processes with the devices oriented where the LCD display could be read naturally by the participant like a watch screen would be. Participants were directly observed throughout each research visit using the Noldus Pocket Observer application for Android tablets (Noldus International Technology, Wageningen, Netherlands). Direct observation was focused on coding activity behaviours, periods of transition, and posture. Participants were also video recorded during each study visit but had the option to opt-out if they were not comfortable with the video component of the observation. The video recordings served as a backup in case the live coded observation records needed to be reviewed and revised for any reasons including adding missing behaviours and adjusting timestamps if an error was made during the live coding process. During the live coded observations, if the observer noticed an error was made regarding an early or late timestamp or a mis-coded activity period, a comment

was made to alert the research staff during post-processing to go back and review that time point in the video to make a correction as needed.

2.3 Data Acquisition

2.3.1 Wearable Device

ActiGraph GT9X: The GT9X is a small (3.5 x 3.5 x 1 cm), light weight (14 grams), water resistant (1 meter, 30 minutes) research-grade activity monitor with a programmable display window that can be worn on the hip, wrist, or ankle. The GT9X contains a primary triaxial accelerometer with a dynamic range of ± 8 gravitational units (g's) and can be initialized to collect data at sampling frequencies of 30 - 100 Hz in increments of 10 Hz. The GT9X also contains an inertial measurement unit (IMU) that samples at a default 100 Hz. The IMU houses a secondary triaxial accelerometer with a dynamic range of ± 16 g's, a triaxial gyroscope with a dynamic range of ± 2000 degrees per second, a magnetometer with a dynamic range of ± 4800 micro-Tesla, and a thermometer that measures the temperature inside the device housing. For the present study, GT9X devices were initialized to collect data at 90 Hz (primary accelerometer) and with the full IMU enabled (100 Hz) and Idle Sleep Mode disabled.

2.3.2 Direct Observation

Noldus Observer XT: The Noldus Observer XT (version 12.5) is a computer software program designed for direct observation. All research visits were live coded in real-time using the Noldus Pocket Observer application (version 3.2) for Android tablets on a Samsung Galaxy Tab 4 tablet (Samsung, Seoul, South Korea). As a secondary record, research visits were also video recorded using a Canon Vixia HFR700 camcorder.

To use the Pocket Observer application, a coding scheme must first be created in the Noldus XT software program. A two-class focal sampling coding scheme was used to first code the activity behaviour being performed and then the posture of the participant during the activity behaviour. A 19 item list of activity behaviours was created: the 16 activities listed in *Appendix B - Table 1* plus an activity label for the start time of the observation period and activity routine (“Start Time”), a label for transitions between activities (“Transition”), and an unknown activity label to use when a participant performed an unlisted activity such as taking a water or bathroom break (“Unknown Activity”). After an activity behaviour was coded, the second layer of the coding scheme included modifier labels for the posture of the participant during an activity including lying, sitting, standing, or stepping. This coding scheme allowed for non-overlapping activity labels to be assigned in real-time by selecting the appropriate activity label button and subsequently the corresponding posture in the Pocket Observer mobile application. This created a continuous record of activities with time stamps that indicated the duration of activities marked by the beginning (e.g. when an activity was marked) and end of an activity label (e.g. when a new activity or transition was marked).

Live-coded observations collected using the Pocket Observer application were imported to a Windows computer into the Noldus Observer XT software program. The criterion labelled direct observation data were aligned with the GT9X monitor data using the Windows system clock as the anchor point for alignment. A comment containing the Windows system clock time was inserted in the Pocket Observer observation notes at the start of the observation period. Additionally, a video of the Windows system clock used to initialize the GT9X devices was taken at the start and end of each data collection period.

2.4 Wearable Device Data

2.4.1 Data Pre-processing and Reduction

All data handling was conducted in R^{13} statistical software. An overview of the data analysis plan is available in *Appendix A – Figure 1*. As a first step in pre-processing, a vector magnitude was calculated from the raw GT9X sensor data collected at 90 Hz for the primary accelerometer ($VM = \sqrt{X_{acc}^2 + Y_{acc}^2 + Z_{acc}^2}$) and 100 Hz for the gyroscope ($GVM = \sqrt{X_{gyro}^2 + Y_{gyro}^2 + Z_{gyro}^2}$). Prior to calculation of GVM, the individual gyroscope axes were low-pass filtered at 35 Hz using a second order Butterworth filter. The individual sensor axes, VM, and GVM were then collapsed to 1-s averages.

2.4.2 Feature Extraction

A series of time and frequency domain features were engineered from the 1-s individual axes and VM and GVM sensor data. Once all features were engineered, the 1-s windows of sensor data were aggregated with labels from the direct observation files for each participant and each monitor attachment site (hip, left wrist, and right wrist) so each 1-s window has a label for individual activity, activity group, and posture.

Based on prior research¹⁴⁻¹⁶, the following list of 15 time domain features were extracted from the 1-s sensor data (9 accelerometer features, 6 gyroscope features). All features were calculated for the accelerometer and gyroscope sensor data unless noted otherwise:

1) *Individual X-, Y-, and Z- axis inclination angles (accelerometer only)*

$$a. \left(ex. \theta Y = \cos^{-1} \frac{Y}{\sqrt{Y^2 + X^2 + Z^2}} = \text{radians} \times \frac{180}{\pi} = \text{axis angle in degrees} \right).$$

Additionally, the following features were calculated from rolling 10-s windows:

- 2) *Minimum VM and GVM value*
- 3) *Maximum VM and GVM value*
- 4) *Mean VM and GVM value*
- 5) *Standard deviation of VM and GVM*
- 6) *Coefficient of variation (CV) of VM and GVM*
- 7) *Variance of VM and GVM*

While this list of features is not comprehensive, the proposed features have been used in multiple machine learning modeling studies using wearable sensor data.¹⁴⁻¹⁸ These feature sets should provide a starting point given the studies are not about novel modeling techniques but rather about adding new sensor data as a means of improving performance. Presently, there is a lack of information in the literature regarding gyroscope features. What little evidence exists in the literature has typically used the same time and frequency domain features that have been used for accelerometer data. Part of the rationale for this being that gyroscopes collect angular velocity data, which is a separate component of movement from linear acceleration captured by the accelerometer. While the data represent different components of movement and the sensor signals are in different units, the data should be able to be processed similarly. A comprehensive feature analysis may need to be explored in future studies.

2.4.3 Feature Sets

For each attachment site dataset (hip, left wrist, and right wrist), three feature sets were created as follows, comprised of the engineered features outlined above: 1) Accelerometer sensor features only ($n = 9$ features), 2) Gyroscope sensor features only ($n = 6$ features), and 3) Accelerometer + Gyroscope sensor features combined ($n = 15$ sensor features). In addition to the sensor features, the following biological variables were included: continuous numerical age (y)

and leg length (cm) making 17 total available features. Previous research in adults has shown that biological variables do not contribute much to the characterization of PA.^{19,20} Thus, the decision was made to exclude height, weight, biological sex, and body mass index (BMI) both as a categorical variable and as a continuous percentile as variables for this study. In the case of the present study in a youth sample spanning a broad age range, the authors elected to include leg length and age as continuous numerical features. This was done in hopes of contributing some information to the classification of locomotive and ambulatory activities as youth of different ages will move differently for walking, running, and stair climbing, much of which has to do with their gross motor coordination but also their leg length which contributes to their step frequency. With this in mind, these biological variables may additionally contribute to distinguishing other upright behaviours not listed here. It is intuitive that both age and leg length may contribute useful information for discriminating walking and running from one another since running for a six-year-old and walking for an eighteen-year-old may have similar sensor signal characteristics. With three feature sets for each attachment site (hip, left wrist, right wrist) and two target classification outcomes (individual activity classes and activity groups) for each feature set, the total number of feature sets is 18.

2.4.4 Model Development

For each attachment site dataset, a series of four model types were developed: 1) decision tree, 2) lasso logistic regression, 3) random forest, and 4) neural networks. These model types have been used previously for classification of PA type using only accelerometer data.^{14,16,21-23} All features were used for the development of each of the four model types.

The sample was randomly split by participant ID into partitions of 75 (training dataset) and 25 (testing dataset) participants. For each model type, the same 75/25 random split of

participants was used to make model results comparable to one another. The ‘*set.seed*’ argument was set to ‘307’ to maintain the same partitions of training data during k-fold cross-validation and to make results reproducible. All four model types were created using the ‘*caret*’²⁴ package in *R*. Model parameters were tuned using the ‘*traincontrol*’ and ‘*expand.grid*’ functions and models were trained using the ‘*train*’ function with the requisite model type used for the method argument.

i. **Decision Tree**

Decision tree models were created using the ‘*rpart*’ method in the ‘*caret*’ package. The complexity parameter (cp) was tuned and optimized. Twenty-five possible cp values between 10^{-10} and 10^{-1} were tested during resampling (10-fold cross validation with no repeats) and the ‘*best.cp*’ value was used to train the final model.

ii. **Logistic Regression**

L1 regularized lasso logistic regression models were created using the ‘*glmnet*’ method in the ‘*caret*’ package. Prior to training, the ‘*lambda*’ parameter was optimized using the ‘*cv.glmnet*’ function with 3-fold cross-validation to identify the ‘*lambda.min*’ and ‘*lambda.1se*’ values (i.e. the values resulting in the lowest error on the training dataset) for each feature set and target classification outcome. Once the ‘*lambda.1se*’ was obtained, the, ‘*alpha*’ parameter was set to 1 and a final ‘*glmnet*’ model was trained using 3-fold cross-validation with no repeats.

iii. **Random Forest**

Random forest models were created using the ‘*rf*’ method in the ‘*caret*’ package. The ‘*mtry*’ argument (i.e. the number of variables available for splitting at each node of a tree) was tuned to be the square root of the number of features used for each model. No restrictions or tuning was performed for the ‘*ntree*’ argument, this was objectively determined by the individual model. Three-fold cross-validation with no repeats was used to train a final model with the single tuned ‘*mtry*’ value.

iv. Neural Network

Neural network models were created using the ‘*nnet*’ method in the ‘*caret*’ package. The ‘*size*’ and ‘*decay*’ arguments were tuned by training 100 iterations using the entire training data set, no partitions, to determine the best combination of parameters to use. Five different values were tested for the ‘*size*’ (1, 5, 10, 15, and 20) and the ‘*decay*’ (0.0001, 0.001, 0.01, 0.1, 0.5) arguments. For the single sensor feature set models (i.e. accelerometer only and gyroscope only) the ‘*size*’ parameter was set to 20 and the ‘*decay*’ parameter was set to 0.1. For the combined sensor feature set models, the ‘*decay*’ parameter remained 0.1 but the ‘*size*’ parameter was changed to 15. Three-fold cross-validation with no repeats using the optimized ‘*size*’ and ‘*decay*’ parameters was used to train the model before fitting a model on the full training dataset. The ‘*maxit*’ value (i.e. the maximum number of iterations allowed to reach convergence at a local minima) was set to 5,000 possible iterations.

2.5 Statistical Analysis

Statistical analyses were performed in R^{13} statistical software. The outcomes of interest for each dataset and model type are classification of: 1) 14 individual activity classes, and 2) six

activity groups. For these analyses, data points with activity labels of “Start Time”, “Transition”, and “Unknown Activity” were removed from the dataset. Final model performance was assessed using a 25% holdout sample to determine generalizability. An overall classification accuracy measure (i.e. the number of correctly classified cases in all classes divided by the total number of cases) was reported for each model type as a general indicator of which feature set results in the highest classification accuracy for each attachment site and target classification outcome. The following summary statistics and individual class performance metrics will be reported for each model type, attachment site, and feature set. These metrics are defined using a standard 2x2 contingency table layout and terminology with the columns representing the criterion values and the rows representing the predicted values.²⁵⁻²⁷ Cell A is true positives (TP), cell B is false positives (FP), cell C is false negatives (FN), and cell D is true negatives (TN).²⁵⁻²⁷ For more than two classes, as is the case with the presented analyses, metric results are calculated using a “one versus all” approach where each factor level is compared to the other factor levels.²⁵⁻²⁷

$$1) \text{ sensitivity} - [sens; TP/(TP+FN)]$$

$$2) \text{ specificity} - [spec; TN/(FP+TN)]$$

$$3) \text{ prevalence} - [prev; (TP+FN)/(TP+FP+FN+TN)]$$

$$4) \text{ positive predictive value} - [PPV; (sens*prev)/((sens*prev) + (1-spec)*(1-prev))]$$

$$5) \text{ negative predictive value} - [NPV; spec*(1-prev)/(((1-sens)*prev)+((spec)*(1-prev)))]$$

$$6) \text{ precision} - [TP/(TP+FP)]$$

$$7) \text{ recall} - [TP/(TP+FN)]$$

$$8) \text{ F1 Score} - [((1+beta^2)*precision*recall)/((beta^2*precision)+recall) - (note: beta = 1)]$$

$$9) \text{ balanced accuracy} - [(sens + spec)/2]$$

Sections 3: Results

There were 100 youth participants enrolled in the study. A complete breakdown of how the data were handled following enrollment is shown in *Appendix A - Figure 2*. One participant withdrew prior to any activity data being collected resulting in the final analytical sample of 99 participants, 74 in the training dataset and 25 in the testing dataset. A breakdown of the number of data points in each activity class is available in *Appendix B – Table 2*. Participant demographics and physical characteristics for the training and testing datasets are available in *Appendix B - Table 3*.

Within-sample cross-validation results are split below based on attachment site (hip, left wrist, and right wrist), target outcome of interest (individual activity classes and activity groups) and model type (decision tree, lasso logistic regression, random forest, and neural network).

3.1 Hip Individual Activity Class Results

For the individual activity classes, *Appendix A - Figure 3A* shows the combined sensor feature set classifiers (range of classification accuracies across all four classifier types: 54.8-67.2%) achieved higher overall classification LOOCV accuracy than all classifiers using either individual sensor feature sets by an average of 5.3% (accelerometer only) and 17.3% (gyroscope only). Of the single sensor classifiers, using the accelerometer only feature set (52.2-61.3%) achieved consistently higher overall LOOCV classification accuracy than using the gyroscope only feature set (35.2-54.0%) for all four classifiers by an average of 12.0%.

Appendix B - Table 4 summarizes the percentage of cases correctly classified for each feature set and classifier type by individual activity class. In general, the combined sensor feature sets correctly classified the most cases with few exceptions where the accelerometer only or

gyroscope only feature sets were superior. Specific examples shown below by individual classifier type.

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 51.4% to 94.2% across all four classifier types with standard deviations ranging from 0.0001 to 0.0026. For all hip individual activity classifiers, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy standard deviations indicating that the training classification accuracies were statistically different.

3.1.1 Hip Individual Activity Class Decision Tree Results

For the decision tree classifiers, the combined sensor feature set was best for 9 of the 14 individual activity classes. Of the remaining five classes, the accelerometer only feature set was best for four activity classes (book reading, lying, reclining, and running), and the gyroscope only feature set was best for one activity class (catch) (*Appendix B - Table 4*). Of the activities the accelerometer only feature set was best for, three of the four were seated and lying activities (book reading, reclining, and lying). On average, across all 14 individual activities, the combined sensor decision tree correctly classified 4.2% more cases than the accelerometer only feature set and 14.9% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 10.7% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for computer by 39.6%, jumping jacks by 19.8%, lying by 30.3%, reclining by 27.6%, running by 16.7%, and stair walking by 22.5%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 22.0%. For all other activities, performance was generally similar. *Appendix B – Table 5* shows performance metrics for each feature set. For all eight additional performance

metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 9.5% higher for the combined sensor feature set (42.6%) compared to the accelerometer only (38.2%) and gyroscope only feature sets (28.0%).

3.1.2 Hip Individual Activity Class Lasso Results

For the lasso models, the combined sensor feature set was best for 11 of 14 activity classes, the accelerometer only feature set was best for two activities (lying and basketball), and the gyroscope only feature set was best for the remaining activity (soccer) (*Appendix B - Table 4*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 8.9% more cases than the accelerometer only feature set and 13.2% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 4.2% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for basketball by 10.1%, computer by 71.7%, jumping jacks by 21.3%, reclining by 18.2%, and stair walking by 13.1%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 50.7%, catch by 29.1%, and dusting by 11.6%. For all other activities, performance was generally similar. *Appendix B – Table 5* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 11.0% higher for the combined sensor feature set (48.3%) compared to the accelerometer only (38.1%) and gyroscope only (36.6%) feature sets.

3.1.3 Hip Individual Activity Class Random Forest Results

For the random forest models, the combined sensor feature set was best for all 14 activity classes (*Appendix B - Table 4*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 8.5% more cases than the accelerometer only feature set and 19.2% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 10.7% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for computer by 45.8%, jumping jacks by 29.7%, lying by 23.7%, reclining by 32.9%, and stair walking by 23.4%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 21.0%. For all other activities, performance was generally similar. *Appendix B – Table 5* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 13.8% higher for the combined sensor feature set (51.8%) compared to the accelerometer only (43.2%) and gyroscope only feature sets (32.7%).

3.1.4 Hip Individual Activity Class Neural Network Results

For the neural network models, the combined sensor feature set was best for 10 of 14 activity classes, the accelerometer only feature set was best for two activities (reclining and running), and the gyroscope only feature set was best for the remaining two activities (soccer and sweeping) (*Appendix B - Table 4*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 6.4% more cases than the accelerometer only feature set and 13.8% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 7.4% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for book reading by

16.2%, computer by 46.6%, jumping jacks by 18.6%, reclining by 46.4%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 23.3%, book reading by 10.4%, and sweeping by 13.1%. For all other activities, performance was generally similar.

Appendix B – Table 5 shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 11.0% higher for the combined sensor feature set (52.2%) compared to the accelerometer only (45.4%) and gyroscope only feature sets (37.2%).

3.2 Hip Activity Group Results

For the activity groups, *Appendix A - Figure 3B* shows the combined sensor feature set classifiers (77.7-84.2%) consistently achieved higher overall classification accuracy than all classifiers using either individual sensor feature sets by an average of 6.2% (accelerometer only) and 5.2% (gyroscope only). Of the single sensor classifiers, using the accelerometer only (71.8-79.2%) and gyroscope only (72.3-80.0%) individual sensor feature sets achieved similar overall classification and differed by an average of only 1.0% in favor of the gyroscope only feature set.

Appendix B - Table 6 summarizes the percentage of cases correctly classified for each feature set by individual activity class. In general, the combined sensor feature set correctly classified the most cases with few exceptions where the accelerometer only or gyroscope only feature sets were superior. Specific examples shown below by individual classifier type.

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 75.2% to 95.7% across all four classifier types with standard deviations ranging from 0.0002 to 0.0022. For all hip activity group classifiers, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy

standard deviations indicating that the training classification accuracies were statistically different.

3.2.1 Hip Activity Group Decision Tree Results

For the decision tree models, the combined sensor feature set was best for five of the six activity groups and the gyroscope only feature set was best for the remaining activity group (wheels) (*Appendix B - Table 6*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 9.9% more cases than the accelerometer only feature set and 7.5% more cases than the gyroscope only feature set. The gyroscope only feature set correctly classified on average 2.3% more cases than the accelerometer only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and leisure time activity by 18.6% and the gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 27.9%. For all other activity groups, performance was generally similar. *Appendix B – Table 7* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 8.7% higher for the combined sensor feature set (62.2%) compared to the accelerometer only (52.6%) and gyroscope only feature sets (54.4%).

3.2.2 Hip Activity Group LASSO Results

For the LASSO models, the combined sensor feature set was best for five of the six activity groups and the gyroscope only feature set was best for the remaining activity group (sports) (*Appendix B - Table 6*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 12.6% more cases than the accelerometer only feature set and

6.1% more cases than the gyroscope only feature set. The gyroscope only feature set correctly classified on average 6.4% more cases than the accelerometer only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and leisure time activity by 15.8% and the gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 34.9% and sports by 11.2%. For all other activity groups, performance was generally similar. *Appendix B – Table 7* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 11.3% higher for the combined sensor feature set (67.0%) compared to the accelerometer only (51.8%) and gyroscope only feature sets (59.6%).

3.2.3 Hip Activity Group Random Forest Results

For the random forest models, the combined sensor feature set was best for all six activity groups (*Appendix B - Table 6*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 11.8% more cases than the accelerometer only feature set and 12.1% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 0.4% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and leisure time activity by 26.4% and the gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 25.1%. For all other activity groups, performance was generally similar. *Appendix B – Table 7* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor

feature set outperformed the single sensor feature sets. F1 was on average 12.9% higher for the combined sensor feature set (72.0%) compared to the accelerometer only (59.5%) and gyroscope only feature sets (58.8%).

3.2.4 Hip Activity Group Neural Network Results

For the neural network models, the combined sensor feature set was best for all six activity groups (*Appendix B - Table 6*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 9.3% more cases than the accelerometer only feature set and 7.2% more cases than the gyroscope only feature set. The gyroscope only feature set correctly classified on average 2.1% more cases than the accelerometer only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and leisure time activity by 19.6% and the gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 23.4%. For all other activity groups, performance was generally similar. *Appendix B – Table 7* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 8.9% higher for the combined sensor feature set (73.3%) compared to the accelerometer only (63.3%) and gyroscope only feature sets (65.6%).

3.3 Left Wrist Individual Activity Class Results

For the individual activity classes, *Appendix A - Figure 3C* shows the combined sensor feature set classifiers (49.2-62.2%) achieved higher overall classification accuracy than all classifiers using either individual sensor feature sets by an average of 8.6% (accelerometer only) and 10.9% (gyroscope only). Of the single sensor classifiers, using the accelerometer only (41.0-

55.8%) and gyroscope only (39.8-53.7%) individual sensor feature sets achieved similar overall classification and differed by an average of only 2.3% in favor of the accelerometer only feature set.

Appendix B - Table 8 summarizes the percentage of cases correctly classified for each feature set and classifier type by individual activity class. In general, the combined sensor feature set correctly classified the most cases with few exceptions where the accelerometer only or gyroscope only feature sets were superior. Specific examples shown below by individual classifier type.

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 50.3% to 91.5% across all four classifier types with standard deviations ranging from 0.0002 to 0.0022. For all left wrist individual activity classifiers, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy standard deviations indicating that the training classification accuracies were statistically different.

3.3.1 Left Wrist Individual Activity Class Decision Tree Results

For the decision tree models, the combined sensor feature set was best for 12 of the 14 activity classes, the accelerometer only feature set was best for one activity (catch), and the gyroscope only feature set was best for one activity (lying) (*Appendix B - Table 8*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 7.3% more cases than the accelerometer only feature set and 15.3% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 8.0% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for running by 36.2%, catch by 10.4%, dusting by 15.2%, stair

walking by 15.2%, sweeping by 15.5%, and walking by 16.9%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 13.7% and lying by 17.6%. For all other activities, performance was generally similar. *Appendix B – Table 9* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 12.8% higher for the combined sensor feature set (47.6%) compared to the accelerometer only (42.2%) and gyroscope only feature sets (27.3%).

3.3.2 Left Wrist Individual Activity Class LASSO Results

For the LASSO models, the combined sensor feature set was best for 10 of 14 activity classes, the accelerometer only feature set was best for two activities (catch and walking), and the gyroscope only feature set was best for the remaining two activities (basketball and lying) (*Appendix B - Table 8*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 11.0% more cases than the accelerometer only feature set and 14.0% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 3.0% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for running by 48.2%, soccer by 15.0%, and walking by 11.7%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 34.6%. For all other activities, performance was generally similar. *Appendix B – Table 9* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 7.0% higher for the combined

sensor feature set (46.3%) compared to the accelerometer only (45.9%) and gyroscope only feature sets (32.6%).

3.3.3 Left Wrist Individual Activity Class Random Forest Results

For the random forest models, the combined sensor feature set was best for all 14 activity classes. (*Appendix B - Table 8*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 9.2% more cases than the accelerometer only feature set and 20.8% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 11.7% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for book reading by 18.3%, catch by 16.3%, computer by 19.1%, jumping jacks by 13.3%, running by 34.1%, soccer by 15.7%, stair walking by 20.3%, sweeping by 18.6%, and walking by 18.8%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 19.6% and lying by 10.6%. For all other activities, performance was generally similar. *Appendix B – Table 9* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 16.0% higher for the combined sensor feature set (56.0%) compared to the accelerometer only (48.8%) and gyroscope only feature sets (31.1%).

3.3.4 Left Wrist Individual Activity Class Neural Network Results

For the neural network models, the combined sensor feature set was best for 10 of 14 activity classes, the accelerometer only feature set was best for two activities (catch and walking), and the gyroscope only feature set was best for the remaining two activities (basketball and lying) (*Appendix B - Table 8*). On average, across all 14 individual activities, the combined

sensor feature set correctly classified 6.5% more cases than the accelerometer only feature set and 12.6% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 6.2% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for book reading by 21.1%, jumping jacks by 13.1%, running by 27.9%, soccer by 22.9%, and walking by 20.8%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 26.9% and lying by 10.1%. For all other activities, performance was generally similar. *Appendix B – Table 9* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 9.1% higher for the combined sensor feature set (50.7%) compared to the accelerometer only (45.3%) and gyroscope only feature sets (38.1%).

3.4 Left Wrist Activity Group Results

For the activity groups, *Appendix A - Figure 3D* shows the combined sensor feature set classifiers (79.0-84.0%) consistently achieved higher overall classification accuracy than all classifiers using either individual sensor feature sets by an average of 5.0% (accelerometer only) and 5.3% (gyroscope only). Of the single sensor classifiers, using the accelerometer only (73.9-78.6%) and gyroscope only (73.5-79.2%) individual sensor feature sets achieved similar overall classification and differed by an average of only 0.2% in favor of the accelerometer only feature set.

Appendix B - Table 10 summarizes the percentage of cases correctly classified for each feature set and classifier type by individual activity class. In general, the combined sensor feature set correctly classified the most cases with few exceptions where the accelerometer only or

gyroscope only feature sets were superior. Specific examples shown below by individual classifier type.

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 75.5% to 94.1% across all four classifier types with standard deviations ranging from 0.0002 to 0.0020. For all left wrist activity group classifiers, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy standard deviations indicating that the training classification accuracies were statistically different.

3.4.1 Left Wrist Activity Group Decision Tree Results

For the decision tree models, the combined sensor feature set was best for all six activity groups (*Appendix B - Table 10*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 10.1% more cases than the accelerometer only feature set and 10.7% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 0.6% more cases than the gyroscope only feature set. The gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 22.3%. For all other activity groups, performance was generally similar. *Appendix B – Table 11* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 10.4% higher for the combined sensor feature set (68.1%) compared to the accelerometer only (57.7%) and gyroscope only feature sets (57.7%).

3.4.2 Left Wrist Activity Group LASSO Results

For the LASSO models, the combined sensor feature set was best for five of the six activity groups and the gyroscope only feature set was best for the remaining activity group (sports) (*Appendix B - Table 10*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 7.4% more cases than the accelerometer only feature set and 6.7% more cases than the gyroscope only feature set. The gyroscope only feature set correctly classified on average 0.7% more cases than the accelerometer only feature set. For all activity groups, performance was generally similar between the accelerometer only and gyroscope only feature sets. *Appendix B – Table 11* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 8.7% higher for the combined sensor feature set (63.4%) compared to the accelerometer only (54.2%) and gyroscope only feature sets (55.2%).

3.4.3 Left Wrist Activity Group Random Forest Results

For the random forest models, the combined sensor feature set was best for all six activity groups (*Appendix B - Table 10*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 11.7% more cases than the accelerometer only feature set and 12.9% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 1.2% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and leisure time activity by 14.1% and the gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 31.7%. For all other activity groups, performance was generally similar. *Appendix B – Table 11* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and

negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 12.5% higher for the combined sensor feature set (74.7%) compared to the accelerometer only (62.6%) and gyroscope only feature sets (62.0%).

3.4.4 Left Wrist Activity Group Neural Network Results

For the neural network models, the combined sensor feature set was best for all six activity groups (*Appendix B - Table 10*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 11.5% more cases than the accelerometer only feature set and 8.5% more cases than the gyroscope only feature set. The gyroscope only feature set correctly classified on average 3.0% more cases than the accelerometer only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and leisure time activity by 12.9% and the gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 36.4%. For all other activity groups, performance was generally similar. *Appendix B – Table 11* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 9.8% higher for the combined sensor feature set (75.2%) compared to the accelerometer only (64.1%) and gyroscope only feature sets (66.6%).

3.5 Right Wrist Individual Activity Class Results

For the individual activity classes, *Appendix A - Figure 3E* shows the combined sensor feature set classifiers (57.3-67.5%) achieved higher overall classification accuracy than all classifiers using either individual sensor feature sets by an average of 9.0% (accelerometer only)

and 11.4% (gyroscope only). Of the single sensor classifiers, using the accelerometer only feature sets (47.6-61.4%) achieved higher overall classification compared to using the gyroscope only feature sets (45.8-59.0%) by an average of 2.5%.

Appendix B - Table 12 summarizes the percentage of cases correctly classified for each feature set and classifier type by individual activity class. In general, the combined sensor feature set correctly classified the most cases with few exceptions where the accelerometer only or gyroscope only feature sets were superior. Specific examples shown below by individual classifier type.

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 51.1% to 91.1% across all four classifier types with standard deviations ranging from 0.0001 to 0.0022. For all right wrist individual activity classifiers, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy standard deviations indicating that the training classification accuracies were statistically different.

3.5.1 Right Wrist Individual Activity Class Decision Tree Results

For the decision tree models, the combined sensor feature set was best for 13 of 14 activity classes and the gyroscope only feature set was best for the remaining activity class (lying) (*Appendix B - Table 12*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 9.2% more cases than the accelerometer only feature set and 15.6% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 6.4% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for book reading by 16.1%, computer by 11.2%, reclining by 12.5%, running by 27.2%, stair walking by 13.4%, and

walking by 22.3%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 28.7% and lying by 13.2%. For all other activities, performance was generally similar. *Appendix B – Table 13* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 15.2% higher for the combined sensor feature set (50.9%) compared to the accelerometer only (42.7%) and gyroscope only feature sets (28.8%).

3.5.2 Right Wrist Individual Activity Class LASSO Results

For the LASSO models, the combined sensor feature set was best for 11 of 14 activity classes, the accelerometer only feature set was best for two activities (catch and walking), and the gyroscope only feature set was best for the remaining activity (lying) (*Appendix B - Table 12*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 10.4% more cases than the accelerometer only feature set and 16.5% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 6.1% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for computer by 27.6%, running 44.8%, soccer by 11.5%, sweeping by 15.7%, and walking by 16.9%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 27.7%. *Appendix B – Table 13* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 15.3% higher for the combined sensor feature set (49.6%) compared to the accelerometer only (40.9%) and gyroscope only feature sets (27.5%).

3.5.3 Right Wrist Individual Activity Class Random Forest Results

For the random forest models, the combined sensor feature set was best for 13 of the 14 activity classes and the gyroscope only feature set was best for the one remaining activity (lying) (*Appendix B - Table 12*). On average, across all 14 individual activities, the combined sensor feature set correctly classified 9.9% more cases than the accelerometer only feature set and 20.2% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 10.3% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for book reading by 26.8%, catch by 13.9%, computer by 12.2%, dusting by 11.3%, jumping jacks by 12.8%, reclining by 11.3%, running by 27.3%, soccer by 14.3%, stair walking by 20.3%, sweeping by 12.8%, and walking by 22.9%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 16.8% and lying by 23.9%. For all other activities, performance was generally similar. *Appendix B – Table 13* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 17.2% higher for the combined sensor feature set (58.5%) compared to the accelerometer only (50.0%) and gyroscope only feature sets (32.5%).

3.5.4 Right Wrist Individual Activity Class Neural Network Results

For the neural network models, the combined sensor feature set was best for 11 of 14 activity classes, the accelerometer only feature set was best for two activities (catch and walking), and the gyroscope only feature set was best for the remaining activity (lying) (*Appendix B - Table 12*). On average, across all 14 individual activities, the combined sensor

feature set correctly classified 6.4% more cases than the accelerometer only feature set and 12.7% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 6.3% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for book reading by 28.7%, running by 15.9%, soccer by 18.6%, sweeping by 16.5%, and walking by 25.4%. The gyroscope only feature set was better than the accelerometer feature set for cycling by 25.5%. For all other activities, performance was generally similar. *Appendix B – Table 13* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 12.7% higher for the combined sensor feature set (54.0%) compared to the accelerometer only (47.8%) and gyroscope only feature sets (34.7%).

3.6 Right Wrist Activity Group Results

For the activity groups, *Appendix A - Figure 3F* shows the combined sensor feature set classifiers (80.2-85.1%) consistently achieved higher overall classification accuracy than all classifiers using either individual sensor feature sets by an average of 4.8% (accelerometer only) and 5.7% (gyroscope only). Of the single sensor classifiers, using the accelerometer only (75.0-80.9%) and gyroscope only (74.6-79.6%) individual sensor feature sets achieved similar overall classification and differed by an average of only 0.9% in favor of the accelerometer only feature set.

Appendix B - Table 14 summarizes the percentage of cases correctly classified for each feature set and classifier type by individual activity class. In general, the combined sensor feature set correctly classified the most cases with few exceptions where the accelerometer only or

gyroscope only feature sets were superior. Specific examples shown below by individual classifier type.

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 75.6% to 94.0% across all four classifier types with standard deviations ranging from 0.0002 to 0.0023. For all right wrist individual activity classifiers, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy standard deviations indicating that the training classification accuracies were statistically different.

3.6.1 Right Wrist Activity Group Decision Tree Results

For the decision tree models, the combined sensor feature set was best for five of the six activity groups and the accelerometer only feature set was best for the remaining activity group (sports) (*Appendix B - Table 14*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 9.5% more cases than the accelerometer only feature set and 9.6% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 0.1% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and leisure time activity by 10.7%. The gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 26.7%. For all other activity groups, performance was generally similar. *Appendix B – Table 15* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 9.8% higher for the combined sensor feature set (69.7%) compared to the accelerometer only (59.9%) and gyroscope only feature sets (60.0%).

3.6.2 Right Wrist Activity Group LASSO Results

For the LASSO models, the accelerometer only feature set was best for four of the six activity groups, the combined sensor feature set was best for one activity group (continuous walking and running), and the gyroscope only feature set was best for the remaining activity groups (seated and lying activity) (*Appendix B - Table 14*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 9.1% more cases than the accelerometer feature set and 7.2% more cases than the gyroscope only feature set. The gyroscope only feature set correctly classified on average 1.9% more cases than the accelerometer only feature set. The gyroscope only feature set was better than the accelerometer only feature set for household chores by 14.9%. For all other activity groups, performance was generally similar. *Appendix B – Table 15* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 10.0% higher for the combined sensor feature set (64.8%) compared to the accelerometer only (53.8%) and gyroscope only feature sets (55.9%).

3.6.3 Right Wrist Activity Group Random Forest Results

For the random forest models, the combined sensor feature set was best for all six activity groups (*Appendix B - Table 14*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 9.8% more cases than the accelerometer only feature set and 12.7% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 2.9% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and

leisure time activity by 16.7%. The gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 23.2%%. For all other activity groups, performance was generally similar. *Appendix B – Table 15* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 11.3% higher for the combined sensor feature set (76.1%) compared to the accelerometer only (66.4%) and gyroscope only feature sets (63.2%).

3.6.4 Right Wrist Activity Group Neural Network Results

For the neural network models, the combined sensor feature set was best for all six activity groups (*Appendix B - Table 14*). On average, across all 6 activity groups, the combined sensor feature set correctly classified 8.0% more cases than the accelerometer only feature set and 9.7% more cases than the gyroscope only feature set. The accelerometer only feature set correctly classified on average 1.1% more cases than the gyroscope only feature set. The accelerometer only feature set was better than the gyroscope only feature set for recreation and leisure time activity by 13.7%. The gyroscope only feature set was better than the accelerometer only feature set for wheeled activities by 20.7%. For all other activity groups, performance was generally similar. *Appendix B – Table 15* shows performance metrics for each feature set. For all eight additional performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor feature set outperformed the single sensor feature sets. F1 was on average 8.5% higher for the combined sensor feature set (76.3%) compared to the accelerometer only (68.6%) and gyroscope only feature sets (67.1%).

For additional detail about individual classifier confusion matrices and model performance summaries, please see *Appendix B – Tables 16-63* for hip, *Appendix B – Tables 64-111* for left wrist, and *Appendix B – Tables 112-159* for right wrist results.

Section 4: Discussion

4.1 Summary

The purpose of this study was two-fold: 1) to compare youth PA classification performance using classifiers trained with only accelerometer sensor data to classifiers trained with only gyroscope sensor data, and 2) to determine if using gyroscope sensor data in addition to accelerometer sensor data improved youth PA classification performance. The results of the study indicate that combining accelerometer and gyroscope data results in superior classification performance compared to using data from either single sensor individually for classifying individual activity type and activity group in youth. The combined sensor feature set was, on average, 10.4% better for classifying individual activities than the accelerometer only (difference in classification accuracy: 7.7%) and gyroscope only (13.2%) feature sets across all three attachment sites and all four classifier types. Similarly, for the activity group classifiers, the combined sensor feature set was on average 5.3% better than the accelerometer only (5.3%) and gyroscope only (5.4%) feature sets across all three attachment sites and all four classifier types. These findings indicate that for semi-structured free-living activities, using data from both the GT9X primary accelerometer and gyroscope will result in universally higher classification accuracy for both individual activity classes and activity groups regardless of attachment site when compared to using either single sensor individually. This study contributes to the limited sub-area of literature pertaining to youth PA assessment and is one of the first studies to

investigate using data from both accelerometer and gyroscope data from the GT9X for PA classification in both youth and adults.

4.2 Performance Overview

The findings in the present study are in line with previous studies in adults.³⁻⁸ Ferrari et al.⁴ showed across three benchmark datasets, using adult data collected using smartphones, that classification accuracy is, on average, 10% higher when machine learning and deep learning models are developed using both accelerometer and gyroscope sensor features compared to using either single sensor data individually. Wu et al.⁷ showed using iPod Touch data collected in adults, that machine learning classifiers developed using gyroscope and accelerometer data together results in a 3.1% to 13.4% improvement in classification accuracy over classifiers using only accelerometer data. The authors note that the magnitude of the difference in performance is dependent on the activity being classified. Shoaib et al.⁸ also showed that using both accelerometer and gyroscope data together results in higher true positive rates for a variety of machine learning classifiers compared to using either data from single sensor alone. This establishes a pattern that using data from multiple inertial sensors at the same attachment site generally results in improved characterization of PA across a variety of activities in both youth and adults. This is the first study, to the knowledge of the authors, exploring using accelerometer and gyroscope data together in youth. Additionally, of the existing studies in adults, many have small sample sizes ($N < 20$) and include a limited number of activities ($N < 6$) that most often include locomotion and gross body movements only.

Shoaib et al.³ showed that the relative improvement in classification using data from both accelerometer and gyroscope sensors together is largely dependent on the activity and in some cases the difference is negligible from the individual sensor performance. Shoaib et al.³ noted

that when performance is high for a single sensor classifier, it becomes increasingly difficult to further improve the classifier performance by adding additional data from a complementary sensor. The results of the present study demonstrate this effect for the seated and lying activity group. Both of the individual accelerometer only and gyroscope only feature sets are excellent for the classification of this activity group ($>92.0\%$ classification accuracy) for all classifier types across all three attachment sites. The combined sensor feature set does outperform the individual sensor feature sets, but the improvement is marginal due to already high classification accuracy with each individual sensor feature set. This story is slightly different when the individual activities contained in the seated and lying activity group are examined independently. For example, the percentage of correctly classified cases of book reading across all four classifiers types ranged (min to max) from 4.0% to 44.0% for the accelerometer only feature set, 0.0% to 22.4% for the gyroscope only feature set and 8.1% to 53.8% for the combined sensor feature set across all attachment sites. For computer work, the percentage of correctly classified cases across all four classifier types ranged from 26.6% to 88.8% for the accelerometer only feature set, 17.1% to 72.4% for the gyroscope only feature set, and 36.6% to 92.4% for the combined sensor feature set across all attachment sites. For lying, the percentage of correctly classified cases across all four classifiers types ranged from 43.6% to 98.8% for the accelerometer only feature set, 52.4% to 94.9% for the gyroscope only attachment site, and 53.1% to 98.6% for the combined sensor feature set across all attachment sites. For reclining, the percentage of correctly classified cases across all four classifiers types ranged from 0.0% to 47.6% for the accelerometer only feature set, 0.0% to 13.6% for the gyroscope only attachment site, and 0.5% to 41.4% for the combined sensor feature set across all attachment sites.

Conversely, when neither single sensor classifier performs particularly well but one is better than the other, the resulting combined sensor classifier performance is improved. In the present study for example, the accelerometer only feature set classified 36.4% of jumping jacks cases correctly on average at the hip compared to 16.6% for the gyroscope only feature set. However, the combined sensor feature set correctly classified 48.4% of cases on average, an improvement over both single sensor models demonstrating that for some classes, the two single sensor signals contain complementary information. This is true for cases where performance is similar between the two single sensor classifiers as well. In the present study for example, the accelerometer only and gyroscope only feature sets classified 48.5% and 45.6% of walking cases correctly on average at the hip while the combined sensor feature set classified 61.9% of walking cases correctly on average. In both examples, the overall combined sensor performance is generally low. However, in the context of the purposes of the present study, the results indicate that using both accelerometer and gyroscope sensor data together generally results in performance improvements across a variety of activities and for both hip and wrist attachment sites.

4.2.1 Differences between Feature Sets for Individual Activity Classes

The results of this study show that for the single sensor models, the accelerometer only classifiers are better than the gyroscope only classifiers for some activities and vice versa. For the individual activity classifiers, the gyroscope only feature set was universally better for the classification of stationary cycling where on average, 25.7% more cycling data points were correctly classified than the accelerometer only feature set across all three attachment sites. Conversely, the accelerometer only feature set was universally better for the classification of stair walking, where on average, 12.8% more stair walking data points were correctly classified

than the gyroscope only feature set across all three attachment sites. Additionally, the accelerometer only classifiers were notably better than the gyroscope only classifiers for book reading (average difference in percentage of correctly classified data points: 12.9%), computer (24.1%), jumping jacks (12.2%), reclining (13.8%), running (24.2%), and walking (15.0%) across all attachment sites.

4.2.2 Effect of GT9X Attachment Site on Individual Sensor Contribution to Classification

It has been shown previously using data from smartphones collected in adults that the contribution of each sensor type varies by attachment site.³ This section highlights a few notable instances in the present study further documenting the role attachment site plays on the individual sensor contribution to classification performance. The 24.1% difference in correctly classified computer work cases between the accelerometer only and gyroscope only classifiers is largely driven by the hip where the accelerometer only model correctly classified 50.9% more computer data points than the gyroscope only model. Similarly, the 12.2% difference in correctly classified cases of jumping jacks is also driven by the large difference between the accelerometer only and gyroscope only hip classifiers which were 22.4% different in the number of correctly classified cases in favor of the accelerometer only classifier. Conversely, the large differences in the number of correctly classified walking (15.0%) and running (24.2%) data points between the accelerometer only and gyroscope only classifiers is driven by the wrists. For walking, the left and right wrist accelerometer only models classified 17.1% and 21.9% more walking data points correctly than the gyroscope only models, respectively. Similarly, for running, the left and right wrist accelerometer only models classified 36.6% and 28.8% more running data points correctly than the gyroscope only models. The differences between the accelerometer only and gyroscope only hip classifiers was only 6.2% for walking and 7.3% for running. For the remaining activity

classes, the trends in favor of one sensor feature set over another were minor ($<5.0\%$) and generally trended in the same direction with the exception of lying. For the hip classifiers, the accelerometer only feature set was better than the gyroscope only feature set but for the wrists the pattern was different where the gyroscope only feature set was better than the accelerometer only feature set indicating the performance of a classifier using one data from one sensor versus the other is specific to the activity being performed and the attachment site confirming the findings of Shoaib and colleagues.³

4.3 Comparison to Previous Research

For the activity groups, the classification accuracies for the accelerometer only feature set (average: 76.7%; range: 71.8-80.9%) were fairly comparable to the lower end of what has been shown in other previous studies using only accelerometer sensor data.^{14,22,28,29} For example, Kerr et al.²⁸ achieved average classification accuracies between 84.0% and 93.0% across five or six activity classes in different three adult populations. Ellis et al.²² achieved balanced accuracies of 88.1% and 83.6% for hip and wrist classifiers across four activity classes in middle-aged adult women. Trost et al.²⁹ achieved accuracies between 80.0 – 81.0% for the hip, 78.0% - 80.0% for the wrist, and 82.0% - 85.0% for a combined hip and wrist model across five activity classes in preschoolers. Chowdhury et al.¹⁴ developed a series of machine learning models using wrist data and achieved F1 scores between 72.0% and 86.0% across eight activity classes in adults, 65.0% and 80.0% across five activity classes in adults, and 76.0% and 86.0% across seven activity classes in youth. The classifiers developed in the present study using the gyroscope only feature set (average classification accuracy 76.6%; range: 72.3-80.0%) can also be considered comparable to the results of previous studies using only accelerometer sensor data. Additionally, classifiers developed in the present study using the combined sensor feature (82.0%; 77.7-

85.1%) are quite comparable to the results of previous studies using only accelerometer sensor data.

One reason for the possible lower performance of the accelerometer only feature set and improved but similar performance of the combined sensor feature set is likely attributable to the activities used for development in the present study. While true the number of activity groups used in the present study is similar to the number of groups used in previous studies, the present study had more variety in the individual activities contained in the collapsed activity groups which is likely why performance is slightly lower but generally still comparable. For example, most other studies have used 3-6 activity groups. The majority of the activities included in those groups include gross body movements (e.g. locomotion, stair walking) and different postures (e.g. sitting, standing, and lying) and less activities of daily living. Some studies, such as those by Mannini et al.^{30,31} have included more specific individual activities as was done in the present study, but the activity groups these activities were collapsed into were broader than the groups used in the present study (ambulation, cycling, other, and sedentary). Additionally, most previous studies have adhered to using a protocol where each activity lasts for a given time period and frequently starts and stops on the minute making for clean activity start and end points. In the present study, participants were coached to start an activity when they were ready and when they could begin to transition between activities and were provided a window in which to do so. Collectively, the more natural end points for the activities included in the present study and the more specific nature of the activity groups evaluated, contribute to the slightly lower overall performance when compared to previous studies.

4.4 Left and Right Wrist Justification

Presently, there is interest in using the wrist as attachment site as both the National Health and Nutrition Examination Survey (NHANES) and U.K. Biobank, two national population surveillance studies, have chosen to have participants wear devices on the wrist over the hip. The NHANES³² procedures have participants wear the device on the non-dominant wrist while the U.K. Biobank³³ procedures have participants wear the device on the dominant wrist. While the wrist has been labelled a preferential attachment site for some study protocols aiming to increase wear compliance and monitor 24-hr behaviours, there is no consensus in the field about which placement site is best and what the implications are for choosing dominant versus non-dominant on individual estimates of PA. The present study chose to create classifiers using the left and right wrist terminology, without regard to hand dominance as there were a low number of left-hand dominant participants enrolled in the study. The prevalence of left-hand dominant individuals in the dataset was 8% which is consistent with the prevalence in the general population.³⁴

Currently, the literature favors wearing the device on the non-dominant wrist so as to avoid potential artefact and noise resulting from unilateral dominant limb movements, especially in youth.³⁵⁻³⁷ Conversely, if placed on the dominant wrist, more discrete movements may be captured in free-living and potentially more discriminative information would be available for those unilateral tasks. Some evidence exists to indicate that the impact of hand dominance on device placement has minimal effect across the entire day where a variety of activity behaviours take place.³⁸⁻⁴¹ Some differences may exist between wrists for specific tasks, but the relative contribution of those tasks to overall PA behaviour across the day is presently unknown. Montoye et al.⁴² created wrist-agnostic machine learning classifiers to estimate EE and trained models using data from one wrist and validated the models using data from the other wrist. In

doing so, models developed using VM features achieved similar root mean squared errors during cross-validation and out-of-sample validation, further indicating wrist placement may not be as critical of an issue as it was once considered to be.⁴²

4.5 Model Recommendation

If one of the models developed in this study were to be recommended for future use or revision, it would be the random forest models based on overall performance being among the best for all feature sets coupled with manageable computational cost and efficiency. Computational cost and efficiency were not empirically measured in this study, however, anecdotally, the computation time was the most competitive outside of the decision trees which ran the quickest. While the approach is intuitive, basic decision tree models are inherently weaker classifiers compared to the other classifiers used in this study. This is evidenced by the generally lower performance for the decision tree models across the board justifying not continuing to use that particular classifier moving forward for these data. The lasso models performed comparably to the random forest models in some instances and may be useful in future investigations for determining feature relevance. The computational cost was somewhat high in comparison to the random forest models but not unreasonable. It is possible that the lasso models could be used as a pre-classification step to determine feature contribution for a particular outcome before modeling with a different more advanced classifier. The neural network models also performed comparably to the random forest models, but the increased computational cost, the non-transparent black box classification approach, and the lack of additional performance gains makes them less likely to be recommended over the random forest models. Additionally, one consideration for future research would be to use a meta-ensemble approach where classifications from several classifier types are pooled using a weighting or

voting procedure to make the final classification. Custom ensemble approaches have been shown to achieve higher performance compared to individual ensemble approaches.^{4,14}

4.6 Strengths

The present study has a number of strengths including a large sample size of 99 youth. This is considerably larger than most similar studies which typically have between 10-30 participants. Additionally, the study protocol included a large number of activities ($n = 16$) and rather than only classifying activities groups, as is most common with other similar studies, this study developed a series of models for both individual activities and activities grouped by domain. Most previous studies have used gross movements such as locomotion (e.g. walking, jogging, running, stair climbing) and postures (e.g. sitting, standing, lying) as the main class labels so the inclusion of more inclusive group labels containing a wider variety of activities is an improvement over many previous studies. The activities, while performed in a laboratory environment were collected under more naturalistic semi-structured conditions designed to simulate free-living behaviour. Participants were allowed to transition more autonomously between activities as they would in free-living rather than have to start and stop abruptly as is typical of structured laboratory activity protocols where participants are instructed to start and stop on the minute or told exactly when to transition between activities. Additionally, participants were allowed to choose the order in which they preferred to perform the available activities making for more naturalistic engagement in activities and transitions between activities. The volume of data available for this study is substantial. There was on average 2.2-hrs of data per participant. Using data at the 1-s level resulted in a total of nearly 645,000 seconds of data, which is considerable for training and testing the chosen classifiers. This study developed a series of machine learning models allowing for direct comparison of performance

between model types for three different feature sets across three different attachment sites. The novelty of the study goes back to the investigation of the gyroscope sensor data as an independent classifier of PA compared to accelerometer sensor data as an independent classifier of PA as has typically been done in other studies in this field as well as comparing the two single sensor data approaches to a combined sensor data method to see which is better head-to-head.

4.7 Limitations

While this study does have a number of strengths, it is not without limitations. The chosen window size of 1-s could also be a potential limitation when considering performance. Classifying PA at the second-by-second level could lead to gross misclassification of certain activities because of the short window length. Activities such as walking, and running are typically able to be reasonably well classified due to the periodic and cyclical nature of a typical gait pattern and the generally continuous nature of the activity. The periodicity of the activity may not be adequately captured as the window length may be too short. Conversely, because of the short window length broader activities that contain walking or running as part of the activity behaviour, such as soccer or basketball, may result in misclassification. For example, those sporting activities contain walking and running as part of the activity and at the second-by-second level, a second of walking or running during basketball may be classified as walking or running when the criterion label is actually the larger activity behaviour of basketball. In this case, the second of data is technically correctly classified but misclassified by the model due to the way the data are labelled. Banos et al.⁴³ show that 1 to 2-s window lengths may provide the best tradeoff for both the speed of recognition and accuracy of the activity recognized which may be critical for detecting periods of transition between activities. However, they do note that for more complex activity behaviours, longer window lengths may be necessary.⁴³ This issue could

potentially be overcome by purposefully segmenting the data before making a classification.^{44,45}

The features used in the present study are not all encompassing and used only as a starting point, which could be considered a limitation. Little is known about what features are most valuable for working with gyroscope data and choosing the “right” features for the task at hand may require further evaluation in future studies as part of the optimization process. Lastly, while the results of this study offer plenty of value to the field, one missing component that deserves acknowledgement is the exclusion of transition periods from the development process. Including transitions in future iterations of the development process should be considered as transitions play an integral role in defining PA behaviours in free-living and may be valuable for segmenting activities as a pre-classification step.

4.8 Next Steps and Practical Considerations

Further refinement of the approaches presented in this study should consider how to incorporate unknown and transition data to make models more generalizable in free-living application. While this study demonstrates that using accelerometer and gyroscope sensor data together generally results in improved classification of PA, the magnitude of this improvement needs to be documented in free-living as this is the critical need for the end-user. When developing a predictive model to characterize PA it is common to use data from structured laboratory or semi-structured simulated free-living activities to create criterion labels for training the model.⁴⁶ To evaluate these models, a within-sample cross-validation is typically performed to determine how well the model performs under the same conditions and to gauge how well it might generalize to new data.⁴⁶ For within-sample cross-validations in youth, classification accuracies have ranged between 76.8% and 91.7%⁴⁷⁻⁵² using activity counts from a hip-worn device. Using raw acceleration, classification accuracies have ranged from 81.3 to 91.0%^{16,29,53}

using a hip-worn device and 80.4% to 96.7%^{14,16,29,31} using a wrist-worn device. At first glance, this level of accuracy seems acceptable, however, the true test comes when models are applied to out-of-sample data, often either simulated free-living or unstructured free-living data. In general, it is classifier performance decreases when applied to out-of-sample and free-living data.⁵³⁻⁵⁵ For example, Mannini et al.³¹ showed that for classifiers developed using structured activity data in youth, a decrease in classification accuracy from 91.0% to 71.8% and 92.4% to 84.9% was seen for hip and wrist classifiers, respectively, when the classifiers were applied to out-of-sample data. Ahmadi et al.⁵⁴ validated hip and wrist classifiers developed by Trost et al.²⁹ in a sample of youth under laboratory conditions. Ahmadi et al.⁵⁴ showed when using out-of-sample data, performance decreased by 11-15% for the hip classifiers and 19-21% for the wrist classifiers resulting in classification accuracies of 66.4-69.4% for the hip and 59.1-59.3% for the wrist. Similar trends exist in adults.¹ This decrease in performance is largely due to the structured laboratory environments where data used for classifier development are collected. The nuances and intricacies of unstructured free-living PA behaviour are complex and challenging to model using information from a single device containing a single sensor type.² One potential answer to this issue is to use additional sensor data, such as gyroscope data, to complement the acceleration data and more completely characterize PA.

Another aspect to consider when evaluating the performance of classifier when it is translated to free-living application is how the model was originally cross-validated. For the within-sample cross-validation process, either a holdout, k-fold, or leave-one-out cross-validation (LOOCV) is used to gauge generalizability and performance. For example, Dehghani et al.⁵⁵ showed that the method of cross-validation chosen plays a critical role in the resulting performance estimates. For example, using k-fold cross-validation with non-overlapping

windows can artificially inflate performance results by more than 10% compared to LOOCV. This difference is not always transparent and indicates careful consideration should be taken when evaluating performance results for both within-sample cross-validations and independent out-of-sample validation comparisons.

While the results of this study indicate that using both accelerometer and gyroscope sensor data together improves the characterization of PA, the ability to do this in application is currently limited. As mentioned by Hibbing et al.¹⁰ and LaMunion et al.⁵⁶, the current status of the GT9X device battery life presents a key limitation to using the GT9X in free-living application. Unpublished observations indicate that when the primary accelerometer and the full IMU are engaged during initialization average battery life is approximately 24-hrs. This margin increases to approximately 46-48-hrs if only the gyroscope sensor is enabled in the IMU. Additionally, the volume of data the GT9X collects when the IMU is enabled creates significantly more device data per day than using the accelerometer only which could be a barrier to the end user and should be considered when planning to collect, store, and work with the device data accordingly.^{10,56}

4.9 Conclusion

In summary, the results of this study suggest that using accelerometer and gyroscope sensor data together improves classification of individual activity classes and activity groups compared to using either sensor alone. This is promising for improving the overall characterization of PA using wearable sensor data. The next step is to see how well the models developed in this study perform in an out-of-sample validation and free-living application to quantify how using data from both sensors universally affects PA classification. It should be noted that this study is a starting point intended to demonstrate the utility of the gyroscope sensor

data both on its own and as a complement to the accelerometer sensor data typically used in this field. The focus of this study was not to emphasize classifier optimization and maximize performance, but those items should be addressed in subsequent studies to refine the process of characterizing PA in youth using wearable device data.

Disclosure

This study was supported by the National Institutes of Health grant (R01HD083431). No financial support was received from any of the activity monitor manufacturers, importers, or retailers. The authors have no conflicts of interest to report.

References

1. Farrahi V, Niemela M, Kangas M, Korpelainen R, Jamsa T. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait Posture*. 2019;68:285-299.
2. Tong C, Tailor SA, Lane ND. Are Accelerometers for Activity Recognition a Dead-end? arXiv preprint arXiv:2001.08111 [cs.CV]. 2020.
3. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ. Fusion of smartphone motion sensors for physical activity recognition. *Sensors (Basel)*. 2014;14(6):10146-10176.
4. Ferrari A, Micucci D, Mobilio M, Napoletano P. Human Activities Recognition Using Accelerometer and Gyroscope. Paper presented at: 2019 European Conference on Ambient Intelligence.
5. Jain A, Kanhangad V. Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sens J*. 2017;18(3):1169-1177.
6. Wang A, Chen G, Yang J, Zhao S, Chang C-Y. A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sens J*. 2016;16(11):4566-4578.
7. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. *J Medical Internet Res*. 2012;14(5):e130.
8. Shoaib M, Scholten H, Havinga PJ. Towards physical activity recognition using smartphone sensors. Paper presented at: 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing.
9. ActiGraph Corporation. ActiGraph GT9X User Manual. 2014.

10. Hibbing PR LS, Kaplan AS, Crouter SE. Estimating Energy Expenditure with ActiGraph GT9X Inertial Measurement Unit. *Med Sci Sports Exerc.* 2018;50(5):1093-1102.
11. LaMunion SR, Blythe AL, Hibbing PR, Kaplan AS, Clendenin BJ, Crouter SE. Use of Consumer Monitors for Estimating Energy Expenditure in Youth. *Appl Physiol Nutr Metab.* 2019.
12. Hibbing PR, Bassett DR, Coe DP, LaMunion SR, Crouter SE. Youth Metabolic Equivalents Differ Depending on Operational Definitions. *Med Sci Sports Exerc.* 2020. Online Ahead of Print.
13. R Core Team. R: A language and environment for statistical computing. 2017; <https://www.R-project.org/>.
14. Chowdhury AK, Tjondronegoro D, Chandran V, Trost SG. Ensemble Methods for Classification of Physical Activities from Wrist Accelerometry. *Med Sci Sports Exerc.* 2017.
15. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol.* 2009;107(4):1300-1307.
16. Trost SG, Zheng Y, Wong WK. Machine learning for activity recognition: hip versus wrist data. *Physiol Meas.* 2014;35(11):2183-2189.
17. Rosati S, Balestra G, Knaflitz M. Comparison of Different Sets of Features for Human Activity Recognition by Wearable Sensors. *Sensors (Basel).* 2018;18(12).
18. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *J Appl Physiol.* 2015;119(4):396-403.

19. Crouter SE, Horton M, Bassett DR, Jr. Validity of ActiGraph child-specific equations during various physical activities. *Med Sci Sports Exerc.* 2013;45(7):1403-1409.
20. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *J Appl Physiol.* 2011;111(6):1804-1812.
21. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas.* 2014;35(11):2191-2203.
22. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and Wrist Accelerometer Algorithms for Free-Living Behavior Classification. *Med Sci Sports Exerc.* 2016;48(5):933-940.
23. Pavey TG, Gilson ND, Gomersall SR, Clark B, Trost SG. Field evaluation of a random forest activity classifier for wrist-worn accelerometer data. *J Sci Med Sport.* 2017;20(1):75-80.
24. Kuhn M. *Classification and Regression Training* [computer program]. Version 6.0-80. Comprehensive R Archive Network. 2018. Package ‘caret’
25. Kuhn M. *Caret ‘confusionMatrix’* [computer program]. Version v6.0-85. RDocumentation. 2018. Caret ‘confusionMatrix’
26. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw.* 2008.
27. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag.* 2009;45(4):427-437.
28. Kerr J, Patterson RE, Ellis K, et al. Objective Assessment of Physical Activity: Classifiers for Public Health. *Med Sci Sports Exerc.* 2016;48(5):951-957.

29. Trost SG, Cliff DP, Ahmadi MN, Tuc NV, Hagenbuchner M. Sensor-enabled Activity Class Recognition in Preschoolers: Hip versus Wrist Data. *Med Sci Sports Exerc.* 2018;50(3):634-641.
30. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. *Med Sci Sports Exerc.* 2013;45(11):2193-2203.
31. Mannini A, Rosenberger M, Haskell WL, Sabatini AM, Intille SS. Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle. *Med Sci Sports Exerc.* 2017;49(4):801-812.
32. U.S. Department of Health and Human Services. About the National Health and Nutrition Examination Survey. 2020; <https://www.cdc.gov/nchs/nhanes/index.htm>. Accessed March 13, 2020.
33. Doherty A, Jackson D, Hammerla N, et al. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLoS ONE.* 2017;12(2):e0169649.
34. McManus IC. The inheritance of left-handedness. *Ciba Found Symp.* 1991;162:251-267; discussion 267-281.
35. Chandler J, Beets M, Saint-Maurice P, et al. Wrist-Based Accelerometer Cut-Points to Identify Sedentary Time in 5(-)11-Year-Old Children. *Children (Basel).* 2018;5(10).
36. Hibbing PR, Kim Y, Saint-Maurice PF, Welk GJ. Impact of activity outcome and measurement instrument on estimates of youth compliance with physical activity guidelines: a cross-sectional study. *BMC Public Health.* 2016;16.

37. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med Sci Sports Exerc.* 2013;45(5):964.
38. Buchan DS, McSeveney F, McLellan G. A comparison of physical activity from Actigraph GT3X+ accelerometers worn on the dominant and non-dominant wrist. *Clin Physiol Funct Imaging.* 2019;39(1):51-56.
39. Dieu O, Mikulovic J, Fardy PS, Bui-Xuan G, Beghin L, Vanhelst J. Physical activity using wrist-worn accelerometers: comparison of dominant and non-dominant wrist. *Clin Physiol Funct Imaging.* 2016.
40. Montoye AH, Mudd LM, Biswas S, Pfeiffer KA. Energy Expenditure Prediction Using Raw Accelerometer Data in Simulated Free Living. *Med Sci Sports Exerc.* 2015;47(8):1735-1746.
41. Zhang S, Murray P, Zillmer R, Eston RG, Catt M, Rowlands AV. Activity classification using the GENE: optimum sampling frequency and number of axes. *Med Sci Sports Exerc.* 2012;44(11):2228-2234.
42. Montoye AH, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Wrist-independent energy expenditure prediction models from raw accelerometer data. *Physiol Meas.* 2016;37(10):1770-1784.
43. Banos O, Galvez JM, Damas M, Pomares H, Rojas I. Window size impact in human activity recognition. *Sensors (Basel).* 2014;14(4):6474-6499.
44. van Kuppevelt D, Heywood J, Hamer M, Sabia S, Fitzsimons E, van Hees V. Segmenting accelerometer data from daily life with unsupervised machine learning. *PLoS ONE.* 2019;14(1).

45. Hibbing PR, LaMunion SR, Hilafu H, Crouter SE. Evaluating the performance of sensor-based bout detection algorithms: The transition pairing method. *J Meas Phys Behav*. 2020.
46. Awais M, Mellone S, Chiari L. Physical activity classification meets daily life: Review on existing methodologies and open challenges. Paper presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
47. de Vries SI, Engels M, Garre FG. Identification of children's activity type with accelerometer-based neural networks. *Med Sci Sports Exerc*. 2011;43(10):1994-1999.
48. Hagenbuchner M, Cliff DP, Trost SG, Van Tuc N, Peoples GE. Prediction of activity type in preschool children using machine learning techniques. *J Sci Med Sport*. 2015;18(4):426-431.
49. Mu Y, Lo HZ, Ding W, Amaral K, Crouter SE. Bipart: Learning Block Structure for Activity Detection. *IEEE Trans Knowl Data Eng*. 2014;26(10):2397-2409.
50. Ren X, Ding W, Crouter SE, Mu Y, Xie R. Activity recognition and intensity estimation in youth from accelerometer data aided by machine learning. *Appl Intell*. 2016;45(2):512-529.
51. Trost SG, Wong WK, Pfeiffer KA, Zheng Y. Artificial neural networks to predict activity type and energy expenditure in youth. *Med Sci Sports Exerc*. 2012;44(9):1801-1809.
52. Zhao W, Adolph AL, Puyau MR, Vohra FA, Butte NF, Zakeri IF. Support vector machines classifiers of physical activities in preschoolers. *Physiol Rep*. 2013;1(1):e00006.

53. Kuhnhausen J, Dirk J, Schmiedek F. Individual classification of elementary school children's physical activity: A time-efficient, group-based approach to reference measurements. *Behav Res Methods*. 2017;49(2):685-697.
54. Ahmadi MN, Brookes D, Chowdhury A, Pavey T, Trost SG. Free-living Evaluation of Laboratory-based Activity Classifiers in Preschoolers. *Med Sci Sports Exerc*. 2019.
55. Dehghani A, Glatard T, Shihab E. Subject cross validation in human activity recognition. arXiv preprint arXiv:1904.02666 [cs.LG]. 2019.
56. LaMunion SR, Fitzhugh EC, Crouter SE. Challenges and Opportunities Related to the Objective Assessment of Physical Activity Within U.S. Health Surveys. *Ann Epidemiol*. 2020.

Chapter 5: Study 2 Methods

The data presented in this chapter are part of a larger study aiming to identify periods of transition between bouts of activity in structured laboratory and unstructured free-living activities. This chapter is focused on presenting the methods for collecting the unstructured free-living activity data and includes only procedural information and how the dataset was assembled.

Section 1: Participants

Forty-two youth participants between the ages of 6 and 18 years old were recruited from the Knoxville, TN, Aiken, SC, Augusta, GA, and Ames, IA areas by word of mouth and flyers (*Appendix G*). Written informed consent (*Appendix H*) and a health history questionnaire (*Appendix E*) were obtained from a parent or legal guardian of each participant prior to enrollment. Each participant provided written informed assent (*Appendix I*) before study participation was initiated. The recruitment strategy aimed to achieve an approximately even distribution of biological sex across the age range in four target age groupings: 1) 6-9 years old, 2) 10-12 years old, 3) 13-15 years old, and 4) 16-18 years old. Participants were excluded if any contraindications to exercise were reported by the parent/legal guardian on the health history questionnaire including: any musculo-skeletal injuries within the past six months that would prevent safe participation in regular physical activity, any metabolic condition or prescription medication that may affect metabolism, and any cardio-respiratory conditions, including asthma, that may limit regular physical activity participation. All study methods and procedures were approved by The University of Tennessee Knoxville Institutional Review Board.

Section 2: Procedures

The data presented in this study was collected as part of a larger study aiming to identify transitions between activity bouts in a structured laboratory and an unstructured free-living

environment. As part of the larger study protocol, participants were asked to complete 30 minutes of supine rest as a measure of resting metabolic rate (RMR) using a Cosmed K4b² or Cosmed K5 portable indirect calorimeter (Cosmed s.r.l., Rome, Italy). The primary focus of this study is solely on physical activity classification and does not include an energy expenditure component, however, the wearable device data from the 30 minutes of supine rest during RMR was included as “lying” activity data for the subsequent analyses.

All research visits were conducted in the community, participant’s homes, local parks, and approved facilities of community partners. Data was collected across two separate days, once in the participant’s homes, and once in an after-school environment.

- i. The first study visit included completion of the informed consent and assent process and participant screening before final enrollment. Once enrolled, participants had anthropometric measurements taken including: standing and sitting height (cm) using a SECA 213 portable stadiometer and body mass (kg) and body fat percentage measured using a Tanita BC-418 bioelectrical impedance analyzer. Anthropometric measurements were taken with shoes and socks removed. Following completion of the anthropometric measurements, the 30-min RMR assessment was made before participants engaged in activities of their choice.
- ii. For the second study visit, an “after-school program” environment was sought out for each participant. The “after-school program” environment was loosely defined as after-school programs and the activities participants engage in during the hours after-school vary for each individual participant. In many cases, a typical after-school program was not available as an option for participants. In this case, an

effort was made to work with the participant in an environment outside of the home to create two distinct environments. Due to some family scheduling constraints and weather limitations, this was not always possible, and a second home visit was required to serve as the proxy after-school program environment. In the event of a second home visit being substituted as the after-school program visit, participants were asked to engage in different activities than what they chose during their home visit so as to create two distinct environments with different activity behaviours.

For both study visits, participants were not given instructions on what activities to perform so long as they met these three general criteria: 1) parental permission to engage in an activity was received, 2) the activity did not involve getting wet or in water for the sake of protecting the equipment, and 3) the activity did not place the participants or anyone else immediately in harm's way.

All participants wore an ActiGraph GT9X (GT9X) activity monitor on the left and right wrists, and the right hip. The hip monitor was worn on an elastic belt over the right iliac crest along the anterior axillary line and secured using a manufacturer specified clip. The wrist-worn devices were worn in manufacturer specific watch bands that the devices clipped into. The watch bands were positioned proximal to the styloid processes with the devices oriented where the LCD display could be read naturally by the participant like a watch screen would be. Participants were directly observed for the duration of each study. Video recordings were made of each participants activity behaviour that was later coded off-line using the Noldus Observer XT software program (Noldus International Technology, Wageningen, Netherlands). Direct observation was primarily focused on coding activity behaviours, periods of transition, posture.

Section 3: Equipment

3.1 Wearable Device

ActiGraph GT9X: The GT9X is a small (3.5 x 3.5 x 1 cm), light weight (14 grams), water resistant (1 meter, 30 minutes) research-grade activity monitor with a programmable display window that can be worn on the hip, wrist, or ankle. The GT9X contains a primary triaxial accelerometer with a dynamic range of ± 8 gravitational units (g's) and can be initialized to collect data at sampling frequencies of 30 - 100 Hz in increments of 10 Hz. The GT9X also contains an inertial measurement unit (IMU) that samples at a default 100 Hz. The IMU houses a secondary triaxial accelerometer with a dynamic range of ± 16 g's, a triaxial gyroscope with a dynamic range of ± 2000 degrees per second, a magnetometer with a dynamic range of ± 4800 micro-Tesla, and thermometer that measures the temperature inside the device housing. For the present study, GT9X devices were initialized to collect data at 90 Hz (primary accelerometer) and with the full IMU enabled (100 Hz) and Idle Sleep Mode disabled.

3.2 Direct Observation

Noldus Observer XT: The Noldus Observer XT (version 12.5) is a computer software program designed for direct observation. All research visits were video recorded using a Canon Vixia HFR700 camcorder and later coded offline using the Noldus Observer XT Direct Observation software program for Windows. If participants chose to opt out of being video recorded, live coded observations were made using the Noldus Pocket Observer application (version 3.2) for Android tablets on a Samsung Galaxy Tab 4 tablet (Samsung, Seoul, South Korea).

A two-class focal sampling coding scheme was used to first code the activity behaviour being performed and then the posture of the participant during the activity behaviour. A list of commonly performed activities and behaviours was assembled and is available in *Appendix B - Table 160*. In addition to the list of activity behaviours, labels were created for: 1) the start time of the observation period and activity routine (“Start Time”), 2) transitions between activities (“Transition”), 3) unknown activities to use when a participant performed an undefined activity (“Unknown Activity”), and 4) an off-camera label to use when the participants were not in the camera frame (“Off-Camera”). After an activity behaviour was coded, posture was coded as lying, sitting, standing, stepping, squatting/kneeling, reclining, mixed posture, or unknown (used only when the activity behaviour was “off-camera”). This two-class focal sampling coding scheme allowed for non-overlapping activity labels to be assigned by selecting the appropriate activity label button and subsequently the corresponding posture in the Noldus Observer XT software program. This created a continuous record of activities with time stamps that indicated the duration of activities marked by the beginning (e.g. when an activity was marked) and end of an activity label (e.g. when a new activity or transition was marked). In the event live-coded observations were collected using the Pocket Observer application and then later imported into the Noldus Observer XT software program on a Windows computer. To align the direct observation data with the ActiGraph GT9X data, a video of the Windows system clock used to initialize the GT9X devices was taken at the start and end of each data collection period. This allows for sufficient alignment of observation activity labels and GT9X monitor data during pre-processing. When live-coding was necessary, the observation start time was recorded in the Pocket Observer observation.

Section 4: Dataset Assembly

There were 42 youth participants enrolled in Study 2 that had at least one day of data available for analyses. Thirty-five participants had two days of data available at the time of analysis, one participant did not complete a second study visit, and six participants did not have their second visit coded and available at the time of analysis. All 42 participants with available data were used for analysis, regardless of having one or two days of data. Criterion-labelled activity data from the direct observation files were synchronized and merged with the second-by-second GT9X files so each second of monitor data had a corresponding behaviour label with the primary activity behaviour as well as posture. This was done for each of the three attachment sites, hip, left wrist, and right wrist making three unique datasets.

An overview of the analysis plan is shown in *Appendix A - Figure 1*. Information about further data processing and statistical analyses are available in Chapter 6.

Chapter 6: Out-of-Sample Validation

Abstract

Purpose: The purpose of this study is three-fold: 1) to independently validate previously developed machine learning models for classifying PA using out-of-sample unstructured PA data, 2) to re-develop previously existing random forest models to incorporate transitions and perform a within-sample cross-validation using structured laboratory activity data, and 3) to perform an out-of-sample validation on the re-developed random forest models using unstructured free-living activity data. Classifiers used three feature sets from the ActiGraph GT9X (GT9X): 1) primary accelerometer data only, 2) gyroscope data only, and 3) accelerometer + gyroscope data combined. **Methods:** A sample of 42 youth aged 6-18 years old performed activities at their discretion on two separate visits, each approximately 2-hrs in duration. Participants wore a GT9X on the right hip, left wrist and right wrist and were directly observed for the duration of each visit to create timestamped criterion labels of the activities performed. GT9X accelerometer (90 Hz) and gyroscope (100 Hz) data were reduced to 1-s averages where 47 statistical features were calculated from the sensor signals. Decision trees, lasso logistic regression, random forest, and neural network classifiers were validated for each feature set and attachment site. Overall model classification accuracy for each participant was aggregated to obtain a single estimate for each classifier for comparison. **Results:** The combined sensor feature set (average classification accuracy: 77.5%; classification accuracy range: 74.2-81.4%) outperformed the accelerometer only feature set (72.4%; 68.9-73.1%) by an average of 5.0% and the gyroscope only feature set (76.8%; 71.5-81.5%) by an average of 0.7% across all attachment sites and classifier types. For the within-sample cross-validation of the redeveloped random forest models including transitions, the combined sensor feature set (72.5-75.1%) outperformed the accelerometer only (65.3-70.0%) and gyroscope only (64.7-65.6%) feature sets

by an average of 6.1% and 8.4%, respectively. For the out-of-sample validation of the random forest models including transitions, classification accuracy decreased for all three feature sets, 67.5% to 60.3% for the accelerometer only, 65.1% to 64.1% for the gyroscope only, and 73.6% to 68.6% for the combined sensor feature sets, respectively. **Conclusion:** The results of this study indicate that using GT9X primary accelerometer and gyroscope data together improves PA classification of both semi-structured and free-living activity data compared to using only accelerometer data.

Section 1: Introduction

Accelerometer-based physical activity (PA) monitors are frequently used to assess free-living PA in a variety of populations. The output from these devices is used to characterize different components of PA behaviours including energy expenditure (EE), activity intensity, and PA type. Historically, the output has been in the form of activity counts, a proprietary unit specific to the device manufacturer. Recently, it has become more common to use raw acceleration data instead of counts as there is greater flexibility when modeling the data as well as improved compatibility and comparability between device brands and study outcomes^{1,2}. Both counts and raw acceleration have been used to characterize the aforementioned components of PA behaviour. Regardless of the data format or statistical technique used to develop the model, the classifiers are specific to: 1) the activities used, 2) the sample population, 3) the setting and conditions in which the activities were performed, and 4) the brand of device used to collect the data (for proprietary metrics such as ActiGraph counts).^{3,4}

The current convention for classifier development typically uses data from a battery of structured laboratory activities or semi-structured simulated free-living activities.⁴ The goal with this development structure is to assemble a dataset and create a predictive model that achieves high performance during training and within-sample cross-validation but also generalizes well to new data when applied in free-living situations. Capturing the diversity and variability in human movement can make this balance difficult to achieve due to the typically limited number of activities, homogenous sample populations, and structured nature of the development protocols.⁵ In the field of PA assessment, it has been routine to report within-sample cross-validation performance results while independent out-of-sample validations including unstructured free-living data have been relatively infrequent. This is largely due to a couple of factors: 1) the time

and cost associated with collecting the data, both a major limitation, and 2) a lack of standardization between study protocols, including the device and device parameters used to collect the data. In the limited data available, a distinct trend exists where classifier performance decreases considerably when classifiers developed using structured or semi-structured laboratory-based activities are validated using out-of-sample or unstructured free-living data.⁶⁻¹⁰ The magnitude of the decrease in performance varies but in some studies have been by greater than 20.0%.⁶

Tong et al.¹¹ recently questioned just how well classifiers could perform given the current conventions of collecting movement data using accelerometer-based devices and using only acceleration data for activity recognition. There are obvious limitations to using inertial sensor data to model human movement, however, there are still possible avenues to improve overall classifier performance, including model generalizability in free-living application. One such possibility is to use additional sensor data, such as gyroscope data, in combination with acceleration data. The rotational data captured by the gyroscope should complement the linear movement captured by the accelerometer. Studies have shown previously that using accelerometer and gyroscope data together improves the classification of PA compared to using data from either sensor individually (unpublished data; Chapter 4).¹²⁻¹⁷ Shoaib et al.¹⁶ have shown that the accelerometer is better for some activities and the gyroscope for others but this is dependent on the attachment site of the sensor and the type of activity being performed. For example, Ferrari et al.¹² showed that classification accuracy is approximately 10.0% higher on average when classifiers are developed using both accelerometer and gyroscope sensor features compared to using data from either single sensor independently. These data were collected in adults using smartphones with limited sample sizes and a small number of activities as is

common with most studies in this area to date. LaMunion et al. (Chapter 4) recently developed several machine learning classifiers (decision tree, lasso logistic regression, random forest, and neural networks) using ActiGraph GT9X (GT9X) 1) primary accelerometer data only, gyroscope data only, and 3) primary accelerometer + gyroscope data together (combined sensor) to classify PA type in youth. The results of the study indicated that for semi-structured simulated free-living activities, a combined sensor approach achieves the highest performance regardless of the classifier type and attachment site for classifying individual activities or activities grouped by domain which supports the findings in previous adult studies using smartphone data.¹²⁻¹⁷ It is, however, unknown if this trend in higher performance exists when translating the combined sensor models from semi-structured laboratory-based activities to out-of-sample data and unstructured free-living activities. Additionally, the originally developed LaMunion et al. (Chapter 4) classifiers did not include an unknown or other category to categorize data from new activity classes or from periods of transition. After surveying the landscape of the literature and assessing the needs of the field of PA assessment, it was determined that future iterations of classifier development should account for unseen data points not belonging to a pre-defined activity class. The inclusion of an unknown/other/transition activity class should be included in development to allow for complete assessment of out-of-sample data without excluding data points.

Thus, the purpose of this study is three-fold: 1) to independently validate previously developed machine learning models for classifying PA using out-of-sample unstructured free-living youth activity data, 2) to re-develop previously existing random forest models to incorporate transitions and perform a within-sample cross-validation using structured laboratory

activity youth data, and 3) to perform an out-of-sample validation on the re-developed random forest models using unstructured free-living youth activity data.

Section 2: Methods

2.1 Participants

A sample of 42 youth participants between the ages of 6 and 18 years old were recruited from the Knoxville, TN, Aiken, SC, Augusta, GA, and Ames, IA area by word of mouth and flyers. Written informed consent and a health history questionnaire were obtained from a parent or legal guardian of each participant and written informed assent was provided by each participant prior to enrollment in the study. The recruitment strategy aimed to achieve an approximately even distribution of biological sex across the age range in four target age groupings: 1) 6-9 years old, 2) 10-12 years old, 3) 13-15 years old, and 4) 16-18 years old. Participants were excluded if any contraindications to exercise were reported by the parent/legal guardian on the health history questionnaire including: any musculo-skeletal injuries within the past six months that would prevent safe participation in regular physical activity, any metabolic condition or prescription medication that may affect metabolism, and any cardio-respiratory conditions, including asthma, that may limit regular physical activity participation. All study methods and procedures were approved by The University of Tennessee Knoxville Institutional Review Board.

The data presented in this study was collected as part of a larger study aiming to identify transitions between activity bouts in a structured laboratory and an unstructured free-living environment. As part of the larger study protocol, participants were asked to complete 30 minutes of supine rest as a measure of resting metabolic rate (RMR) using a Cosmed K4b² or

Cosmed K5 portable indirect calorimeter (Cosmed s.r.l., Rome, Italy). The primary focus of this study is solely on physical activity classification and does not include an energy expenditure component, however, the wearable device data from the 30 minutes of supine rest during RMR was included as “lying” activity data for the subsequent analyses.

2.2 Procedures

All research visits were conducted in the community, participant’s homes, local parks, and approved facilities of community partners. Data was collected across two visits on two separate days, once in the participant’s homes, and once in an after-school environment.

i. Study Visit 1

The first study visit included completion of the informed consent and assent process and participant screening before final enrollment. Once enrolled, participants had anthropometric measurements taken including: standing and sitting height (cm) using a SECA 213 portable stadiometer and body mass (kg) and body fat percentage measured using a Tanita BC-418 bioelectrical impedance analyzer. Anthropometric measurements were taken with shoes and socks removed. Following completion of the anthropometric measurements, the 30-min RMR assessment was made before participants engaged in activities of their choice.

ii. Study Visit 2

For the second study visit, an “after-school program” environment was sought out for each participant. The “after-school program” environment was loosely defined as after-school programs and the activities participants engage in during the hours after-school can be vastly different for each individual

participant. In many cases, a typical after-school program was not available as an option for participants. In this case, an effort was made to work with the participant in an environment outside of the home to create two distinct environments. Due to some family scheduling constraints and weather limitations, this was not always possible, and a second home visit was required to serve as the proxy after-school program environment. In the event of a second home visit being substituted as the after-school program visit, participants were asked to engage in different activities than what they chose during their home visit so as to create two distinct environments with different activity behaviours.

For both study visits, participants were not given instructions on what activities to perform so long as they met these three general criteria: 1) parental permission to engage in an activity was received, 2) the activity did not involve getting wet or in water for the sake of protecting the equipment, and 3) the activity did not place the participants or anyone else immediately in harm's way.

All participants wore an ActiGraph GT9X (GT9X) activity monitor on the left and right wrists, and the right hip. The hip monitor was worn on an elastic belt over the right iliac crest along the anterior axillary line and secured using a manufacturer specified clip. The wrist-worn devices were worn in manufacturer specific watch bands that the devices clipped into. The watch bands were positioned proximal to the styloid processes with the devices oriented where the LCD display could be read naturally by the participant like a watch screen would be. Participants were directly observed for the duration of each study. Video recordings were made of each participants activity behaviour that was later coded off-line using the Noldus Observer XT software program (Noldus International Technology, Wageningen, Netherlands). If participants

chose to opt out of being video recorded, live coded observations were made using the Noldus Pocket Observer application (version 3.2) for Android tablets. Direct observation was primarily focused on coding activity behaviours, periods of transition, posture.

2.3 Data Acquisition

2.3.1 Wearable Device

ActiGraph GT9X: The GT9X is a small (3.5 x 3.5 x 1 cm), light weight (14 grams), water resistant (1 meter, 30 minutes) research-grade activity monitor with a programmable display window that can be worn on the hip, wrist, or ankle. The GT9X contains a primary triaxial accelerometer with a dynamic range of ± 8 gravitational units (g's) and can be initialized to collect data at sampling frequencies of 30 - 100 Hz in increments of 10 Hz. The GT9X also contains an inertial measurement unit (IMU) that samples at a default 100 Hz. The IMU houses a secondary triaxial accelerometer with a dynamic range of ± 16 g's, a triaxial gyroscope with a dynamic range of ± 2000 degrees per second, a magnetometer with a dynamic range of ± 4800 micro-Tesla, and thermometer that measures the temperature inside the device housing. For the present study, GT9X devices were initialized to collect data at 90 Hz (primary accelerometer) and with the full IMU enabled (100 Hz) and Idle Sleep Mode disabled.

2.3.2 Direct Observation

Noldus Observer XT: The Noldus Observer XT (version 12.5) is a computer software program designed for direct observation. All research visits were video recorded using a Canon Vixia HFR700 camcorder and later coded offline using the Noldus Observer XT Direct Observation software program for Windows. If participants chose to opt out of being video recorded, live coded observations were made using the Noldus Pocket Observer application

(version 3.2) for Android tablets on a Samsung Galaxy Tab 4 tablet (Samsung, Seoul, South Korea).

A two-class focal sampling coding scheme was used to first code the activity behaviour being performed and then the posture of the participant during the activity behaviour. A list of commonly performed activities and behaviours was assembled and is available in *Appendix B – Table 160*. In addition to the list of activity behaviours, labels were created for: the start time of the observation period and activity routine (“Start Time”), transitions between activities (“Transition”), unknown activities to use when a participant performed an undefined activity (“Unknown Activity”), and an off-camera label to use when the participants were not in the camera frame (“Off-Camera”). After an activity behaviour was coded, posture was coded as lying, sitting, standing, stepping, squatting/kneeling, reclining, mixed posture, or unknown (used only when the activity behaviour was “unknown”). This two-class focal sampling coding scheme allowed for non-overlapping activity labels to be assigned by selecting the appropriate activity label button and subsequently the corresponding posture in the Noldus Observer XT software program. This created a continuous record of activities with time stamps that indicated the duration of activities marked by the beginning (e.g. when an activity was marked) and end of an activity label (e.g. when a new activity or transition was marked). In the event live-coded observations were collected using the Pocket Observer application and then later imported into the Noldus Observer XT software program on a Windows computer. To align the direct observation data with the ActiGraph GT9X data, a video of the Windows system clock used to initialize the GT9X devices was taken at the start and end of each data collection period. This allows for sufficient alignment of observation activity labels and GT9X monitor data during pre-

processing. When live-coding was necessary, the observation start time was recorded in the Pocket Observer observation.

2.4 Wearable Device Data

2.4.1 Data Pre-processing and Reduction

All data handling was conducted in R^{18} statistical software. An overview of the data analysis plan is available in *Appendix A – Figure 1*. A vector magnitude was calculated from the raw GT9X sensor data collected at 90 Hz for the primary accelerometer ($VM = \sqrt{X_{acc}^2 + Y_{acc}^2 + Z_{acc}^2}$) and 100 Hz for the IMU gyroscope ($GVM = \sqrt{X_{gyro}^2 + Y_{gyro}^2 + Z_{gyro}^2}$). Prior to calculation of GVM, the individual gyroscope axes were low-pass filtered at 35 Hz using a second order Butterworth filter. The individual sensor axes, VM, and GVM were then collapsed to 1-s averages.

2.4.2 Features

The following 47 time and frequency domain sensor features were engineered from the 1-s individual axes and VM and GVM sensor data (25 accelerometer features, 22 gyroscope features). All features were calculated in rolling 10-s windows for accelerometer and gyroscope sensor data unless otherwise noted: 1) individual X-, Y-, and Z- axis inclination angles with respect to the vertical axis (accelerometer only), 2) Minimum VM and GVM value, 3) Maximum VM and GVM value, 4) Mean VM and GVM value, 5) Standard deviation of VM and GVM, 6) Coefficient of variation (CV) of VM and GVM, 7) Variance of VM and GVM.

Once all features were engineered, the 1-s windows of sensor data were aggregated with labels from the direct observation files for each participant and each monitor attachment site (hip, left wrist, and right wrist) so each 1-s window has a label for activity group and posture.

2.4.3 Feature Sets

For each attachment site dataset (hip, left wrist, and right wrist), three feature sets were created as follows comprised of the engineered features outlined above: 1) Accelerometer sensor features only ($n = 9$ features), 2) Gyroscope sensor features only ($n = 6$ features), and 3) Accelerometer + Gyroscope sensor features combined ($n = 15$ sensor features). In addition to the sensor features, continuous numerical age (y) and leg length (cm) were used from the demographic dataset making 17 total available features. Most studies in adults indicate that biological variables do not contribute much to the characterization of PA, however, this may be different in youth. However, with the broad age range of the sample population in mind, the decision was made to include leg length and age as features. There are notable differences in how some activities may be performed based on participant age as motor coordination will vary across a broad age range. Additionally, leg length was included to aid in predicting locomotion since youth with different leg lengths will walk and run with different step frequencies corresponding to their leg length so this could impact the classification of ambulatory activities.

2.5 Statistical Analysis

Statistical analyses were performed in *R* statistical software¹⁸. The analyses are split by outcome to correspond with the purposes as follows: **Part I**) perform an out-of-sample free-living validation of the six-class machine learning classifiers previously developed by LaMunion et al. (Chapter 4) using free-living data from 42 youth participants, **Part II**) re-develop the six-class LaMunion et al. random forest classifiers to include a seventh activity group for transition

data and conduct a within-sample cross-validation using semi-structured simulated free-living data from 99 youth participants, and **Part III**) to perform an out-of-sample free-living validation of the seven-class random forest classifiers developed in Part II using free-living data from 42 youth participants.

i. Part I: Out-of-Sample Validation of LaMunion et al. Activity Group Models

LaMunion et al. (Chapter 4) previously developed four machine learning classifiers (decision tree, lasso logistic regression, random forest, and neural network) for classifying individual activities ($n = 14$) and activity groups ($n = 6$). The activity group classifiers are more common in the field as they contain individual activities but are categorized more broadly and are more generalizable when applied to new free-living data. In the original LaMunion et al. (Chapter 4) study, classification accuracy for the activity group classification task was far superior to the classification accuracy for the individual activity classes (78.4% versus 54.6%) across all classifier types, feature sets, and attachment sites. Thus, for the purposes of the present study, only the original six-class activity group classifiers will be evaluated using out-of-sample unstructured free-living data. The six activity groups include the following classes: 1) household chores, 2) seated and lying activities, 3) continuous walking and running, 4) sports, 5) recreation and leisure-time activity, and 6) wheeled activities. The original classifiers were developed using semi-structured simulated free-living activity data from 99 youth participants. The purpose of Part 1 of the present study is to validate the original classifiers using the out-of-sample free-living PA dataset from 42 youth participants. Participants in the free-living dataset performed a total

of 44 different PA behaviours which were collapsed into the original six activity groups as outlined in *Appendix B – Table 161*. The “Transition”, “Unknown Activity” and “Off-camera” labelled data points were excluded from this analysis. Classification accuracy for classifying activity groups was calculated for each individual participant visit and then aggregated to obtain an overall classification accuracy for each model type, attachment site, and feature set.

ii. Part II: Within-Sample and Cross-Validation of Random Forest Models with Transitions

The LaMunion et al. six class random forest models from Chapter 4 originally excluded transition data points. The purpose of Part II of the analyses in the present study was to re-develop the random forest classifiers to include a seventh activity group as follows 1) household chores, 2) seated and lying activities, 3) continuous walking and running, 4) sports, 5) recreation and leisure-time activities, 6) wheeled activities, and 7) transitions. A breakdown of how the data points were distributed between activity groups is outlined in *Appendix B – Table 2*. An overall classification accuracy measure was reported for each model type as a general indicator of which feature set results in the highest classification accuracy for each attachment site. The following summary statistics and individual class performance metrics were calculated for each attachment site, and feature set: 1) sensitivity, 2) specificity, 3) prevalence, 4) positive predictive value, 5) negative predictive value, 6) precision, 7) recall, 8) F1 Score, and 9) balanced accuracy.

iii. Part III: Free-living Validation of Random Forest Models with Transitions

The re-developed seven-class random forest models from Part II were validated using free-living PA data from 42 youth participants including transitions. The 44-individual free-living PA behaviour labels were collapsed into one of the following seven categories: 1) household chores, 2) seated and lying activities, 3) continuous walking and running, 4) sports, 5) recreation and leisure time activity, 6) wheeled activity, and 7) transition as outlined in *Appendix B – Table 161*. All data points labelled as “transition” data points were collapsed into the new “transition” category. Data points labelled as “unknown” and “off-camera” were excluded from analyses due to no criterion label assigned to those seconds of data. Classification accuracy was calculated for each individual participant visit and then aggregated to obtain an overall classification accuracy for each attachment site and feature set.

Sections 3: Results

Participant demographics are shown in *Appendix B – Table 162*.

3.1 Out-of-Sample Validation of LaMunion et al. (Chapter 4) Six-class Activity Group Models

3.1.1 Hip Results

Appendix A – Figure 4A shows that when the six class activity group models were applied to the out-of-sample free-living dataset, the combined sensor models (76.6-81.4%) outperformed the accelerometer only (68.9-73.5%) and gyroscope only (77.8-81.5%) models by an average of 8.7% and 0.2%, respectively. The highest overall classifier was the gyroscope only lasso model (81.5%) and the worst performing classifier was the accelerometer only decision tree (68.9%).

3.1.2 Left Wrist Results

Appendix A – Figure 4B shows that when the six class activity group models were applied to the out-of-sample free-living dataset, the combined sensor models (74.5-79.1%) outperformed the accelerometer only (69.4-76.9%) and gyroscope only (72.2-79.0%) models by an average of 3.6% and 1.8%, respectively. The highest overall classifier was the combined sensor random forest model (79.1%) and the worst performing classifier was the accelerometer only decision tree (69.4%).

3.1.3 Right Wrist Results

Appendix A – Figure 4C shows that when the six class activity group models were applied to the out-of-sample free-living dataset, the combined sensor models (76.6-81.4%) outperformed the accelerometer only (68.9-73.5%) and gyroscope only (77.8-81.5%) models by an average of 8.7% and 0.2%, respectively. The highest overall classifier was the combined sensor lasso model (79.1%) and the worst performing classifier was the accelerometer only decision tree (69.3%).

3.2 Within-Sample Cross-Validation of Random Forest Models with Transitions

3.2.1 Hip Results

Appendix A – Figure 5A shows the combined sensor feature set model (72.5%) achieved higher overall classification accuracy than either model using the individual sensor feature sets by 7.2% (accelerometer only) and 7.8% (gyroscope only). Models using the accelerometer only (65.3%) and gyroscope only (64.7%) individual sensor feature sets achieved similar overall classification and differed by 0.6% in favor of the accelerometer only feature set models.

Appendix B - Table 163 summarizes the percentage of cases correctly classified for each feature set and model type by activity group. The combined sensor model had consistently the highest number of correctly classified cases. Across all seven activity classes, the combined sensor model correctly classified 12.3% and 11.9% more cases on average than the accelerometer only and gyroscope only models, respectively. Across all three feature sets classification of seated and lying activity was excellent with the combined sensor model classifying the most cases correctly (95.4%) followed by the gyroscope only model (93.3%) and accelerometer only model (92.4%). The gyroscope only model (32.3%) classified the wheels activity class better than the accelerometer only model (10.3%). Conversely, the accelerometer only model (48.1%) classified recreation and leisure-time activity better than the gyroscope only model (22.2%). For all other activity classes (continuous walking and running, household chores, sports, and transitions), accelerometer only and gyroscope only model performance was within 8.2% of one another.

Appendix B – Table 164 shows performance metrics for each feature set. For all eight performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor model outperformed the single sensor models. Precision was on average 13.3% higher for the combined sensor model (65.0%) compared to the accelerometer only (51.4%) and gyroscope only models (51.9%). Recall was on average 12.1% higher for the combined sensor model (60.9%) compared to the accelerometer only (48.6%) and gyroscope only models (49.0%). Similarly, since F1 is the harmonic mean of precision and recall, the same trend exists for F1 where the combined sensor model (62.4%) was on average 12.8% higher compared to the accelerometer only (49.3%) and gyroscope only models (50.0%).

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 87.1% to 91.5% across all four classifier types with standard deviations ranging from 0.0004 to 0.0011. For the hip activity group classifiers including transitions, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy standard deviations indicating that the training classification accuracies were statistically different.

3.2.2 Left Wrist Results

Appendix A – Figure 5A shows the combined sensor feature set model (73.2%) achieved higher overall classification accuracy than either model using the individual sensor feature sets by 5.8% (accelerometer only) and 8.0% (gyroscope only). Models using the accelerometer only (67.3%) and gyroscope only (65.2%) individual sensor feature sets achieved similar overall classification and differed by 2.1% in favor of the accelerometer only feature set models.

Appendix B - Table 165 summarizes the percentage of cases correctly classified for each feature set and model type by activity group. The combined sensor model had consistently the highest number of correctly classified cases, with the exception of the sports activity class, where the accelerometer only model was superior (72.1% versus 72.0%), but only marginally so. Across all seven activity classes, the combined sensor model correctly classified 11.7% and 12.3% more cases on average than the accelerometer only and gyroscope only models, respectively. Across all three feature sets classification of seated and lying activity was excellent with the combined sensor model classifying the most cases correctly (94.4%) followed by the accelerometer only model (91.7%) and gyroscope only model (91.6%). The gyroscope only model (47.5%) classified the wheels activity class better than the accelerometer only model (15.6%). Conversely, the accelerometer only model (72.1%) classified recreation and leisure-

time activity better than the gyroscope only model (56.5%). For all other activity classes (continuous walking and running, household chores, and transitions), accelerometer only and gyroscope only model performance was within 7.7% of one another.

Appendix B – Table 166 shows performance metrics for each feature set. For all eight performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor model outperformed the single sensor models. Precision was on average 11.1% higher for the combined sensor model (68.0%) compared to the accelerometer only (58.3%) and gyroscope only models (55.6%). Recall was on average 12.0% higher for the combined sensor model (63.2%) compared to the accelerometer only (51.4%) and gyroscope only models (50.9%). Similarly, the same trend exists for F1 where the combined sensor model (64.9%) was on average 12.4% higher compared to the accelerometer only (52.7%) and gyroscope only models (52.4%).

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 86.4% to 89.2% across all four classifier types with standard deviations ranging from 0.0001 to 0.0010. For the left wrist activity group classifiers including transitions, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy standard deviations indicating that the training classification accuracies were statistically different.

3.2.3 Right Wrist Results

Appendix A – Figure 5A shows the combined sensor feature set model (75.1%) achieved higher overall classification accuracy than either model using the individual sensor feature sets by 5.1% (accelerometer only) and 9.5% (gyroscope only). Models using the accelerometer only

(70.0%) and gyroscope only (65.6%) individual sensor feature sets achieved similar overall classification and differed by 4.4% in favor of the accelerometer only feature set models.

Appendix B - Table 167 summarizes the percentage of cases correctly classified for each feature set and model type by activity group. The combined sensor model had consistently the highest number of correctly classified cases. Across all seven activity classes, the combined sensor model correctly classified 10.0% and 13.0% more cases on average than the accelerometer only and gyroscope only models, respectively. Across all three feature sets classification of seated and lying activity was excellent with the combined sensor model classifying the most cases correctly (95.2%) followed by the accelerometer only model (92.9%) and gyroscope only model (91.3%). The gyroscope only model (52.1%) classified the wheels activity class better than the accelerometer only model (27.7%). Conversely, the accelerometer only model (37.0%) classified recreation and leisure-time activity better than the gyroscope only model (24.6%). For all other activity classes (continuous walking and running, household chores, and transitions), accelerometer only and gyroscope only model performance was within 10.0% of one another.

Appendix B – Table 168 shows performance metrics for each feature set. For all eight performance metrics (sensitivity, specificity, positive and negative predictive value, precision, recall, F1, and balanced accuracy), the combined sensor model outperformed the single sensor models. Precision was on average 11.3% higher for the combined sensor model (69.3%) compared to the accelerometer only (60.6%) and gyroscope only models (55.6%). Recall was on average 12.0% higher for the combined sensor model (63.2%) compared to the accelerometer only (51.4%) and gyroscope only models (50.9%). Similarly, the same trend exists for F1 where

the combined sensor model (64.9%) was on average 12.4% higher compared to the accelerometer only (52.7%) and gyroscope only models (52.4%).

For training classification accuracy using k-fold cross-validation, classifier accuracies ranged from 86.1% to 89.1% across all four classifier types with standard deviations ranging from 0.0006 to 0.0015. For the right wrist activity group classifiers including transitions, the difference in classifier training accuracies between feature sets was greater than the maximum classifier training accuracy standard deviations indicating that the training classification accuracies were statistically different.

For additional detail about the individual model confusion matrices and model performance summaries, please see *Appendix B – Tables 169-174* for hip, *Appendix B – Tables 175-180* for left wrist, and *Appendix B – Tables 181-186* for right wrist results.

3.3 Out-of-Sample Validation of Random Forest Models with Transitions

Appendix A- Figure 5B shows that overall, the best classification accuracy was achieved using the combined sensor left wrist classifier. Across all three attachment sites, the accelerometer only models averaged 60.3%. In contrast, the gyroscope only and combine sensor models averaged 64.1% and 68.6% classification accuracy across all three attachment sites, respectively. For the hip, the gyroscope only feature set (68.6%) outperformed the accelerometer only feature set (57.1%) and combined sensor feature set (68.4%) by 11.5% and 0.2%, respectively. For the left wrist, the combined sensor feature set (70.2%) outperformed the accelerometer only feature set (62.1%) and gyroscope only feature set (62.6%) by 8.1% and 7.6%, respectively. For the right wrist, the combined sensor feature set (67.2%) outperformed the accelerometer only feature set (61.7%) and gyroscope only feature set (60.9%) by 5.5% and 6.3%, respectively.

Section 4: Discussion

4.1 Summary

The results of this study indicate that classification accuracy is generally maintained for models developed using gyroscope sensor data across all attachment sites. Conversely, the accelerometer only models see a decrease in classification accuracy across all attachment sites and classifier types in out-of-sample conditions. This is consistent with the literature in youth and adults for predictive models developed using only accelerometer sensor data.⁶⁻¹⁰ The combined sensor feature set follows the same trend as the accelerometer only models, however, the combined sensor feature sets were superior to either single sensor feature set in free-living conditions.

4.2 Performance Overview

4.2.1 Within-Sample Validation versus Out-of-Sample Validation Performance Differences for the LaMunion et al. (Chapter 4 – Study 1) Six-Class Models

A major point of interest in this study is how the gyroscope and combined sensor classifiers developed using semi-structured simulated free-living data perform when applied to out-of-sample unstructured free-living data. LaMunion et al. (Chapter 4) found in a previous study that within-sample cross-validation performance using only accelerometer sensor data, classification accuracies ranged from 71.8% to 80.9% were achieved for hip, left wrist, and right wrist classifiers, respectively, across six activity classes. Similarly, classification accuracies ranged from 72.3% to 80.0% using only gyroscope sensor data and 77.7% to 85.1% for accelerometer and gyroscope data combined. When these classifiers were applied to out-of-sample unstructured free-living data in the present study classification accuracies ranged from

68.9% to 76.9% for the accelerometer only feature set, 71.5.9% to 81.5% for the gyroscope only feature set, and 74.2% to 81.4% for the combined sensor feature set. This resulted no change for the gyroscope only feature set and decreases in average classification accuracy of 4.0% for the accelerometer only feature set (76.7% to 72.7%) and 4.0% for the combined sensor feature set (82.0% to 78.0%) across all three attachment sites. This decrease in performance for the accelerometer only feature set is in line with what other studies have shown. This is generally consistent with the limited literature in this area. For example, Trost et al.¹⁹ originally developed five class hip and wrist classifiers using youth data from structured activities. Ahmadi et al.⁶ recently performed an out-of-sample validation of the original Trost et al.¹⁹ classifiers and showed that performance decreased by 11-15% for the hip classifiers and 19-21% for the wrist classifiers resulting in final classification accuracies of 66.4-69.4% for the hip and 59.1-59.3% for the wrist. Similarly, Mannini et al.⁷ showed that for classifiers developed using structured activity data in youth, a decrease in classification accuracy from 91.0% to 71.8% (difference of 19.2%) and 92.4% to 84.9% (difference of 7.5%) was seen for hip and wrist classifiers, respectively, when the classifiers were tested on out-of-sample data.

The maintenance in performance for the gyroscope only feature set is favorable as there was very little change in performance when averaging across all four classifier types and attachment sites and in some cases the classifier performance improved from within-sample to out-of-sample for the gyroscope only feature set. The performance decrease observed for the combined sensor feature set is being driven by the accelerometer sensor data. However, it should be noted that while a 4.0% decrease in average classification accuracy was observed for the acceleration only and combined sensor feature sets, the final out-of-sample classification accuracies for the combined sensor feature set were on average 1.4% higher than the gyroscope

only feature set indicating that a combined sensor approach is best in both within-sample and out-of-sample situations. One potential avenue to overcome this issue with the accelerometer is to use a different development strategy. For example, Ermes et al.²⁰ suggest that using a mix of structured, semi-structured, and unstructured data for classifier development may improve classifier generalizability. This may be especially beneficial for the accelerometer only approach as performance is comparable to or better than the gyroscope only approach in within-sample cross-validations using structured or semi-structured data. The deterioration in classifier performance only occurs when the classifiers are applied to out-of-sample data and unstructured free-living data.

4.2.2 Within-Sample Validation versus Out-of-Sample Validation Performance Differences for the Seven-Class Random Forest Models including Transitions

Another key component of this study was the re-development of the six-class random forest models to include a seventh class to incorporate transitions and unknown data points. This allows for a true free-living validation of the classifier where no data are cleaned from the activity periods that were observed which is not commonly done due to the difficulty (cost and time) to obtain criterion activity labels in free-living. The within-sample cross-validation performance for the re-developed random forest classifiers ranged from 65.3% to 70.0% for the accelerometer only feature set, 64.7% to 65.6% for the gyroscope only feature set, and 72.5% to 75.1% for the combined sensor feature set. In contrast, the out-of-sample validation performance for the re-developed random forest classifiers ranged from 57.1% to 62.1% for the accelerometer only feature set, 60.9% to 68.6% for the gyroscope only feature set, and 67.2% to 70.2% for the combined sensor feature set. Classification accuracy decreased on average, 7.2% for the

accelerometer only feature set, 1.1% for the gyroscope only feature set, and 4.9% for the combined sensor feature set.

The impact of including transitions in the development and validation process drove down classifier performance across all attachment sites and feature sets (see *Appendix A - Figure 6*). This is not unsurprising given how ambiguous the transition class is. On average, across all attachment sites and feature sets, classification accuracy decreased by 4.4% when transitions were added as a seventh activity group for the random forest classifiers. To better understand the implications and impact of adding the transition class a side-by-side comparison of the six-class random forest results and the seven-class random forest results is necessary. For the original six activity groups, chores, continuous walking and running, recreation and leisure-time activities, seated and lying activities, sports, and wheeled activities, the number of correctly classified cases decreased in each activity group across all attachment sites when transitions were added as a seventh group. The magnitude of decrease was relatively small ($< 10.0\%$) for all groups except household chores where the number of correctly classified cases decreased on average 29.9% across all attachment sites when the transition group was added. This indicates that cases from each group are being misclassified as transitions, but the household chores group may be the most susceptible to this misclassification as indicated by the sharp decrease in number of cases correctly classified.

4.3 Contribution and Recommendation

While true that the gyroscope only models minimize performance loss when translating from laboratory-based activity to free-living activity, it may not be feasible long term to recommend use of gyroscope only models. Several previous studies have shown that using accelerometer and gyroscope data together generally results in an improvement in predictive

performance compared to using data from either single sensor on its own and the results of this study indicate that superior performance can be achieved in both within-sample and out-of-sample conditions using data from both sensors as opposed to using data from either single sensor individually.¹²⁻¹⁷ The field of PA assessment has been using accelerometer sensor data to model PA for more than two decades and while it has its limitations, accelerometer sensor data should not be left out of classifier development and optimization moving forward.

While this study does improve the general knowledge base about PA classification in youth, one larger point of interest for overall PA characterization is about better capturing the individual variability in PA behaviours. One reason classifier error is high at the individual level is there are a multitude of factors that contribute to an individual's PA behaviour making it difficult to use a single general classifier to robustly characterize individual PA behaviours using information from a single sensor type at a single attachment site. As previously mentioned, one potential way to improve both the overall model performance while simultaneously minimizing performance loss and improving generalizability in application is to use a mixed set of data for development. Ermes et al.²⁰ recommends using a mix of structured laboratory activities and free-living activities to develop more generalizable classifiers. When using both structured and unstructured data for training and cross-validation, a classification accuracy of 89.0% was achieved. When using only structured data for training and cross-validation a classification accuracy of 90.0% was achieved. These results indicate that it is feasible to use both structured and unstructured data to develop a classifier that performs similarly in cross-validation but should generalize to new data better since more diverse activities and information are modeled during development with the inclusion of the unstructured data.

Another possible avenue to maintain the general course of classifier development using inertial sensor data is to add a personalization component.²¹ Currently, model performance from within-sample cross-validation is reasonably high at the group level as suggested by Mannini et al.²² who developed a support vector machine classifier achieved average classification accuracies ranging from 91.0% (wrist) and 92.4% (ankle) in youth⁷ and 87.0% (wrist) and 94.8% (ankle) in adults²² across four activity classes (ambulation, cycling, sedentary, and other) in a within-sample cross-validation. Mannini and Intille²¹ have since shown in a separate study that the SVM classifiers developed under structured laboratory conditions ranged in performance from 80.8% to 96.5% in youth and 70.7% to 95.0% in adults across the four activity classes indicating a large amount of individual variability.²¹ In order to improve the overall classification of PA at both the individual and group levels, Mannini and Intille²¹ have suggested personalizing classifiers. This interest and motivation comes from more specific classification tasks such as identifying handwriting, speech, and gesture-based movements that have very individualized components. The idea is to take a previously developed generalizable classifier and adapt it using individual data points to improve performance across the board.²¹ The method proposed by Mannini and Intille²¹ uses a classification uncertainty parameter to determine where individual data may need to be supplied to make a classification and adapt the algorithm. When the uncertainty parameter exceeds the established threshold criteria, the algorithm requests information about the criterion activity label be supplied for that individual during that window of time and then subsequently updates the model adapting it using the new information.²¹ It is possible that a blending of the two methods discussed here may result in the greatest improvements for PA characterization.²¹ This method improved overall within-sample cross-validation accuracy 11.0% on average and improved individual classification ranging from -

2.0% to 36.0%.²¹ If a generalizable classifier is developed using a mix of structured and unstructured activity data like Ermes et al.²⁰ suggests and then the classifier is adapted using personalization as Mannini and Intille²¹ suggest, it may be possible to improve both the group and individual level performance of classifiers being developed. It is presently unknown how the contribution of using accelerometer and gyroscope data together may impact these two endeavors but it should be explored in future studies as the results of the present study suggest using both sensor signals together may improve the overall characterization of PA and should contribute to improved success of these other strategies.

4.4 Strengths

This study has a number of strengths including a large sample size, a large and diverse set of activities, and a large volume of data. There were 99 participants available for the within-sample cross-validation and 42 participants available for the out-of-sample cross validation, which is considerable. The availability of semi-structured and unstructured free-living PA data allow for a true free-living out-of-sample validation of developed models. This is a highly valuable component to this study as there are very few out-of-sample validations available in the literature, especially in youth. The volume of criterion labelled activity data collected per participant, approximately 3 hours of semi-structured and approximately 4 hours of unstructured free-living data is a notable strength of the study as both the number of participants and the duration of observation both exceed what is typically done in other studies. The inclusion of transitions adds to the value of this study as little is currently known about transition periods but they are a critical component to free-living assessment. This study provides a starting point for beginning to be able to more completely assess youth free-living PA behaviour. Additionally, this is the first study in youth to investigate the head-to-head within- and out-of-sample

performance of classifiers developed using accelerometer sensor data only, gyroscope sensor data only, and data from both sensors combined from the ActiGraph GT9X. Lastly, several machine learning model types were evaluated using the aforementioned sensor data feature sets data for multiple attachment sites which provides useful information about modeling similar data in future studies.

4.5 Limitations

While this study does add considerably to the current state of the literature, there are several limitations. The statistical features used in this study could be considered a limitation. The features used have been used previously but in studies focused on using accelerometer sensor data. There is currently little information available in the literature about which features are best for gyroscope sensor data. Future studies should consider additional or different features as part of optimizing the characterization of PA using accelerometer and gyroscope sensor data. It is possible different feature sets could yield performance gains. While working with 1-s data points provides a large dataset, classifying each 1-s data point independently may be a limitation to model performance in this study. At the 1-s level, movements that are characteristic of a specific activity such as locomotion, may be contained in a broader PA behaviour context. An example of this would be periods of walking or running while playing soccer or basketball. There are often times where participants can be engaged in one of these sporting activities and time off the ball is frequently spent either walking, jogging, or running. In the case of the developed classifiers, the 1-s of walking or running in the sporting activities could be misclassified as walking or running. Technically, the classifier is correct but based on context, the data are misclassified.

4.6 Next Steps and Practical Considerations

The results of this study should help to inform next steps for modeling wearable device data in both youth and adults. It seems there is now sufficient evidence to support using accelerometer and gyroscope together for characterizing PA in both populations. However, there are current limitations specific to the deployment of the GT9X in free-living application. Hibbing et al.²³ and LaMunion et al.²⁴ have noted through informal testing and unpublished data that the GT9X cannot collect more than 48-hrs of data when the primary accelerometer and IMU gyroscope are enabled. This window grows smaller (~24-hrs) if the full IMU is enabled at initialization. Using both sensors results in additional data accumulation of approximately ~1 GB per 24-hrs of data collection when the full IMU is engaged. This issue should be improved or resolved in the future but presently, it is a limitation to deploying the GT9X devices for multiple days of free-living assessment without requiring charging and subsequently data loss if the goal is to capture continuous 24-hr behaviour.

Future studies should consider evaluating different feature sets as a starting point for improving classifier performance. Additionally, future studies should consider, when possible, using a mix of structured, semi-structured, and unstructured data for model development to improve model generalizability. The field needs more generalizable models for using at the population level but also needs to improve the individual precision of the classifiers being developed as a more individualized approach to health and wellness emerges in healthcare for diagnostic and clinical purposes. Both aspects of PA characterization should remain targets for future studies

4.7 Conclusion

In summary, this study demonstrates that loss in PA classification accuracy can be minimized when translating from structured laboratory or semi-structured simulated free-living activities to unstructured free-living activity by using data from both the accelerometer and gyroscope sensor together.

Disclosure

This study was supported by the National Institutes of Health grant (R01HD083431). No financial support was received from any of the activity monitor manufacturers, importers, or retailers. The authors have no conflicts of interest to report.

References

1. Rowlands AV, Mirkes EM, Yates T, et al. Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent? *Med Sci Sports Exerc.* 2018;50(2):257-265.
2. Rowlands AV, Yates T, Davies M, Khunti K, Edwardson CL. Raw Accelerometer Data Analysis with GGIR R-package: Does Accelerometer Brand Matter? *Med Sci Sports Exerc.* 2016;48(10):1935-1941.
3. Strath SJ, Kaminsky LA, Ainsworth BE, et al. Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. *Circulation.* 2013;128(20):2259-2279.
4. Awais M, Mellone S, Chiari L. Physical activity classification meets daily life: Review on existing methodologies and open challenges. Paper presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
5. Dehghani A, Glatard T, Shihab E. Subject cross validation in human activity recognition. arXiv preprint arXiv:1904.02666 [cs.LG]. 2019.
6. Ahmadi MN, Brookes D, Chowdhury A, Pavey T, Trost SG. Free-living Evaluation of Laboratory-based Activity Classifiers in Preschoolers. *Med Sci Sports Exerc.* 2019.
7. Mannini A, Rosenberger M, Haskell WL, Sabatini AM, Intille SS. Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle. *Med Sci Sports Exerc.* 2017;49(4):801-812.
8. Bastian T, Maire A, Dugas J, et al. Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough. *J Appl Physiol.* 2015;118(6):716-722.

9. Gyllensten IC, Bonomi AG. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. *IEEE T Biomed Eng.* 2011;58(9):2656-2663.
10. Montoye AHK, Westgate BS, Fonley MR, Pfeiffer KA. Cross-validation and out-of-sample testing of physical activity intensity predictions with a wrist-worn accelerometer. *J Appl Physiol.* 2018;124(5):1284-1293.
11. Tong C, Tailor SA, Lane ND. Are Accelerometers for Activity Recognition a Dead-end? arXiv preprint arXiv:2001.08111 [cs.CV]. 2020.
12. Ferrari A, Micucci D, Mobilio M, Napoletano P. Human Activities Recognition Using Accelerometer and Gyroscope. Paper presented at: European Conference on Ambient Intelligence 2019.
13. Jain A, Kanhangad V. Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sens J.* 2017;18(3):1169-1177.
14. Wang A, Chen G, Yang J, Zhao S, Chang C-Y. A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sens J.* 2016;16(11):4566-4578.
15. Wu W, Dasgupta S, Ramirez EE, Peterson C, Norman GJ. Classification accuracies of physical activities using smartphone motion sensors. *J Med Internet Res.* 2012;14(5):e130.
16. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ. Fusion of smartphone motion sensors for physical activity recognition. *Sensors (Basel).* 2014;14(6):10146-10176.
17. Shoaib M, Scholten H, Havinga PJ. Towards physical activity recognition using smartphone sensors. Paper presented at: 2013 IEEE 10th International Conference on

- Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing 2013.
18. R Core Team. R: A language and environment for statistical computing. 2017;
<https://www.R-project.org/>.
 19. Trost SG, Cliff DP, Ahmadi MN, Tuc NV, Hagenbuchner M. Sensor-enabled Activity Class Recognition in Preschoolers: Hip versus Wrist Data. *Med Sci Sports Exerc.* 2018;50(3):634-641.
 20. Ermes M, Parkka J, Mantyjarvi J, Korhonen I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Trans Inf Technol Biomed.* 2008;12(1):20-26.
 21. Mannini A, Intille SS. Classifier Personalization for Activity Recognition Using Wrist Accelerometers. *IEEE J Biomed Health Inform.* 2019;23(4):1585-1594.
 22. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. *Med Sci Sports Exerc.* 2013;45(11):2193-2203.
 23. Hibbing PR LS, Kaplan AS, Crouter SE. Estimating Energy Expenditure with ActiGraph GT9X Inertial Measurement Unit. *Med Sci Sports Exerc.* 2018;50(5):1093-1102.
 24. LaMunion SR, Fitzhugh EC, Crouter SE. Challenges and Opportunities Related to the Objective Assessment of Physical Activity Within U.S. Health Surveys. *Ann Epidemiol.* 2020.

Appendices

Appendix A: Figures

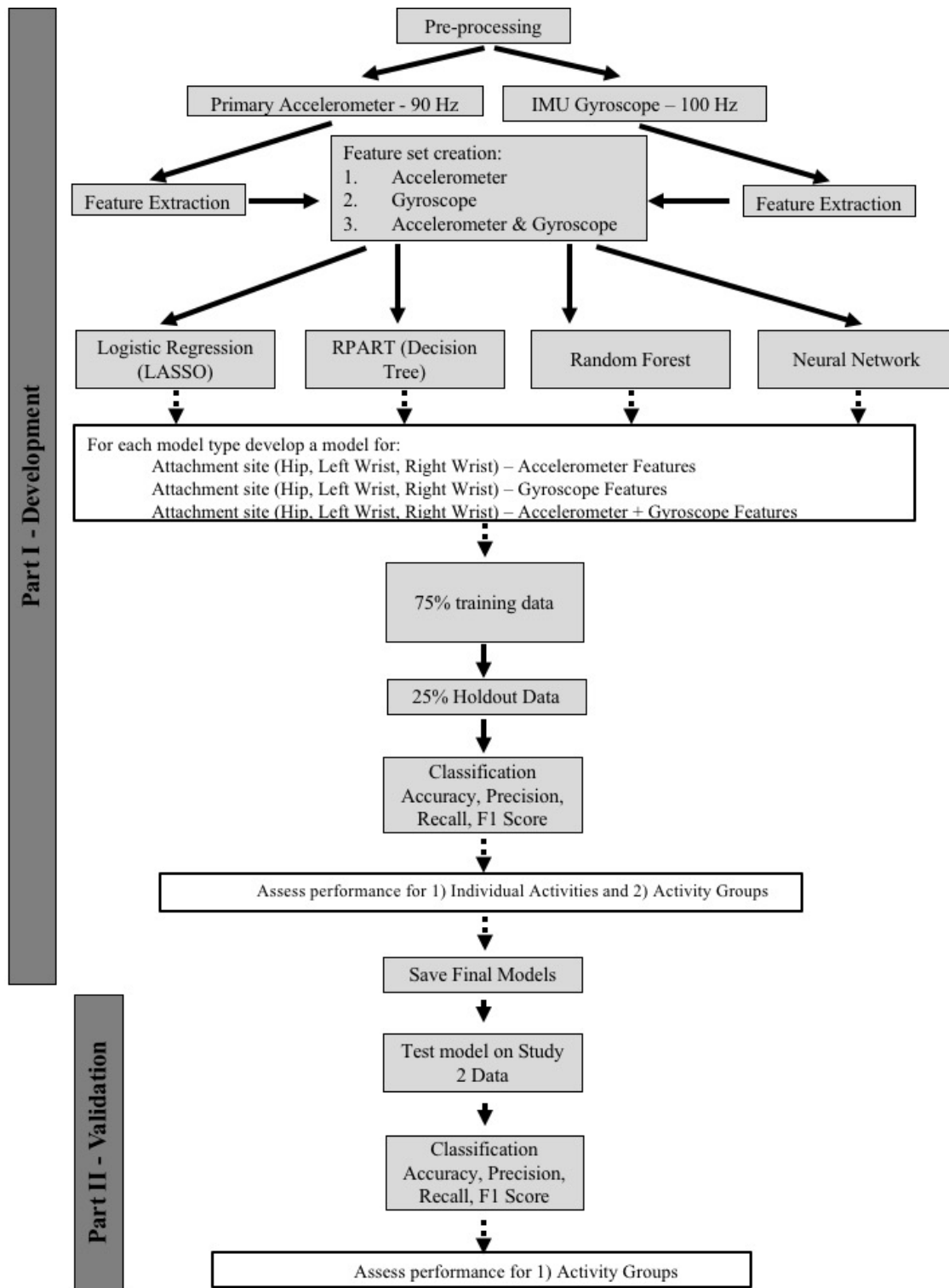


Figure 1. Project overview flow chart.

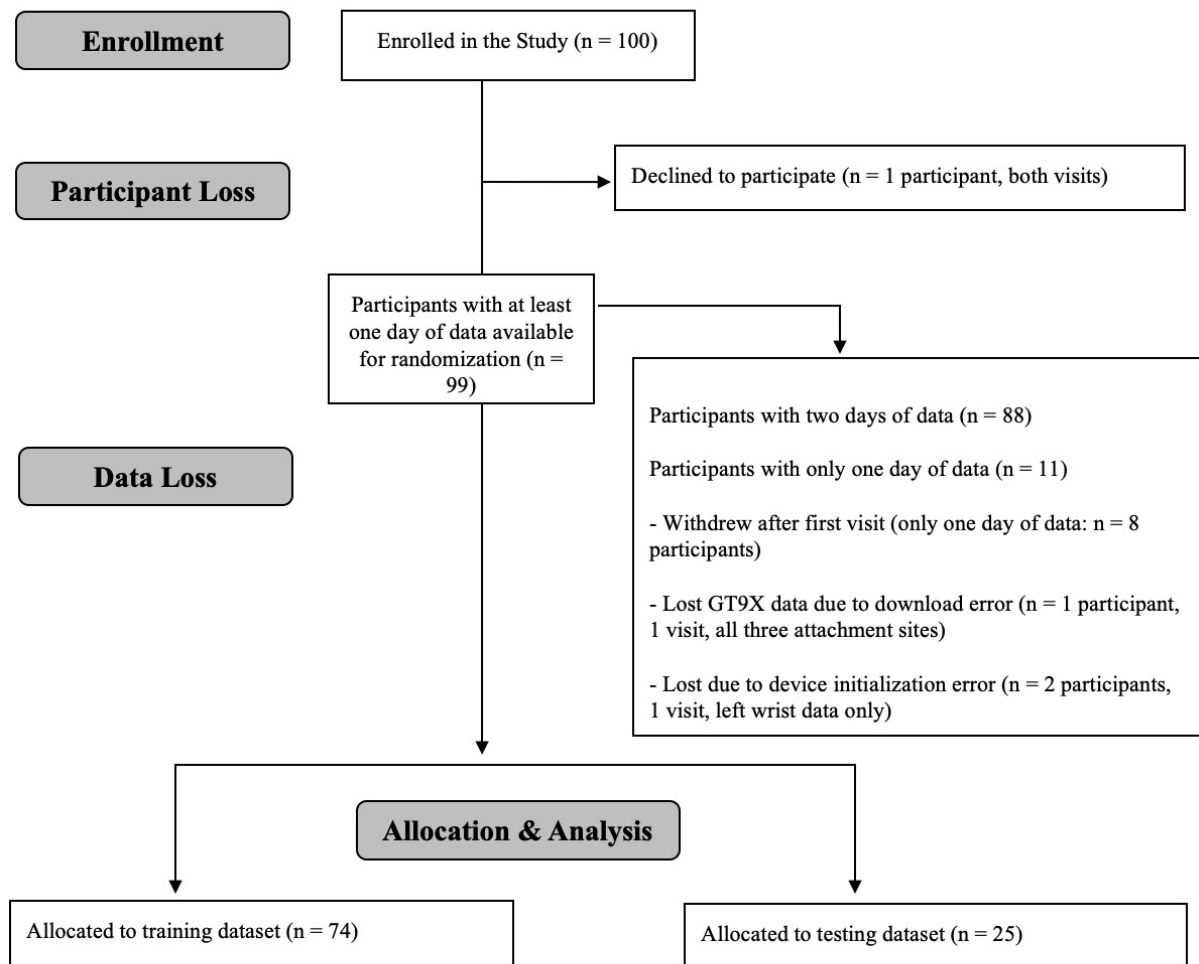


Figure 2. Study 1 participant enrollment, description of data loss, and participant allocation for analyses.

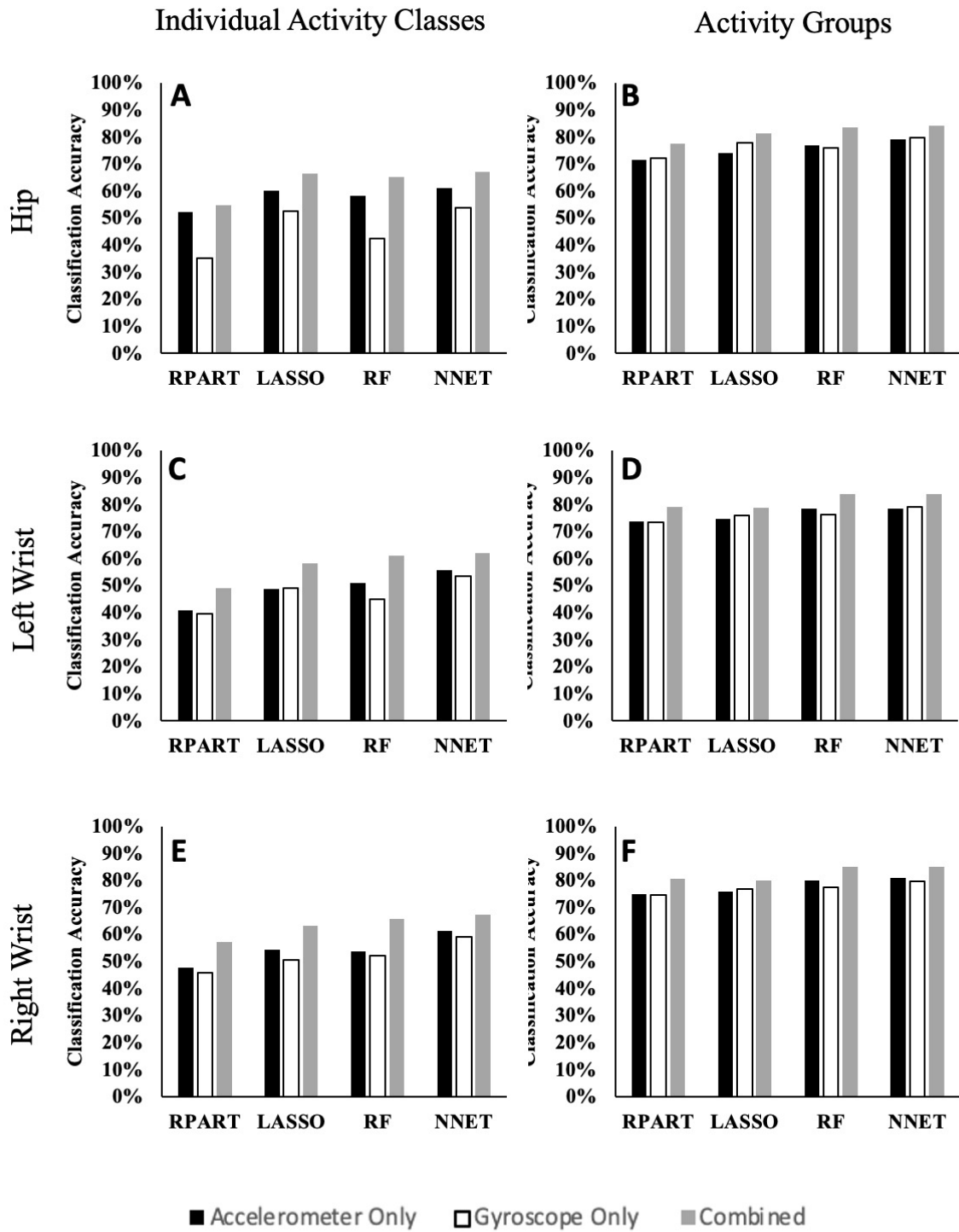


Figure 3. Overall classification accuracy results from 1) decision tree (RPART), 2) logistic regression (LASSO), 3) random forest (RF), and 4) neural network (NNET) models for hip A) individual activity classes, and B) activity groups, left wrist C) individual activity classes, and D) activity groups, and right wrist E) individual activity classes, and F) activity groups.

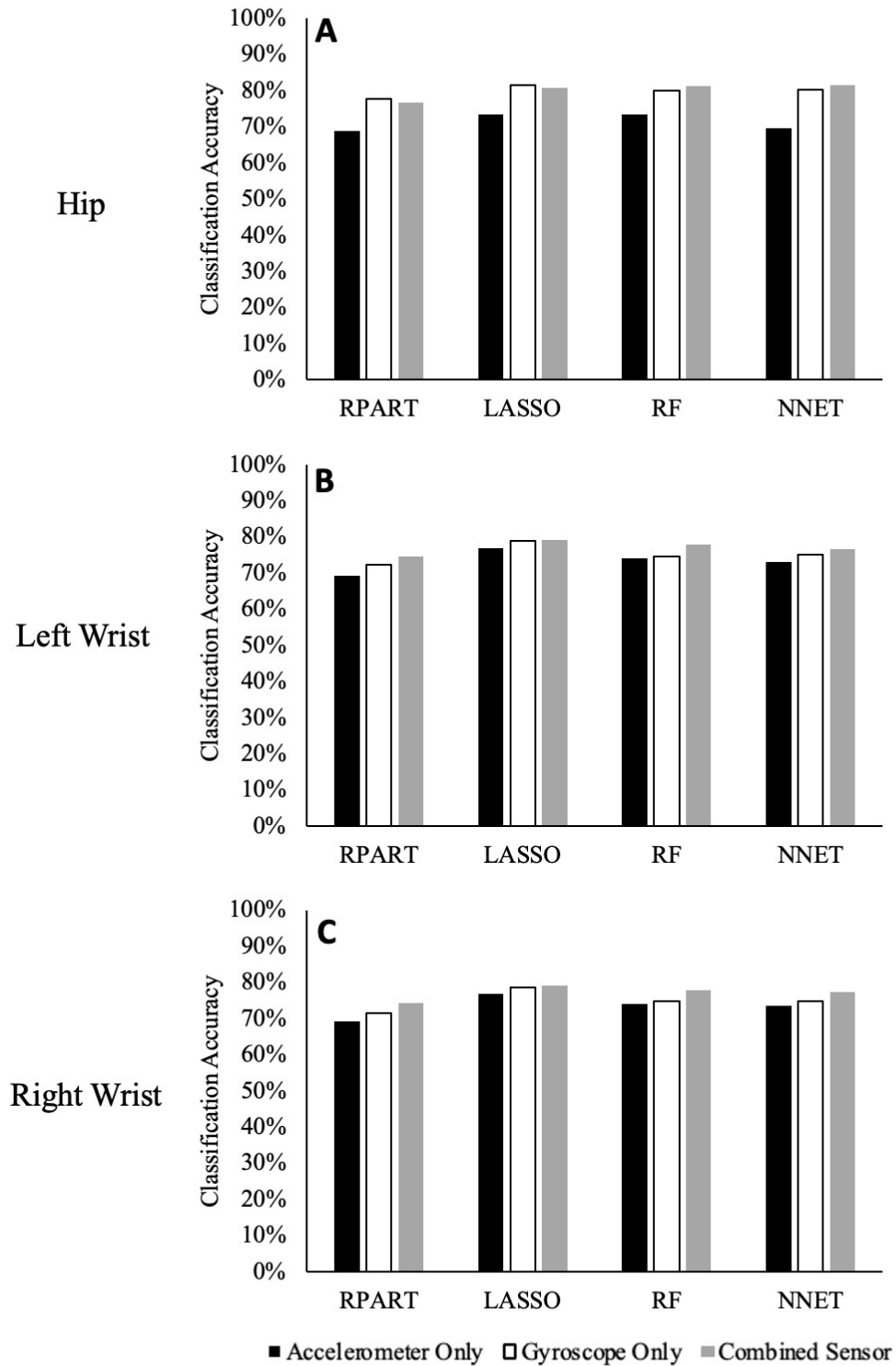


Figure 4. Out-of-sample validation model classification accuracy of LaMunion et al. (Chapter 4) six-class activity group classifiers 1) decision tree (RPART), 2) logistic regression (LASSO), 3) random forest (RF), and 4) neural network (NNET) models for A) hip, B) left wrist, C) right wrist data.

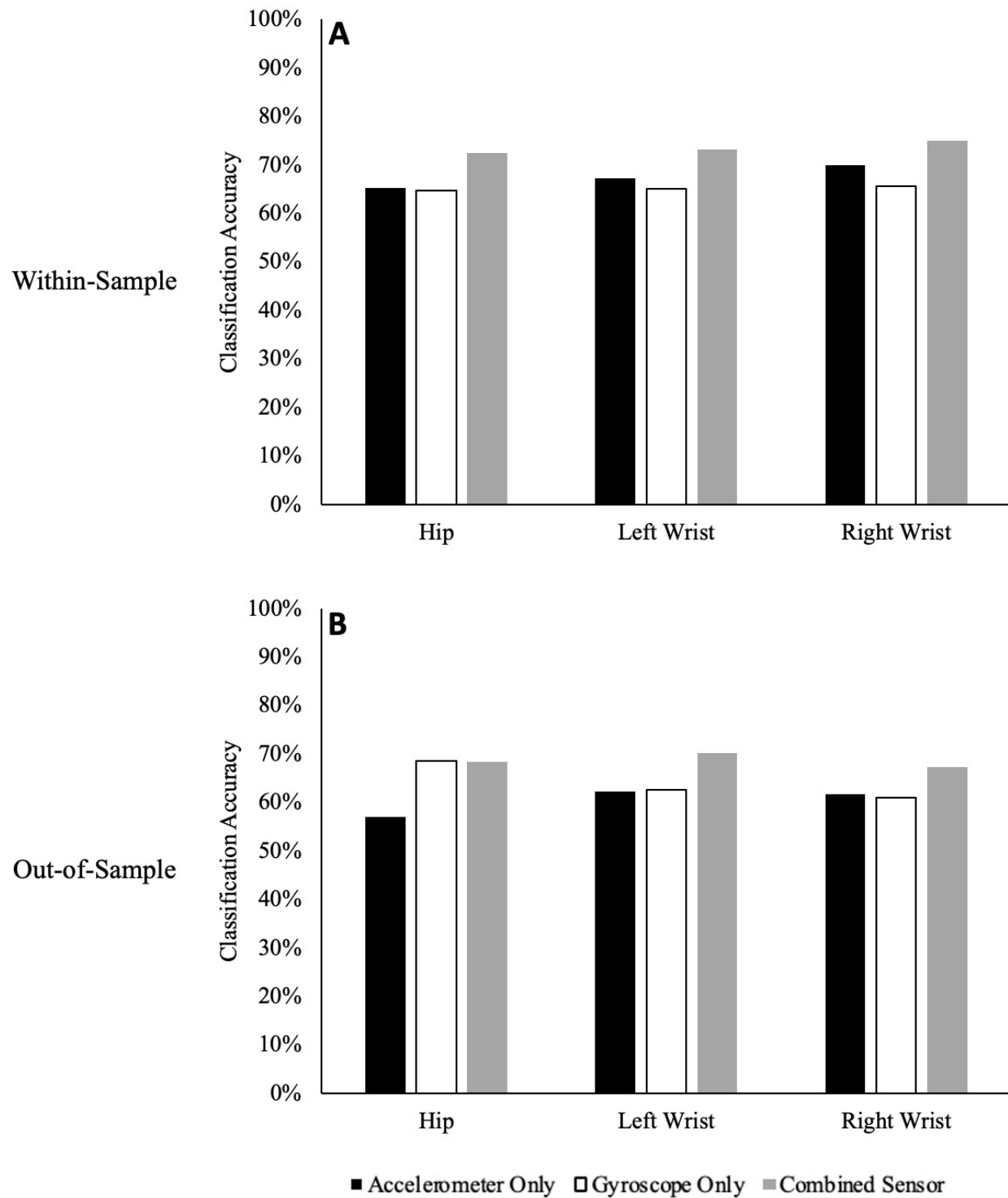


Figure 5. Overall classification accuracy results for seven-class random forest classifiers including transitions for a A) within-sample cross-validation using semi-structured simulated free-living activity data and B) out-of-sample validation using unstructured free-living data.

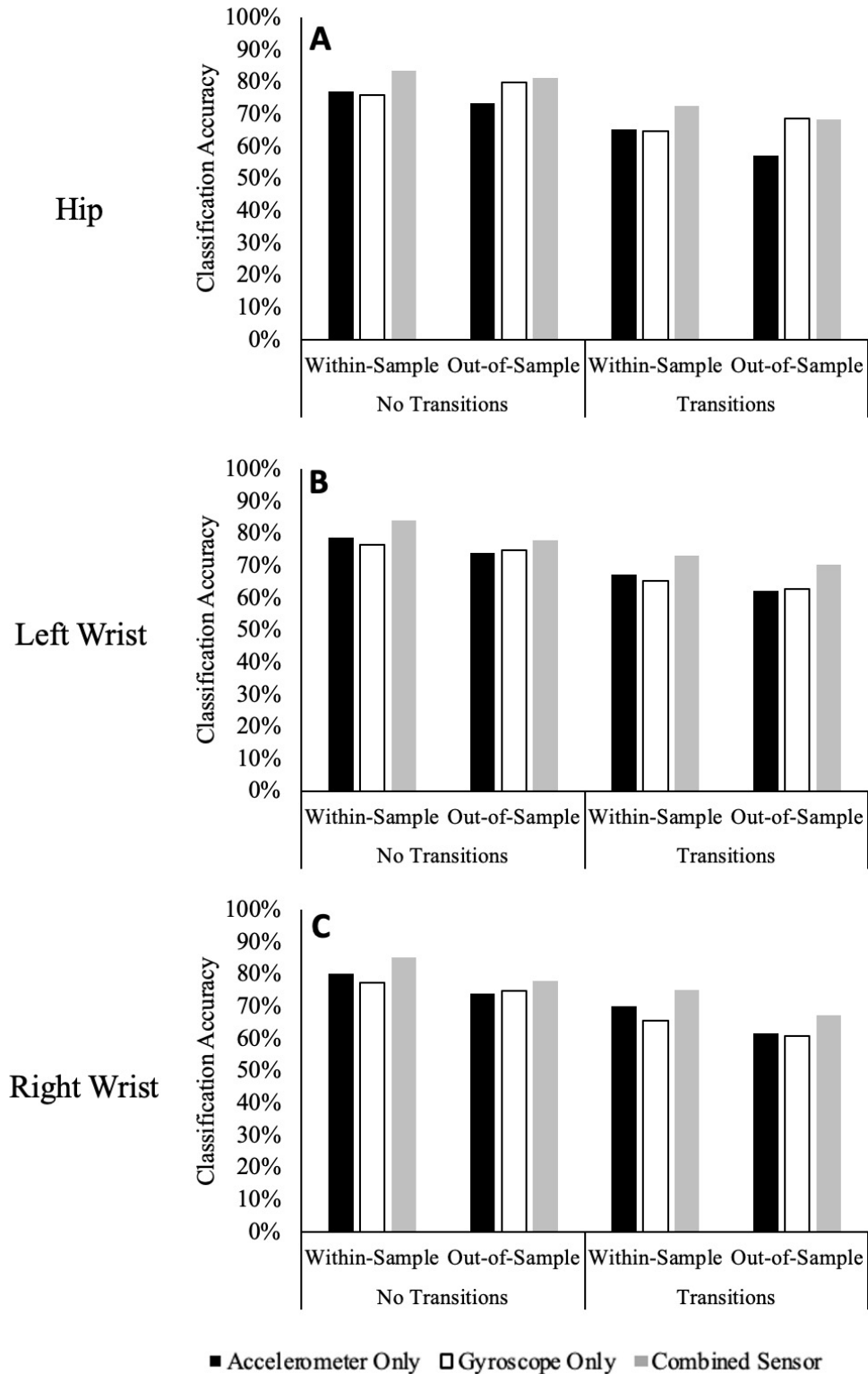


Figure 6. Overall classification accuracy results from within-sample cross-validation and out-of-sample validation of random forest models developed to classify six activity classes excluding transitions and seven activity classes including transitions using A) hip, B) left wrist, and C) right wrist data.

Appendix B: Tables

Table 1. Individual activity class labels, activity group labels, with the original activity labels and operational definitions.

Activity Group Labels (n = 6)	Individual Activity Class Labels (n = 14)	Activity (n = 16)	Operational Definition of Activity
Seated/Lying Activity (SLA)	Lying	Lying	Lying supine with arms by the side in a quiet room with no stimuli
	Computer	Computer Gaming	Sitting at a tabletop or desk playing a self-selected computer game (e.g. Tetris, Slither.io, etc.)
		Internet Browsing	Sitting at a tabletop or desk using the internet to browse websites and perform computer-based tasks (e.g. e-mail, reading, typing, etc.)
		Reclining	Sitting in an office chair in a reclined posture
	Book Reading (Book)	Book Reading	Reading a physical book
Chores	Dusting	Dusting	Wiping down a tabletop or desk surface with a paper towel and spray bottle
	Sweeping	Sweeping	Sweeping paper shreds with a broom and dustpan from a hallway floor
Continuous Walking and Running (CWR)	Overground Walking (Walk)	Slow Overground Walking	Walking at a self-selected slow and leisurely pace around the perimeter of a basketball or tennis court
		Brisk Overground Walking	Walking at a self-selected brisk pace around the perimeter of a basketball or tennis court as if they were in a hurry or running late for something
	Overground Running (Run)	Overground Running	Running at a self-selected pace that could be maintained for approximately 5 minutes around the perimeter of a basketball or tennis court
Sports	Catch	Catch	Passing a football back and forth with a partner at a self-selected comfortable distance
	Soccer	Soccer	Simulated soccer drills (e.g. one-vs-one gameplay, dribbling, passing, shooting, defending, etc.)
	Basketball	Basketball	Simulated basketball drills (e.g. one-vs-one gameplay, dribbling, passing, shooting, defending, etc.)
Wheels	Stationary Cycling	Stationary Cycling	Riding a stationary cycle ergometer at a self-selected resistance (minimum 25 Watts) and cadence (RPM)
Recreation and Leisure Time Activity (Rec_LTA)	Stair Walking	Stair Walking	Ascending and descending a staircase at a self-selected pace
	Jumping Jacks	Jumping Jacks	Continuous self-paced jumping jacks

Table 2. Study 1 dataset breakdown by number of data points in each individual activity class and activity group.

Hip	Training	Individual Activity Classes	Sweep	Dust	Lying	Reclining	Book	Computer	Walk	Run	Basketball	Soccer	Catch	Cycling	Stairs	JJ	Transition	Total
			23,542	22,630	146,462	22,107	24,105	48,492	46,731	18,894	19,425	22,180	21,421	23,717	23,418	21,441	106,276	590,841
		Activity Groups	Chores		SLA				CWR		Sports			Wheels	Rec LTA		Transition	Total
	46,172		241,166				65,625		63,026			23,717	44,859		106,276	590,841		
	Testing	Individual Activity Classes	Sweep	Dust	Lying	Reclining	Book	Computer	Walk	Run	Basketball	Soccer	Catch	Cycling	Stairs	JJ	Transition	Total
			7,570	7,670	50,584	7,941	7,336	15,625	14,340	6,158	5,891	7,308	7,316	8,017	7,826	6,624	38,866	199,072
Activity Groups		Chores		SLA				CWR		Sports			Wheels	Rec LTA		Transition	Total	
	15,240	81,486				20,498		20,515			8,017	14,450		38,866	199,072			
LW	Training	Individual Activity Classes	Sweep	Dust	Lying	Reclining	Book	Computer	Walk	Run	Basketball	Soccer	Catch	Cycling	Stairs	JJ	Transition	Total
			23,207	22,315	146,096	21,449	24,105	47,749	46,123	18,600	18,671	22,180	21,094	23,337	23,089	21,135	104,719	583,869
		Activity Groups	Chores		SLA				CWR		Sports			Wheels	Rec LTA		Transition	Total
	45,522		239,399				64,723		61,945			23,337	44,224		104,719	583,869		
	Testing	Individual Activity Classes	Sweep	Dust	Lying	Reclining	Book	Computer	Walk	Run	Basketball	Soccer	Catch	Cycling	Stairs	JJ	Transition	Total
			7,570	7,670	50,584	7,941	7,336	15,625	14,340	6,158	5,891	7,308	7,316	8,017	7,826	6,624	38,868	199,074
Activity Groups		Chores		SLA				CWR		Sports			Wheels	Rec LTA		Transition	Total	
	15,240	81,486				20,498		20,515			8,017	14,450		38,868	199,074			
RW	Training	Individual Activity Classes	Sweep	Dust	Lying	Reclining	Book	Computer	Walk	Run	Basketball	Soccer	Catch	Cycling	Stairs	JJ	Transition	Total
			23,542	22,630	146,462	22,107	24,105	48,492	46,731	18,894	19,425	22,180	21,421	23,717	23,418	21,441	106,276	590,841
		Activity Groups	Chores		SLA				CWR		Sports			Wheels	Rec LTA		Transition	Total
	46,172		241,166				65,625		63,026			23,717	44,859		106,276	590,841		
	Testing	Individual Activity Classes	Sweep	Dust	Lying	Reclining	Book	Computer	Walk	Run	Basketball	Soccer	Catch	Cycling	Stairs	JJ	Transition	Total
			7,570	7,670	50,584	7,941	7,336	15,625	14,340	6,158	5,891	7,308	7,316	8,017	7,826	6,624	38,866	199,072
Activity Groups		Chores		SLA				CWR		Sports			Wheels	Rec LTA		Transition	Total	
	15,240	81,486				20,498		20,515			8,017	14,450		38,866	199,072			

LW = Left Wrist, RW = Right Wrist, JJ = jumping jacks, SLA = seated/lying activity, CWR = continuous walking and running, Rec_LTA = recreation/leisure-time activity

Table 3. *Physical characteristics of participants for development and within-sample validation.*

	Training Sample (n = 74)	Testing Sample (n = 25)
Age [years; mean (SD)]	12.1 (3.4)	12.5 (3.7)
Female [n (%)]	39 (52.7%)	12 (48.0%)
Height [cm; mean (SD)]	150.3 (19.4)	150.1 (18.4)
Weight [kg; mean (SD)]	44.9 (18.4)	45.3 (19.8)
Handedness [n (% left-handed)]	6 (8.1%)	2 (8.0%)
BMI classification [n (%)]		
< 5th percentile [n (% underweight)]	3 (4.1%)	2 (8.0%)
5th - < 85th percentile [n (% normal weight)]	59 (79.7%)	20 (80.0%)
85th - < 95th percentile [n (% overweight)]	8 (10.8%)	1 (4.0%)
≥ 95th percentile [n (% obese)]	4 (5.4%)	2 (8.0%)
Ethnicity		
Hispanic [n (%)]	3 (4.1%)	1 (4.0%)
Race		
White [n (%)]	67 (90.5%)	23 (92.0%)
African American [n (%)]	7 (9.5%)	2 (8.0%)

Table 4. Confusion Matrix Performance Summary for Individual Activity Class for the Hip classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Basketball	21.7%	18.9%	22.3%	10.5%	0.4%	10.0%	22.4%	17.4%	26.0%	10.2%	1.2%	11.2%
Book	24.1%	14.2%	18.7%	5.9%	0.0%	8.1%	15.9%	10.1%	18.2%	17.0%	0.9%	19.3%
Catch	17.7%	25.8%	24.2%	6.6%	35.7%	39.1%	23.1%	29.3%	38.8%	29.7%	40.1%	44.4%
Computer	62.4%	22.8%	70.7%	88.8%	17.1%	92.4%	75.9%	30.1%	84.7%	75.6%	29.0%	83.0%
Cycling	18.0%	40.0%	40.5%	3.7%	54.4%	55.1%	23.2%	44.2%	52.5%	28.6%	51.9%	60.6%
Dust	29.3%	25.2%	34.0%	25.5%	37.1%	40.0%	35.9%	35.4%	45.3%	33.5%	33.0%	41.2%
Jumping Jacks	36.4%	16.6%	48.4%	41.6%	20.3%	45.3%	44.8%	15.1%	54.6%	43.5%	24.8%	57.6%
Lying	82.8%	52.4%	78.2%	98.8%	94.9%	98.6%	89.1%	65.4%	89.9%	90.6%	90.4%	94.6%
Reclining	41.2%	13.6%	36.0%	18.2%	0.0%	22.9%	40.4%	7.5%	41.4%	47.6%	1.2%	40.2%
Run	56.9%	40.2%	48.8%	54.8%	51.8%	57.8%	51.2%	47.6%	55.0%	61.2%	55.5%	52.4%
Soccer	26.2%	23.3%	29.2%	32.2%	33.4%	30.7%	32.0%	30.1%	33.3%	33.2%	40.5%	35.8%
Stairs	48.9%	26.4%	55.8%	63.0%	50.0%	68.3%	61.7%	38.3%	69.5%	64.2%	56.0%	73.4%
Sweep	28.5%	27.6%	32.1%	24.6%	28.0%	30.3%	32.1%	35.0%	45.1%	37.3%	50.4%	45.3%
Walk	48.5%	45.6%	61.9%	84.9%	77.0%	85.6%	62.5%	54.7%	75.4%	76.9%	70.7%	79.7%

RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 5. Performance Summary for Individual Activity Classes for the Hip Classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Sensitivity	38.8%	28.0%	42.9%	39.9%	35.7%	48.9%	43.6%	32.9%	52.1%	46.4%	39.0%	52.8%
Specificity	96.4%	94.9%	96.6%	96.9%	96.0%	97.4%	96.8%	95.4%	97.4%	97.1%	96.2%	97.5%
PPV	38.2%	28.4%	42.9%	43.3%	39.3%	51.9%	43.4%	33.0%	52.1%	45.7%	39.7%	52.8%
NPV	96.3%	94.8%	96.4%	97.0%	96.4%	97.5%	96.8%	95.3%	97.3%	97.0%	96.4%	97.5%
Precision	38.2%	28.4%	42.9%	43.3%	39.3%	51.9%	43.4%	33.0%	52.1%	45.7%	39.7%	52.8%
Recall	38.8%	28.0%	42.9%	39.9%	35.7%	48.9%	43.6%	32.9%	52.1%	46.4%	39.0%	52.8%
F1	38.2%	28.0%	42.6%	38.1%	36.6%	48.3%	43.2%	32.7%	51.8%	45.4%	37.2%	52.2%
Balanced Accuracy	67.6%	61.4%	69.7%	68.4%	65.8%	73.1%	70.2%	64.1%	74.7%	71.7%	67.6%	75.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 6. Confusion Matrix Performance Summary by Activity Group for the Hip classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Chores	48.9%	54.5%	62.1%	62.4%	64.1%	75.2%	64.2%	64.0%	74.6%	71.7%	71.9%	75.8%
CWR	49.8%	48.7%	58.5%	72.7%	76.0%	79.0%	57.7%	60.9%	69.2%	68.8%	72.7%	73.1%
Rec_LTA	47.5%	28.9%	51.5%	29.6%	13.8%	37.9%	52.0%	25.5%	57.4%	54.9%	35.3%	61.2%
SLA	95.6%	96.6%	97.5%	95.1%	98.1%	98.4%	97.2%	97.4%	98.9%	96.5%	97.3%	98.4%
Sports	53.8%	52.8%	61.8%	60.2%	71.4%	69.1%	68.0%	63.8%	75.7%	66.0%	69.8%	73.0%
Wheels	19.0%	46.9%	42.3%	0.8%	35.7%	36.5%	16.1%	41.2%	50.0%	21.2%	44.5%	53.4%

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 7. Performance Summary by Activity Group for the Hip Classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Sensitivity	52.4%	54.7%	62.3%	53.5%	59.9%	66.0%	59.2%	58.8%	71.0%	63.2%	65.3%	72.5%
Specificity	94.3%	94.7%	95.7%	94.3%	95.5%	96.2%	95.2%	95.4%	96.8%	95.7%	96.1%	96.8%
PPV	53.1%	54.4%	62.5%	56.1%	64.3%	72.1%	61.1%	59.7%	74.4%	64.1%	67.6%	74.8%
NPV	94.4%	94.6%	95.6%	94.8%	95.8%	96.4%	95.5%	95.3%	96.9%	95.9%	96.1%	96.9%
Precision	53.1%	54.4%	62.5%	56.1%	64.3%	72.1%	61.1%	59.7%	74.4%	64.1%	67.6%	74.8%
Recall	52.4%	54.7%	62.3%	53.5%	59.9%	66.0%	59.2%	58.8%	71.0%	63.2%	65.3%	72.5%
F1	52.6%	54.4%	62.2%	51.8%	59.6%	67.0%	59.5%	58.8%	72.0%	63.3%	65.6%	73.3%
Balanced Accuracy	73.3%	74.7%	79.0%	73.9%	77.7%	81.1%	77.2%	77.1%	83.9%	79.4%	80.7%	84.7%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 8. Confusion Matrix Performance Summary for Individual Activity Class for the Left Wrist classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Basketball	25.6%	22.4%	25.9%	24.4%	28.4%	28.3%	29.9%	24.8%	35.1%	22.5%	29.7%	26.5%
Book	29.9%	22.2%	38.0%	4.0%	5.5%	20.8%	40.7%	22.4%	49.0%	32.4%	11.3%	40.0%
Catch	31.9%	21.5%	30.5%	36.7%	27.7%	36.1%	42.2%	25.9%	46.7%	42.9%	37.2%	40.5%
Computer	38.9%	31.7%	51.0%	26.6%	21.2%	36.6%	52.1%	33.0%	60.3%	48.5%	44.2%	53.3%
Cycling	33.9%	47.6%	68.5%	8.2%	42.8%	59.9%	34.6%	54.2%	76.6%	44.7%	71.6%	77.3%
Dust	39.6%	24.4%	43.9%	27.7%	31.8%	54.5%	47.5%	40.0%	57.7%	43.2%	46.2%	53.0%
Jumping Jacks	50.2%	47.5%	56.7%	61.4%	55.9%	63.8%	58.9%	45.5%	66.3%	64.8%	51.7%	67.7%
Lying	43.6%	61.3%	53.1%	80.4%	89.8%	87.4%	58.4%	69.1%	72.2%	77.0%	87.1%	84.6%
Reclining	17.9%	10.6%	23.2%	0.0%	0.0%	0.5%	15.4%	8.9%	18.5%	1.3%	0.0%	2.8%
Run	67.7%	31.5%	71.2%	71.0%	22.8%	77.7%	70.2%	36.1%	79.4%	75.1%	47.2%	80.4%
Soccer	27.4%	21.4%	27.8%	23.3%	8.2%	26.3%	35.3%	19.6%	35.8%	30.7%	7.8%	33.6%
Stairs	35.9%	20.8%	41.7%	18.4%	20.0%	33.0%	46.4%	26.1%	52.0%	44.7%	37.7%	54.2%
Sweep	56.2%	40.7%	66.7%	55.1%	53.1%	69.8%	69.2%	50.6%	75.9%	66.8%	57.3%	74.4%
Walk	54.3%	37.4%	57.1%	72.6%	60.8%	69.6%	61.8%	43.0%	65.6%	64.8%	43.9%	61.8%

RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 9. Performance Summary for Individual Activity Classes for the Left Wrist Classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Sensitivity	39.5%	31.5%	46.8%	36.4%	33.4%	47.4%	47.3%	35.7%	56.5%	47.1%	40.9%	53.6%
Specificity	95.4%	95.2%	96.0%	95.6%	95.7%	96.5%	96.2%	95.6%	96.9%	96.4%	96.3%	97.0%
PPV	39.6%	32.5%	46.3%	39.1%	38.7%	50.0%	47.2%	36.5%	56.7%	47.9%	42.4%	53.7%
NPV	95.3%	95.2%	95.9%	95.9%	96.0%	96.7%	96.0%	95.6%	96.9%	96.4%	96.4%	97.0%
Precision	39.6%	32.5%	46.3%	39.1%	38.7%	50.0%	47.2%	36.5%	56.7%	47.9%	42.4%	53.7%
Recall	39.5%	31.5%	46.8%	36.4%	33.4%	47.4%	47.3%	35.7%	56.5%	47.1%	40.9%	53.6%
F1	38.7%	31.7%	45.9%	38.8%	35.9%	47.9%	46.4%	35.9%	55.9%	46.2%	42.7%	52.6%
Balanced Accuracy	67.5%	63.4%	71.4%	66.0%	64.6%	72.0%	71.7%	65.6%	76.7%	71.8%	68.6%	75.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 10. Confusion Matrix Performance Summary by Activity Group for the Left Wrist classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Chores	60.8%	52.6%	66.9%	60.7%	61.5%	71.1%	71.0%	61.5%	76.5%	69.5%	68.2%	75.6%
CWR	58.0%	53.2%	60.6%	64.5%	65.5%	68.7%	64.6%	57.7%	67.5%	69.2%	62.4%	70.1%
Rec_LTA	41.4%	32.8%	47.2%	30.2%	21.8%	34.1%	45.1%	31.0%	51.9%	42.8%	29.8%	52.6%
SLA	92.8%	95.7%	96.2%	96.0%	97.8%	97.8%	95.7%	96.6%	97.6%	94.6%	96.4%	96.9%
Sports	67.2%	60.3%	68.3%	70.9%	74.3%	71.7%	77.8%	68.6%	79.3%	75.0%	76.2%	77.4%
Wheels	23.8%	46.1%	65.3%	4.9%	10.7%	27.9%	17.9%	49.6%	69.8%	29.2%	65.6%	76.8%

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 11. Performance Summary by Activity Group for the Left Wrist Classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Sensitivity	57.3%	56.8%	67.4%	54.5%	55.2%	61.9%	62.0%	60.8%	73.8%	63.4%	66.4%	74.9%
Specificity	94.4%	94.5%	95.8%	94.3%	94.6%	95.4%	95.3%	95.1%	96.7%	95.3%	95.8%	96.8%
PPV	59.1%	59.4%	69.1%	63.0%	68.4%	69.7%	68.2%	64.6%	76.6%	66.8%	68.6%	75.8%
NPV	94.6%	94.7%	95.9%	95.0%	95.4%	95.9%	95.7%	95.4%	96.8%	95.6%	95.9%	96.8%
Precision	59.1%	59.4%	69.1%	63.0%	68.4%	69.7%	68.2%	64.6%	76.6%	66.8%	68.6%	75.8%
Recall	57.3%	56.8%	67.4%	54.5%	55.2%	61.9%	62.0%	60.8%	73.8%	63.4%	66.4%	74.9%
F1	57.7%	57.7%	68.1%	54.2%	55.2%	63.4%	62.6%	62.0%	74.7%	64.1%	66.6%	75.2%
Balanced Accuracy	75.9%	75.7%	81.6%	74.4%	74.9%	78.6%	78.7%	78.0%	85.2%	79.4%	81.1%	85.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 12. Confusion Matrix Performance Summary for Individual Activity Class for the Right Wrist classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Basketball	28.5%	28.5%	31.3%	22.3%	25.9%	31.7%	33.9%	35.1%	40.1%	21.9%	27.1%	29.1%
Book	37.2%	21.1%	43.1%	7.4%	7.0%	28.1%	43.3%	16.5%	53.8%	44.0%	15.2%	47.1%
Catch	36.2%	27.2%	36.4%	56.3%	48.4%	51.2%	44.8%	31.0%	47.5%	57.1%	52.9%	55.0%
Computer	60.6%	49.4%	68.8%	52.1%	24.5%	65.0%	74.4%	62.2%	78.8%	70.9%	72.4%	80.8%
Cycling	35.1%	63.9%	68.4%	14.6%	42.3%	55.8%	43.2%	60.1%	73.4%	49.5%	72.4%	73.9%
Dust	33.1%	23.1%	39.7%	20.9%	30.3%	36.4%	41.9%	30.6%	50.5%	39.7%	33.6%	44.3%
Jumping Jacks	48.2%	45.2%	53.8%	59.1%	56.8%	64.4%	58.6%	45.9%	66.2%	63.1%	59.2%	68.3%
Lying	56.8%	70.1%	67.6%	84.0%	90.7%	88.8%	57.7%	81.6%	75.8%	81.5%	91.0%	86.7%
Reclining	23.3%	10.8%	23.8%	0.3%	0.1%	4.0%	17.8%	6.5%	26.2%	1.8%	0.1%	8.3%
Run	67.6%	40.4%	75.8%	75.6%	30.8%	81.3%	74.1%	46.8%	81.6%	73.5%	57.7%	80.3%
Soccer	28.7%	20.2%	36.5%	23.7%	12.2%	35.5%	35.5%	21.3%	44.1%	34.9%	16.3%	41.9%
Stairs	38.0%	24.6%	45.3%	18.0%	12.3%	32.8%	48.6%	28.4%	56.3%	42.3%	36.2%	51.8%
Sweep	39.3%	41.1%	64.1%	61.5%	45.8%	69.9%	60.0%	47.2%	74.3%	68.0%	51.5%	71.7%
Walk	58.8%	36.5%	66.1%	80.6%	63.7%	77.2%	67.1%	44.1%	71.2%	73.2%	47.7%	72.7%

RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 13. Performance Summary for Individual Activity Classes for the Right Wrist Classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Sensitivity	42.3%	35.9%	51.5%	41.2%	35.1%	51.6%	50.1%	39.8%	60.0%	51.5%	45.2%	58.0%
Specificity	95.9%	95.8%	96.7%	96.1%	95.8%	96.9%	96.4%	96.2%	97.3%	96.9%	96.7%	97.4%
PPV	42.6%	36.1%	51.3%	44.7%	38.9%	55.0%	49.3%	39.9%	59.2%	51.1%	43.9%	58.3%
NPV	95.8%	95.7%	96.6%	96.4%	96.2%	97.1%	96.3%	96.2%	97.2%	96.9%	96.8%	97.4%
Precision	42.6%	36.1%	51.3%	44.7%	38.9%	55.0%	49.3%	39.9%	59.2%	51.1%	43.9%	58.3%
Recall	42.3%	35.9%	51.5%	41.2%	35.1%	51.6%	50.1%	39.8%	60.0%	51.5%	45.2%	58.0%
F1	41.7%	35.8%	50.9%	40.4%	35.0%	52.2%	48.9%	39.6%	59.2%	50.4%	43.9%	57.4%
Balanced Accuracy	69.1%	65.8%	74.1%	68.6%	65.5%	74.3%	73.2%	68.0%	78.7%	74.2%	71.0%	77.7%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 14. Confusion Matrix Performance Summary by Activity Group for the Right Wrist classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Chores	55.0%	50.4%	62.8%	45.6%	60.6%	65.4%	66.5%	58.5%	74.3%	67.1%	62.5%	72.5%
CWR	61.7%	55.8%	67.3%	76.3%	68.7%	74.8%	68.8%	59.5%	73.5%	72.4%	63.7%	74.3%
Rec_LTA	43.4%	32.6%	49.8%	29.6%	20.9%	36.3%	47.5%	30.8%	55.0%	48.5%	34.8%	57.3%
SLA	92.1%	95.5%	96.5%	97.3%	98.2%	98.1%	96.0%	97.0%	97.9%	95.6%	96.8%	97.5%
Sports	73.9%	64.2%	72.9%	75.0%	77.5%	75.9%	80.6%	73.3%	81.8%	78.8%	77.0%	79.7%
Wheels	32.1%	58.9%	65.7%	1.7%	11.1%	29.8%	31.9%	55.1%	67.9%	44.1%	64.8%	73.2%

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 15. Performance Summary by Activity Group for the Right Wrist Classifiers by Feature Set.

	RPART			LASSO			RF			NNET		
	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined	Accelerometer	Gyroscope	Combined
Sensitivity	59.7%	59.6%	69.2%	54.3%	56.2%	63.4%	65.2%	62.4%	75.1%	67.7%	66.6%	75.7%
Specificity	94.6%	94.8%	96.1%	94.5%	94.9%	95.7%	95.7%	95.4%	96.9%	95.9%	95.8%	97.0%
PPV	60.3%	60.6%	70.4%	57.1%	68.1%	70.1%	69.8%	65.2%	77.7%	70.4%	68.7%	77.1%
NPV	94.7%	94.9%	96.2%	95.4%	95.6%	96.2%	96.0%	95.6%	97.1%	96.1%	96.0%	97.0%
Precision	60.3%	60.6%	70.4%	57.1%	68.1%	70.1%	69.8%	65.2%	77.7%	70.4%	68.7%	77.1%
Recall	59.7%	59.6%	69.2%	54.3%	56.2%	63.4%	65.2%	62.4%	75.1%	67.7%	66.6%	75.7%
F1	59.9%	60.0%	69.7%	53.8%	55.9%	64.8%	66.4%	63.2%	76.1%	68.6%	67.1%	76.3%
Balanced Accuracy	77.2%	77.2%	82.7%	74.4%	75.5%	79.5%	80.4%	78.9%	86.0%	81.8%	81.2%	86.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; RPART = decision tree; LASSO = logistic regression; RF = random forest; NNET = neural network; the color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type

Table 16. Individual Class Confusion Matrix Results for the Hip Accelerometer Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1278	17	982	29	104	200	708	0	10	341	1448	521	204	1201
Book	9	1766	103	1722	544	137	1	2790	2216	0	9	14	115	96
Catch	611	104	1293	132	291	483	160	43	37	43	1010	945	872	486
Computer	22	2447	196	9746	1798	401	25	569	1292	3	19	138	494	172
Cycling	79	195	445	701	1446	1233	35	24	119	21	223	109	1000	888
Dust	236	368	712	1084	1427	2247	72	78	98	25	241	299	1491	454
Jumping Jacks	466	2	178	1	44	63	2411	28	0	1112	505	286	26	762
Lying	0	119	12	25	4	18	0	41883	738	0	0	9	23	12
Reclining	4	2085	34	1801	555	63	3	4900	3271	1	2	17	33	19
Run	76	7	7	18	1	21	1536	0	7	3501	177	6	1	381
Soccer	1423	10	824	14	63	171	604	11	21	386	1918	501	391	726
Stairs	883	49	1131	44	151	385	499	200	34	45	828	3827	467	1309
Sweep	206	142	764	282	694	1950	59	52	79	30	240	432	2159	879
Walk	598	25	635	26	895	298	511	6	19	650	688	722	294	6955

Table 17. Individual Class Performance Summaries for the Hip Accelerometer Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	21.7%	24.1%	17.7%	62.4%	18.0%	29.3%	36.4%	82.8%	41.2%	56.9%	26.2%	48.9%	28.5%	48.5%
Specificity	96.3%	94.9%	96.6%	94.8%	96.7%	95.7%	97.7%	99.1%	93.7%	98.5%	96.6%	96.0%	96.2%	96.3%
PPV	18.1%	18.5%	19.9%	56.3%	22.2%	25.4%	41.0%	97.8%	25.6%	61.0%	27.2%	38.8%	27.1%	56.4%
NPV	97.0%	96.3%	96.1%	95.9%	95.7%	96.4%	97.3%	92.6%	96.8%	98.3%	96.5%	97.3%	96.4%	95.0%
Precision	18.1%	18.5%	19.9%	56.3%	22.2%	25.4%	41.0%	97.8%	25.6%	61.0%	27.2%	38.8%	27.1%	56.4%
Recall	21.7%	24.1%	17.7%	62.4%	18.0%	29.3%	36.4%	82.8%	41.2%	56.9%	26.2%	48.9%	28.5%	48.5%
F1	19.8%	21.0%	18.7%	59.2%	19.9%	27.2%	38.6%	89.7%	31.6%	58.9%	26.7%	43.3%	27.8%	52.2%
Balanced Accuracy	59.0%	59.5%	57.1%	78.6%	57.4%	62.5%	67.1%	91.0%	67.5%	77.7%	61.4%	72.5%	62.4%	72.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 18. Individual Class Confusion Matrix Results for the Hip Gyroscope Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1114	47	1031	49	182	430	638	59	58	259	1245	1118	553	582
Book	13	1045	67	2410	46	137	8	5862	1200	9	16	5	197	30
Catch	794	74	1884	122	269	1112	388	111	154	111	1030	747	932	307
Computer	16	1944	21	3559	86	166	24	13285	1797	2	8	8	68	19
Cycling	117	40	277	138	3206	535	324	38	182	321	114	195	723	1859
Dust	280	134	759	351	1441	1930	136	110	520	34	304	556	1705	628
Jumping Jacks	437	9	541	6	270	120	1097	16	31	1212	945	437	128	1193
Lying	19	3015	54	6899	14	63	30	26531	2574	3	17	8	97	20
Reclining	19	921	23	1919	154	237	21	4379	1078	1	20	10	120	41
Run	121	0	54	8	58	28	1593	3	4	2475	275	132	25	770
Soccer	1438	30	1157	41	197	411	896	71	96	365	1701	1085	387	664
Stairs	839	14	518	3	142	293	352	13	21	130	920	2064	279	700
Sweep	374	57	690	103	782	1830	133	82	192	15	338	626	2090	982
Walk	310	6	240	17	1170	378	984	24	34	1221	375	835	266	6545

Table 19. Individual Class Performance Summaries for the Hip Gyroscope Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	18.9%	14.2%	25.8%	22.8%	40.0%	25.2%	16.6%	52.4%	13.6%	40.2%	23.3%	26.4%	27.6%	45.6%
Specificity	95.9%	93.5%	96.0%	87.9%	96.8%	95.4%	96.5%	88.3%	94.8%	98.0%	95.5%	97.2%	95.9%	96.0%
PPV	15.1%	9.5%	23.4%	16.9%	39.7%	21.7%	17.0%	67.4%	12.1%	44.6%	19.9%	32.8%	25.2%	52.8%
NPV	96.9%	95.8%	96.4%	91.3%	96.8%	96.2%	96.4%	80.1%	95.5%	97.6%	96.3%	96.3%	96.4%	94.7%
Precision	15.1%	9.5%	23.4%	16.9%	39.7%	21.7%	17.0%	67.4%	12.1%	44.6%	19.9%	32.8%	25.2%	52.8%
Recall	18.9%	14.2%	25.8%	22.8%	40.0%	25.2%	16.6%	52.4%	13.6%	40.2%	23.3%	26.4%	27.6%	45.6%
F1	16.8%	11.4%	24.5%	19.4%	39.9%	23.3%	16.8%	59.0%	12.8%	42.3%	21.5%	29.2%	26.3%	48.9%
Balanced Accuracy	57.4%	53.9%	60.9%	55.4%	68.4%	60.3%	56.5%	70.4%	54.2%	69.1%	59.4%	61.8%	61.8%	70.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 20. Individual Class Confusion Matrix Results for the Hip Combined Sensor Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1315	25	1224	49	143	380	521	16	19	223	1734	716	315	517
Book	7	1375	23	2267	92	76	5	4264	2242	0	3	9	21	7
Catch	781	68	1770	92	368	786	348	36	58	86	978	784	766	292
Computer	14	3501	170	11045	235	259	33	732	1696	1	24	10	146	23
Cycling	95	51	226	91	3245	790	55	28	115	40	85	174	664	1435
Dust	224	84	930	317	1480	2604	54	32	93	28	273	289	2309	431
Jumping Jacks	384	2	144	126	173	51	3207	293	0	1962	509	137	24	676
Lying	0	261	5	116	14	64	0	39580	760	0	1	1	14	10
Reclining	13	1865	37	1345	477	60	6	5552	2857	6	3	3	30	0
Run	97	6	41	7	25	19	1223	7	14	3006	213	30	3	292
Soccer	1490	14	1078	33	245	176	589	13	19	323	2137	499	360	719
Stairs	756	11	661	14	83	173	203	13	16	31	638	4364	231	659
Sweep	219	57	686	101	757	1966	33	14	31	22	213	378	2431	407
Walk	496	16	321	22	680	266	347	4	21	430	497	432	256	8872

Table 21. Individual Class Performance Summaries for the Hip Combined Sensor Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	22.3%	18.7%	24.2%	70.7%	40.5%	34.0%	48.4%	78.2%	36.0%	48.8%	29.2%	55.8%	32.1%	61.9%
Specificity	96.2%	94.1%	96.4%	95.3%	97.5%	95.7%	97.1%	98.9%	93.8%	98.7%	96.4%	97.7%	96.8%	97.4%
PPV	18.3%	13.2%	24.5%	61.7%	45.7%	28.5%	41.7%	96.9%	23.3%	60.3%	27.8%	55.6%	33.2%	70.1%
NPV	97.0%	96.0%	96.4%	96.8%	96.9%	96.6%	97.8%	90.8%	96.6%	98.0%	96.6%	97.7%	96.6%	96.3%
Precision	18.3%	13.2%	24.5%	61.7%	45.7%	28.5%	41.7%	96.9%	23.3%	60.3%	27.8%	55.6%	33.2%	70.1%
Recall	22.3%	18.7%	24.2%	70.7%	40.5%	34.0%	48.4%	78.2%	36.0%	48.8%	29.2%	55.8%	32.1%	61.9%
F1	20.1%	15.5%	24.4%	65.9%	42.9%	31.0%	44.8%	86.6%	28.3%	54.0%	28.5%	55.7%	32.7%	65.7%
Balanced Accuracy	59.3%	56.4%	60.3%	83.0%	69.0%	64.8%	72.7%	88.6%	64.9%	73.8%	62.8%	76.7%	64.5%	79.6%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 22. Individual Class Confusion Matrix Results for the Hip Accelerometer Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	619	0	452	6	62	76	194	2	0	38	591	263	61	105
Book	0	434	3	362	33	30	0	2	362	0	0	0	3	0
Catch	311	11	484	11	31	63	46	37	30	23	471	117	142	51
Computer	10	4543	323	13876	3947	2152	54	33	3181	5	30	11	1949	53
Cycling	22	2	53	113	294	193	10	41	11	1	29	9	219	223
Dust	131	171	443	190	950	1954	29	14	58	28	82	172	1343	252
Jumping Jacks	605	0	78	0	62	70	2755	0	0	1577	803	156	17	617
Lying	0	1104	1	293	159	0	0	50001	2643	0	0	0	3	0
Reclining	23	825	38	323	362	159	0	55	1448	0	0	1	5	33
Run	68	0	0	0	1	9	1309	0	0	3374	167	13	0	29
Soccer	1472	0	721	1	16	64	478	8	0	253	2353	223	158	241
Stairs	974	56	1363	82	116	136	999	44	43	26	656	4933	130	413
Sweep	62	120	510	182	274	1286	26	301	68	15	55	183	1863	153
Walk	1594	70	2847	186	1710	1478	724	46	97	818	2071	1745	1677	12170

Table 23. Individual Class Performance Summaries for the Hip Accelerometer Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	10.5%	5.9%	6.6%	88.8%	3.7%	25.5%	41.6%	98.8%	18.2%	54.8%	32.2%	63.0%	24.6%	84.9%
Specificity	98.8%	99.5%	99.1%	88.7%	99.4%	97.5%	97.4%	96.2%	98.8%	99.0%	97.6%	96.7%	97.9%	89.7%
PPV	25.1%	35.3%	26.5%	46.0%	24.1%	33.6%	40.9%	92.2%	44.3%	67.9%	39.3%	49.5%	36.5%	44.7%
NPV	96.7%	95.7%	95.7%	98.7%	95.1%	96.3%	97.5%	99.5%	95.9%	98.2%	96.8%	98.1%	96.3%	98.4%
Precision	25.1%	35.3%	26.5%	46.0%	24.1%	33.6%	40.9%	92.2%	44.3%	67.9%	39.3%	49.5%	36.5%	44.7%
Recall	10.5%	5.9%	6.6%	88.8%	3.7%	25.5%	41.6%	98.8%	18.2%	54.8%	32.2%	63.0%	24.6%	84.9%
F1	14.8%	10.1%	10.6%	60.6%	6.4%	29.0%	41.2%	95.4%	25.8%	60.6%	35.4%	55.4%	29.4%	58.5%
Balanced Accuracy	54.7%	52.7%	52.9%	88.8%	51.5%	61.5%	69.5%	97.5%	58.5%	76.9%	64.9%	79.9%	61.2%	87.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 24. Individual Class Confusion Matrix Results for the Hip Gyroscope Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	22	0	32	0	4	3	50	0	0	4	25	12	6	1
Book	0	1	0	1	0	0	0	0	2	0	0	0	0	0
Catch	957	98	2612	128	369	518	698	233	199	116	1183	457	559	216
Computer	40	1239	108	2678	200	724	31	2107	1565	11	41	10	196	61
Cycling	25	12	72	37	4360	578	22	5	88	4	27	44	710	668
Dust	252	96	1321	171	340	2847	103	146	309	34	346	294	2544	196
Jumping Jacks	235	4	68	0	66	28	1345	1	3	935	491	493	28	357
Lying	2	5824	2	12522	268	105	22	47985	5547	0	0	0	5	0
Reclining	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Run	88	0	0	0	20	6	1595	0	0	3192	104	33	0	663
Soccer	1727	21	1020	1	92	66	726	2	1	282	2440	334	139	127
Stairs	1704	8	1295	4	106	266	439	29	21	106	1765	3910	520	765
Sweep	132	20	425	66	463	1689	33	41	169	11	144	350	2123	248
Walk	707	13	361	17	1729	840	1560	35	37	1463	742	1889	740	11038

Table 25. Individual Class Performance Summaries for the Hip Gyroscope Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	0.4%	0.0%	35.7%	17.1%	54.4%	37.1%	20.3%	94.9%	0.0%	51.8%	33.4%	50.0%	28.0%	77.0%
Specificity	99.9%	100.0%	96.3%	95.6%	98.5%	96.0%	98.2%	77.8%	100.0%	98.4%	97.0%	95.4%	97.5%	93.1%
PPV	13.8%	25.0%	31.3%	29.7%	65.5%	31.6%	33.2%	66.4%	NA	56.0%	35.0%	35.7%	35.9%	52.1%
NPV	96.3%	95.4%	96.9%	91.4%	97.6%	96.8%	96.6%	97.0%	95.0%	98.1%	96.8%	97.4%	96.5%	97.6%
Precision	13.8%	25.0%	31.3%	29.7%	65.5%	31.6%	33.2%	66.4%	NA	56.0%	35.0%	35.7%	35.9%	52.1%
Recall	0.4%	0.0%	35.7%	17.1%	54.4%	37.1%	20.3%	94.9%	0.0%	51.8%	33.4%	50.0%	28.0%	77.0%
F1	0.7%	0.0%	33.4%	21.7%	59.4%	34.2%	25.2%	78.1%	NA	53.8%	34.2%	41.7%	31.5%	62.2%
Balanced Accuracy	50.1%	50.0%	66.0%	56.4%	76.4%	66.5%	59.3%	86.3%	50.0%	75.1%	65.2%	72.7%	62.8%	85.0%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 26. Individual Class Confusion Matrix Results for the Hip Combined Sensor Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	587	0	274	4	25	86	122	3	0	30	412	256	55	69
Book	0	594	2	253	27	6	0	30	583	0	0	2	1	2
Catch	1206	72	2861	109	421	322	327	61	95	85	1695	362	608	319
Computer	41	4688	126	14442	357	698	47	369	2683	3	46	7	198	53
Cycling	39	8	64	38	4417	490	9	0	72	0	32	11	746	368
Dust	272	88	1103	169	535	3068	49	28	66	40	332	189	2744	179
Jumping Jacks	504	0	48	0	34	91	3000	0	0	1446	682	174	13	321
Lying	0	794	0	157	35	0	0	49895	2457	0	0	0	1	0
Reclining	10	983	5	355	355	186	0	99	1822	0	0	2	5	4
Run	90	0	0	0	5	7	1278	0	0	3557	170	17	0	39
Soccer	1457	0	654	1	21	50	389	5	0	225	2245	172	120	242
Stairs	837	35	1000	16	19	133	493	38	20	17	614	5347	213	335
Sweep	68	57	601	47	254	1792	16	28	105	13	80	196	2296	131
Walk	780	17	578	34	1512	741	894	28	38	742	1000	1091	570	12278

Table 27. Individual Class Performance Summaries for the Hip Combined Sensor Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	10.0%	8.1%	39.1%	92.4%	55.1%	40.0%	45.3%	98.6%	22.9%	57.8%	30.7%	68.3%	30.3%	85.6%
Specificity	99.1%	99.4%	96.3%	93.6%	98.8%	96.2%	97.8%	96.9%	98.7%	99.0%	97.8%	97.5%	97.8%	94.5%
PPV	30.5%	39.6%	33.5%	60.8%	70.2%	34.6%	47.5%	93.5%	47.6%	68.9%	40.2%	58.6%	40.4%	60.5%
NPV	96.6%	95.8%	97.1%	99.1%	97.7%	97.0%	97.6%	99.4%	96.1%	98.3%	96.7%	98.4%	96.6%	98.5%
Precision	30.5%	39.6%	33.5%	60.8%	70.2%	34.6%	47.5%	93.5%	47.6%	68.9%	40.2%	58.6%	40.4%	60.5%
Recall	10.0%	8.1%	39.1%	92.4%	55.1%	40.0%	45.3%	98.6%	22.9%	57.8%	30.7%	68.3%	30.3%	85.6%
F1	15.0%	13.4%	36.1%	73.3%	61.7%	37.1%	46.4%	96.0%	31.0%	62.8%	34.8%	63.1%	34.6%	70.9%
Balanced Accuracy	54.5%	53.8%	67.7%	93.0%	76.9%	68.1%	71.6%	97.7%	60.8%	78.4%	64.3%	82.9%	64.1%	90.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 28. Individual Class Confusion Matrix Results for the Hip Accelerometer Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1320	5	785	9	74	124	521	2	4	199	1433	426	172	556
Book	1	1165	48	1007	615	122	0	2115	1786	0	0	3	26	3
Catch	700	37	1692	92	157	372	101	28	42	39	1041	880	814	401
Computer	9	3747	160	11859	2011	351	29	379	1744	2	22	39	492	278
Cycling	107	33	352	449	1860	1068	37	12	75	32	215	112	775	1348
Dust	160	278	794	871	1393	2756	57	27	116	29	142	136	1943	311
Jumping Jacks	382	0	117	0	80	59	2967	0	0	1781	468	162	7	703
Lying	16	35	0	14	14	21	0	45078	867	0	0	0	0	1
Reclining	7	1866	6	1017	506	72	3	2884	3207	0	3	12	12	0
Run	50	0	0	0	7	15	1349	0	0	3154	77	5	0	62
Soccer	1523	2	934	5	50	123	529	0	8	324	2338	334	380	414
Stairs	822	52	1186	38	114	138	569	27	13	14	602	4828	207	714
Sweep	168	97	669	227	417	2083	30	26	71	28	192	268	2432	586
Walk	626	19	573	37	719	366	432	6	8	556	775	621	310	8963

Table 29. Individual Class Performance Summaries for the Hip Accelerometer Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	22.4%	15.9%	23.1%	75.9%	23.2%	35.9%	44.8%	89.1%	40.4%	51.2%	32.0%	61.7%	32.1%	62.5%
Specificity	97.2%	96.3%	96.9%	93.6%	97.0%	95.9%	97.6%	99.1%	95.8%	99.0%	97.0%	97.0%	96.8%	96.5%
PPV	23.4%	16.9%	26.5%	56.1%	28.7%	30.6%	44.1%	97.9%	33.4%	66.8%	33.6%	51.8%	33.3%	64.0%
NPV	97.0%	96.0%	96.3%	97.3%	96.0%	96.7%	97.6%	95.2%	96.9%	98.1%	96.8%	98.0%	96.6%	96.3%
Precision	23.4%	16.9%	26.5%	56.1%	28.7%	30.6%	44.1%	97.9%	33.4%	66.8%	33.6%	51.8%	33.3%	64.0%
Recall	22.4%	15.9%	23.1%	75.9%	23.2%	35.9%	44.8%	89.1%	40.4%	51.2%	32.0%	61.7%	32.1%	62.5%
F1	22.9%	16.4%	24.7%	64.5%	25.7%	33.0%	44.4%	93.3%	36.6%	58.0%	32.8%	56.3%	32.7%	63.2%
Balanced Accuracy	59.8%	56.1%	60.0%	84.7%	60.1%	65.9%	71.2%	94.1%	68.1%	75.1%	64.5%	79.4%	64.5%	79.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 30. Individual Class Confusion Matrix Results for the Hip Gyroscope Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1025	21	761	31	115	189	322	42	41	139	986	717	337	267
Book	4	739	12	1345	39	118	2	3233	945	1	8	2	92	9
Catch	832	72	2146	124	333	851	585	112	176	85	1201	553	808	243
Computer	12	2468	24	4699	61	134	23	12105	2234	2	12	12	61	27
Cycling	67	40	106	59	3543	308	231	22	150	23	59	107	394	1881
Dust	306	96	877	367	977	2715	155	121	587	26	268	413	2094	335
Jumping Jacks	342	6	345	0	264	64	997	1	17	1384	573	236	94	1293
Lying	6	3416	17	7825	16	16	13	33086	2932	1	9	6	34	17
Reclining	18	395	32	1053	294	168	10	1722	598	2	9	2	35	12
Run	110	1	20	0	35	5	1742	0	0	2930	161	105	10	594
Soccer	1622	26	1394	10	130	196	996	40	54	412	2203	1091	302	337
Stairs	991	10	607	3	153	321	449	17	16	57	1146	2999	326	853
Sweep	254	39	819	98	742	2219	106	61	162	17	294	642	2650	622
Walk	302	7	156	11	1315	366	993	22	29	1079	379	941	333	7850

Table 31. Individual Class Performance Summaries for the Hip Gyroscope Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	17.4%	10.1%	29.3%	30.1%	44.2%	35.4%	15.1%	65.4%	7.5%	47.6%	30.1%	38.3%	35.0%	54.7%
Specificity	97.4%	96.2%	96.1%	88.1%	97.7%	95.7%	97.0%	86.9%	97.5%	98.2%	95.7%	96.8%	96.0%	95.9%
PPV	20.5%	11.3%	26.4%	21.5%	50.7%	29.1%	17.8%	69.8%	13.7%	51.3%	25.0%	37.7%	30.4%	57.0%
NPV	96.9%	95.7%	96.6%	92.1%	97.1%	96.7%	96.4%	84.5%	95.3%	97.9%	96.6%	96.8%	96.8%	95.6%
Precision	20.5%	11.3%	26.4%	21.5%	50.7%	29.1%	17.8%	69.8%	13.7%	51.3%	25.0%	37.7%	30.4%	57.0%
Recall	17.4%	10.1%	29.3%	30.1%	44.2%	35.4%	15.1%	65.4%	7.5%	47.6%	30.1%	38.3%	35.0%	54.7%
F1	18.8%	10.6%	27.8%	25.1%	47.2%	31.9%	16.3%	67.5%	9.7%	49.4%	27.3%	38.0%	32.5%	55.8%
Balanced Accuracy	57.4%	53.1%	62.7%	59.1%	71.0%	65.5%	56.0%	76.2%	52.5%	72.9%	62.9%	67.5%	65.5%	75.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 32. Individual Class Confusion Matrix Results for the Hip Combined Sensor Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1529	8	655	7	86	171	363	4	5	198	1618	414	227	222
Book	5	1336	10	910	24	25	0	942	1770	0	0	0	0	0
Catch	835	84	2842	106	348	534	131	35	71	42	1212	586	704	279
Computer	7	3850	72	13241	212	179	26	379	1756	1	25	5	78	100
Cycling	20	16	51	40	4212	312	27	8	37	5	48	38	244	1122
Dust	194	116	961	245	1178	3471	55	34	100	27	197	143	2321	226
Jumping Jacks	351	0	106	0	82	24	3616	0	0	1646	415	105	6	280
Lying	1	44	1	13	11	6	0	45486	859	0	0	0	0	1
Reclining	5	1828	6	985	413	21	12	3650	3284	0	0	3	2	3
Run	70	1	1	0	5	19	1445	0	1	3387	96	5	0	61
Soccer	1632	0	1027	10	87	146	531	3	3	364	2435	415	226	413
Stairs	689	12	664	9	40	162	186	17	13	15	631	5439	188	458
Sweep	152	32	689	47	564	2356	20	20	34	26	168	317	3415	367
Walk	401	9	231	12	755	244	212	6	8	447	463	356	159	10808

Table 33. Individual Class Performance Summaries for the Hip Combined Sensor Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	26.0%	18.2%	38.8%	84.7%	52.5%	45.3%	54.6%	89.9%	41.4%	55.0%	33.3%	69.5%	45.1%	75.4%
Specificity	97.4%	97.6%	96.8%	95.4%	98.7%	96.2%	98.0%	99.1%	95.5%	98.9%	96.8%	98.0%	96.9%	97.7%
PPV	27.8%	26.6%	36.4%	66.4%	68.2%	37.5%	54.5%	98.0%	32.2%	66.5%	33.4%	63.8%	41.6%	76.6%
NPV	97.2%	96.1%	97.1%	98.3%	97.5%	97.2%	98.0%	95.5%	96.9%	98.2%	96.8%	98.4%	97.3%	97.6%
Precision	27.8%	26.6%	36.4%	66.4%	68.2%	37.5%	54.5%	98.0%	32.2%	66.5%	33.4%	63.8%	41.6%	76.6%
Recall	26.0%	18.2%	38.8%	84.7%	52.5%	45.3%	54.6%	89.9%	41.4%	55.0%	33.3%	69.5%	45.1%	75.4%
F1	26.8%	21.6%	37.6%	74.5%	59.3%	41.0%	54.6%	93.8%	36.2%	60.2%	33.4%	66.5%	43.3%	76.0%
Balanced Accuracy	61.7%	57.9%	67.8%	90.1%	75.6%	70.7%	76.3%	94.5%	68.4%	76.9%	65.1%	83.7%	71.0%	86.6%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 34. Individual Class Confusion Matrix Results for the Hip Accelerometer Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	598	1	403	8	86	68	255	1	0	65	571	210	44	131
Book	2	1250	57	986	358	9	0	223	1425	0	0	9	16	32
Catch	845	27	2173	75	117	322	137	33	21	57	1233	830	566	344
Computer	6	3624	158	11815	1411	333	23	29	1667	2	4	20	345	86
Cycling	49	98	263	883	2292	834	38	76	166	4	53	45	739	515
Dust	108	198	655	556	1341	2567	40	44	125	25	137	99	2058	98
Jumping Jacks	537	0	104	0	43	55	2880	0	0	1390	627	158	19	439
Lying	0	87	5	21	3	4	0	45813	648	0	3	1	0	0
Reclining	40	1793	12	945	607	194	0	3826	3780	0	0	14	8	111
Run	44	0	0	0	34	10	1446	240	0	3771	130	9	1	339
Soccer	1628	3	751	0	28	77	658	8	1	371	2428	265	172	321
Stairs	1034	43	1085	24	64	135	548	35	15	21	821	5024	140	606
Sweep	126	190	681	263	394	2419	21	242	69	21	63	209	2823	286
Walk	874	22	969	49	1239	643	578	14	24	431	1238	933	639	11032

Table 35. Individual Class Performance Summaries for the Hip Accelerometer Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	10.2%	17.0%	29.7%	75.6%	28.6%	33.5%	43.5%	90.6%	47.6%	61.2%	33.2%	64.2%	37.3%	76.9%
Specificity	98.8%	98.0%	97.0%	94.7%	97.5%	96.4%	97.8%	99.3%	95.0%	98.5%	97.2%	97.0%	96.7%	94.8%
PPV	24.5%	28.6%	32.1%	60.5%	37.9%	31.9%	46.1%	98.3%	33.4%	62.6%	36.2%	52.4%	36.2%	59.0%
NPV	96.6%	96.1%	96.6%	97.3%	96.3%	96.6%	97.6%	95.8%	97.2%	98.5%	96.8%	98.1%	96.9%	97.7%
Precision	24.5%	28.6%	32.1%	60.5%	37.9%	31.9%	46.1%	98.3%	33.4%	62.6%	36.2%	52.4%	36.2%	59.0%
Recall	10.2%	17.0%	29.7%	75.6%	28.6%	33.5%	43.5%	90.6%	47.6%	61.2%	33.2%	64.2%	37.3%	76.9%
F1	14.4%	21.4%	30.8%	67.2%	32.6%	32.7%	44.7%	94.3%	39.2%	61.9%	34.6%	57.7%	36.7%	66.8%
Balanced Accuracy	54.5%	57.5%	63.3%	85.1%	63.1%	64.9%	70.6%	94.9%	71.3%	79.9%	65.2%	80.6%	67.0%	85.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 36. Individual Class Confusion Matrix Results for the Hip Gyroscope Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	69	0	138	0	3	29	13	4	0	0	96	108	32	29
Book	0	63	0	164	4	16	0	29	73	0	0	0	0	0
Catch	1103	96	2931	117	351	782	637	163	183	113	1282	501	778	272
Computer	11	2496	25	4531	64	171	39	4374	2069	2	11	0	46	23
Cycling	29	31	30	104	4163	288	41	14	224	33	16	44	226	660
Dust	129	187	752	330	433	2530	42	134	621	20	215	94	1594	143
Jumping Jacks	301	7	234	2	167	47	1646	0	13	1222	489	381	38	369
Lying	6	4332	9	10176	5	7	5	45703	4396	0	4	0	4	13
Reclining	1	58	2	98	14	52	0	43	92	0	2	0	29	0
Run	68	0	0	0	53	5	1506	0	0	3416	143	20	0	743
Soccer	2228	15	1218	0	64	66	776	6	0	337	2963	675	186	210
Stairs	1273	10	836	6	164	350	654	25	25	118	1396	4383	477	1106
Sweep	347	29	1002	90	913	2892	147	73	227	25	336	620	3812	629
Walk	326	12	139	7	1619	435	1118	16	18	872	355	1000	348	10143

Table 37. Individual Class Performance Summaries for the Hip Gyroscope Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	1.2%	0.9%	40.1%	29.0%	51.9%	33.0%	24.8%	90.4%	1.2%	55.5%	40.5%	56.0%	50.4%	70.7%
Specificity	99.7%	99.8%	95.8%	93.5%	98.9%	96.9%	97.9%	82.7%	99.8%	98.4%	96.2%	95.8%	95.2%	95.7%
PPV	13.2%	18.1%	31.5%	32.7%	70.5%	35.0%	33.5%	70.7%	23.5%	57.4%	33.9%	40.5%	34.2%	61.8%
NPV	96.4%	95.5%	97.1%	92.4%	97.5%	96.6%	96.8%	94.9%	95.1%	98.2%	97.1%	97.7%	97.5%	97.1%
Precision	13.2%	18.1%	31.5%	32.7%	70.5%	35.0%	33.5%	70.7%	23.5%	57.4%	33.9%	40.5%	34.2%	61.8%
Recall	1.2%	0.9%	40.1%	29.0%	51.9%	33.0%	24.8%	90.4%	1.2%	55.5%	40.5%	56.0%	50.4%	70.7%
F1	2.2%	1.6%	35.3%	30.7%	59.8%	34.0%	28.5%	79.3%	2.2%	56.4%	36.9%	47.0%	40.7%	66.0%
Balanced Accuracy	50.4%	50.3%	67.9%	61.3%	75.4%	65.0%	61.4%	86.5%	50.5%	76.9%	68.4%	75.9%	72.8%	83.2%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 38. Individual Class Confusion Matrix Results for the Hip Combined Sensor Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	661	1	361	4	41	101	147	2	1	37	571	308	114	104
Book	14	1415	15	1377	71	73	0	174	1910	0	2	0	0	12
Catch	1256	65	3248	105	490	578	109	39	34	49	1686	522	592	319
Computer	6	3830	121	12966	113	133	28	61	1680	0	15	7	52	26
Cycling	28	23	34	57	4861	427	26	12	23	4	27	48	282	771
Dust	217	113	1094	288	695	3162	60	35	107	34	238	157	2396	119
Jumping Jacks	464	0	57	2	73	40	3815	0	0	1873	398	130	1	211
Lying	0	21	2	18	0	4	1	47873	900	0	1	0	19	0
Reclining	26	1737	11	728	407	117	0	2336	3192	0	8	3	7	210
Run	55	0	0	0	8	21	1289	0	0	3227	132	10	1	62
Soccer	1655	1	801	2	56	126	451	2	1	392	2616	231	225	361
Stairs	955	13	778	8	24	155	302	23	17	17	928	5745	210	535
Sweep	87	97	607	52	260	2468	25	24	55	19	135	239	3426	175
Walk	467	20	187	18	918	265	371	3	21	506	551	426	245	11435

Table 39. Individual Class Performance Summaries for the Hip Combined Sensor Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	11.2%	19.3%	44.4%	83.0%	60.6%	41.2%	57.6%	94.6%	40.2%	52.4%	35.8%	73.4%	45.3%	79.7%
Specificity	98.8%	97.6%	96.2%	95.8%	98.8%	96.4%	97.9%	99.1%	96.3%	99.0%	97.2%	97.4%	97.2%	97.3%
PPV	26.9%	27.9%	35.7%	68.1%	73.4%	36.3%	54.0%	98.0%	36.3%	67.2%	37.8%	59.2%	44.7%	74.1%
NPV	96.7%	96.2%	97.3%	98.1%	97.9%	97.0%	98.2%	97.6%	96.9%	98.1%	96.9%	98.6%	97.3%	98.0%
Precision	26.9%	27.9%	35.7%	68.1%	73.4%	36.3%	54.0%	98.0%	36.3%	67.2%	37.8%	59.2%	44.7%	74.1%
Recall	11.2%	19.3%	44.4%	83.0%	60.6%	41.2%	57.6%	94.6%	40.2%	52.4%	35.8%	73.4%	45.3%	79.7%
F1	15.8%	22.8%	39.6%	74.8%	66.4%	38.6%	55.7%	96.3%	38.2%	58.9%	36.8%	65.5%	45.0%	76.8%
Balanced Accuracy	55.0%	58.5%	70.3%	89.4%	79.7%	68.8%	77.7%	96.9%	68.3%	75.7%	66.5%	85.4%	71.2%	88.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 40. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	7458	1141	649	1845	2222	1923
CWR	788	10199	2537	117	1983	882
Rec_LTA	614	3626	6867	361	3884	228
SLA	1363	156	298	77881	478	2955
Sports	2664	3839	3863	298	11031	509
Wheels	2353	1537	236	984	917	1520

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 41. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	48.9%	49.8%	47.5%	95.6%	53.8%	19.0%
Specificity	94.6%	95.5%	94.0%	93.3%	92.0%	96.0%
PPV	48.9%	61.8%	44.1%	93.7%	49.7%	20.1%
NPV	94.6%	92.8%	94.8%	95.3%	93.1%	95.7%
Precision	48.9%	61.8%	44.1%	93.7%	49.7%	20.1%
Recall	48.9%	49.8%	47.5%	95.6%	53.8%	19.0%
F1	48.9%	55.1%	45.7%	94.6%	51.6%	19.5%
Balanced Accuracy	71.8%	72.6%	70.8%	94.5%	72.9%	57.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 42. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	8312	1599	1677	1379	2903	1845
CWR	759	9988	3277	118	1228	1119
Rec_LTA	695	4003	4175	102	4667	389
SLA	1060	119	95	78751	360	249
Sports	3297	2459	4746	826	10828	659
Wheels	1117	2330	480	310	529	3756

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 43. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	54.5%	48.7%	28.9%	96.6%	52.8%	46.9%
Specificity	93.5%	95.3%	93.2%	97.6%	91.4%	96.9%
PPV	46.9%	60.6%	29.8%	97.7%	47.5%	44.1%
NPV	95.1%	92.7%	93.0%	96.6%	92.9%	97.2%
Precision	46.9%	60.6%	29.8%	97.7%	47.5%	44.1%
Recall	54.5%	48.7%	28.9%	96.6%	52.8%	46.9%
F1	50.4%	54.0%	29.3%	97.2%	50.0%	45.4%
Balanced Accuracy	74.0%	72.0%	61.1%	97.1%	72.1%	71.9%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 44. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9460	1029	751	690	2462	2537
CWR	511	11997	2166	143	1382	779
Rec_LTA	434	3092	7438	83	3310	136
SLA	679	113	51	79452	254	538
Sports	2937	2479	3825	513	12674	635
Wheels	1219	1788	219	605	433	3392

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 45. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	62.1%	58.5%	51.5%	97.5%	61.8%	42.3%
Specificity	94.8%	96.4%	95.2%	97.9%	92.6%	97.2%
PPV	55.9%	70.7%	51.3%	98.0%	55.0%	44.3%
NPV	96.0%	94.1%	95.2%	97.4%	94.3%	97.0%
Precision	55.9%	70.7%	51.3%	98.0%	55.0%	44.3%
Recall	62.1%	58.5%	51.5%	97.5%	61.8%	42.3%
F1	58.8%	64.0%	51.4%	97.7%	58.2%	43.3%
Balanced Accuracy	78.5%	77.5%	73.3%	97.7%	77.2%	69.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 46. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9505	1446	1215	3503	2514	1967
CWR	1402	14911	3152	214	3520	1124
Rec_LTA	119	2366	4278	3	1706	183
SLA	3122	81	66	77489	419	4485
Sports	1087	1557	5737	276	12345	191
Wheels	5	137	2	1	11	67

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 47. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	62.4%	72.7%	29.6%	95.1%	60.2%	0.8%
Specificity	92.7%	93.3%	97.0%	89.6%	93.7%	99.9%
PPV	47.2%	61.3%	49.4%	90.5%	58.3%	30.0%
NPV	95.9%	95.9%	93.3%	94.6%	94.1%	95.0%
Precision	47.2%	61.3%	49.4%	90.5%	58.3%	30.0%
Recall	62.4%	72.7%	29.6%	95.1%	60.2%	0.8%
F1	53.7%	66.5%	37.0%	92.7%	59.2%	1.6%
Balanced Accuracy	77.5%	83.0%	63.3%	92.4%	76.9%	50.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 48. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9774	1087	846	819	2652	1557
CWR	1087	15584	4998	63	1619	1566
Rec_LTA	48	1467	2000	5	1195	50
SLA	1827	125	88	79970	365	1422
Sports	2286	1360	6494	613	14643	559
Wheels	218	875	24	16	41	2863

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 49. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	64.1%	76.0%	13.8%	98.1%	71.4%	35.7%
Specificity	95.2%	93.3%	98.1%	95.1%	91.9%	99.2%
PPV	58.4%	62.5%	42.0%	95.4%	56.4%	70.9%
NPV	96.2%	96.4%	92.0%	98.0%	95.6%	96.7%
Precision	58.4%	62.5%	42.0%	95.4%	56.4%	70.9%
Recall	64.1%	76.0%	13.8%	98.1%	71.4%	35.7%
F1	61.1%	68.6%	20.8%	96.8%	63.0%	47.5%
Balanced Accuracy	79.7%	84.7%	56.0%	96.6%	81.6%	67.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 50. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	11467	617	766	769	2525	1714
CWR	877	16184	3978	115	1789	1371
Rec_LTA	126	1906	5479	5	1777	36
SLA	924	72	63	80176	233	1476
Sports	1647	1386	4158	366	14171	490
Wheels	199	333	6	55	20	2930

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 51. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	75.2%	79.0%	37.9%	98.4%	69.1%	36.5%
Specificity	95.6%	94.2%	97.4%	96.5%	94.2%	99.6%
PPV	64.2%	66.6%	58.7%	96.7%	63.8%	82.7%
NPV	97.3%	96.8%	94.1%	98.3%	95.4%	96.8%
Precision	64.2%	66.6%	58.7%	96.7%	63.8%	82.7%
Recall	75.2%	79.0%	37.9%	98.4%	69.1%	36.5%
F1	69.3%	72.2%	46.1%	97.5%	66.3%	50.7%
Balanced Accuracy	85.4%	86.6%	67.6%	97.4%	81.7%	68.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 52. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9783	1091	496	1653	2130	2174
CWR	453	11832	2189	38	1515	699
Rec_LTA	258	3075	7509	78	2168	126
SLA	1304	332	155	79173	358	3326
Sports	2326	3206	4009	277	13958	398
Wheels	1116	962	92	267	386	1294

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 53. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	64.2%	57.7%	52.0%	97.2%	68.0%	16.1%
Specificity	94.8%	96.5%	96.1%	93.0%	92.7%	98.1%
PPV	56.5%	70.7%	56.8%	93.5%	57.7%	31.4%
NPV	96.2%	94.0%	95.3%	96.9%	95.2%	95.7%
Precision	56.5%	70.7%	56.8%	93.5%	57.7%	31.4%
Recall	64.2%	57.7%	52.0%	97.2%	68.0%	16.1%
F1	60.1%	63.6%	54.3%	95.3%	62.5%	21.3%
Balanced Accuracy	79.5%	77.1%	74.0%	95.1%	80.4%	57.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 54. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9757	1272	1310	1178	2699	1862
CWR	555	12491	3737	38	898	1352
Rec_LTA	581	3069	3688	49	3462	392
SLA	785	97	74	79374	227	465
Sports	3072	1848	5433	723	13097	642
Wheels	490	1721	208	124	132	3304

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 55. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	64.0%	60.9%	25.5%	97.4%	63.8%	41.2%
Specificity	94.3%	95.3%	94.8%	97.9%	91.6%	98.2%
PPV	54.0%	65.5%	32.8%	98.0%	52.8%	55.3%
NPV	96.1%	94.3%	92.8%	97.3%	94.5%	96.9%
Precision	54.0%	65.5%	32.8%	98.0%	52.8%	55.3%
Recall	64.0%	60.9%	25.5%	97.4%	63.8%	41.2%
F1	58.6%	63.1%	28.7%	97.7%	57.8%	47.2%
Balanced Accuracy	79.1%	78.1%	60.2%	97.7%	77.7%	69.7%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 56. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	11366	766	520	475	2111	1812
CWR	291	14194	2028	18	912	745
Rec_LTA	205	2363	8300	35	1671	80
SLA	496	196	62	80574	247	738
Sports	2514	2187	3506	323	15521	631
Wheels	368	792	34	61	53	4011

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 57. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	74.6%	69.2%	57.4%	98.9%	75.7%	50.0%
Specificity	96.1%	97.1%	97.0%	97.8%	93.4%	99.1%
PPV	66.7%	78.0%	65.6%	97.9%	62.9%	75.4%
NPV	97.3%	95.6%	95.8%	98.8%	96.3%	97.4%
Precision	66.7%	78.0%	65.6%	97.9%	62.9%	75.4%
Recall	74.6%	69.2%	57.4%	98.9%	75.7%	50.0%
F1	70.4%	73.4%	61.2%	98.4%	68.7%	60.2%
Balanced Accuracy	85.3%	83.2%	77.2%	98.3%	84.5%	74.6%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 58. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10925	672	491	1493	2223	2315
CWR	680	14095	2441	158	2057	786
Rec_LTA	200	2650	7931	76	2077	170
SLA	994	229	87	78655	281	2750
Sports	1582	1806	3441	187	13539	299
Wheels	859	1046	59	917	338	1697

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 59. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	71.7%	68.8%	54.9%	96.5%	66.0%	21.2%
Specificity	95.0%	95.6%	96.5%	94.5%	94.8%	97.9%
PPV	60.3%	69.7%	60.5%	94.8%	64.9%	34.5%
NPV	97.0%	95.4%	95.6%	96.3%	95.0%	95.9%
Precision	60.3%	69.7%	60.5%	94.8%	64.9%	34.5%
Recall	71.7%	68.8%	54.9%	96.5%	66.0%	21.2%
F1	65.5%	69.2%	57.6%	95.6%	65.5%	26.2%
Balanced Accuracy	83.4%	82.2%	75.7%	95.5%	80.4%	59.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 60. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10960	1054	1066	1289	2603	1475
CWR	649	14903	3806	48	973	1853
Rec_LTA	404	2101	5094	58	2464	220
SLA	547	73	66	79302	125	278
Sports	2428	1397	4396	617	14320	621
Wheels	252	970	22	172	30	3570

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 61. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	71.9%	72.7%	35.3%	97.3%	69.8%	44.5%
Specificity	94.8%	94.8%	96.4%	98.6%	93.2%	99.0%
PPV	59.4%	67.0%	49.3%	98.6%	60.2%	71.2%
NPV	97.0%	95.9%	93.8%	97.3%	95.5%	97.1%
Precision	59.4%	67.0%	49.3%	98.6%	60.2%	71.2%
Recall	71.9%	72.7%	35.3%	97.3%	69.8%	44.5%
F1	65.1%	69.8%	41.1%	98.0%	64.7%	54.8%
Balanced Accuracy	83.4%	83.7%	65.8%	98.0%	81.5%	71.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 62. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	11553	523	566	714	2111	1190
CWR	354	14975	1966	41	1186	733
Rec_LTA	194	2354	8850	34	1901	145
SLA	558	65	70	80210	274	1049
Sports	2107	1694	2941	353	14976	622
Wheels	474	887	57	134	67	4278

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 63. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	75.8%	73.1%	61.2%	98.4%	73.0%	53.4%
Specificity	96.5%	96.9%	96.8%	97.4%	94.5%	98.9%
PPV	69.4%	77.8%	65.7%	97.5%	66.0%	72.5%
NPV	97.4%	96.1%	96.2%	98.4%	96.0%	97.6%
Precision	69.4%	77.8%	65.7%	97.5%	66.0%	72.5%
Recall	75.8%	73.1%	61.2%	98.4%	73.0%	53.4%
F1	72.4%	75.3%	63.4%	98.0%	69.3%	61.5%
Balanced Accuracy	86.1%	85.0%	79.0%	97.9%	83.7%	76.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 64. Individual Class Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1509	124	1078	1	24	172	535	18	4	260	1080	387	67	554
Book	11	2196	6	2347	815	290	8	8632	1635	1	8	8	308	39
Catch	1296	20	2335	16	73	237	308	69	22	180	1503	809	335	786
Computer	9	1494	13	6079	1365	306	121	14333	2109	0	3	33	193	63
Cycling	8	366	32	1221	2719	338	27	633	566	2	19	39	84	75
Dust	132	435	283	427	773	3039	75	418	342	39	235	773	992	889
Jumping Jacks	439	0	274	5	8	37	3327	0	0	828	361	95	7	472
Lying	9	1174	37	2628	309	227	29	22077	1420	0	10	9	178	42
Reclining	4	862	14	2516	1471	257	31	3325	1421	0	3	23	110	54
Run	195	1	237	27	4	25	621	3	21	4171	154	42	0	238
Soccer	1369	5	1018	41	29	121	760	47	81	280	2005	613	72	880
Stairs	382	26	928	43	104	917	48	170	63	35	875	2812	586	1958
Sweep	163	340	374	228	230	1064	67	767	212	45	214	685	4256	503
Walk	365	293	687	46	93	640	667	92	45	317	838	1498	382	7787

Table 65. Individual Class Performance Summaries for the Left Wrist Accelerometer Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	25.6%	29.9%	31.9%	38.9%	33.9%	39.6%	50.2%	43.6%	17.9%	67.7%	27.4%	35.9%	56.2%	54.3%
Specificity	97.2%	90.8%	96.3%	86.1%	97.8%	96.2%	98.4%	94.5%	94.3%	99.0%	96.5%	96.0%	96.8%	95.9%
PPV	26.0%	13.5%	29.2%	23.3%	44.4%	34.3%	56.8%	78.4%	14.1%	72.7%	27.4%	31.4%	46.5%	56.6%
NPV	97.2%	96.4%	96.7%	92.9%	96.6%	96.9%	97.9%	78.4%	95.7%	98.7%	96.5%	96.7%	97.8%	95.5%
Precision	26.0%	13.5%	29.2%	23.3%	44.4%	34.3%	56.8%	78.4%	14.1%	72.7%	27.4%	31.4%	46.5%	56.6%
Recall	25.6%	29.9%	31.9%	38.9%	33.9%	39.6%	50.2%	43.6%	17.9%	67.7%	27.4%	35.9%	56.2%	54.3%
F1	25.8%	18.6%	30.5%	29.1%	38.4%	36.8%	53.3%	56.1%	15.8%	70.1%	27.4%	33.5%	50.9%	55.4%
Balanced Accuracy	61.4%	60.4%	64.1%	62.5%	65.8%	67.9%	74.3%	69.1%	56.1%	83.4%	62.0%	66.0%	76.5%	75.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 66. Individual Class Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1319	111	1117	31	48	337	515	54	62	175	1153	599	132	680
Book	19	1632	52	2468	1302	487	48	3875	1356	11	44	74	164	79
Catch	1032	42	1574	31	179	514	336	74	72	121	1137	998	222	885
Computer	20	2149	16	4952	408	367	26	11619	2176	0	11	5	44	16
Cycling	14	210	56	306	3815	794	17	226	563	1	32	12	303	87
Dust	214	238	465	220	949	1875	34	217	428	99	365	666	1427	552
Jumping Jacks	401	7	292	0	29	35	3146	2	1	908	535	198	14	777
Lying	32	1653	25	5407	159	145	58	31003	2108	8	79	58	65	25
Reclining	10	917	25	1977	630	487	12	3264	844	2	17	31	171	47
Run	106	0	188	0	14	39	679	3	5	1942	103	132	12	1527
Soccer	1382	80	1457	100	64	330	477	62	95	364	1564	1050	382	1202
Stairs	672	31	896	11	40	395	160	33	25	327	1091	1626	526	1819
Sweep	70	228	127	94	260	1296	135	85	168	28	228	643	3084	1283
Walk	600	38	1026	28	120	569	981	67	38	2172	949	1734	1024	5361

Table 67. Individual Class Performance Summaries for the Left Wrist Gyroscope Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	22.4%	22.2%	21.5%	31.7%	47.6%	24.4%	47.5%	61.3%	10.6%	31.5%	21.4%	20.8%	40.7%	37.4%
Specificity	96.8%	93.5%	96.3%	88.3%	98.3%	96.1%	97.9%	91.0%	95.0%	98.2%	95.4%	96.0%	97.0%	93.6%
PPV	20.8%	14.1%	21.8%	22.7%	59.3%	24.2%	49.6%	75.9%	10.0%	40.9%	18.2%	21.2%	39.9%	36.5%
NPV	97.0%	96.2%	96.2%	92.3%	97.3%	96.2%	97.7%	83.6%	95.3%	97.3%	96.2%	95.9%	97.1%	93.8%
Precision	20.8%	14.1%	21.8%	22.7%	59.3%	24.2%	49.6%	75.9%	10.0%	40.9%	18.2%	21.2%	39.9%	36.5%
Recall	22.4%	22.2%	21.5%	31.7%	47.6%	24.4%	47.5%	61.3%	10.6%	31.5%	21.4%	20.8%	40.7%	37.4%
F1	21.6%	17.2%	21.7%	26.5%	52.8%	24.3%	48.5%	67.8%	10.3%	35.6%	19.7%	21.0%	40.3%	36.9%
Balanced Accuracy	59.6%	57.9%	58.9%	60.0%	72.9%	60.3%	72.7%	76.2%	52.8%	64.9%	58.4%	58.4%	68.8%	65.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 68. Individual Class Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1528	2	1241	9	41	180	674	18	5	182	1161	423	84	497
Book	17	2785	31	2085	438	272	18	7315	1952	1	8	27	129	26
Catch	1310	24	2229	5	90	329	272	71	47	209	1449	631	289	759
Computer	14	1749	16	7976	432	230	92	11316	1838	1	12	14	32	17
Cycling	10	124	16	307	5495	615	23	124	552	1	10	25	58	29
Dust	122	294	359	314	619	3364	70	179	316	53	296	665	957	552
Jumping Jacks	444	4	306	8	5	37	3756	4	3	582	397	155	22	706
Lying	11	1038	32	2496	85	110	82	26883	1197	0	5	8	33	11
Reclining	48	1061	38	2277	534	188	13	4365	1841	0	8	3	47	10
Run	143	10	155	1	1	64	454	1	0	4382	151	33	17	126
Soccer	1300	23	898	17	24	319	571	7	14	289	2033	498	108	735
Stairs	414	28	811	24	42	602	80	39	40	33	702	3267	301	2095
Sweep	101	163	298	61	94	883	47	225	102	49	164	350	5052	591
Walk	429	31	886	45	117	477	472	37	34	376	912	1727	441	8186

Table 69. Individual Class Performance Summaries for the Left Wrist Combined Sensor Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	25.9%	38.0%	30.5%	51.0%	68.5%	43.9%	56.7%	53.1%	23.2%	71.2%	27.8%	41.7%	66.7%	57.1%
Specificity	97.1%	91.9%	96.4%	89.1%	98.8%	96.9%	98.3%	95.3%	94.4%	99.2%	96.9%	96.6%	98.0%	95.9%
PPV	25.3%	18.4%	28.9%	33.6%	74.4%	41.2%	58.4%	84.0%	17.6%	79.1%	29.7%	38.5%	61.8%	57.8%
NPV	97.2%	96.9%	96.7%	94.4%	98.3%	97.2%	98.1%	81.5%	95.9%	98.9%	96.6%	97.0%	98.3%	95.8%
Precision	25.3%	18.4%	28.9%	33.6%	74.4%	41.2%	58.4%	84.0%	17.6%	79.1%	29.7%	38.5%	61.8%	57.8%
Recall	25.9%	38.0%	30.5%	51.0%	68.5%	43.9%	56.7%	53.1%	23.2%	71.2%	27.8%	41.7%	66.7%	57.1%
F1	25.6%	24.8%	29.7%	40.5%	71.3%	42.5%	57.5%	65.1%	20.0%	74.9%	28.7%	40.1%	64.2%	57.4%
Balanced Accuracy	61.5%	65.0%	63.4%	70.1%	83.6%	70.4%	77.5%	74.2%	58.8%	85.2%	62.3%	69.2%	82.3%	76.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 70. Individual Class Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1435	0	887	0	0	63	389	1	0	186	874	129	27	150
Book	4	293	14	184	16	84	12	588	159	0	4	11	553	28
Catch	1688	16	2687	15	65	157	219	41	23	171	1404	416	148	442
Computer	5	894	16	4158	2066	289	159	7288	1295	0	3	35	120	37
Cycling	10	53	3	231	656	313	10	79	208	0	12	47	81	38
Dust	35	263	112	188	430	2127	22	197	203	12	96	389	690	182
Jumping Jacks	419	0	71	0	1	53	4070	0	0	737	481	116	4	589
Lying	15	5403	5	10344	4382	615	39	40668	5630	0	3	7	209	111
Reclining	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Run	153	0	160	0	0	8	341	0	0	4370	75	14	2	70
Soccer	892	0	646	5	11	85	446	10	2	131	1700	237	30	469
Stairs	368	10	946	28	105	336	110	36	52	30	597	1442	247	1115
Sweep	276	352	498	249	152	954	43	481	282	24	280	820	4169	703
Walk	591	52	1271	222	133	2586	764	1195	87	497	1779	4163	1290	10406

Table 71. Individual Class Performance Summaries for the Left Wrist Accelerometer Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	24.4%	4.0%	36.7%	26.6%	8.2%	27.7%	61.4%	80.4%	0.0%	71.0%	23.3%	18.4%	55.1%	72.6%
Specificity	98.2%	98.9%	96.9%	91.6%	99.3%	98.2%	98.4%	75.6%	100.0%	99.5%	98.1%	97.4%	96.6%	90.0%
PPV	34.7%	15.0%	35.9%	25.4%	37.7%	43.0%	62.2%	60.3%	0.0%	84.2%	36.4%	26.6%	44.9%	41.6%
NPV	97.1%	95.5%	97.0%	92.0%	95.4%	96.4%	98.3%	89.3%	95.0%	98.8%	96.4%	95.9%	97.7%	97.1%
Precision	34.7%	15.0%	35.9%	25.4%	37.7%	43.0%	62.2%	60.3%	0.0%	84.2%	36.4%	26.6%	44.9%	41.6%
Recall	24.4%	4.0%	36.7%	26.6%	8.2%	27.7%	61.4%	80.4%	0.0%	71.0%	23.3%	18.4%	55.1%	72.6%
F1	28.6%	6.3%	36.3%	26.0%	13.4%	33.7%	61.8%	68.9%	NA	77.0%	28.4%	21.8%	49.5%	52.9%
Balanced Accuracy	61.3%	51.5%	66.8%	59.1%	53.7%	62.9%	79.9%	78.0%	50.0%	85.2%	60.7%	57.9%	75.9%	81.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 72. Individual Class Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1674	0	1168	1	4	60	540	0	0	244	1344	422	36	522
Book	5	407	14	519	509	952	31	304	488	1	28	4	58	45
Catch	1386	50	2025	52	220	469	205	84	71	143	1503	739	341	990
Computer	2	1809	2	3312	335	206	44	4193	1371	0	10	0	38	13
Cycling	0	83	1	120	3433	645	1	61	321	0	1	0	213	44
Dust	151	462	378	211	594	2438	79	317	356	36	261	405	969	237
Jumping Jacks	476	1	150	10	0	17	3704	23	4	864	570	163	4	978
Lying	12	4352	8	11323	2542	270	150	45448	5129	2	25	3	3	23
Reclining	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Run	28	0	1	0	0	13	743	0	0	1404	41	14	0	567
Soccer	521	1	422	7	11	30	255	2	0	68	601	242	27	299
Stairs	698	27	1164	17	18	357	63	44	33	76	1221	1565	295	982
Sweep	34	116	51	45	307	1590	15	73	149	14	182	641	4023	915
Walk	904	28	1932	8	44	623	794	35	19	3306	1521	3628	1563	8725

Table 73. Individual Class Performance Summaries for the Left Wrist Gyroscope Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	28.4%	5.5%	27.7%	21.2%	42.8%	31.8%	55.9%	89.8%	0.0%	22.8%	8.2%	20.0%	53.1%	60.8%
Specificity	97.2%	98.1%	95.9%	94.5%	99.0%	97.1%	97.9%	78.3%	100.0%	99.1%	98.8%	96.7%	97.3%	90.1%
PPV	27.8%	12.1%	24.5%	29.2%	69.7%	35.4%	53.2%	65.6%	NA	49.9%	24.2%	23.9%	49.3%	37.7%
NPV	97.3%	95.6%	96.5%	91.7%	97.0%	96.6%	98.1%	94.4%	95.0%	97.0%	95.7%	95.9%	97.7%	95.9%
Precision	27.8%	12.1%	24.5%	29.2%	69.7%	35.4%	53.2%	65.6%	NA	49.9%	24.2%	23.9%	49.3%	37.7%
Recall	28.4%	5.5%	27.7%	21.2%	42.8%	31.8%	55.9%	89.8%	0.0%	22.8%	8.2%	20.0%	53.1%	60.8%
F1	28.1%	7.6%	26.0%	24.6%	53.1%	33.5%	54.5%	75.8%	NA	31.3%	12.3%	21.8%	51.2%	46.6%
Balanced Accuracy	62.8%	51.8%	61.8%	57.8%	70.9%	64.4%	76.9%	84.0%	50.0%	60.9%	53.5%	58.4%	75.2%	75.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 74. Individual Class Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1666	0	1155	0	0	35	442	1	0	118	932	161	21	270
Book	17	1525	23	842	182	382	33	2107	851	0	1	19	201	27
Catch	1601	22	2640	25	96	199	237	62	15	221	1412	474	159	653
Computer	5	1292	2	5720	770	179	85	3661	2124	0	14	14	8	6
Cycling	14	99	5	158	4799	323	12	28	500	0	2	25	94	65
Dust	110	223	236	177	502	4181	53	235	205	38	285	539	856	310
Jumping Jacks	356	0	82	8	11	42	4224	26	0	351	357	112	11	588
Lying	16	3948	21	8562	1423	274	103	44217	4076	1	7	2	4	5
Reclining	0	41	0	56	122	86	3	5	39	0	0	2	4	1
Run	82	0	15	0	0	24	281	0	0	4782	150	11	0	13
Soccer	883	10	507	2	1	130	316	16	10	196	1921	126	63	242
Stairs	498	31	1058	27	24	446	112	40	46	14	658	2580	321	1498
Sweep	142	126	256	36	62	620	11	128	62	27	287	481	5286	677
Walk	501	19	1316	12	25	749	712	58	13	410	1282	3280	542	9985

Table 75. Individual Class Performance Summaries for the Left Wrist Combined Sensor Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	28.3%	20.8%	36.1%	36.6%	59.9%	54.5%	63.8%	87.4%	0.5%	77.7%	26.3%	33.0%	69.8%	69.6%
Specificity	98.0%	96.9%	96.6%	94.4%	99.1%	97.5%	98.7%	83.2%	99.8%	99.6%	98.4%	96.9%	98.1%	93.9%
PPV	34.7%	24.6%	33.8%	41.2%	78.4%	52.6%	68.5%	70.6%	10.9%	89.2%	43.4%	35.1%	64.5%	52.8%
NPV	97.3%	96.2%	96.9%	93.2%	97.9%	97.7%	98.4%	93.5%	95.1%	99.1%	96.5%	96.6%	98.5%	96.9%
Precision	34.7%	24.6%	33.8%	41.2%	78.4%	52.6%	68.5%	70.6%	10.9%	89.2%	43.4%	35.1%	64.5%	52.8%
Recall	28.3%	20.8%	36.1%	36.6%	59.9%	54.5%	63.8%	87.4%	0.5%	77.7%	26.3%	33.0%	69.8%	69.6%
F1	31.2%	22.5%	34.9%	38.8%	67.9%	53.5%	66.0%	78.1%	0.9%	83.0%	32.8%	34.0%	67.0%	60.1%
Balanced Accuracy	63.1%	58.9%	66.3%	65.5%	79.5%	76.0%	81.3%	85.3%	50.1%	88.6%	62.3%	64.9%	84.0%	81.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 76. Individual Class Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1761	10	1001	0	12	100	611	9	9	228	990	230	37	354
Book	13	2989	0	1946	715	231	10	4135	1709	0	0	5	264	63
Catch	1545	19	3087	23	90	203	225	83	28	179	1411	702	358	759
Computer	4	1223	12	8144	1732	237	137	13177	2043	0	3	17	44	23
Cycling	6	232	3	808	2772	183	26	397	573	0	7	12	26	10
Dust	71	504	191	634	822	3647	34	219	381	20	174	459	612	620
Jumping Jacks	311	0	196	0	2	44	3899	0	0	732	255	69	7	415
Lying	6	1368	14	2599	184	214	21	29556	1567	0	4	5	87	35
Reclining	2	619	1	1126	1175	111	19	2294	1225	0	0	8	34	25
Run	168	1	199	5	3	16	357	1	3	4324	128	18	0	143
Soccer	1228	2	974	6	18	127	584	12	11	288	2580	485	53	656
Stairs	331	38	801	43	218	865	36	76	49	29	728	3630	535	1864
Sweep	149	300	275	251	245	1202	76	560	311	22	208	625	5238	514
Walk	296	31	562	40	29	490	589	65	32	336	820	1561	275	8859

Table 77. Individual Class Performance Summaries for the Left Wrist Accelerometer Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	29.9%	40.7%	42.2%	52.1%	34.6%	47.5%	58.9%	58.4%	15.4%	70.2%	35.3%	46.4%	69.2%	61.8%
Specificity	97.7%	94.1%	96.3%	87.1%	98.5%	96.9%	98.7%	94.4%	96.4%	99.3%	97.1%	96.3%	96.9%	96.5%
PPV	32.9%	24.7%	35.4%	30.4%	54.8%	43.5%	65.8%	82.9%	18.5%	80.6%	36.7%	39.3%	52.5%	63.3%
NPV	97.3%	97.1%	97.2%	94.4%	96.6%	97.4%	98.2%	83.1%	95.6%	98.8%	96.9%	97.2%	98.4%	96.3%
Precision	32.9%	24.7%	35.4%	30.4%	54.8%	43.5%	65.8%	82.9%	18.5%	80.6%	36.7%	39.3%	52.5%	63.3%
Recall	29.9%	40.7%	42.2%	52.1%	34.6%	47.5%	58.9%	58.4%	15.4%	70.2%	35.3%	46.4%	69.2%	61.8%
F1	31.3%	30.8%	38.5%	38.4%	42.4%	45.4%	62.1%	68.5%	16.8%	75.0%	36.0%	42.5%	59.7%	62.6%
Balanced Accuracy	63.8%	67.4%	69.3%	69.6%	66.5%	72.2%	78.8%	76.4%	55.9%	84.8%	66.2%	71.4%	83.0%	79.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 78. Individual Class Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1462	24	1128	25	22	136	769	18	23	227	1192	464	107	576
Book	12	1646	39	2017	1275	439	26	1853	1277	9	28	15	100	32
Catch	1245	61	1897	47	163	533	227	78	64	134	1288	873	193	891
Computer	7	2174	11	5158	278	329	22	10519	1847	7	14	17	59	35
Cycling	6	155	16	291	4347	612	19	158	545	1	15	9	206	46
Dust	178	445	348	272	987	3065	37	325	556	46	355	581	1383	401
Jumping Jacks	358	0	194	0	17	21	3017	1	2	875	497	162	33	756
Lying	4	2035	6	6659	121	53	84	34940	2699	0	24	12	10	5
Reclining	0	571	6	1022	368	160	8	2543	708	0	5	9	52	6
Run	58	2	106	4	15	21	761	0	1	2222	83	59	6	1577
Soccer	1304	23	1359	39	52	249	489	25	22	267	1435	974	238	951
Stairs	725	25	972	16	13	366	73	27	31	237	1204	2042	521	1644
Sweep	42	152	91	67	240	1184	123	67	139	13	240	665	3833	1252
Walk	490	23	1143	8	119	502	969	30	27	2120	928	1944	829	6168

Table 79. Individual Class Performance Summaries for the Left Wrist Gyroscope Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	24.8%	22.4%	25.9%	33.0%	54.2%	40.0%	45.5%	69.1%	8.9%	36.1%	19.6%	26.1%	50.6%	43.0%
Specificity	96.9%	95.3%	96.2%	89.4%	98.6%	96.1%	98.1%	89.3%	96.9%	98.3%	96.1%	96.2%	97.2%	93.7%
PPV	23.7%	18.8%	24.7%	25.2%	67.6%	34.1%	50.9%	74.9%	13.0%	45.2%	19.3%	25.9%	47.3%	40.3%
NPV	97.1%	96.2%	96.4%	92.5%	97.6%	97.0%	97.7%	86.2%	95.3%	97.5%	96.2%	96.2%	97.5%	94.4%
Precision	23.7%	18.8%	24.7%	25.2%	67.6%	34.1%	50.9%	74.9%	13.0%	45.2%	19.3%	25.9%	47.3%	40.3%
Recall	24.8%	22.4%	25.9%	33.0%	54.2%	40.0%	45.5%	69.1%	8.9%	36.1%	19.6%	26.1%	50.6%	43.0%
F1	24.2%	20.4%	25.3%	28.6%	60.2%	36.8%	48.1%	71.9%	10.6%	40.1%	19.5%	26.0%	48.9%	41.6%
Balanced Accuracy	60.9%	58.9%	61.1%	61.2%	76.4%	68.0%	71.8%	79.2%	52.9%	67.2%	57.9%	61.1%	73.9%	68.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 80. Individual Class Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	2070	2	1102	17	13	63	598	2	5	147	1030	221	24	382
Book	11	3593	5	1768	217	273	17	3694	1885	0	2	4	40	14
Catch	1519	38	3416	22	120	321	210	69	35	185	1529	623	350	702
Computer	4	1391	5	9420	302	131	67	9386	2297	0	5	9	5	10
Cycling	6	68	2	256	6140	333	11	55	426	0	6	19	13	30
Dust	110	345	214	192	759	4426	52	225	289	30	216	505	759	402
Jumping Jacks	252	1	184	0	3	40	4390	0	0	347	234	86	8	545
Lying	2	1154	9	3090	35	95	92	36535	1404	0	7	2	5	1
Reclining	0	594	1	773	190	66	23	429	1472	0	1	3	1	3
Run	87	1	89	0	4	31	223	0	0	4892	162	16	7	61
Soccer	1070	1	715	9	4	202	372	3	9	347	2614	378	70	496
Stairs	385	28	799	21	55	608	34	35	37	16	614	4068	285	1787
Sweep	85	96	125	47	136	690	49	122	70	33	144	261	5745	502
Walk	290	24	650	10	39	391	486	29	12	161	744	1631	258	9405

Table 81. Individual Class Performance Summaries for the Left Wrist Combined Sensor Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	35.1%	49.0%	46.7%	60.3%	76.6%	57.7%	66.3%	72.2%	18.5%	79.4%	35.8%	52.0%	75.9%	65.6%
Specificity	97.7%	94.8%	96.3%	90.6%	99.2%	97.3%	98.9%	94.6%	98.6%	99.6%	97.6%	96.9%	98.5%	96.8%
PPV	36.5%	31.2%	37.4%	40.9%	83.4%	51.9%	72.1%	86.1%	41.4%	87.8%	41.6%	46.4%	70.9%	66.6%
NPV	97.5%	97.5%	97.4%	95.5%	98.8%	97.9%	98.6%	88.1%	95.9%	99.2%	97.0%	97.5%	98.8%	96.6%
Precision	36.5%	31.2%	37.4%	40.9%	83.4%	51.9%	72.1%	86.1%	41.4%	87.8%	41.6%	46.4%	70.9%	66.6%
Recall	35.1%	49.0%	46.7%	60.3%	76.6%	57.7%	66.3%	72.2%	18.5%	79.4%	35.8%	52.0%	75.9%	65.6%
F1	35.8%	38.1%	41.5%	48.7%	79.8%	54.7%	69.1%	78.6%	25.6%	83.4%	38.4%	49.0%	73.3%	66.1%
Balanced Accuracy	66.4%	71.9%	71.5%	75.4%	87.9%	77.5%	82.6%	83.4%	58.6%	89.5%	66.7%	74.4%	87.2%	81.2%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 82. Individual Class Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1323	0	802	0	1	69	362	1	0	122	769	101	42	118
Book	14	2376	1	1191	385	252	13	2571	931	0	0	5	482	99
Catch	1612	24	3137	32	111	274	150	64	48	183	1321	645	354	584
Computer	10	1717	4	7576	2527	212	148	7091	2550	0	2	15	67	31
Cycling	5	313	4	887	3587	199	21	718	829	0	6	11	13	11
Dust	25	262	95	457	725	3310	29	195	254	13	96	313	438	309
Jumping Jacks	414	0	116	0	3	50	4294	0	0	704	484	105	5	625
Lying	17	2129	24	5003	270	295	28	38958	2788	1	24	14	72	40
Reclining	0	14	0	80	73	21	1	62	101	0	0	0	6	0
Run	204	0	235	0	4	16	406	0	0	4625	72	14	0	155
Soccer	1309	1	864	3	0	92	393	8	5	222	2246	330	29	680
Stairs	396	41	949	35	83	713	68	68	40	33	849	3495	457	1746
Sweep	197	418	366	307	208	1310	48	696	349	18	210	782	5058	653
Walk	365	41	719	54	40	857	663	152	46	237	1229	1996	547	9289

Table 83. Individual Class Performance Summaries for the Left Wrist Accelerometer Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	22.5%	32.4%	42.9%	48.5%	44.7%	43.2%	64.8%	77.0%	1.3%	75.1%	30.7%	44.7%	66.8%	64.8%
Specificity	98.5%	96.1%	96.5%	90.1%	98.0%	97.9%	98.4%	90.2%	99.8%	99.3%	97.4%	96.4%	96.4%	95.2%
PPV	35.7%	28.6%	36.7%	34.5%	54.3%	50.8%	63.1%	78.4%	28.2%	80.7%	36.3%	39.0%	47.6%	57.2%
NPV	97.1%	96.7%	97.2%	94.2%	97.1%	97.2%	98.5%	89.5%	95.1%	99.0%	96.7%	97.1%	98.3%	96.5%
Precision	35.7%	28.6%	36.7%	34.5%	54.3%	50.8%	63.1%	78.4%	28.2%	80.7%	36.3%	39.0%	47.6%	57.2%
Recall	22.5%	32.4%	42.9%	48.5%	44.7%	43.2%	64.8%	77.0%	1.3%	75.1%	30.7%	44.7%	66.8%	64.8%
F1	27.6%	30.4%	39.6%	40.3%	49.1%	46.6%	64.0%	77.7%	2.4%	77.8%	33.3%	41.6%	55.6%	60.8%
Balanced Accuracy	60.5%	64.2%	69.7%	69.3%	71.4%	70.5%	81.6%	83.6%	50.6%	87.2%	64.1%	70.5%	81.6%	80.0%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 84. Individual Class Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1751	6	840	6	10	49	617	26	11	251	1462	283	81	483
Book	1	831	6	683	262	315	20	755	607	1	15	0	17	25
Catch	1774	69	2719	52	175	641	229	76	54	190	2044	1311	288	1287
Computer	2	3256	1	6908	373	129	16	4906	2405	0	8	6	1	11
Cycling	10	201	16	413	5740	629	14	196	801	3	20	4	151	53
Dust	130	675	319	342	881	3546	85	410	680	26	213	343	1349	253
Jumping Jacks	370	0	193	10	17	20	3426	0	0	584	529	183	14	735
Lying	0	2144	4	7139	55	31	103	44072	3161	0	17	7	15	11
Reclining	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Run	74	0	47	0	2	16	1036	0	0	2906	84	26	0	1750
Soccer	576	2	749	3	23	73	213	8	9	120	570	409	91	408
Stairs	698	31	1419	18	15	565	99	47	31	184	1481	2954	548	1850
Sweep	29	105	54	48	407	1362	71	75	175	15	179	650	4341	1172
Walk	476	16	949	3	57	294	695	13	7	1878	686	1650	674	6302

Table 85. Individual Class Performance Summaries for the Left Wrist Gyroscope Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	29.7%	11.3%	37.2%	44.2%	71.6%	46.2%	51.7%	87.1%	0.0%	47.2%	7.8%	37.7%	57.3%	43.9%
Specificity	97.3%	98.2%	94.6%	92.3%	98.4%	96.3%	98.3%	88.4%	100.0%	98.0%	98.2%	95.4%	97.2%	94.9%
PPV	29.8%	23.5%	24.9%	38.3%	69.6%	38.3%	56.3%	77.6%	NA	48.9%	17.5%	29.7%	50.0%	46.0%
NPV	97.3%	95.8%	96.9%	93.9%	98.5%	97.3%	97.9%	93.7%	95.0%	97.9%	95.7%	96.8%	97.9%	94.5%
Precision	29.8%	23.5%	24.9%	38.3%	69.6%	38.3%	56.3%	77.6%	NA	48.9%	17.5%	29.7%	50.0%	46.0%
Recall	29.7%	11.3%	37.2%	44.2%	71.6%	46.2%	51.7%	87.1%	0.0%	47.2%	7.8%	37.7%	57.3%	43.9%
F1	29.8%	15.3%	29.8%	41.1%	70.6%	41.9%	53.9%	82.1%	NA	48.0%	10.8%	33.3%	53.4%	45.0%
Balanced Accuracy	63.5%	54.8%	65.9%	68.3%	85.0%	71.2%	75.0%	87.8%	50.0%	72.6%	53.0%	66.6%	77.3%	69.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 86. Individual Class Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1563	1	1141	2	1	66	457	1	0	108	887	165	13	271
Book	22	2938	35	1232	290	368	21	3436	1437	1	10	6	88	15
Catch	1605	32	2962	37	102	238	192	59	28	168	1323	550	264	644
Computer	4	1398	2	8329	354	99	53	3616	2873	0	5	6	0	4
Cycling	6	175	1	423	6200	403	21	85	821	0	3	5	13	47
Dust	98	261	180	192	730	4064	60	208	312	35	212	450	629	309
Jumping Jacks	352	0	135	0	1	42	4483	0	0	388	416	98	10	841
Lying	9	2209	19	4970	43	98	101	42807	2091	1	5	12	62	21
Reclining	0	77	0	314	92	25	12	25	224	0	9	1	4	8
Run	111	0	50	0	3	26	258	0	0	4952	128	4	5	43
Soccer	1160	1	849	4	5	152	348	11	16	218	2454	324	91	397
Stairs	518	30	1051	23	19	683	99	51	32	24	1006	4239	399	2462
Sweep	111	176	119	91	154	937	30	256	96	49	198	443	5633	420
Walk	332	38	772	8	23	469	489	29	11	214	652	1523	359	8858

Table 87. Individual Class Performance Summaries for the Left Wrist Combined Sensor Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	26.5%	40.0%	40.5%	53.3%	77.3%	53.0%	67.7%	84.6%	2.8%	80.4%	33.6%	54.2%	74.4%	61.8%
Specificity	98.0%	95.4%	96.6%	94.2%	98.7%	97.6%	98.5%	91.2%	99.6%	99.6%	97.7%	95.8%	98.0%	96.6%
PPV	33.4%	29.7%	36.1%	49.7%	75.6%	52.5%	66.3%	81.6%	28.3%	88.7%	40.7%	39.9%	64.7%	64.3%
NPV	97.2%	97.1%	97.1%	94.9%	98.8%	97.6%	98.6%	92.8%	95.2%	99.2%	96.9%	97.6%	98.7%	96.3%
Precision	33.4%	29.7%	36.1%	49.7%	75.6%	52.5%	66.3%	81.6%	28.3%	88.7%	40.7%	39.9%	64.7%	64.3%
Recall	26.5%	40.0%	40.5%	53.3%	77.3%	53.0%	67.7%	84.6%	2.8%	80.4%	33.6%	54.2%	74.4%	61.8%
F1	29.6%	34.1%	38.2%	51.5%	76.4%	52.7%	67.0%	83.1%	5.1%	84.4%	36.8%	45.9%	69.2%	63.0%
Balanced Accuracy	62.3%	67.7%	68.5%	73.7%	88.0%	75.3%	83.1%	87.9%	51.2%	90.0%	65.6%	75.0%	86.2%	79.2%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 88. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9267	1391	1678	2879	1360	886
CWR	1080	11898	2737	292	2406	52
Rec_LTA	1252	3749	5989	290	2774	249
SLA	1892	276	286	75595	159	4794
Sports	1407	3152	3717	326	13778	130
Wheels	342	32	43	2104	38	1906

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 89. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	60.8%	58.0%	41.4%	92.8%	67.2%	23.8%
Specificity	94.3%	95.3%	94.3%	90.6%	93.7%	98.3%
PPV	53.1%	64.4%	41.9%	91.1%	61.2%	42.7%
NPV	95.8%	93.9%	94.2%	92.4%	95.1%	96.1%
Precision	53.1%	64.4%	41.9%	91.1%	61.2%	42.7%
Recall	60.8%	58.0%	41.4%	92.8%	67.2%	23.8%
F1	56.7%	61.1%	41.7%	91.9%	64.0%	30.5%
Balanced Accuracy	77.6%	76.7%	67.9%	91.7%	80.5%	61.0%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 90. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	8017	1957	1517	1591	1303	1000
CWR	1322	10906	3465	130	2839	126
Rec_LTA	1088	3535	4736	112	3708	59
SLA	1844	190	279	77996	255	2862
Sports	2076	3845	4438	636	12372	276
Wheels	893	65	15	1021	38	3694

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 91. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	52.6%	53.2%	32.8%	95.7%	60.3%	46.1%
Specificity	94.9%	94.4%	94.2%	93.1%	91.9%	98.7%
PPV	52.1%	58.0%	35.8%	93.5%	52.3%	64.5%
NPV	95.0%	93.2%	93.4%	95.5%	94.0%	97.2%
Precision	52.1%	58.0%	35.8%	93.5%	52.3%	64.5%
Recall	52.6%	53.2%	32.8%	95.7%	60.3%	46.1%
F1	52.4%	55.5%	34.2%	94.6%	56.0%	53.8%
Balanced Accuracy	73.8%	73.8%	63.5%	94.4%	76.1%	72.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 92. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10203	1527	1088	1339	1218	952
CWR	928	12421	2699	163	2338	126
Rec_LTA	915	3401	6827	309	2768	87
SLA	1331	93	345	78416	143	1469
Sports	1492	3023	3450	455	14005	148
Wheels	371	33	41	804	43	5235

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 93. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	66.9%	60.6%	47.2%	96.2%	68.3%	65.3%
Specificity	95.8%	95.5%	94.9%	95.7%	93.9%	99.2%
PPV	62.5%	66.5%	47.7%	95.9%	62.0%	80.2%
NPV	96.5%	94.3%	94.8%	96.1%	95.3%	98.2%
Precision	62.5%	66.5%	47.7%	95.9%	62.0%	80.2%
Recall	66.9%	60.6%	47.2%	96.2%	68.3%	65.3%
F1	64.6%	63.4%	47.5%	96.0%	65.0%	72.0%
Balanced Accuracy	81.4%	78.1%	71.1%	96.0%	81.1%	82.2%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 94. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9246	2218	2289	1713	1579	621
CWR	1876	13216	4343	1209	3094	27
Rec_LTA	182	2207	4371	28	1200	73
SLA	3092	281	280	78222	96	6793
Sports	762	2571	3136	173	14539	111
Wheels	82	5	31	141	7	392

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 95. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	60.7%	64.5%	30.2%	96.0%	70.9%	4.9%
Specificity	94.2%	92.4%	97.5%	86.6%	95.2%	99.8%
PPV	52.3%	55.6%	54.2%	88.1%	68.3%	59.6%
NPV	95.8%	94.7%	93.4%	95.4%	95.7%	95.2%
Precision	52.3%	55.6%	54.2%	88.1%	68.3%	59.6%
Recall	60.7%	64.5%	30.2%	96.0%	70.9%	4.9%
F1	56.2%	59.7%	38.8%	91.9%	69.6%	9.0%
Balanced Accuracy	77.4%	78.5%	63.9%	91.3%	83.0%	52.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 96. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9379	1545	1346	1330	1020	1052
CWR	1654	13426	4608	55	3306	42
Rec_LTA	8	1472	3144	0	767	0
SLA	2345	138	311	79663	187	5733
Sports	1754	3904	5040	420	15235	335
Wheels	100	13	1	18	0	855

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 97. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	61.5%	65.5%	21.8%	97.8%	74.3%	10.7%
Specificity	95.7%	93.1%	98.5%	88.9%	91.8%	99.9%
PPV	59.8%	58.1%	58.3%	90.1%	57.1%	86.6%
NPV	95.9%	94.8%	92.7%	97.5%	96.0%	95.5%
Precision	59.8%	58.1%	58.3%	90.1%	57.1%	86.6%
Recall	61.5%	65.5%	21.8%	97.8%	74.3%	10.7%
F1	60.7%	61.6%	31.7%	93.8%	64.6%	19.0%
Balanced Accuracy	78.6%	79.3%	60.1%	93.3%	83.0%	55.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 98. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10839	1336	1334	995	1246	528
CWR	1241	14087	4567	82	2915	22
Rec_LTA	269	2181	4929	82	1459	31
SLA	1518	52	298	79678	162	5053
Sports	1159	2806	3299	297	14718	143
Wheels	214	36	23	352	15	2240

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 99. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	71.1%	68.7%	34.1%	97.8%	71.7%	27.9%
Specificity	96.2%	93.7%	97.2%	91.0%	94.5%	99.6%
PPV	66.6%	61.5%	55.1%	91.8%	65.6%	77.8%
NPV	96.9%	95.3%	93.7%	97.5%	95.8%	96.3%
Precision	66.6%	61.5%	55.1%	91.8%	65.6%	77.8%
Recall	71.1%	68.7%	34.1%	97.8%	71.7%	27.9%
F1	68.8%	64.9%	42.1%	94.7%	68.6%	41.1%
Balanced Accuracy	83.7%	81.2%	65.7%	94.4%	83.1%	63.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 100. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10820	1336	1489	2317	1171	1041
CWR	595	13245	2306	95	1721	22
Rec_LTA	941	2423	6517	124	1566	146
SLA	1689	215	268	77973	81	5239
Sports	1131	3278	3862	252	15970	133
Wheels	64	1	8	725	6	1436

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 101. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	71.0%	64.6%	45.1%	95.7%	77.8%	17.9%
Specificity	94.9%	96.6%	96.4%	90.5%	93.8%	99.5%
PPV	59.5%	73.6%	55.6%	91.2%	64.9%	64.1%
NPV	96.9%	94.9%	94.7%	95.3%	96.6%	95.8%
Precision	59.5%	73.6%	55.6%	91.2%	64.9%	64.1%
Recall	71.0%	64.6%	45.1%	95.7%	77.8%	17.9%
F1	64.8%	68.8%	49.8%	93.4%	70.8%	28.0%
Balanced Accuracy	83.0%	80.6%	70.8%	93.1%	85.8%	58.7%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 102. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9374	1873	1515	1533	1108	1102
CWR	1074	11835	3444	53	2250	86
Rec_LTA	602	2768	4481	56	2811	23
SLA	1749	129	236	78709	261	2539
Sports	1889	3861	4764	518	14070	287
Wheels	552	32	10	617	15	3980

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 103. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	61.5%	57.7%	31.0%	96.6%	68.6%	49.6%
Specificity	95.1%	95.1%	95.7%	93.8%	91.9%	99.2%
PPV	56.8%	63.1%	41.7%	94.1%	55.4%	76.5%
NPV	95.9%	93.9%	93.3%	96.4%	95.2%	97.4%
Precision	56.8%	63.1%	41.7%	94.1%	55.4%	76.5%
Recall	61.5%	57.7%	31.0%	96.6%	68.6%	49.6%
F1	59.1%	60.3%	35.6%	95.3%	61.3%	60.2%
Balanced Accuracy	78.3%	76.4%	63.4%	95.2%	80.2%	74.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 104. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	11666	1129	1029	1195	959	889
CWR	472	13835	2126	36	1629	34
Rec_LTA	629	2286	7496	87	1565	26
SLA	983	38	254	79527	86	1309
Sports	1300	3182	3533	227	16274	165
Wheels	190	28	12	414	2	5594

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 105. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	76.5%	67.5%	51.9%	97.6%	79.3%	69.8%
Specificity	96.4%	96.9%	96.8%	96.6%	94.0%	99.6%
PPV	69.2%	76.3%	62.0%	96.8%	65.9%	89.6%
NPV	97.5%	95.3%	95.3%	97.5%	96.9%	98.4%
Precision	69.2%	76.3%	62.0%	96.8%	65.9%	89.6%
Recall	76.5%	67.5%	51.9%	97.6%	79.3%	69.8%
F1	72.7%	71.6%	56.5%	97.2%	72.0%	78.5%
Balanced Accuracy	86.5%	82.2%	74.4%	97.1%	86.7%	84.7%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 106. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10595	1205	1439	2276	1368	687
CWR	996	14193	3138	126	2017	23
Rec_LTA	759	2231	6179	92	1645	64
SLA	1669	108	256	77101	96	4796
Sports	1079	2751	3418	266	15386	106
Wheels	142	10	20	1625	3	2341

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 107. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	69.5%	69.2%	42.8%	94.6%	75.0%	29.2%
Specificity	95.2%	95.5%	96.7%	91.2%	94.5%	98.8%
PPV	60.3%	69.3%	56.3%	91.8%	66.9%	56.5%
NPV	96.7%	95.5%	94.5%	94.2%	96.3%	96.4%
Precision	60.3%	69.3%	56.3%	91.8%	66.9%	56.5%
Recall	69.5%	69.2%	42.8%	94.6%	75.0%	29.2%
F1	64.6%	69.2%	48.6%	93.2%	70.7%	38.5%
Balanced Accuracy	82.4%	82.4%	69.7%	92.9%	84.8%	64.0%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 108. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10390	1636	1436	1659	895	1039
CWR	864	12786	3399	40	2084	90
Rec_LTA	397	1915	4312	32	1766	1
SLA	1238	76	190	78551	129	1376
Sports	1787	4044	5103	461	15626	255
Wheels	564	41	10	743	15	5256

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 109. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	68.2%	62.4%	29.8%	96.4%	76.2%	65.6%
Specificity	95.4%	95.4%	97.2%	96.2%	91.7%	99.1%
PPV	60.9%	66.4%	51.2%	96.3%	57.3%	79.3%
NPV	96.6%	94.5%	93.3%	96.3%	96.3%	98.2%
Precision	60.9%	66.4%	51.2%	96.3%	57.3%	79.3%
Recall	68.2%	62.4%	29.8%	96.4%	76.2%	65.6%
F1	64.3%	64.3%	37.7%	96.4%	65.4%	71.8%
Balanced Accuracy	81.8%	78.9%	63.5%	96.3%	83.9%	82.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 110. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	11514	940	925	1402	1024	682
CWR	578	14363	2414	51	1475	45
Rec_LTA	770	2398	7604	117	2008	29
SLA	967	47	261	78972	116	965
Sports	1143	2712	3233	206	15885	139
Wheels	268	38	13	738	7	6157

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 111. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	75.6%	70.1%	52.6%	96.9%	77.4%	76.8%
Specificity	96.6%	96.7%	96.3%	97.0%	94.7%	99.3%
PPV	69.8%	75.9%	58.8%	97.1%	68.1%	85.3%
NPV	97.4%	95.7%	95.4%	96.8%	96.6%	98.8%
Precision	69.8%	75.9%	58.8%	97.1%	68.1%	85.3%
Recall	75.6%	70.1%	52.6%	96.9%	77.4%	76.8%
F1	72.6%	72.9%	55.6%	97.0%	72.5%	80.8%
Balanced Accuracy	86.1%	83.4%	74.5%	97.0%	86.1%	88.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 112. Individual Class Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1680	2	1775	9	7	273	963	21	3	330	1021	409	68	363
Book	6	2730	8	1353	788	255	14	9067	1965	7	5	63	835	172
Catch	1465	23	2651	35	12	278	336	31	8	139	1364	236	114	273
Computer	3	796	1	9468	1492	213	5	6615	1246	9	8	67	111	44
Cycling	4	555	9	1146	2815	424	12	729	603	2	7	36	73	44
Dust	308	342	369	226	688	2538	94	438	284	59	334	957	1264	916
Jumping Jacks	436	2	227	2	6	100	3194	0	0	722	356	64	5	509
Lying	12	1595	21	1797	289	204	17	28753	1527	2	14	51	305	114
Reclining	10	832	5	1384	1388	219	17	4337	1852	4	6	40	162	52
Run	214	0	111	0	0	27	626	0	0	4164	119	15	7	91
Soccer	1028	7	1297	3	22	341	607	16	9	285	2099	604	139	1003
Stairs	224	39	280	48	275	1087	54	95	63	36	829	2976	878	1847
Sweep	109	291	155	115	107	1096	40	374	225	30	185	674	2975	482
Walk	392	122	407	39	128	615	645	108	156	369	961	1634	634	8430

Table 113. Individual Class Performance Summaries for the Right Wrist Accelerometer Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	28.5%	37.2%	36.2%	60.6%	35.1%	33.1%	48.2%	56.8%	23.3%	67.6%	28.7%	38.0%	39.3%	58.8%
Specificity	96.6%	90.5%	97.2%	92.7%	97.6%	95.9%	98.4%	94.6%	94.4%	99.2%	96.5%	96.2%	97.5%	95.7%
PPV	24.3%	15.8%	38.1%	47.2%	43.6%	28.8%	56.8%	82.9%	18.0%	77.5%	28.1%	34.1%	43.4%	57.6%
NPV	97.3%	96.8%	97.0%	95.6%	96.6%	96.6%	97.8%	82.6%	95.9%	98.7%	96.6%	96.8%	97.0%	95.9%
Precision	24.3%	15.8%	38.1%	47.2%	43.6%	28.8%	56.8%	82.9%	18.0%	77.5%	28.1%	34.1%	43.4%	57.6%
Recall	28.5%	37.2%	36.2%	60.6%	35.1%	33.1%	48.2%	56.8%	23.3%	67.6%	28.7%	38.0%	39.3%	58.8%
F1	26.2%	22.2%	37.1%	53.0%	38.9%	30.8%	52.2%	67.4%	20.3%	72.2%	28.4%	35.9%	41.2%	58.2%
Balanced Accuracy	62.6%	63.9%	66.7%	76.6%	66.4%	64.5%	73.3%	75.7%	58.9%	83.4%	62.6%	67.1%	68.4%	77.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 114. Individual Class Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1681	39	1624	44	44	300	727	27	30	324	1139	413	77	639
Book	11	1551	32	2416	765	250	12	5720	1423	7	31	23	158	100
Catch	1121	41	1990	21	68	379	428	66	46	109	1025	469	169	583
Computer	21	2386	29	7718	919	230	10	4344	1884	5	34	62	191	35
Cycling	16	320	42	1151	5119	424	17	130	528	7	20	35	229	134
Dust	151	203	370	283	251	1774	99	153	188	105	578	1169	1223	1178
Jumping Jacks	500	6	395	1	8	43	2997	5	0	706	396	74	5	472
Lying	8	1599	57	1980	103	78	6	35447	2724	3	26	44	64	12
Reclining	2	932	10	1816	270	135	13	4506	858	3	25	16	146	29
Run	231	0	83	3	4	60	878	1	0	2487	121	89	3	1491
Soccer	1175	72	1351	46	48	710	646	39	37	244	1474	908	230	1029
Stairs	344	20	602	14	32	1140	73	25	21	219	1190	1927	740	1852
Sweep	101	145	104	107	300	1278	6	89	143	45	335	823	3109	1553
Walk	529	22	627	25	86	869	712	32	59	1894	914	1774	1226	5233

Table 115. Individual Class Performance Summaries for the Right Wrist Gyroscope Only Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	28.5%	21.1%	27.2%	49.4%	63.9%	23.1%	45.2%	70.1%	10.8%	40.4%	20.2%	24.6%	41.1%	36.5%
Specificity	96.5%	92.8%	97.0%	93.0%	98.0%	96.1%	98.3%	93.9%	94.8%	98.1%	95.7%	95.9%	96.7%	94.0%
PPV	23.6%	12.4%	30.5%	43.2%	62.6%	23.0%	53.4%	84.1%	9.8%	45.6%	18.4%	23.5%	38.2%	37.4%
NPV	97.3%	96.1%	96.5%	94.4%	98.1%	96.1%	97.7%	87.2%	95.3%	97.6%	96.2%	96.1%	97.1%	93.8%
Precision	23.6%	12.4%	30.5%	43.2%	62.6%	23.0%	53.4%	84.1%	9.8%	45.6%	18.4%	23.5%	38.2%	37.4%
Recall	28.5%	21.1%	27.2%	49.4%	63.9%	23.1%	45.2%	70.1%	10.8%	40.4%	20.2%	24.6%	41.1%	36.5%
F1	25.9%	15.6%	28.8%	46.1%	63.2%	23.0%	49.0%	76.4%	10.3%	42.8%	19.2%	24.0%	39.6%	36.9%
Balanced Accuracy	62.5%	57.0%	62.1%	71.2%	80.9%	59.6%	71.8%	82.0%	52.8%	69.2%	57.9%	60.3%	68.9%	65.2%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 116. Individual Class Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1846	16	1978	7	10	289	830	18	5	192	968	212	91	455
Book	10	3162	12	1009	221	229	10	7283	2174	2	7	16	226	33
Catch	1407	17	2665	37	35	226	346	15	12	106	1055	195	123	240
Computer	6	1115	6	10747	770	164	3	4225	1369	0	12	30	16	20
Cycling	10	158	5	618	5486	266	7	67	531	1	11	22	48	30
Dust	177	212	254	128	585	3043	72	171	215	95	326	1035	982	661
Jumping Jacks	485	0	220	2	6	69	3563	0	8	356	244	41	2	332
Lying	3	1322	4	955	56	61	2	34191	1567	3	7	14	46	24
Reclining	7	1025	19	1995	427	255	13	4386	1886	4	13	7	45	19
Run	219	8	92	1	18	75	404	2	1	4665	181	31	12	73
Soccer	1001	17	1234	16	8	446	481	44	12	379	2666	455	108	736
Stairs	212	27	324	20	52	1159	64	51	25	44	744	3547	600	1718
Sweep	62	194	97	57	131	873	22	103	100	48	142	396	4850	525
Walk	446	63	406	33	212	515	807	28	36	263	932	1825	421	9474

Table 117. Individual Class Performance Summaries for the Right Wrist Combined Sensor Feature Set Decision Tree Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	31.3%	43.1%	36.4%	68.8%	68.4%	39.7%	53.8%	67.6%	23.8%	75.8%	36.5%	45.3%	64.1%	66.1%
Specificity	96.7%	92.7%	97.5%	94.6%	98.8%	96.8%	98.9%	96.3%	94.6%	99.3%	96.8%	96.7%	98.2%	95.9%
PPV	26.7%	22.0%	41.1%	58.1%	75.6%	38.2%	66.9%	89.4%	18.7%	80.7%	35.1%	41.3%	63.8%	61.3%
NPV	97.4%	97.1%	97.0%	96.6%	98.3%	97.0%	98.0%	86.6%	96.0%	99.0%	97.0%	97.2%	98.2%	96.6%
Precision	26.7%	22.0%	41.1%	58.1%	75.6%	38.2%	66.9%	89.4%	18.7%	80.7%	35.1%	41.3%	63.8%	61.3%
Recall	31.3%	43.1%	36.4%	68.8%	68.4%	39.7%	53.8%	67.6%	23.8%	75.8%	36.5%	45.3%	64.1%	66.1%
F1	28.8%	29.1%	38.6%	63.0%	71.8%	38.9%	59.6%	77.0%	20.9%	78.1%	35.8%	43.2%	63.9%	63.6%
Balanced Accuracy	64.0%	67.9%	67.0%	81.7%	83.6%	68.2%	76.3%	81.9%	59.2%	87.5%	66.6%	71.0%	81.1%	81.0%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 118. Individual Class Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1311	7	592	0	1	134	352	0	0	138	409	77	4	89
Book	12	542	6	54	23	212	7	547	234	1	1	10	489	91
Catch	1953	3	4118	1	1	199	295	15	0	199	1767	74	49	133
Computer	11	727	11	8135	2311	350	10	5081	1049	0	19	69	59	75
Cycling	1	183	6	310	1170	828	17	1417	462	3	16	31	134	52
Dust	108	56	149	98	451	1600	60	182	91	19	91	1043	377	289
Jumping Jacks	506	0	122	0	0	178	3918	0	0	643	334	29	0	274
Lying	5	5356	4	6850	3653	234	8	42485	5640	0	13	11	234	88
Reclining	0	14	0	4	2	4	0	27	23	1	0	0	7	0
Run	314	0	170	0	0	19	577	4	0	4655	102	47	8	82
Soccer	857	1	1023	0	2	148	381	2	0	111	1733	199	21	486
Stairs	168	7	220	23	17	740	135	36	25	11	289	1408	321	465
Sweep	158	382	228	79	103	1039	68	414	311	18	145	633	4653	665
Walk	487	58	667	71	283	1985	796	374	106	359	2389	4195	1214	11551

Table 119. Individual Class Performance Summaries for the Right Wrist Accelerometer Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	22.3%	7.4%	56.3%	52.1%	14.6%	20.9%	59.1%	84.0%	0.3%	75.6%	23.7%	18.0%	61.5%	80.6%
Specificity	98.8%	98.9%	96.9%	93.2%	97.7%	98.0%	98.6%	79.8%	100.0%	99.1%	97.9%	98.4%	97.2%	91.1%
PPV	42.1%	24.3%	46.8%	45.4%	25.3%	34.7%	65.3%	65.8%	28.0%	77.9%	34.9%	36.4%	52.3%	47.1%
NPV	97.1%	95.7%	97.9%	94.7%	95.6%	96.1%	98.2%	91.5%	95.1%	99.0%	96.4%	95.9%	98.1%	97.9%
Precision	42.1%	24.3%	46.8%	45.4%	25.3%	34.7%	65.3%	65.8%	28.0%	77.9%	34.9%	36.4%	52.3%	47.1%
Recall	22.3%	7.4%	56.3%	52.1%	14.6%	20.9%	59.1%	84.0%	0.3%	75.6%	23.7%	18.0%	61.5%	80.6%
F1	29.1%	11.3%	51.1%	48.5%	18.5%	26.1%	62.1%	73.8%	0.6%	76.7%	28.2%	24.1%	56.5%	59.4%
Balanced Accuracy	60.5%	53.1%	76.6%	72.7%	56.2%	59.4%	78.9%	81.9%	50.1%	87.4%	60.8%	58.2%	79.3%	85.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 120. Individual Class Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1526	0	1210	0	2	52	638	0	0	241	799	164	1	391
Book	20	514	23	437	358	309	11	278	364	7	63	13	113	88
Catch	1810	16	3542	31	118	264	352	47	17	83	1651	465	74	376
Computer	10	2389	23	3823	1891	318	26	3860	1945	1	39	8	191	44
Cycling	4	138	0	254	3391	353	5	65	261	0	0	51	252	150
Dust	186	226	394	182	315	2324	53	216	144	55	774	1065	922	677
Jumping Jacks	898	3	96	5	0	61	3762	58	6	642	452	98	0	596
Lying	6	3880	2	10785	1617	23	5	45879	4989	1	12	0	10	16
Reclining	0	1	0	9	6	2	0	19	9	1	2	0	0	3
Run	117	0	7	0	0	14	896	0	0	1899	76	14	0	918
Soccer	584	8	892	0	8	230	188	0	7	110	893	339	41	418
Stairs	179	14	284	2	18	493	22	31	15	34	600	962	219	544
Sweep	33	118	27	72	258	1181	9	85	164	15	90	615	3466	991
Walk	518	29	816	25	35	2046	657	46	20	3069	1857	4032	2281	9128

Table 121. Individual Class Performance Summaries for the Right Wrist Gyroscope Only Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	25.9%	7.0%	48.4%	24.5%	42.3%	30.3%	56.8%	90.7%	0.1%	30.8%	12.2%	12.3%	45.8%	63.7%
Specificity	97.7%	98.6%	96.5%	92.6%	99.0%	96.6%	98.1%	80.5%	100.0%	98.7%	98.2%	98.4%	97.6%	89.4%
PPV	30.4%	19.8%	40.0%	26.2%	68.9%	30.9%	56.3%	68.2%	17.3%	48.2%	24.0%	28.2%	48.7%	37.2%
NPV	97.2%	95.7%	97.5%	91.9%	97.0%	96.5%	98.1%	94.9%	95.0%	97.3%	95.9%	95.6%	97.3%	96.2%
Precision	30.4%	19.8%	40.0%	26.2%	68.9%	30.9%	56.3%	68.2%	17.3%	48.2%	24.0%	28.2%	48.7%	37.2%
Recall	25.9%	7.0%	48.4%	24.5%	42.3%	30.3%	56.8%	90.7%	0.1%	30.8%	12.2%	12.3%	45.8%	63.7%
F1	28.0%	10.3%	43.8%	25.3%	52.4%	30.6%	56.6%	77.9%	0.2%	37.6%	16.2%	17.1%	47.2%	46.9%
Balanced Accuracy	61.8%	52.8%	72.5%	58.5%	70.6%	63.4%	77.4%	85.6%	50.0%	64.8%	55.2%	55.3%	71.7%	76.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 122. Individual Class Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1869	2	1170	1	4	75	424	2	4	58	573	118	1	214
Book	11	2065	16	199	178	289	15	1994	926	3	28	10	256	65
Catch	1643	8	3749	1	3	108	279	26	1	139	1104	102	41	129
Computer	4	1020	4	10151	1347	146	2	2971	1499	0	5	6	7	4
Cycling	1	52	3	235	4473	621	12	45	390	0	4	20	89	31
Dust	108	140	181	137	319	2793	47	190	152	72	268	873	614	357
Jumping Jacks	621	0	127	1	0	97	4265	21	0	288	237	23	0	314
Lying	14	3743	9	4754	1152	20	12	44911	4522	0	4	0	12	11
Reclining	2	148	4	69	268	63	6	236	314	0	7	4	11	11
Run	112	0	51	0	0	50	286	0	0	5008	120	18	8	28
Soccer	740	0	1024	3	8	412	359	11	3	249	2594	182	96	345
Stairs	262	17	350	13	44	788	195	42	24	19	530	2570	318	1089
Sweep	56	107	133	33	145	874	10	71	72	29	152	653	5290	672
Walk	448	34	495	28	76	1334	712	64	34	293	1682	3247	827	11070

Table 123. Individual Class Performance Summaries for the Right Wrist Combined Sensor Feature Set Lasso Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	31.7%	28.1%	51.2%	65.0%	55.8%	36.4%	64.4%	88.8%	4.0%	81.3%	35.5%	32.8%	69.9%	77.2%
Specificity	98.3%	97.4%	97.7%	95.1%	99.0%	97.7%	98.9%	87.0%	99.5%	99.6%	97.8%	97.6%	98.0%	93.6%
PPV	41.4%	34.1%	51.1%	59.1%	74.8%	44.7%	71.2%	75.9%	27.5%	88.2%	43.0%	41.0%	63.8%	54.4%
NPV	97.4%	96.6%	97.7%	96.2%	97.7%	96.8%	98.5%	94.4%	95.2%	99.3%	96.9%	96.6%	98.5%	97.7%
Precision	41.4%	34.1%	51.1%	59.1%	74.8%	44.7%	71.2%	75.9%	27.5%	88.2%	43.0%	41.0%	63.8%	54.4%
Recall	31.7%	28.1%	51.2%	65.0%	55.8%	36.4%	64.4%	88.8%	4.0%	81.3%	35.5%	32.8%	69.9%	77.2%
F1	35.9%	30.8%	51.2%	61.9%	63.9%	40.1%	67.6%	81.8%	6.9%	84.6%	38.9%	36.5%	66.7%	63.8%
Balanced Accuracy	65.0%	62.8%	74.4%	80.1%	77.4%	67.1%	81.6%	87.9%	51.7%	90.4%	66.6%	65.2%	84.0%	85.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 124. Individual Class Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1995	2	1556	2	7	224	601	20	4	300	817	117	36	245
Book	4	3174	12	804	309	274	11	8319	2060	1	1	8	577	147
Catch	1588	11	3281	8	6	255	374	13	1	169	1373	152	86	216
Computer	1	887	0	11628	1977	168	2	7996	1384	0	4	42	46	24
Cycling	0	539	1	685	3467	307	3	676	855	0	1	7	13	5
Dust	178	254	226	150	718	3215	49	263	246	35	222	995	984	748
Jumping Jacks	354	0	106	0	3	116	3884	0	0	579	219	24	0	343
Lying	3	1594	12	1236	268	106	6	29170	1649	3	1	16	101	64
Reclining	3	432	2	900	857	84	12	3607	1411	0	4	11	25	17
Run	211	2	61	1	0	10	531	0	0	4562	111	17	1	60
Soccer	991	3	1317	1	11	273	492	7	3	227	2597	400	83	641
Stairs	185	22	287	50	180	928	70	67	52	43	796	3805	654	1593
Sweep	121	361	135	124	104	1211	33	418	253	22	168	670	4540	619
Walk	257	55	320	36	110	499	556	28	23	217	994	1562	424	9618

Table 125. Individual Class Performance Summaries for the Right Wrist Accelerometer Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	33.9%	43.3%	44.8%	74.4%	43.2%	41.9%	58.6%	57.7%	17.8%	74.1%	35.5%	48.6%	60.0%	67.1%
Specificity	97.5%	91.8%	97.2%	91.3%	98.0%	96.7%	98.9%	95.4%	96.1%	99.3%	97.1%	96.8%	97.2%	96.5%
PPV	33.7%	20.2%	43.6%	48.1%	52.9%	38.8%	69.0%	85.2%	19.2%	81.9%	36.9%	43.6%	51.7%	65.4%
NPV	97.5%	97.1%	97.4%	97.1%	97.0%	97.1%	98.2%	83.0%	95.7%	99.0%	96.9%	97.3%	98.0%	96.8%
Precision	33.7%	20.2%	43.6%	48.1%	52.9%	38.8%	69.0%	85.2%	19.2%	81.9%	36.9%	43.6%	51.7%	65.4%
Recall	33.9%	43.3%	44.8%	74.4%	43.2%	41.9%	58.6%	57.7%	17.8%	74.1%	35.5%	48.6%	60.0%	67.1%
F1	33.8%	27.6%	44.2%	58.5%	47.6%	40.3%	63.4%	68.8%	18.4%	77.8%	36.2%	46.0%	55.5%	66.2%
Balanced Accuracy	65.7%	67.5%	71.0%	82.9%	70.6%	69.3%	78.7%	76.5%	56.9%	86.7%	66.3%	72.7%	78.6%	81.8%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 126. Individual Class Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	2065	28	1825	19	29	230	877	26	15	312	1182	283	67	651
Book	16	1209	22	1674	889	227	5	2679	1022	6	37	26	142	66
Catch	1180	34	2266	40	64	292	484	51	30	66	974	354	71	413
Computer	9	2916	17	9721	1082	208	20	4421	2387	5	44	15	49	30
Cycling	10	297	12	790	4815	290	16	91	523	1	14	9	202	130
Dust	182	207	334	202	304	2347	60	176	187	62	637	1220	1307	833
Jumping Jacks	494	0	271	0	3	33	3039	0	0	772	404	69	4	451
Lying	2	1913	8	2278	78	32	4	41259	3011	7	8	5	52	4
Reclining	3	523	9	721	284	93	6	1692	515	3	10	4	42	14
Run	212	3	31	0	6	29	961	1	3	2880	121	56	5	1457
Soccer	1091	39	1553	29	67	527	396	41	18	232	1554	931	155	865
Stairs	198	15	482	16	17	1155	29	20	13	166	1195	2219	575	1595
Sweep	61	125	67	121	305	1187	6	87	182	19	241	734	3570	1502
Walk	368	27	419	14	74	1020	721	40	35	1627	887	1901	1329	6329

Table 127. Individual Class Performance Summaries for the Right Wrist Gyroscope Only Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	35.1%	16.5%	31.0%	62.2%	60.1%	30.6%	45.9%	81.6%	6.5%	46.8%	21.3%	28.4%	47.2%	44.1%
Specificity	96.4%	95.5%	97.3%	92.3%	98.4%	96.3%	98.4%	93.2%	97.8%	98.1%	96.1%	96.4%	97.0%	94.2%
PPV	27.1%	15.1%	35.9%	46.5%	66.9%	29.1%	54.9%	84.8%	13.1%	50.0%	20.7%	28.8%	43.5%	42.8%
NPV	97.5%	96.0%	96.7%	95.8%	97.9%	96.5%	97.7%	91.6%	95.2%	97.9%	96.2%	96.3%	97.4%	94.5%
Precision	27.1%	15.1%	35.9%	46.5%	66.9%	29.1%	54.9%	84.8%	13.1%	50.0%	20.7%	28.8%	43.5%	42.8%
Recall	35.1%	16.5%	31.0%	62.2%	60.1%	30.6%	45.9%	81.6%	6.5%	46.8%	21.3%	28.4%	47.2%	44.1%
F1	30.6%	15.7%	33.2%	53.2%	63.3%	29.8%	50.0%	83.1%	8.7%	48.3%	21.0%	28.6%	45.3%	43.5%
Balanced Accuracy	65.7%	56.0%	64.2%	77.2%	79.2%	63.4%	72.1%	87.4%	52.1%	72.4%	58.7%	62.4%	72.1%	69.2%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 128. Individual Class Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	2361	0	1705	0	9	150	666	8	1	107	706	99	25	339
Book	6	3944	17	697	159	185	8	4693	2371	1	3	4	171	35
Catch	1429	13	3475	4	14	198	290	17	0	121	1010	183	70	153
Computer	0	887	1	12320	800	87	6	3644	1022	0	1	8	6	3
Cycling	10	91	0	544	5881	260	2	29	417	0	4	6	22	29
Dust	131	225	215	173	649	3876	51	166	205	75	218	873	834	557
Jumping Jacks	394	0	123	0	0	48	4387	0	0	263	191	18	3	329
Lying	0	1462	3	849	14	11	1	38363	1725	3	6	2	14	4
Reclining	3	573	0	956	319	100	15	3490	2081	0	8	3	9	2
Run	132	0	15	0	1	33	301	0	0	5024	137	14	0	50
Soccer	800	4	986	1	6	441	321	7	13	384	3226	321	115	580
Stairs	222	24	307	19	48	967	58	51	21	29	736	4404	394	1565
Sweep	55	93	94	35	91	854	32	85	65	25	162	329	5624	491
Walk	348	20	375	27	26	460	486	31	20	126	900	1562	283	10203

Table 129. Individual Class Performance Summaries for the Right Wrist Combined Sensor Feature Set Random Forest Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	40.1%	53.8%	47.5%	78.8%	73.4%	50.5%	66.2%	75.8%	26.2%	81.6%	44.1%	56.3%	74.3%	71.2%
Specificity	97.5%	94.5%	97.7%	95.5%	99.1%	97.1%	99.1%	96.3%	96.4%	99.6%	97.4%	97.1%	98.4%	96.8%
PPV	38.2%	32.1%	49.8%	65.6%	80.6%	47.0%	76.2%	90.4%	27.5%	88.0%	44.8%	49.8%	70.0%	68.6%
NPV	97.7%	97.7%	97.5%	97.7%	98.6%	97.5%	98.6%	89.6%	96.2%	99.3%	97.3%	97.7%	98.7%	97.2%
Precision	38.2%	32.1%	49.8%	65.6%	80.6%	47.0%	76.2%	90.4%	27.5%	88.0%	44.8%	49.8%	70.0%	68.6%
Recall	40.1%	53.8%	47.5%	78.8%	73.4%	50.5%	66.2%	75.8%	26.2%	81.6%	44.1%	56.3%	74.3%	71.2%
F1	39.1%	40.2%	48.6%	71.6%	76.8%	48.7%	70.9%	82.5%	26.9%	84.7%	44.5%	52.8%	72.1%	69.9%
Balanced Accuracy	68.8%	74.2%	72.6%	87.2%	86.2%	73.8%	82.7%	86.1%	61.3%	90.6%	70.8%	76.7%	86.4%	84.0%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 130. Individual Class Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1290	0	636	0	0	156	340	0	0	219	570	77	4	152
Book	9	3225	5	294	187	270	10	3996	1647	2	2	8	674	142
Catch	2090	8	4177	1	1	268	349	24	2	181	1409	115	81	185
Computer	0	936	3	11080	2202	157	6	3618	1304	0	2	13	14	14
Cycling	2	614	5	833	3965	360	12	465	852	0	4	31	22	6
Dust	123	181	144	160	535	3046	58	264	226	50	189	1078	573	507
Jumping Jacks	507	0	101	0	1	177	4182	0	0	681	285	34	0	287
Lying	6	1805	20	2951	615	212	11	41240	3458	1	12	16	40	29
Reclining	0	101	0	101	10	7	1	496	145	0	0	0	1	0
Run	226	0	61	0	0	6	563	0	0	4529	105	34	2	93
Soccer	941	2	1287	1	8	215	444	6	3	201	2551	339	56	598
Stairs	185	18	292	59	131	771	66	43	16	37	829	3313	496	1084
Sweep	173	408	237	110	147	1242	36	385	250	27	173	718	5148	751
Walk	339	38	348	35	215	783	546	47	38	230	1177	2050	459	10492

Table 131. Individual Class Performance Summaries for the Right Wrist Accelerometer Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	21.9%	44.0%	57.1%	70.9%	49.5%	39.7%	63.1%	81.5%	1.8%	73.5%	34.9%	42.3%	68.0%	73.2%
Specificity	98.6%	95.3%	96.9%	94.3%	97.9%	97.3%	98.7%	91.6%	99.5%	99.3%	97.3%	97.4%	96.9%	95.7%
PPV	37.5%	30.8%	47.0%	57.3%	55.3%	42.7%	66.9%	81.8%	16.8%	80.6%	38.3%	45.1%	52.5%	62.5%
NPV	97.1%	97.3%	97.9%	96.8%	97.4%	97.0%	98.4%	91.5%	95.1%	98.9%	96.9%	97.0%	98.4%	97.3%
Precision	37.5%	30.8%	47.0%	57.3%	55.3%	42.7%	66.9%	81.8%	16.8%	80.6%	38.3%	45.1%	52.5%	62.5%
Recall	21.9%	44.0%	57.1%	70.9%	49.5%	39.7%	63.1%	81.5%	1.8%	73.5%	34.9%	42.3%	68.0%	73.2%
F1	27.6%	36.2%	51.5%	63.4%	52.2%	41.2%	64.9%	81.7%	3.3%	76.9%	36.5%	43.7%	59.3%	67.4%
Balanced Accuracy	60.3%	69.6%	77.0%	82.6%	73.7%	68.5%	80.9%	86.6%	50.7%	86.4%	66.1%	69.8%	82.5%	84.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 132. Individual Class Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1599	2	1182	0	3	66	590	0	0	220	754	166	0	452
Book	11	1117	31	792	331	222	9	721	591	2	59	9	184	38
Catch	1888	35	3867	47	104	341	347	46	13	85	1866	570	75	526
Computer	10	3354	3	11312	979	164	27	3129	2696	3	22	5	25	33
Cycling	6	252	6	820	5803	347	16	79	588	1	10	14	135	121
Dust	163	319	317	219	380	2580	39	214	235	60	685	1218	1169	726
Jumping Jacks	672	0	129	0	1	66	3919	0	0	604	387	80	1	422
Lying	0	2082	0	2292	27	0	0	46048	3551	0	3	0	0	5
Reclining	0	5	0	26	0	0	0	44	5	0	0	0	0	0
Run	159	0	17	0	0	32	672	0	0	3551	138	37	0	1533
Soccer	787	11	838	10	14	468	317	17	10	167	1194	505	79	589
Stairs	273	21	556	7	15	1253	49	35	12	112	1363	2835	523	1852
Sweep	46	126	27	89	322	1390	8	218	235	9	139	658	3900	1197
Walk	277	12	343	11	38	741	631	33	5	1344	688	1729	1479	6846

Table 133. Individual Class Performance Summaries for the Right Wrist Gyroscope Only Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	27.1%	15.2%	52.9%	72.4%	72.4%	33.6%	59.2%	91.0%	0.1%	57.7%	16.3%	36.2%	51.5%	47.7%
Specificity	97.8%	98.0%	96.1%	92.8%	98.4%	96.2%	98.5%	92.7%	100.0%	98.3%	97.5%	96.0%	97.1%	95.0%
PPV	31.8%	27.1%	39.4%	52.0%	70.8%	31.0%	62.4%	85.3%	6.3%	57.8%	23.9%	31.8%	46.6%	48.3%
NPV	97.2%	96.0%	97.7%	96.9%	98.5%	96.6%	98.2%	95.7%	95.0%	98.3%	96.1%	96.7%	97.6%	94.9%
Precision	31.8%	27.1%	39.4%	52.0%	70.8%	31.0%	62.4%	85.3%	6.3%	57.8%	23.9%	31.8%	46.6%	48.3%
Recall	27.1%	15.2%	52.9%	72.4%	72.4%	33.6%	59.2%	91.0%	0.1%	57.7%	16.3%	36.2%	51.5%	47.7%
F1	29.3%	19.5%	45.2%	60.5%	71.6%	32.3%	60.7%	88.1%	0.1%	57.8%	19.4%	33.9%	49.0%	48.0%
Balanced Accuracy	62.5%	56.6%	74.5%	82.6%	85.4%	64.9%	78.8%	91.9%	50.0%	78.0%	56.9%	66.1%	74.3%	71.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 134. Individual Class Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Basketball	1714	0	861	1	3	112	435	1	0	104	455	108	0	258
Book	11	3454	19	400	194	227	10	3292	2059	1	13	4	162	34
Catch	1831	13	4023	5	30	213	246	28	8	150	1055	152	86	134
Computer	2	1181	3	12626	922	81	9	2281	1312	0	6	6	8	1
Cycling	6	95	1	744	5921	252	8	60	647	0	9	3	11	34
Dust	96	212	154	166	425	3394	33	180	204	72	211	721	829	379
Jumping Jacks	545	0	91	3	7	46	4526	0	0	319	244	25	1	483
Lying	2	1915	6	1237	22	39	3	43833	2920	2	8	4	9	8
Reclining	0	249	1	343	119	47	12	719	657	1	3	1	3	7
Run	81	0	21	0	0	38	382	0	0	4947	80	0	0	20
Soccer	861	1	1243	4	4	417	317	5	5	331	3059	287	108	465
Stairs	316	48	395	27	61	1214	190	72	36	42	952	4051	489	1552
Sweep	42	151	108	45	202	913	6	77	76	23	172	330	5426	535
Walk	384	17	390	24	107	677	447	36	17	166	1041	2134	438	10430

Table 135. Individual Class Performance Summaries for the Right Wrist Combined Sensor Feature Set Neural Network Model.

	Basketball	Book	Catch	Computer	Cycling	Dust	Jumping Jacks	Lying	Reclining	Run	Soccer	Stairs	Sweep	Walk
Sensitivity	29.1%	47.1%	55.0%	80.8%	73.9%	44.3%	68.3%	86.7%	8.3%	80.3%	41.9%	51.8%	71.7%	72.7%
Specificity	98.5%	95.8%	97.4%	96.0%	98.8%	97.6%	98.9%	94.4%	99.0%	99.6%	97.4%	96.5%	98.2%	96.0%
PPV	42.3%	35.0%	50.5%	68.5%	76.0%	48.0%	72.0%	87.7%	30.4%	88.8%	43.0%	42.9%	66.9%	64.0%
NPV	97.3%	97.4%	97.8%	97.9%	98.6%	97.2%	98.6%	93.9%	95.4%	99.2%	97.2%	97.5%	98.6%	97.3%
Precision	42.3%	35.0%	50.5%	68.5%	76.0%	48.0%	72.0%	87.7%	30.4%	88.8%	43.0%	42.9%	66.9%	64.0%
Recall	29.1%	47.1%	55.0%	80.8%	73.9%	44.3%	68.3%	86.7%	8.3%	80.3%	41.9%	51.8%	71.7%	72.7%
F1	34.5%	40.1%	52.6%	74.1%	74.9%	46.0%	70.1%	87.2%	13.0%	84.4%	42.4%	46.9%	69.2%	68.1%
Balanced Accuracy	63.8%	71.4%	76.2%	88.4%	86.3%	70.9%	83.6%	90.5%	53.6%	90.0%	69.6%	74.1%	85.0%	84.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall

Table 136. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	8380	1683	1935	1902	1121	711
CWR	1131	12657	2728	216	2037	131
Rec_LTA	1676	2942	6265	169	2055	197
SLA	2396	398	294	75064	132	4354
Sports	1308	2807	3204	175	15153	49
Wheels	349	11	24	3960	17	2575

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 137. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	55.0%	61.7%	43.4%	92.1%	73.9%	32.1%
Specificity	94.9%	95.5%	95.2%	90.4%	94.6%	97.1%
PPV	53.3%	67.0%	47.1%	90.8%	66.8%	37.1%
NPV	95.3%	94.5%	94.4%	91.7%	96.1%	96.4%
Precision	53.3%	67.0%	47.1%	90.8%	66.8%	37.1%
Recall	55.0%	61.7%	43.4%	92.1%	73.9%	32.1%
F1	54.1%	64.3%	45.1%	91.5%	70.1%	34.4%
Balanced Accuracy	75.0%	78.6%	69.3%	91.2%	84.2%	64.6%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 138. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	7683	2951	2229	1237	1467	611
CWR	2112	11438	3660	181	2539	103
Rec_LTA	1670	2801	4715	103	2969	47
SLA	1271	220	148	77809	339	2362
Sports	2062	3000	3650	512	13170	175
Wheels	442	88	48	1644	31	4719

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 139. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	50.4%	55.8%	32.6%	95.5%	64.2%	58.9%
Specificity	94.1%	93.8%	94.8%	94.5%	93.3%	98.5%
PPV	47.5%	57.1%	38.3%	94.7%	58.4%	67.7%
NPV	94.8%	93.5%	93.4%	95.3%	94.7%	97.8%
Precision	47.5%	57.1%	38.3%	94.7%	58.4%	67.7%
Recall	50.4%	55.8%	32.6%	95.5%	64.2%	58.9%
F1	48.9%	56.4%	35.2%	95.1%	61.1%	63.0%
Balanced Accuracy	72.3%	74.8%	63.7%	95.0%	78.7%	78.7%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 140. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9575	1415	1638	1151	1127	703
CWR	897	13804	2696	130	2095	66
Rec_LTA	2083	2857	7203	148	2174	157
SLA	1079	105	155	78671	132	1737
Sports	1257	2282	2726	200	14952	83
Wheels	349	35	32	1186	35	5271

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 141. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Decision Tree Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	62.8%	67.3%	49.8%	96.5%	72.9%	65.7%
Specificity	95.8%	95.8%	94.9%	95.9%	95.3%	98.9%
PPV	61.3%	70.1%	49.3%	96.1%	69.5%	76.3%
NPV	96.1%	95.2%	95.0%	96.4%	96.0%	98.2%
Precision	61.3%	70.1%	49.3%	96.1%	69.5%	76.3%
Recall	62.8%	67.3%	49.8%	96.5%	72.9%	65.7%
F1	62.1%	68.7%	49.6%	96.3%	71.2%	70.6%
Balanced Accuracy	79.3%	81.6%	72.4%	96.2%	84.1%	82.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 142. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	6952	937	2046	1184	808	616
CWR	3017	15640	5659	432	3257	236
Rec_LTA	1002	1667	4279	63	950	26
SLA	3056	426	116	79321	105	6959
Sports	945	1803	2310	53	15388	40
Wheels	268	25	40	433	7	140

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 143. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	45.6%	76.3%	29.6%	97.3%	75.0%	1.7%
Specificity	96.1%	91.0%	97.5%	86.5%	96.3%	99.5%
PPV	55.4%	55.4%	53.6%	88.2%	74.9%	15.3%
NPV	94.4%	96.3%	93.3%	96.9%	96.3%	95.1%
Precision	55.4%	55.4%	53.6%	88.2%	74.9%	15.3%
Recall	45.6%	76.3%	29.6%	97.3%	75.0%	1.7%
F1	50.0%	64.2%	38.1%	92.5%	75.0%	3.1%
Balanced Accuracy	70.9%	83.6%	63.5%	91.9%	85.7%	50.6%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 144. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9228	2222	2413	1122	1334	651
CWR	2784	14088	4585	62	2196	19
Rec_LTA	52	897	3016	1	806	0
SLA	1285	195	96	80014	289	6248
Sports	1773	3019	4338	239	15890	208
Wheels	118	77	2	48	0	891

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 145. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	60.6%	68.7%	20.9%	98.2%	77.5%	11.1%
Specificity	94.7%	93.1%	98.8%	89.7%	93.1%	99.8%
PPV	54.4%	59.4%	63.2%	90.8%	62.4%	78.4%
NPV	95.8%	95.3%	92.6%	98.0%	96.6%	95.5%
Precision	54.4%	59.4%	63.2%	90.8%	62.4%	78.4%
Recall	60.6%	68.7%	20.9%	98.2%	77.5%	11.1%
F1	57.3%	63.7%	31.4%	94.3%	69.1%	19.5%
Balanced Accuracy	77.6%	80.9%	59.8%	93.9%	85.3%	55.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 146. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9974	1280	1998	839	988	550
CWR	1990	15328	4712	81	2368	23
Rec_LTA	893	1940	5250	111	1470	44
SLA	931	128	68	79913	118	4958
Sports	1021	1815	2404	108	15568	50
Wheels	431	7	18	434	3	2392

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 147. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Lasso Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	65.4%	74.8%	36.3%	98.1%	75.9%	29.8%
Specificity	96.1%	93.4%	96.9%	92.1%	96.1%	99.4%
PPV	63.8%	62.6%	54.1%	92.8%	74.3%	72.8%
NPV	96.4%	96.2%	93.9%	97.9%	96.4%	96.4%
Precision	63.8%	62.6%	54.1%	92.8%	74.3%	72.8%
Recall	65.4%	74.8%	36.3%	98.1%	75.9%	29.8%
F1	64.6%	68.1%	43.5%	95.4%	75.1%	42.3%
Balanced Accuracy	80.8%	84.1%	66.6%	95.1%	86.0%	64.6%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 148. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10128	1515	1952	1717	1000	885
CWR	672	14106	2587	87	1602	57
Rec_LTA	1223	2174	6869	143	1310	125
SLA	1743	301	159	78247	66	4352
Sports	1291	2401	2879	103	16536	39
Wheels	183	1	4	1189	1	2559

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 149. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	66.5%	68.8%	47.5%	96.0%	80.6%	31.9%
Specificity	95.1%	96.4%	96.6%	91.6%	95.2%	99.1%
PPV	58.9%	73.8%	58.0%	92.2%	71.1%	65.0%
NPV	96.4%	95.5%	94.9%	95.7%	97.1%	96.5%
Precision	58.9%	73.8%	58.0%	92.2%	71.1%	65.0%
Recall	66.5%	68.8%	47.5%	96.0%	80.6%	31.9%
F1	62.4%	71.2%	52.2%	94.1%	75.6%	42.8%
Balanced Accuracy	80.8%	82.6%	72.1%	93.8%	87.9%	65.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 150. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	8909	2666	2422	1048	1435	602
CWR	1967	12202	3393	67	1671	61
Rec_LTA	1232	2351	4451	42	2088	13
SLA	1127	198	99	79029	267	2734
Sports	1689	2981	4079	326	15040	192
Wheels	316	100	6	974	14	4415

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 151. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	58.5%	59.5%	30.8%	97.0%	73.3%	55.1%
Specificity	94.4%	94.9%	96.1%	94.4%	93.4%	99.1%
PPV	52.2%	63.0%	43.7%	94.7%	61.9%	75.8%
NPV	95.6%	94.1%	93.3%	96.8%	96.0%	97.7%
Precision	52.2%	63.0%	43.7%	94.7%	61.9%	75.8%
Recall	58.5%	59.5%	30.8%	97.0%	73.3%	55.1%
F1	55.1%	61.2%	36.1%	95.8%	67.1%	63.8%
Balanced Accuracy	76.4%	77.2%	63.4%	95.7%	83.3%	77.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 152. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	11326	1220	1574	894	832	732
CWR	559	15076	2262	58	1450	19
Rec_LTA	1070	1873	7950	69	1366	42
SLA	802	74	68	79754	89	1736
Sports	1293	2228	2592	97	16772	46
Wheels	190	27	4	614	6	5442

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 153. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	74.3%	73.5%	55.0%	97.9%	81.8%	67.9%
Specificity	96.4%	96.9%	97.0%	96.5%	95.5%	99.4%
PPV	68.3%	77.6%	64.3%	96.6%	72.8%	86.6%
NPV	97.3%	96.1%	95.6%	97.8%	97.3%	98.3%
Precision	68.3%	77.6%	64.3%	96.6%	72.8%	86.6%
Recall	74.3%	73.5%	55.0%	97.9%	81.8%	67.9%
F1	71.2%	75.5%	59.3%	97.3%	77.0%	76.1%
Balanced Accuracy	85.3%	85.2%	76.0%	97.2%	88.6%	83.7%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 154. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	10227	1647	2022	1669	1117	689
CWR	948	14837	2865	101	1654	99
Rec_LTA	1104	1842	7014	84	1516	138
SLA	1489	173	129	77879	64	3528
Sports	1124	1995	2404	74	16160	28
Wheels	348	4	16	1679	4	3535

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 155. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	67.1%	72.4%	48.5%	95.6%	78.8%	44.1%
Specificity	95.1%	95.9%	96.8%	93.2%	96.0%	98.7%
PPV	58.9%	72.4%	60.0%	93.5%	74.2%	63.3%
NPV	96.5%	95.9%	95.0%	95.3%	96.9%	97.1%
Precision	58.9%	72.4%	60.0%	93.5%	74.2%	63.3%
Recall	67.1%	72.4%	48.5%	95.6%	78.8%	44.1%
F1	62.7%	72.4%	53.6%	94.5%	76.4%	52.0%
Balanced Accuracy	81.1%	84.2%	72.7%	94.4%	87.4%	71.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 156. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	9531	2359	2509	1235	1478	678
CWR	1671	13056	3108	43	1501	29
Rec_LTA	1114	1839	5035	39	1512	8
SLA	1060	147	71	78841	210	1940
Sports	1591	3003	3718	273	15804	167
Wheels	273	94	9	1055	10	5195

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 157. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	62.5%	63.7%	34.8%	96.8%	77.0%	64.8%
Specificity	94.3%	95.5%	96.9%	95.6%	93.7%	99.1%
PPV	53.6%	67.3%	52.7%	95.8%	64.4%	78.3%
NPV	96.0%	94.7%	93.8%	96.6%	96.5%	98.2%
Precision	53.6%	67.3%	52.7%	95.8%	64.4%	78.3%
Recall	62.5%	63.7%	34.8%	96.8%	77.0%	64.8%
F1	57.7%	65.4%	42.0%	96.3%	70.1%	70.9%
Balanced Accuracy	78.4%	79.6%	65.9%	96.2%	85.4%	81.9%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 158. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Chores	11048	1178	1224	881	869	608
CWR	707	15237	2454	64	1500	56
Rec_LTA	1219	2019	8274	96	1675	38
SLA	826	87	74	79472	117	1374
Sports	1189	1957	2418	109	16347	75
Wheels	251	20	6	864	7	5866

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 159. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Neural Network Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Wheels
Sensitivity	72.5%	74.3%	57.3%	97.5%	79.7%	73.2%
Specificity	96.7%	96.6%	96.5%	96.9%	95.9%	99.2%
PPV	69.9%	76.1%	62.1%	97.0%	74.0%	83.6%
NPV	97.1%	96.2%	95.8%	97.4%	97.0%	98.6%
Precision	69.9%	76.1%	62.1%	97.0%	74.0%	83.6%
Recall	72.5%	74.3%	57.3%	97.5%	79.7%	73.2%
F1	71.2%	75.2%	59.6%	97.3%	76.7%	78.1%
Balanced Accuracy	84.6%	85.5%	76.9%	97.2%	87.8%	86.2%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 160. *List of Study 2 activities and their operational definitions.*

Activity	Operational Definition
Arts/Crafts	activities might include drawing, painting, coloring, sketching, Playdough, clay, etc.
Baseball/Softball	playing a baseball/softball game
Baseball - Drills	simulated baseball tasks; swinging, hitting, catching, pitching, fielding, base running, etc.
Basketball	playing basketball or simulated basketball game scenario against an opponent
Basketball - Drills	playing a game with a basketball and hoop (e.g. Horse), general shooting around, dribbling etc.
Book Reading	reading or interacting with a physical book
Blocks	includes Legos, building blocks, K'nex, Lincoln Logs, Magnetic building sets etc.
Card/Board Games	includes playing card games as well as general board game play such as Monopoly, Life, Sorry, etc.
Catch	playing catch with a ball, not associated with playing the sport the ball is intended for, just throwing the ball back and forth in a pair or group.
Cell Phone	general use of cell phone. Includes social media, games, reading, texting, emailing, etc.
Cleaning/Wiping	involves using a cloth or rag to dust or wipe down windows, tabletops, counters, etc.
Playground - Climbing	climbing on gym or playground equipment
Computer	general use of a desktop computer with a mouse. Includes social media, viewing videos, doing homework, e-mailing, reading, etc.
Cooking	includes cooking, baking, food prep tasks in the kitchen
Dishes	includes loading and unloading the dishwasher as well as washing and drying dishes
Dog Walking (on a leash)	continuous walking with a dog on a leash
Football	playing a game of touch football
Football - Drills	simulated football includes running, passing, catching, kicking, etc. primarily individual or in a pair
Gardening	general gardening tasks including weeding, planting, picking fruit/vegetables, <u>etc</u>
Golf	playing a round of golf
Golf - Drills	simulated golf tasks include swinging a golf club and hitting a golf ball, putting, driving range, etc.
Homework/Screen	general button for tasks involving mixed homework activities involving using a screen, pencil/paper and handwritten activities or book reading and digital response/interaction with a screen.
Homework/Writing	general button for tasks involving writing on paper by hand and reading a book without a screen
Laptop	general use of a laptop computer. Includes social media, viewing videos, doing homework, e-mailing, reading, etc.
Laundry	includes folding and putting away laundry

Table 160 continued

Lying ONLY	to be used when lying is the main activity such as RMR. Not to be used when the participant is lying and watching TV. In that case the participant is Watching TV in a lying posture.
Mowing - Push	cutting the grass, pushing a lawnmower
Mowing - Riding	cutting the grass, riding a lawnmower
Over-ground Cycling	riding a bike outdoors, over ground
Playground - Other	general button for playground activities involving a jungle gym or similar type equipment. This is a general, non-specific button to be used when the participant is on the playground but not doing any discernable activity.
Raking	raking leaves or sand
Running (Treadmill)	continuous running on a treadmill
Running (Over-ground)	continuous running activity
Scooter	riding a self-propelled scooter over ground
Sitting ONLY	to be used when sitting is the main activity. Not to be used when the participant is sitting and watching TV. In that case the participant is Watching TV in a seated posture.
Roller Skating	includes in-line skating, rollerblading, etc.
Skate Boarding	riding a skateboard
Playground - Slide	involves the participant using the slide, as it is intended, for sliding down from the top to the bottom.
Soccer	playing a game of soccer or simulated soccer game activities against an opponent includes running, kicking, defending, shooting, etc.
Soccer - Drills	passing, shooting, dribbling, juggling tasks, primarily individual or in a pair
Stationary Cycling	riding a stationary cycle
Stair Climbing	ascending and descending stairs...should be used for an activity where this behaviour is continuous, not when stairs are ascended or descended during a transition period
Sweeping/Mopping	sweeping or mopping the floor with a broom or mop or another floor cleaning device (e.g. Swiffer)
Playground - Swing	can include a standard swing or a tree swing
Tablet	general use of a tablet - includes social media, viewing videos, doing homework, e-mailing, reading, etc.
Tag/Hide & Seek	playing a game of tag or hide & seek
Tennis	playing tennis with an opponent
Tennis - Drills	simulated tennis drills including serving, returning, and retrieving balls on a tennis court. Can be done individually.
Transition	intended to be used when there is a clear break from one activity before beginning another activity. Note: Exercise good judgement and caution when using this button. Try not to overuse it.
Unknown Activity	intended to be used when an activity/behaviour occurs that is not included or defined in this list. Note: Be sure to carefully review this list before using this button to make sure what you are coding is not already defined in another behaviour description.
Video Gaming	includes using a controller to play a game on a screen (TV, computer, etc.)
Volleyball	playing volleyball or simulated volleyball with teammates and an opponent

Table 160 continued

Volleyball - Drills	simulated volleyball drills including serving, returning, and retrieving balls on a volleyball court. Can be done individually.
Walking (Treadmill)	continuous walking on a treadmill
Walking (Over-ground)	continuous running over ground
Watching TV	viewing TV on a TV not watching video on another screen type

Table 161. Study 2 dataset breakdown by number of data points in each activity class and group corresponding to the recoded activity group labels.

Activity Group Label	Original Activity Behavior Label	Number of Data Points
Chores	Cleaning/Wiping	2,370
	Cooking	4,671
	Dishes	40
	Laundry	2,732
	Mowing - Push	1,664
	Raking	531
	Tidying up	10,112
CWR	Dog Walking	1,059
	Overground Running	585
	Overground Walking	15,756
Rec_LTPA	Weight Lifting	1,947
	Playground	9,293
	Playground - Slide	363
	Playground - Climb	195
	Playground - Swing	6,617
	Tag/Hide&Seek	834
SLA	Arts/Crafts	18,688
	Arts/Crafts - Media	2,043
	Blocks	16,061
	Book	4,885
	Card/Board Games	20,532
	Cell Phone	20,729
	Computer	4,861
	Homework-Screen	1,580
	Homework-Writing	20,703
	Laptop	7,666
	Lying Only	49,076
	Sitting Only	21,813
	Tablet	7,506
	Video Gaming	47,787
	Watching TV	73,697
Sports	Baseball - D	1,670
	Basketball	1,785
	Basketball - D	11,007
	Catch	2,698
	Football - D	1,142
	Golf - D	1,828
	Soccer - D	502
	Tennis	943
Wheels	Overground Cycle	552
	Roller Skating	248
	Scooter	2,254
	Skate Boarding	66
	Stationary Cycle	141
Transition	Off-Camera	9,649
	Transition	72,892
	Unknown Activity	47,048
Total	-	530,821

Table 162. *Physical characteristics of Study 2 participants.*

	Free-Living Sample (n = 42)
Age [years; mean (SD)]	11.6 (3.3)
Female [n (%)]	15 (35.7%)
Height [cm; mean (SD)]	148.0 (19.4)
Weight [kg; mean (SD)]	45.4 (21.1)
Handedness [n (% left-handed)]	1 (2.4%)
BMI classification [n (%)]	
< 5th percentile [n (% underweight)]	2 (4.8%)
5th - < 85th percentile [n (% normal weight)]	32 (76.2%)
85th - < 95th percentile [n (% overweight)]	2 (4.8%)
≥ 95th percentile [n (% obese)]	6 (14.2%)
Ethnicity	
Hispanic	1 (2.4%)
Race	
White	39 (92.8%)
African American	0 (0.0%)
Asian	2 (4.8%)
Hawaiian/Pacific Islander	1 (2.4%)

Table 163. Confusion Matrix Performance Summary by Activity Group for the Hip Random Forest Model by Feature Set.

	Accelerometer	Gyroscope	Combined
Chores	32.1%	39.0%	52.2%
CWR	45.6%	53.7%	54.5%
Rec_LTA	48.1%	22.2%	54.1%
SLA	92.4%	93.3%	95.4%
Sports	58.8%	53.6%	62.8%
Transition	53.0%	49.2%	58.8%
Wheels	10.3%	32.3%	48.6%

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities; The color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 164. Performance Summary by Activity Group for the Hip Random Forest Model by Feature Set.

	Accelerometer	Gyroscope	Combined
Sensitivity	48.6%	49.0%	60.9%
Specificity	94.0%	94.0%	95.3%
PPV	51.4%	51.9%	65.0%
NPV	94.1%	94.0%	95.3%
Precision	51.4%	51.9%	65.0%
Recall	48.6%	49.0%	60.9%
F1	49.3%	50.0%	62.4%
Balanced Accuracy	71.3%	71.5%	78.1%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; The color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 165. Confusion Matrix Performance Summary by Activity Group for the Left Wrist Random Forest Model by Feature Set.

	Accelerometer	Gyroscope	Combined
Chores	33.4%	26.1%	46.0%
CWR	57.4%	53.7%	60.9%
Rec_LTA	35.0%	27.3%	42.7%
SLA	91.7%	91.6%	94.4%
Sports	72.1%	56.5%	72.0%
Transition	55.0%	53.3%	58.7%
Wheels	15.6%	47.5%	67.4%

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities; The color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 166. Performance Summary by Activity Group for the Left Wrist Random Forest Model by Feature Set.

	Accelerometer	Gyroscope	Combined
Sensitivity	51.4%	50.9%	63.2%
Specificity	94.2%	94.0%	95.3%
PPV	58.3%	55.6%	68.0%
NPV	94.4%	94.0%	95.4%
Precision	58.3%	55.6%	68.0%
Recall	51.4%	50.9%	63.2%
F1	52.7%	52.4%	64.9%
Balanced Accuracy	72.8%	72.4%	79.3%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; The color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 167. Confusion Matrix Performance Summary by Activity Group for the Right Wrist Random Forest Model by Feature Set.

	Accelerometer	Gyroscope	Combined
Chores	37.0%	24.6%	51.5%
CWR	58.1%	56.7%	64.7%
Rec_LTA	35.8%	26.1%	44.9%
SLA	92.9%	91.3%	95.2%
Sports	75.9%	65.3%	76.3%
Transition	59.2%	50.0%	60.5%
Wheels	27.7%	52.1%	63.9%

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities; The color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 168. Performance Summary by Activity Group for the Right Wrist Random Forest Model by Feature Set.

	Accelerometer	Gyroscope	Combined
Sensitivity	55.2%	52.3%	65.3%
Specificity	94.7%	94.0%	95.7%
PPV	60.6%	55.3%	69.3%
NPV	94.9%	94.1%	95.8%
Precision	60.6%	55.3%	69.3%
Recall	55.2%	52.3%	65.3%
F1	56.7%	53.2%	66.8%
Balanced Accuracy	75.0%	73.1%	80.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; The color formatting corresponds to which feature set is best (dark) and worst (light) for each activity class within a single model type.

Table 169. Activity Group Confusion Matrix Results for the Hip Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	4889	309	204	740	863	3693	813
CWR	194	9339	2030	14	1107	2705	499
Rec_LTA	197	2927	6948	52	1990	1595	108
SLA	368	131	34	75256	123	4295	1826
Sports	1147	2145	3351	109	12055	4984	295
Transition	7862	4879	1830	5230	4163	20593	3649
Wheels	583	768	53	85	214	1001	827

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 170. Activity Group Performance Summaries for the Hip Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	32.1%	45.6%	48.1%	92.4%	58.8%	53.0%	10.3%
Specificity	96.4%	96.3%	96.3%	94.2%	93.3%	82.8%	98.6%
PPV	42.5%	58.8%	50.3%	91.7%	50.0%	42.7%	23.4%
NPV	94.5%	93.9%	96.0%	94.7%	95.2%	87.9%	96.3%
Precision	42.5%	58.8%	50.3%	91.7%	50.0%	42.7%	23.4%
Recall	32.1%	45.6%	48.1%	92.4%	58.8%	53.0%	10.3%
F1	36.6%	51.3%	49.2%	92.0%	54.1%	47.3%	14.3%
Balanced Accuracy	64.2%	70.9%	72.2%	93.3%	76.0%	67.9%	54.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 171. Activity Group Confusion Matrix Results for the Hip Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	5939	477	559	401	1060	3706	1233
CWR	358	11011	3307	16	713	3118	1013
Rec_LTA	411	2961	3213	24	2550	1929	344
SLA	223	5	24	76005	10	4945	354
Sports	1579	1434	4050	227	10996	5450	394
Transition	6434	3465	3173	4714	5135	19109	2092
Wheels	296	1145	124	99	51	609	2587

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 172. Activity Group Performance Summaries for the Hip Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	39.0%	53.7%	22.2%	93.3%	53.6%	49.2%	32.3%
Specificity	96.0%	95.2%	95.5%	95.3%	92.6%	84.4%	98.8%
PPV	44.4%	56.4%	28.1%	93.2%	45.6%	43.3%	52.7%
NPV	95.0%	94.7%	94.0%	95.3%	94.6%	87.2%	97.2%
Precision	44.4%	56.4%	28.1%	93.2%	45.6%	43.3%	52.7%
Recall	39.0%	53.7%	22.2%	93.3%	53.6%	49.2%	32.3%
F1	41.5%	55.0%	24.8%	93.2%	49.3%	46.1%	40.0%
Balanced Accuracy	67.5%	74.5%	58.9%	94.3%	73.1%	66.8%	65.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 173. Activity Group Confusion Matrix Results for the Hip Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	7952	240	244	117	781	3141	1193
CWR	148	11176	1860	3	673	2785	572
Rec_LTA	162	2306	7814	24	1525	1315	53
SLA	56	5	15	77713	40	3483	557
Sports	1238	1620	2922	132	12882	4828	382
Transition	5390	4453	1569	3456	4590	22862	1367
Wheels	294	698	26	41	24	452	3893

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 174. Activity Group Performance Summaries for the Hip Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	52.2%	54.5%	54.1%	95.4%	62.8%	58.8%	48.6%
Specificity	96.9%	96.6%	97.1%	96.5%	93.8%	87.0%	99.2%
PPV	58.2%	64.9%	59.2%	94.9%	53.7%	52.3%	71.7%
NPV	96.1%	94.9%	96.4%	96.8%	95.6%	89.7%	97.9%
Precision	58.2%	64.9%	59.2%	94.9%	53.7%	52.3%	71.7%
Recall	52.2%	54.5%	54.1%	95.4%	62.8%	58.8%	48.6%
F1	55.0%	59.3%	56.5%	95.1%	57.9%	55.4%	57.9%
Balanced Accuracy	74.5%	75.6%	75.6%	95.9%	78.3%	72.9%	73.9%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 175. Activity Group Confusion Matrix Results for the Left Wrist Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	5090	346	644	820	450	3306	452
CWR	204	11760	1724	26	1280	2745	14
Rec_LTA	387	1535	5053	39	1002	1604	107
SLA	867	86	200	74690	43	4492	4669
Sports	721	2754	3291	105	14801	5033	85
Transition	7937	4017	3530	5143	2936	21359	1440
Wheels	34	0	8	663	3	329	1250

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 176. Activity Group Performance Summaries for the Left Wrist Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	33.4%	57.4%	35.0%	91.7%	72.1%	55.0%	15.6%
Specificity	96.7%	96.6%	97.5%	91.2%	93.3%	84.4%	99.5%
PPV	45.8%	66.2%	51.9%	87.8%	55.2%	46.1%	54.7%
NPV	94.6%	95.2%	95.0%	94.0%	96.7%	88.5%	96.6%
Precision	45.8%	66.2%	51.9%	87.8%	55.2%	46.1%	54.7%
Recall	33.4%	57.4%	35.0%	91.7%	72.1%	55.0%	15.6%
F1	38.6%	61.5%	41.8%	89.7%	62.6%	50.1%	24.3%
Balanced Accuracy	65.1%	77.0%	66.2%	91.4%	82.7%	69.7%	57.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 177. Activity Group Confusion Matrix Results for the Left Wrist Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	3975	957	569	379	292	3218	543
CWR	756	11017	3101	22	1918	2887	56
Rec_LTA	268	2266	3945	18	2360	1770	5
SLA	388	10	115	74625	20	4430	1700
Sports	693	3032	3613	129	11594	5311	59
Transition	8773	3198	3102	5861	4328	20736	1849
Wheels	387	18	5	452	3	516	3805

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 178. Activity Group Performance Summaries for the Left Wrist Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	26.1%	53.7%	27.3%	91.6%	56.5%	53.3%	47.5%
Specificity	96.8%	95.1%	96.4%	94.3%	92.8%	83.1%	99.3%
PPV	40.0%	55.8%	37.1%	91.8%	47.5%	43.3%	73.4%
NPV	94.0%	94.7%	94.4%	94.2%	94.9%	88.0%	97.8%
Precision	40.0%	55.8%	37.1%	91.8%	47.5%	43.3%	73.4%
Recall	26.1%	53.7%	27.3%	91.6%	56.5%	53.3%	47.5%
F1	31.6%	54.7%	31.5%	91.7%	51.6%	47.8%	57.6%
Balanced Accuracy	61.4%	74.4%	61.8%	93.0%	74.7%	68.2%	73.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 179. Activity Group Confusion Matrix Results for the Left Wrist Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	7004	432	343	270	262	2791	464
CWR	179	12491	1541	7	1153	2813	17
Rec_LTA	293	1651	6177	28	1084	1669	20
SLA	362	11	124	76963	14	3710	900
Sports	653	2488	2860	65	14770	4777	44
Transition	6634	3400	3403	3794	3232	22833	1169
Wheels	115	25	2	359	0	275	5403

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 180. Activity Group Performance Summaries for the Left Wrist Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	46.0%	60.9%	42.7%	94.4%	72.0%	58.7%	67.4%
Specificity	97.5%	96.8%	97.4%	95.6%	93.9%	86.5%	99.6%
PPV	60.6%	68.6%	56.6%	93.8%	57.6%	51.4%	87.4%
NPV	95.6%	95.6%	95.6%	96.1%	96.7%	89.6%	98.6%
Precision	60.6%	68.6%	56.6%	93.8%	57.6%	51.4%	87.4%
Recall	46.0%	60.9%	42.7%	94.4%	72.0%	58.7%	67.4%
F1	52.3%	64.6%	48.7%	94.1%	64.0%	54.8%	76.1%
Balanced Accuracy	71.7%	78.9%	70.1%	95.0%	82.9%	72.6%	83.5%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 181. Activity Group Confusion Matrix Results for the Right Wrist Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	5646	615	1263	658	445	2871	397
CWR	184	11901	1667	13	1012	2743	6
Rec_LTA	722	1331	5178	68	839	1673	80
SLA	824	95	99	75719	15	3872	3912
Sports	997	2011	2472	45	15570	4231	8
Transition	6723	4544	3767	3808	2633	23026	1397
Wheels	144	1	4	1175	1	450	2217

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 182. Activity Group Performance Summaries for the Right Wrist Accelerometer Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	37.0%	58.1%	35.8%	92.9%	75.9%	59.2%	27.7%
Specificity	96.6%	96.9%	97.4%	92.5%	94.5%	85.7%	99.1%
PPV	47.5%	67.9%	52.4%	89.6%	61.5%	50.2%	55.5%
NPV	94.9%	95.3%	95.1%	95.0%	97.2%	89.7%	97.0%
Precision	47.5%	67.9%	52.4%	89.6%	61.5%	50.2%	55.5%
Recall	37.0%	58.1%	35.8%	92.9%	75.9%	59.2%	27.7%
F1	41.6%	62.6%	42.5%	91.2%	67.9%	54.3%	36.9%
Balanced Accuracy	66.8%	77.5%	66.6%	92.7%	85.2%	72.5%	63.4%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 183. Activity Group Confusion Matrix Results for the Right Wrist Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	3744	1406	831	269	446	3702	245
CWR	1314	11627	2926	37	1402	3260	41
Rec_LTA	765	2005	3769	16	1660	1829	7
SLA	227	29	25	74384	14	5343	2073
Sports	845	2355	3329	96	13400	4646	47
Transition	8112	2997	3565	5866	3590	19427	1431
Wheels	233	79	5	818	3	659	4173

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 184. Activity Group Performance Summaries for the Right Wrist Gyroscope Only Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	24.6%	56.7%	26.1%	91.3%	65.3%	50.0%	52.1%
Specificity	96.2%	95.0%	96.6%	93.4%	93.7%	84.0%	99.1%
PPV	35.2%	56.4%	37.5%	90.6%	54.2%	43.2%	69.9%
NPV	93.9%	95.0%	94.3%	93.9%	95.9%	87.4%	98.0%
Precision	35.2%	56.4%	37.5%	90.6%	54.2%	43.2%	69.9%
Recall	24.6%	56.7%	26.1%	91.3%	65.3%	50.0%	52.1%
F1	28.9%	56.6%	30.8%	90.9%	59.2%	46.3%	59.7%
Balanced Accuracy	60.4%	75.8%	61.3%	92.4%	79.5%	67.0%	75.6%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 185. Activity Group Confusion Matrix Results for the Right Wrist Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Chores	7849	547	791	288	280	2392	388
CWR	197	13258	1562	14	943	3350	3
Rec_LTA	625	1264	6481	19	888	1738	16
SLA	280	6	29	77580	3	3481	1443
Sports	968	1693	2174	28	15652	4052	23
Transition	5202	3707	3410	2967	2748	23507	1024
Wheels	119	23	3	590	1	346	5120

CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Table 186. Activity Group Performance Summaries for the Right Wrist Combined Sensor Feature Set Random Forest Model.

	Chores	CWR	Rec_LTA	SLA	Sports	Transition	Wheels
Sensitivity	51.5%	64.7%	44.9%	95.2%	76.3%	60.5%	63.9%
Specificity	97.5%	96.6%	97.5%	95.5%	95.0%	88.1%	99.4%
PPV	62.6%	68.6%	58.8%	93.7%	63.7%	55.2%	82.6%
NPV	96.0%	96.0%	95.8%	96.6%	97.2%	90.2%	98.5%
Precision	62.6%	68.6%	58.8%	93.7%	63.7%	55.2%	82.6%
Recall	51.5%	64.7%	44.9%	95.2%	76.3%	60.5%	63.9%
F1	56.5%	66.6%	50.9%	94.4%	69.4%	57.7%	72.0%
Balanced Accuracy	74.5%	80.6%	71.2%	95.4%	85.6%	74.3%	81.6%

PPV = positive predictive value; NPV = negative predictive value; F1 = the harmonic mean of precision and recall; CWR = continuous walking and running; SLA = seated and lying activities; Rec_LTA = recreational and leisure time activities

Appendix C: Study 1 Recruitment Flyer

Seeking Research Participants

Study Details

- » Perform common daily activities, including household chores and sports, while wearing activity monitors
- » Two visits (~2.5 hours total)
- » Compensation up to \$75
- » Excellent opportunity for research and technology exposure!



Who can Participate?

- » Boys and girls ages 6-18 years old
- » Active or non-active youth eligible
- » All testing will take place at the University of Tennessee campus



**Are you Interested?
Contact us below**

Email: APLStudy@utk.edu
Phone: (865) 974-5651

Appendix D: Study 1 Informed Consent

Informed Consent Statement for Parent or Guardian

Title of Research: Novel Approaches for Predicting Unstructured Short Periods of Physical Activities in Youth: Study 1 – Development of New Techniques

Principal Investigator: Dr. Scott Crouter

Location: Applied Physiology Laboratory, 1914 Andy Holt Ave., University of Tennessee, Knoxville, TN 37996

PURPOSE

You are being asked to give permission for your child to participate in a research project that will examine new methods to measure the amount of physical activity youth obtain using small devices called accelerometers. These devices are attached to a belt which is worn around the waist, wrist, or ankle similar to a pedometer. Your child will be eligible if he or she is: 1) 6 to 18 years of age, and 2) you and your child can read and speak English. Your child will not be eligible if: 1) he or she has a serious health problem or injury that would prevent them from participating safely in regular physical activity, 2) he or she has a Pacemaker or other internal medical device.

The study will take place at the Applied Physiology Laboratory in the Health, Physical Education, and Recreation (HPER) building at The University of Tennessee. The lead researcher is Scott Crouter, PhD, Assistant Professor in the Department of Kinesiology, Recreation, and Sport Studies. Please read this form and feel free to ask questions. If you have further questions, Dr. Crouter will discuss them with you. You can contact him by telephone (617-287-7509) or email (scrouter@utk.edu).

PARTICIPANT'S INVOLVEMENT IN THIS STUDY

Here is what the study involves:

- 1) You will be asked to fill out a health history questionnaire pertaining to your child's health status to determine eligibility for the study.
- 2) Your child will be asked to visit the Applied Physiology Laboratory in the HPER building on 2 separate days for approximately 100 minutes on visit 1 and 60 minutes on visit 2, for a total time commitment of 160 minutes (2 hrs 40 minutes).
 - a. During visit 1 your child will have his/her height, weight, percent body fat measured, resting metabolic rate measured, and will perform 8 activities randomly selected from a list of 16 activities.
 - b. During visit 2 your child will perform the remaining 8 activities, from the list of 16 that they did not perform during visit 1. The 2 testing days will occur within 2 weeks of each other.
- 3) The percent body fat will be measured using bio-electrical impedance (BIA). BIA is a non-invasive technique that involves standing on a scale while also holding on to a handlebar. During this time an undetectable and harmless electrical current is passed through the body.
- 4) The resting metabolic rate test will take 35 minutes. This test will involve having your child rest in a lying position for 25 minutes and it will take 10 minutes to connect the equipment and prepare for the test. During this time he/she will have 3 accelerometers placed on their waist, 3 on each wrist, 2 on each ankle, and 1 on each shoe using an elastic belt. Another device will also be attached to each thigh using a sticky pad. Your child will also be attached to a Cosmed portable system, which will measure the air that they breathe. This device is light weight and straps onto the body. It also requires the use of a facemask, over the nose and mouth. While wearing the facemask your child will be able to breathe normally and talk, and the mask does not block or restrict airflow.

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/16/2018
IRB EXPIRATION DATE: 09/13/2019

_____ Participant's Initials

- 5) Your child will perform 16 different activities. During each visit 8 activities will be performed which include:

1) self-paced slow walk, 2) self-paced brisk walk, 3) self-paced running, 4) self-paced stair climbing, 5) shooting a basketball, 6) playing catch with a ball, 7) playing soccer, 8) sitting in a reclined position, 9) sitting upright while reading, 10) sitting upright while searching the internet, 11) sitting upright while playing computer games, 12) lying, 13) cycling at a moderate pace, 14) performing jumping jacks, 15) sweeping, and 16) dusting

- a. While performing the activities your child will be wearing the elastic belts with the accelerometers and the Cosmed metabolic cart as described above. Each visit your child will receive a list of 8 activities from which to choose from and they will be instructed to perform each activity for two times; once for between 60-90 seconds and once for between 4-5 minutes. A trained research assistant will work with your child to ensure they perform the activities correctly.
- b. While performing the activities a trained research assistant will also video record your child so that exact transition times between activities can be obtained. This is an essential part of this project for development of better tracking methods in real-world situations.

RISKS

There are minimal risks associated with participation in the research study. Since the study involves children engaging in physical activities, there is the possibility of minor injuries such as cuts, bruises, and muscle soreness, and a slight risk of sprains. The risk is similar to that involved in a physical education class or a community sport program. There is also a risk of skin irritation from wearing the sticky pad on the thigh, which will be minimized by cleaning the area after it is removed. If irritation does occur during testing it will be removed immediately. Additionally, there is also a potential for social and psychological risks associated with wearing the Cosmed K4b² and being followed during the activities. Lastly, there is a risk of the identity of your child being known through the use of the video recording.

BENEFITS

The benefits of the proposed research include improved methods for measuring free-living physical activity. This has implications for intervention researchers and those looking at the health of children and adolescents. Participants will learn about their general health characteristics (e.g. BMI) and strategies to become more active.

CONFIDENTIALITY

Information and records included in this study will be kept confidential. Data will be stored in a secure location and will only be made available to the people conducting the study, unless you specifically give permission in writing to do otherwise. The video recordings will be kept on a secure password protected server and only investigators involved with this study will have access to these files. The results of the study will be published, but no reference will be made in oral or written reports that could link participants to the study.

The data files will also be sent to study investigators at the University of Massachusetts Boston and the University of California Riverside. Only non-identifiable data will be sent to these investigators and they will not receive the videos of your child, thus they would not be able to identify who your child is. The Office of Human Research Protections (OHRP) may also review the study records at any time.

COMPENSATION

For participation in the study, your child will receive a \$30 gift card for completion of visit 1 and a \$45 gift card for completion of visit 2, for a total of \$75 in gift cards if both days are completed.

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/16/2018
IRB EXPIRATION DATE: 09/13/2019

_____ Participant's Initials

EMERGENCY MEDICAL TREATMENT

The University of Tennessee does not "automatically" reimburse subjects for medical claims or other compensation. If physical injury is suffered in the course of research, or for more information, please notify the investigator in charge, Dr. Scott Crouter, at (865) 974-1272.

CONTACT INFORMATION

If you have questions at any time about the study or the procedures (or you experience adverse effects as a result of participating in this study), you should immediately contact the principal investigator, Dr. Scott Crouter, 334 HPER Building, The University of Tennessee, Knoxville, TN 37996, (865) 974-1272. If you have questions concerning your rights as a participant, contact the Office of Research Compliance Officer at (865) 974-7697.

PARTICIPATION

Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed.

STATEMENT OF CONSENT

I have read the above information, and I have received a copy of this form. I consent for my child to participate in this study. I also certify that I am 18 years of age or older.

Printed name of your child _____

Parent or Legal Guardian signature _____ Date _____

Investigator's signature _____ Date _____

STATEMENT OF VIDEO CONSENT

I consent for my child to be videotaped while completing the activity described above. Not signing, does not exclude your child from the study and they can still participate, however they will not be videotaped.

Parent or Legal Guardian signature _____ Date _____

I consent for my child's videotape to be used for training purposes in the future. This might include use for demonstration in an undergraduate or graduate kinesiology course or training for future studies where investigators are using the same procedures. By not signing below, your child can still be videotaped for use in the current study, but their file will not be used for future training or educational purposes.

Parent or Legal Guardian signature _____ Date _____

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/16/2018
IRB EXPIRATION DATE: 09/13/2019

Appendix E: Health History Questionnaire

HEALTH HISTORY QUESTIONNAIRE

Today's Date: _____

Name of parent/legal guardian completing the form: _____

ABOUT THE PARTICIPANT

Please answer the following questions about your child.

Name: _____

Address: _____

City: _____ Zip Code: _____

Parent's Phone: _____ Date of Birth (month/day/year): _____

Age: _____ Gender: ____ M ____ F

Current School (if summer, School entering in Fall): _____

Current Grade (if summer, Grade entering in the Fall): _____

How do you identify your child?

- ____ Asian, Non-Hispanic
- ____ Asian, Hispanic
- ____ Black/African American, Non-Hispanic
- ____ Black/African American, Hispanic
- ____ Native Hawaiian/Pacific Islander, Non-Hispanic
- ____ Native Hawaiian/Pacific Islander, Hispanic
- ____ Native American/Alaskan, Non-Hispanic
- ____ Native American/Alaskan, Hispanic
- ____ White, Non-Hispanic
- ____ White, Hispanic

Is this child Multiracial? ____ Y ____ N

Who resides in the household with your child?

____ Both parents ____ Single Mother ____ Single Father ____ Grandparent(s) ____ Other

Please Turn Over

Health History

Has your child ever been diagnosed with any of the follow conditions? If yes, please explain.

	NO	YES	Current Condition	Explanation
Heart Disease				
Diabetes/High Blood Sugar				
High Blood Pressure				
Seizure				
Asthma or Other Lung Condition				
Cancer				
Musculoskeletal Condition or Injury				
Other Serious Illness				
Allergy				

Please describe any additional medical conditions that may affect your child's participation in physical activity.

Is your child taking any medication (including prescription and non-prescription? If yes, please state below.

YES _____ NO _____

Name of Medication Reason for Taking For How Long?

Emergency Contact

Name: _____

Relationship: _____ Phone: Work: _____

Home: _____

Appendix F: Study 1 Informed Assent

Children's Assent form

Title of Research: Novel Approaches for Predicting Unstructured Short Periods of Physical Activities in Youth: Study 1 – Development of New Techniques

Principal Investigator: Dr. Scott Crouter

Location: Applied Physiology Laboratory, 1914 Andy Holt Ave., University of Tennessee, Knoxville, TN 37996

PURPOSE

You are being asked to participate in a research project that will use small devices called accelerometers to measure how much you move. These devices are attached to a belt and can be worn around the waist, hip, or ankle. The study will take place at the Applied Physiology Laboratory in the Health, Physical Education, and Recreation (HPER) building at The University of Tennessee. The lead researcher is Scott Crouter, PhD, Assistant Professor in the Department of Kinesiology, Recreation, and Sport Studies. Please read this form and feel free to ask questions. If you have further questions, Dr. Crouter or a research assistant will discuss them with you. You can contact him by telephone (617-287-7509) or email (scrouter@utk.edu).

PARTICIPANT'S INVOLVEMENT IN THIS STUDY

If you decide that you want to be in this study, this is what you will be asked to do:

- 1) Visit the Applied Physiology Laboratory in the HPER building on 2 different days for approximately 1.5 hours on visit 1 and 1 hour on visit 2.
- 2) Have your height, weight, and percent body fat measured, which requires you to stand barefoot on a scale.
 - a. During your 2 visits you will be asked to do the following tests:
 - b. Resting metabolic rate (visit 1): During this test you will be asked to rest quietly while lying on a table. During the test you will have 3 small devices placed on your waist, 3 on each wrist, 2 on each ankle, and 1 on each shoe using an elastic belt. Another device will also be attached to each thigh using a sticky pad. You will also have devices placed on your shoes and You will also be asked to wear a mask over your nose and mouth which will be connected to a device that is worn on your chest. The mask over your face will not affect how you breathe and you will be able to talk. This test will take 35 minutes to complete.
- 3) Perform 16 different activities (visits 1 and 2): During each visit 8 activities from the list below will be performed:
 - 1) walk slowly, 2) walk quickly, 3) run as fast or as slow as you want, 4) walk up and down stairs, 5) play basketball, 6) play catch with a ball, 7) playing soccer, 8) sit in a reclined position, 9) sit in a chair and read, 10) sit in a chair and play on the internet, 11) sit in a chair and play computer games, 12) lay on your back, 13) indoor cycling, 14) jumping jacks, 15) sweeping a floor, and 16) dusting a table
 - a. While performing the activities you will be asked to wear the elastic belts facemask as you will do for the resting metabolic measurement test. You will do each activity two for either 60-90 seconds or 4-5 minutes. A research assistant will work with you to perform each activity.
 - b. While doing the activities a research assistant will also video record what you are doing.

CAN ANYTHING BAD HAPPEN TO ME?

Although there are very few risks to taking part in this study, you could get hurt just like with any type of physical activity such as what you do in PE class. Dr. Crouter and the other research team members will help you to avoid being hurt or upset. There is also the risk that your skin will become itchy while wearing the sticky pad on your thigh.

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/16/2018
IRB EXPIRATION DATE: 09/13/2019

_____ Participant's Initials

CAN ANYTHING GOOD HAPPEN TO ME?

By participating in this study, you will learn about other physical activities that will be good for your health while also having fun.

DO I HAVE OTHER CHOICES?

You can choose not to be in this study if you don't want to.

WILL ANYONE KNOW I AM IN THE STUDY?

We won't tell anyone you took part in this study. Other people who are in the Applied Physiology Laboratory will know that you are being tested but we will not tell them any of your personal information. When we are done with the study, we will write a report about what we found out. We won't use your name in the report.

WHAT WILL YOU RECEIVE?

You will receive a \$30 gift card for completing visit 1 and a \$45 gift card for completing visit 2. If you complete both days of testing you will get a total of \$75 in gift cards.

WHAT IF I DO NOT WANT TO DO THIS?

You don't have to be in this study. It's up to you. If you say yes now, but you change your mind later, that's okay too. All you have to do is tell us.

Before you tell us your decision about being a part of this study; be sure to ask Dr. Crouter or one of his research assistants to tell you more about anything you don't understand.

STATEMENT OF CONSENT

I have read the above information, and I have received a copy of this form. By signing this form I give assent to participate in this study.

Participant's signature _____ Date _____

Investigator's signature _____ Date _____

STATEMENT OF VIDEO CONSENT

I give assent to be videotaped while completing the activity described above. Not signing, does not exclude you from the study and you can still participate, however you will not be videotaped.

Participant's signature _____ Date _____

I give assent for my videotape to be used for training purposes in the future. This might include use for demonstration in an undergraduate or graduate kinesiology course or training for future studies where investigators are using the same procedures. By not signing below, you can still be videotaped for use in the current study, but there file will not be used for future training or educational purposes.

Participant's signature _____ Date _____

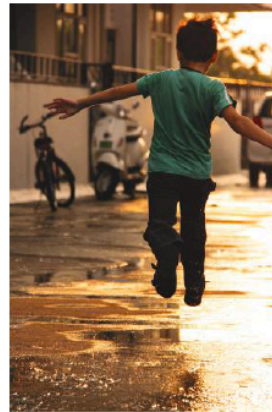
IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 08/16/2018
IRB EXPIRATION DATE: 09/13/2019

Appendix G: Study 2 Recruitment Flyer

Seeking Research Participants

Study Details

- » Perform common daily activities, including household chores and sports, while wearing activity monitors
- » Three visits (~6 hours total)
- » Compensation up to \$75
- » Excellent opportunity for research and technology exposure!



Who can Participate?

- » Boys and girls ages 6-18 years old
- » Active or non-active youth eligible
- » All testing will take place at participants home and after school programs



**Are you Interested?
Contact us below**

Email: APLStudy@utk.edu

Phone: (865) 974-5651

Appendix H: Study 2 Informed Consent

Informed Consent Statement for Parent or Guardian

Title of Research: Novel Approaches for Predicting Unstructured Short Periods of Physical Activities in Youth: Study 2 – Validation of New Techniques

Principal Investigator: Dr. Scott Crouter

Location: Applied Physiology Laboratory, 1914 Andy Holt Ave., University of Tennessee, Knoxville, TN 37996

PURPOSE

You are being asked to give permission for your child to participate in a research project that will examine new methods to measure the amount of physical activity youth obtain using small devices called accelerometers. These devices are attached to a belt which is worn around the waist, wrist, or ankle similar to a pedometer. Your child will be eligible if he or she is: 1) 6 to 18 years of age, and 2) you and your child can read and speak English. Your child will not be eligible if: 1) he or she has a serious health problem or injury that would prevent them from participating safely in regular physical activity, 2) he or she has a Pacemaker or other internal medical device.

The study will take place at the Applied Physiology Laboratory in the Health, Physical Education, and Recreation (HPER) building at The University of Tennessee. The lead researcher is Scott Crouter, PhD, Assistant Professor in the Department of Kinesiology, Recreation, and Sport Studies. Please read this form and feel free to ask questions. If you have further questions, Dr. Crouter will discuss them with you. You can contact him by telephone (617-287-7509) or email (scrouter@utk.edu).

PARTICIPANT'S INVOLVEMENT IN THIS STUDY

Here is what the study involves:

- 1) You will be asked to fill out a health history questionnaire pertaining to your child's health status to determine eligibility for the study.
- 2) Your child will be asked to be tested on 3 different days. Day 1 will take place at the Applied Physiology Laboratory in the HPER building (approximately 60 minute visit) and days 2 and 3 will take place afterschool at a local facility (e.g. YMCA, the school they attend), public park, or in their home. Days 2 and 3 will take approximately 2 hrs and 10 minutes to complete on each day. The three testing days will be completed over a 2 week period. If your child is enrolled in an afterschool program where we are testing, day 1 will be completed at the program facility. In addition, we may combine days 1 and 2 into a single testing day, if time allows.
 - a. **Day 1:** Your child will have his/her height, weight, percent body fat and resting metabolic rate measured.
 - i. The percent body fat will be measured using bio-electrical impedance (BIA). BIA is a non-invasive technique that involves standing on a scale while also holding on to a handlebar. During this time an undetectable and harmless electrical current is passed through the body.
 - ii. The resting metabolic rate test will take 35 minutes. This test will involve having your child rest in a lying position for 25 minutes and it will take 10 minutes to connect the equipment and prepare for the test. During this time he/she will have 3 accelerometers placed on their waist, 3 on each wrist, 2 on each ankle, and 1 on each shoe using an elastic belt. Another device will be attached to each thigh using a sticky pad. There will also be a device attached to a pair of eyeglasses that will be worn. Your child will also be attached to a Cosmed portable system, which will measure the air that they breathe. This device is light weight and straps onto the body. It also requires the use of a facemask, over the nose and mouth. While wearing the facemask your child will be able to breathe normally and talk, and the mask does not block or restrict airflow.

IRB NUMBER: UTK IRB-15-02467-PB
IRB APPROVAL DATE: 10/25/2019
IRB EXPIRATION DATE: 09/12/2020

_____ **Participant's Initials**

- b. **Days 2 and 3:** Your child will be asked to wear the Cosmed portable system and belt with the accelerometers for a 2-hour period. On one of the days this testing will take place at an afterschool facility and on the other day it will take place in their home. If testing in the home or an afterschool facility is not possible, alternative testing sites may be found (e.g. public park). The home visit will be scheduled in advance so that a parent or guardian can be there. During this time he/she will be free to choose the activities they perform using what is available at the testing location. During all times a trained research assistant will be with your child. While performing the activities a trained research assistant will also video record your child so that exact transition times between activities can be obtained. This is an essential part of this project for development of better tracking methods in real-world situations.

RISKS

There are minimal risks associated with participation in the research study. Since the study involves children engaging in physical activities, there is the possibility of minor injuries such as cuts, bruises, and muscle soreness, and a slight risk of sprains. The risk is similar to that involved in a physical education class or a community sport program. There is also a risk of skin irritation from wearing the sticky pad on the thigh, which will be minimized by cleaning the area after it is removed. If irritation does occur during testing it will be removed immediately. Additionally, there is also a potential for social and psychological risks associated with wearing the Cosmed K4b² and being followed during the activities. Lastly, there is a risk of the identity of your child being known through the use of the video recording.

BENEFITS

The benefits of the proposed research include improved methods for measuring free-living physical activity. This has implications for intervention researchers and those looking at the health of children and adolescents. Participants will learn about their general health characteristics (e.g. BMI) and strategies to become more active.

CONFIDENTIALITY

Information and records included in this study will be kept confidential. Data will be stored in a secure location and will only be made available to the people conducting the study, unless you specifically give permission in writing to do otherwise. The video recordings will be kept on a secure password protected server and only investigators involved with this study will have access to these files. The results of the study will be published, but no reference will be made in oral or written reports that could link participants to the study.

The data files will also be sent to study investigators at the University of Massachusetts Boston, University of California Riverside, University of Alabama, and University of West Florida. Data files will be sent through a secure computer connection to reduce the potential of data being viewed by others and the files will be linked to a unique study ID so no names will be included. The Office of Human Research Protections (OHRP) may also review the study records at any time.

WHAT WILL HAPPEN TO INFORMATION AFTER THE STUDY IS OVER?

We will keep your child's information to use for future research ideas and training for future studies. Your child's name and other information that can directly identify them will be kept secure and stored separately from their research data collected as part of the study. All video files will have the faces blurred out and names removed. However, there is still the possibility that your child could be recognized by someone reviewing the video files.

We may share your child's research data with other researchers without asking for your permission again, but it will not contain information that could directly identify your child. This sharing of data may include being deposited into a data repository as required by publishers for publication or public repository.

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 10/25/2019
IRB EXPIRATION DATE: 09/12/2020

_____ **Participant's Initials**

COMPENSATION

For participation in the study, your child will receive a \$25 gift card for completing visit one, a \$25 gift card for completing visit 2, and a \$25 gift card for completing visit 3; for a total of \$75 in gift cards if all 3 days are completed. If visit 1 and 2 testing is completed on the same day then your child will receive \$50 for the visit and \$25 for the second visit.

EMERGENCY MEDICAL TREATMENT

The University of Tennessee does not "automatically" reimburse subjects for medical claims or other compensation. If physical injury is suffered in the course of research, or for more information, please notify the investigator in charge, Dr. Scott Crouter, at (865) 974-1272.

CONTACT INFORMATION

If you have questions at any time about the study or the procedures (or you experience adverse effects as a result of participating in this study), you should immediately contact the principal investigator, Dr. Scott Crouter, 334 HPER Building, The University of Tennessee, Knoxville, TN 37996, (865) 974-1272. If you have questions concerning your rights as a participant, contact the Office of Research Compliance Officer at (865) 974-7697.

PARTICIPATION

Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed.

STATEMENT OF CONSENT

I have read the above information, and I have received a copy of this form. I consent for my child to participate in this study. I also certify that I am 18 years of age or older.

Printed name of your child _____

Parent or Legal Guardian signature _____ Date _____

Investigator's signature _____ Date _____

STATEMENT OF VIDEO CONSENT

I consent for my child to be videotaped while completing the activity described above. Not signing, does not exclude your child from the study and they can still participate, however they will not be videotaped.

Parent or Legal Guardian signature _____ Date _____

I consent for my child's videotape to be used for training purposes in the future. This might include use for demonstration in an undergraduate or graduate kinesiology course or training for future studies where investigators are using the same procedures. By not signing below, your child can still be videotaped for use in the current study, but their file will not be used for future training or educational purposes.

Parent or Legal Guardian signature _____ Date _____

IRB NUMBER: UTK-IRB-15-02487-FB
IRB APPROVAL DATE: 10/25/2019
IRB EXPIRATION DATE: 09/12/2020

Appendix I: Study 2 Informed Assent

Children's Assent form

Title of Research: Novel Approaches for Predicting Unstructured Short Periods of Physical Activities in Youth: Study 2 – Development of New Techniques

Principal Investigator: Dr. Scott Crouter

Location: Applied Physiology Laboratory, 1914 Andy Holt Ave., University of Tennessee, Knoxville, TN 37996

PURPOSE

You are being asked to participate in a research project that will use small devices called accelerometers to measure how much you move. These devices are attached to a belt and can be worn around the waist, hip, or ankle. The study will take place at the Applied Physiology Laboratory in the Health, Physical Education, and Recreation (HPER) building at The University of Tennessee. The lead researcher is Scott Crouter, PhD, Assistant Professor in the Department of Kinesiology, Recreation, and Sport Studies. Please read this form and feel free to ask questions. If you have further questions, Dr. Crouter or a research assistant will discuss them with you. You can contact him by telephone (617-287-7509) or email (scrouter@utk.edu).

PARTICIPANT'S INVOLVEMENT IN THIS STUDY

If you decide that you want to be in this study, this is what you will be asked to do:

- 1) Complete testing on 3 different days. Day 1 will take place at the Applied Physiology Laboratory in the HPER building (approximately 1 hour visit) and days 2 and 3 will take place afterschool at a local facility (e.g. YMCA, the school they attend), public park, or your home. Days 2 and 3 will take approximately 2 hrs to complete. If you are enrolled in an afterschool program that we are testing at we will complete day 1 at the program facility and may combine days 1 and 2 into one day of testing.
 - a. **Day 1:** You will have your height, weight, percent body fat and resting metabolic rate measured.
 - i. The percent body fat measurement will require you to stand barefoot on a scale.
 - ii. For the resting metabolic rate test you will be asked to rest quietly while lying on a table. During the test you will have 3 small devices placed on your waist, 3 on each wrist, 2 on each ankle, and 1 on each shoe using an elastic belt. Another device will also be attached to each thigh using a sticky pad. There will also be a device attached to a pair of eyeglasses that you will wear. You will also be asked to wear a mask over your nose and mouth which will be connected to a device that is worn on your chest. The mask over your face will not affect how you breathe and you will be able to talk. This test will take 35 minutes to complete.
 - b. **Days 2 and 3:** You will be asked to wear the facemask and elastic belt as described above for 2 hours on each day. During this time you will be able to freely choose what activities you do based on what is available. On one of the days this testing will take place at an afterschool facility and on the other day it will take place in their home. If this is not possible, alternative testing sites might be found (e.g. public park). You will have a research assistant with you at all times to ensure your safety. While performing the activities a research assistant will also video record what you are doing.

CAN ANYTHING BAD HAPPEN TO ME?

Although there are very few risks to taking part in this study, you could get hurt just like with any type of physical activity such as what you do in PE class. Dr. Crouter and the other research team members will help you to avoid being hurt or upset. There is also the risk that your skin will become itchy while wearing the sticky pad on your thigh.

CAN ANYTHING GOOD HAPPEN TO ME?

By participating in this study, you will learn about other physical activities that will be good for your health while also having fun.

IRB NUMBER: UTK IRB-15-02487-FB
IRB APPROVAL DATE: 10/25/2019
IRB EXPIRATION DATE: 09/12/2020

_____ Participant's Initials

DO I HAVE OTHER CHOICES?

You can choose not to be in this study if you don't want to.

WILL ANYONE KNOW I AM IN THE STUDY?

We won't tell anyone you took part in this study. Other people who are in the Applied Physiology Laboratory will know that you are being tested but we will not tell them any of your personal information. When we are done with the study, we will write a report about what we found out. We won't use your name in the report.

WHAT WILL HAPPEN TO MY INFORMATION AFTER THE STUDY IS OVER?

We will keep your information to use for future research ideas and training. Your name and other information that can directly identify you will be kept in a safe location separate from your research data.

We may share your research data with other researchers without asking for your permission again, but we will not give them your name or any information that would allow them to identify who you are. Your research data might also be placed into a data repository when we publish the report.

WHAT WILL YOU RECEIVE?

You will receive a \$25 gift card for completing visit 1, a \$25 gift card for completing visit 2, and a \$25 gift card for completing visit 3. If you complete all 3 days of testing you will get a total of \$75 in gift cards. If you do day 1 and 2 testing on the same day then you will receive \$50 for the visit and \$25 for the second visit.

WHAT IF I DO NOT WANT TO DO THIS?

You don't have to be in this study. It's up to you. If you say yes now, but you change your mind later, that's okay too. All you have to do is tell us.

Before you tell us your decision about being a part of this study; be sure to ask Dr. Crouter or one of his research assistants to tell you more about anything you don't understand.

STATEMENT OF CONSENT

I have read the above information, and I have received a copy of this form. By signing this form I give assent to participate in this study.

Participant's signature _____ Date _____

Investigator's signature _____ Date _____

STATEMENT OF VIDEO CONSENT

I give assent to be videotaped while completing the activity described above. Not signing, does not exclude you from the study and you can still participate, however you will not be videotaped.

Participant's signature _____ Date _____

I give assent for my videotape to be used for training purposes in the future. This might include use for demonstration in an undergraduate or graduate kinesiology course or training for future studies where investigators are using the same procedures. By not signing below, you can still be videotaped for use in the current study, but there file will not be used for future training or educational purposes.

Participant's signature _____ Date _____

IRB NUMBER: UTK IRB-15-0248 R-PP

IRB APPROVAL DATE: 10/25/2019

IRB EXPIRATION DATE: 09/12/2020

Vita

Samuel Robert LaMunion was born January 29, 1992 to Hurshel Van LaMunion and Sara Catherine “Cathy” Corley LaMunion in Augusta, GA. He was raised in North Augusta, South Carolina where he attended and graduated from August Christian Schools in 2010. Samuel completed his Bachelor of Science degree with magna cum laude honors in Exercise and Sports Science with a Basic Sciences concentration at the University of South Carolina Aiken in August of 2014. In Fall of 2014 he enrolled in the Master of Science program in Kinesiology with a concentration in Exercise Physiology at The University of Tennessee Knoxville. Samuel Graduated from the Master’s program in the August of 2016 at which time he re-enrolled at The University of Tennessee Knoxville for Fall 2016 to begin work on a doctorate in Kinesiology and Sports Studies with a concentration in Kinesiology and a specialization in Exercise Physiology. Samuel graduated with a Doctor of Philosophy in Kinesiology and Sports Studies and minors in Epidemiology and Statistics in Spring 2020. Samuel has accepted a post-doctoral fellowship under the mentorship of Dr. Kong Y. Chen at the National Institute of Diabetes, Digestion, and Kidney Diseases in Bethesda, MD.