HUMAN DECOMPOSITION EVALUATION: A STANDARDIZED APPROACH FOR STAGING AND SCORING MORPHOLOGICAL FEATURES USING ARTIFICIAL INTELLIGENCE

A Thesis Presented for the
Master of Arts
Degree
The University of Tennessee, Knoxville

Phillip D. Ditto
August 2024

DEDICATION

This thesis is dedicated to Dr. David Aftandilian. Dave, you have always been a shining example of humanity. The course of my life was changed in your Intro to Osteology course at Texas Christian University. Thank you for instilling a love of anthropology, supporting my endeavors, and sharing your wisdom over the years.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and guidance of my advisor and committee chair, Dr. Amy Mundorff, who provided for my graduate experience in ways few will ever know. I want to thank my committee, Drs. Dawnie Steadman and Audris Mockus for their expertise and commitment. I cannot thank Anna-Maria Nau enough for her unwavering support, guidance, and partnership. I want to acknowledge my daughters, Abby and Katie, for their patience and love while I completed my degrees. Lastly, I want to thank my wife, Tasha, for their faith in me. Without her, none of this would have been possible.

ABSTRACT

Anthropological estimates of the post-mortem interval (PMI) or the time since an individual died depend on understanding the morphological features of the body present at the time of examination. Such changes may include skin color, bloating, or mummification that are assumed to occur sequentially from the time of death until skeletonization, broadly indicating how long a person may have been deceased. However, there are no standards or even agreed-upon stages in which these morphological changes are observed, given the number of factors influencing human decomposition over time. This lack of standards makes the observer's reliability of morphological decomposition traits and stages tenuous and an important research subject. Since photography is widely used in forensic cases and is an affordable and replicable method of capturing the morphological changes in human decomposition, pairing two-dimensional photographs of human decomposition with the growing field of artificial intelligence (AI) may have the ability to decrease observer error and bias in estimating PMI and produce an automated process for human decomposition assessment.

Over 20,000 photographs of human decomposition from the University of Tennessee, Knoxville Anthropological Research Facility (ARF) digital catalog were categorized and utilized as training data for AI machine learning algorithms. These algorithms reflected AI's ability to accurately and repetitively categorize human decomposition utilizing two published methods and one novel method. AI

performed well, correctly identifying states of decomposition using three different methods, as evidenced by average MaF1 scores above 0.8 and precision score averages of more than 85%. Results indicate that AI can produce an automated and accessible method of accurately and repetitively categorizing human decomposition. Additional raters with diverse experience and skills and an increased and diverse sample size are needed to develop AI training and performance further.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Medico-legal investigators have long sought to establish an accurate time since death (TSD) or post-mortem interval (PMI), which may be necessary when investigating the identity of decedents (Shrestha, Kanchan, and Kirshan, 2022). Extensive research has been conducted utilizing morphological changes that occur during human decomposition to create a standardized process to estimate PMI (Galloway et al., 1989; Gelderman et al., 2018; Megyesi et al., 2005; Vidoli, Kenyhercz, and Steadman, 2020). Researchers have identified a general progression of morphological changes in human decomposition, which were categorized into stages such as fresh, early decomposition, advanced decomposition, and skeletonization, generally, with few differences among methods (e.g., Galloway et al., 1989). However, because decomposition is a complex process dependent upon numerous intrinsic and extrinsic variables (e.g., Mann et al. 1990), there is also a considerable error rate in categorizing human decomposition and estimating PMI among expert observers (Gelderman et al., 2019; Sauerwein, 2018). Challenges persist in developing a decomposition assessment method that can be universally applied.

Photography has long been a staple of forensic investigations worldwide (Gouse et al., 2018) and is also used to record human decomposition. Photography is an affordable and efficient method of capturing visual data that can be stored for future use. Photography is also used at human decomposition facilities, such as the Anthropology Research Facility at the University of

Tennessee, Knoxville, to track and record variables impacting human decomposition rates, from deposition to skeletonization.

Artificial intelligence is computer programming developed by learning algorithms to assist machines in accomplishing trained tasks (Cheng-Tek Tai 2020). The developing field of AI, coupled with photography of human decay, may provide an alternate and more precise method of PMI estimation (Nau, Mockus, and Steadman, 2023; Tai, 2020). AI has become more sophisticated and easier to use. It provides a unique opportunity to automate and assist investigators, who may be nonspecialists, in accurately evaluating the morphological features present in human decomposition and decreasing inter-observer error, more accurate predictions of PMI may be made. Additionally, AI analysis provides access to examiners with disabilities, such as color blindness, which affects approximately 8% of the population (Fakorede et al., 2022). However, AI requires training data that is accurate and consistent among raters to create an accurate model. High levels of inter-observer error when utilizing PMI methods such as Megyesi et al. (2005) and Gelderman et al. (2018) emphasize a need for further examination of these scoring methods, evaluation of the method's morphological categories, and a deeper understanding of the progression of human decay (Gelderman et al., 2019; Sauerwein, 2018).

Using AI to assist in assessing an accurate stage of decomposition depends upon repetitive object recognition and labeling through machine learning algorithms based on a standardized scoring method. The process of training AI models, however, is largely dependent upon the utilization of learning models that function using quality input data, which, in this case, are the stages of human decomposition categorized by visual morphological features. Previous research by computer and data scientists, as well as anthropologists, identified features in human decomposition, such as "maggots" or anatomical features such as a "leg," and placed labels, or "tags" in the photo for the AI algorithm to learn. The tagged images were then placed in the web-based application known as Image Cloud Platform for Use in Tagging and Research on Decomposition (ICPUTD) for recognition of visual morphological features, such as anatomical features and decomposition changes like coloration and entomological activity (Mousavi et al. 2019). AI testing is required to better understand its potential in categorizing human decomposition.

Objectives

The primary goal of this research was to test three anthropological PMI methods to determine which, if any, best informs AI of the visual stages of human decomposition. AI's ability to distinguish visual morphological changes was tested for two published PMI assessment methods (Megyesi et al. 2005 and Gelderman et al. 2018). Additionally, one new method was tested, DittoNau. This method, designed with AI learning models in mind, was developed after categorizing thousands of human decomposition photographs based on Megyesi et al. (2005) and Gelderman et al. (2018), and hypothesizing which morphological features in which stages would best inform AI. Determining which human decomposition assessment method is most optimal will allow for a more objective AI system to be developed. Independent variables such as temperature, soil composition, whether a body is deposited above or below ground, insect activity, scavenging, microbes, and biome are variables that can affect the decomposition process (Javan et al., 2019; Pittner et al., 2020; Simmons et al., 2010; Zapico and Adserias-Garriga, 2022). Independent variables may produce morphological features that could be confused for advanced stages in current scoring methods, such as bone appearance from trauma, insect activity, or scavenging earlier than expected, which may produce results in assessments inconsistent with the true PMI of the individual. The optimal method trains AI machine learning algorithms to score human decomposition and all its variables with less error.

Research Questions

The following questions will be addressed to assess which methodology best informs AI in categorizing the visual morphological features present in human decomposition.

- Which method best informs and trains AI algorithms to correctly categorize human decomposition from photos?
- How are the three categorical methods regarding human decomposition different in their assessments, and how might they influence AI?

- What, if any, morphological features in each of the methods provide poor training for AI deep learning and why?

Thesis Outline

Chapter Two, Background, will introduce the basic concepts of human decomposition, focusing on visual morphological characteristics, intrinsic and extrinsic modifiers, and the methods used to capture and categorize decomposition characteristics.

Chapter Three, Materials and Methods, will provide detailed information on how this research was conducted, including the materials utilized and the methodology used for investigating the research questions. Additionally, quality assurance and improvement controls and inter- and intra-observer testing (IOT) procedures will be discussed.

Chapter Four, Results, will present quantitative and qualitative results followed by a discussion chapter.

Chapter Five, Discussion, will discuss the results of this study. The discussion will focus on how each decomposition method performed after the AI algorithm was trained to assess a photo population of human decomposition, emphasizing potential reasoning and differences between the three methods. The discussion section will also discuss potential bias influences and their effects.

Finally, Chapter Six, Conclusion, will summarize the research results and the study's limitations and elucidate the need for further research.

CHAPTER TWO

BACKGROUND

This chapter discusses the basic concepts of the human decomposition process, common anthropological methods used to evaluate the visual morphological features present during decomposition, some of the commonly associated stages of decomposition, and challenges in reliably categorizing these visual morphological features that may affect the accurate estimation of PMI.

The decomposition process begins as soon as the individual dies. Through observation and research, this process has become generally understood. However, many questions remain regarding how individual factors and their general timing influence this process. The body may eventually skeletonize due to multiple factors, such as microbe activity, insect involvement, climate, and weather (Marks, Love, and Dadour, 2016; Zapico and Adserias-Garriga, 2022).

Human decomposition is loosely categorized into four stages, progressing from fresh to skeletonization (Galloway et al., 1989; Megyesi et al., 2005). Common stages associated with published methods on human decomposition are Fresh, Active or Early, Advanced, and Skeletonization (Galloway et al., 1989; Megyesi et al., 2005). Some authors use criteria to expand and amend these stages and categorize morphological features that appear during decomposition (Gelderman et al., 2018; Vidoli, Kenyhercz, and Steadman, 2020). Additionally, the DittoNau method developed for this study categorizes morphological features into stages with different labeling, such as active decomposition, and notably, Gelderman et al. 2018 do not use stages at all.

Fresh Stage

The fresh stage is shortly after an individual dies and is defined by the absence of morphological features visually present shortly after death (Gelderman et al., 2018; Megyesi et al., 2005). This stage is typically short, depending on the prevailing temperatures and season, with morphological changes of early decomposition happening within hours of death when autolysis of cells begins, as discussed in the next section (Mann, Bass, and Meadows, 1990; Marks, Love, and Dadour, 2016; Vass, 2001).

Early or Active Decomposition

The early or active decomposition stage, as utilized in Megyesi et al. (2005) and DittoNau, is characterized by the process of autolysis as enzymatic activity begins the process of self-digestion of the cells, producing morphological changes that can be visually observed such as color changes in the skin of the individual (Janaway, Percival, and Wilson, 2009). Typically, the first morphological change observed after death is livor mortis, characterized by blood pooling in the capillary beds, producing large areas of red or purple skin coloration that blanches when touched. Marbling also occurs during early morphological changes. As blood vessels dilate during autolysis, higher ferrous sulfide produces a marble-like pattern on the skin (Marks, Love and Dadour

2016). These morphological features are categorized as early decomposition in early decomposition in Megyesi et al. (2005) and DittoNau.

Discoloration is one of the more remarkable visual morphological changes during early or active decomposition. Discoloration occurs in all body regions and is caused by putrefaction. As microbes destroy tissues and cells release gases and other byproducts of self-digestion, color changes can be seen in the body, transitioning these color changes from grey to green and progressing to black (Vass, 2001). Other color changes include the darkening around the edges of the nose, eyes, ears, and lips due to tissue drying (Gelderman et al., 2018).

The accumulation of gasses can cause visible swelling of the body, referred to as bloat. Deflation occurs when gasses or fluid is purged through orifices in the remains. (Dautartas, 2009; Janaway, Percival, and Wilson, 2009; Keough, Myburgh, and Steyn, 2017; Lennartz, Hamilton, and Weaver, 2020). This stage of decay is commonly referred to as the "wet" stage due to the sloughing of tissues and the discharge of fluids as membranes degrade. While sloughing of tissues and fluid discharge happen later in the early or active stage of decomposition, they produce prominent visual changes.

Advanced or Halted Decomposition

The advanced or halted decomposition, as labeled in Megyesi et al. (2005) and DittoNau, is characterized by common morphological changes such as sagging and caving in of remaining flesh, bone exposure, and mummification (Lennartz, Hamilton, and Weaver 2020; Mann, Bass and Meadows 1990; Marks, Love and Dadour 2016). The advanced decomposition stage is commonly referred to as the "dry" stage of decomposition. During this stage of decay, microbe activity begins to subside as the putrefaction process has slowed, and skeletonization becomes more prominent as the tissues have been scavenged, weathered, or decomposed (Galloway et al., 1989; Gelderman et al., 2018; Marks, Love, and Dadour, 2016; Megyesi et al., 2005).

Geographic location greatly influences advanced decomposition, as climatological modifiers such as temperature and humidity are related to mummification in extremely hot or cold climates (Bugelli et al., 2018; Lennartz, Hamilton, and Weaver, 2020). Mummification presents quickly and reliably in arid regions (Galloway et al. 1989).

Skeletonization

The simplest stage of human decomposition to assess, skeletonization, occurs after the body's tissue and fluids have decomposed, leaving behind the skeleton. However, it is common for skeletal remains to retain mummified or desiccated tissue (Galloway et al., 1989; Lennartz, Hamilton, and Weaver, 2020; Vass, 2001).

Decomposition Variables and Modifiers

Issues in observing and subsequent scoring of human decomposition are grounded in the assumption that decomposition occurs in a consistently sequential manner. Human decomposition is affected by many independent and dependent factors, so much so that it may be impossible for human observers to precisely predict the timing of morphological feature appearance in human decomposition without the assistance of AI (Dautartas 2009; Gelderman et al. 2019). Visual morphological changes in human decomposition are dependent on multiple factors such as weather, soil composition, the biome, scavenging, insects, microbes, and deposition (Dautartas, 2009; Janaway, Percival, and Wilson, 2009; Marks, Love, and Dadour, 2016). For example, insect activity increases decomposition rates by directly liquifying and consuming tissues and even producing ancillary heat sources (Mann, Bass, and Meadows, 1990). Entomological PMI estimation is based on the life cycles of the insects present and the temperature data for the area where the remains are found. Insect life cycles are measured by accumulated degree days and hours that account for temperature (Franceschetti et al., 2021; Gelderman et al., 2018; Megyesi et al., 2005; Wells, 2019). Decomposition estimation methods such as Megyesi et al. (2005) and Gelderman et al. (2018) integrate accumulated degree days into their PMI formulas but do not include entomology.

As decomposition progresses, multiple intrinsic and extrinsic variables and modifiers of human remains influence the rate of decomposition, such as weather patterns, heat, humidity, insect activity, scavenging, soil composition, and body composition (Galloway et al., 1989; Lennartz, Hamilton, and Weaver, 2020; Simmons et al., 2010). These variables may lead to morphological changes more quickly than expected or slow the process than expected. For example, excessive scavenging activity may expose bone sooner than natural decomposition, while excessively cold temperatures essentially stall decomposition (Jeong, Jantz, and Smith, 2016; Steadman et al., 2018). Soil acidity levels and whether remains are buried can also influence decomposition, such as increasing decomposition rates in which soil acidity is high, which accelerates protein degradation (Pittner et al., 2020; Simmons et al., 2010).

Understanding and Categorizing Human Decomposition

Morphological changes during decomposition show variations in visible presentation and time of presentation. These variations are known to influence PMI estimations. Galloway et al. (1989) sought to standardize the process of categorizing human decomposition. They provided a scoring method of commonly observed characteristics in five stages of progressive morphological changes observed in human decomposition. These stages were fresh, early, advanced, skeletonization, and extreme decomposition. The extreme decomposition stage in Galloway et al. (1989) describes the morphological features of weathering and degradation of skeletal remains. The Galloway et al. (1989) method is specific to Arizona, in which mummification and skeletonization occur more quickly than in temperate climates due to its hot and arid climate. Notably, Galloway et al. (1989) developed a categorical method that does not assign point values to morphological features visually present; rather, it provides a general timeline for decomposition in arid environments based on the presence of these features as an estimator of PMI. This process of categorizing morphological features and applied weather data implies the linearity of human decomposition as it pertains to PMI estimation, in which it is assumed that given a set of parameters such as weather, time, and location, certain morphological features in the absence of scavenging or insect activity, should be progressive from time of death to skeletonization (Galloway et al., 1989).

As the influence of temperature and humidity on decomposition became better understood, Megyesi et al. (2005) modified Galloway's model and produced an enhanced scoring method for temperate climates by eliminating the extreme decomposition stage and morphological features such as the "flesh burned" in the fresh stage from Galloway et al. (1989). Megyesi et al. (2005) amended Galloway's technique for more temperate climates by utilizing accumulated degree days (ADD) to calculate PMI. The Megyesi et al. (2005) method adaptations highlight how variables, such as temperatures, coverings, and insect activity, can accelerate or decelerate decomposition (Dautartas, 2009). These variables have caused researchers to question the general applicability of scoring methods in different geographical locations with varied climates (Dautartas, 2009; Gelderman et al. 2018; Ribéreau-Gayon et al., 2018).

Megyesi et al. (2005) further modified Galloway et al. (1989) by assigning point values to visually present morphological features as they progressed from each stage, with higher values placed on features in the skeletonization stage. Additionally, Megyesi et al. (2005) divided the visual morphological features of human decay into three specific anatomical regions - the head/neck, torso, and limbs. Points were assigned for the features present in these regions, comprising the Total Body Score (TBS) when totaled. A PMI estimation was made by

Megyesi et al. (2005) using a logarithmic calculation with accumulated degree days and the TBS as inputs.

Gelderman et al. (2018) created a method using a scoring method focused on similar morphological features as Megyesi et al. (2005) but eliminated unclear scoring standards, such as fifty percent bone appearance. Utilizing Megyesi et al. (2005) as a basis, Gelderman et al. (2018) assessed morphological features to assign point values when present based on three regions: (a) facial, (b) body, and (c) limb scores. These scores are totaled to create a total decomposition score (TDS), similar to Megyesi et al. (2005). The Gelderman et al. (2018) method omitted stages, focusing on assigning values to the morphological features present in each body region. While removing staging simplifies the scoring process, point scales are still weighted based on morphological features thought to present later in the decomposition process. Gelderman et al. (2018) produces two formulas to predict PMI, one utilizing TDS and ADD as inputs and the other using only TDS.

Temperature, entomology, trauma, and scavenging activity could easily expose bone or mummify tissue before the early/active decomposition stage ceases (Janaway, Percival, and Wilson, 2009; Lennartz, Hamilton, and Weaver, 2020; Wescott, 2018). Notably absent from Galloway et al. (1989), Gelderman et al. (2018), and Megyesi et al. (2005) is the observation of insect and scavenging activity, both of which are known to accelerate decomposition rates and hence change the morphological appearance of the body (Franceschetti et al., 2021; Steadman et al. 2018).

Cross and Simmons (2010) studied penetrative trauma and subsequent insect activity on decay rates in human analogs. They found that this type of trauma had little to no influence on decay rates and that modifiers such as insects preferred natural orifices to openings created by penetrative trauma. Additionally, the environment in which remains are found and their position will influence insect colonization and minimize or increase the rate of human decomposition. For example, if a body is suspended from hanging, there is a notable decrease in insect activity, and the body has a higher incidence of irregular decomposition and partial mummification (Bugelli et al., 2018).

Animal scavenging also influences the progression of human decomposition. Scavengers target soft tissues such as the eyes and genitals, as well as limbs, and have been shown to both hasten the progression of decomposition as well as produce mummification. (Jeong, Jantz, and Smith 2016; Steadman et al. 2018). A persistent challenge in categorizing morphological features is accounting for the number of variables, their influence on human decomposition, and their effect on determining PMI (Cockle and Bell, 2015).

Which morphological features of human decay should be included, when they typically appear, and how important they are in the assessment process when assessing human decomposition continues to be a challenge in the development of evaluative methods. There is conflicting information on how individual body regions should be examined and their weighted importance in determining the individual's total decomposition stage in decomposition assessment methods. These conflicting methods of evaluation are evident in how Megyesi et al. 2005 and Gelderman et al. 2018 approach assessing limb decomposition, the appearance of bone, and the percentage of appearance required to appropriately categorize the stage of decomposition (Galloway et al., 1989; Gelderman et al., 2018; Marks, Love, and Dadour, 2016; Megyesi et al., 2005; Vidoli, Kenyhercz, and Steadman, 2020). Additional issues in assessing human decomposition are rater subjectivity and error, which demonstrates potential confusion in the application of the methods and assessment by the rater on the morphological features of decomposition present (Gelderman et al., 2019; Sauerwein, 2018).

Galloway et al. (1989) categorized human decomposition into stages represented by common morphological features. However, practitioners differ in identifying the liminal point of human decomposition, in which the body advances from one stage to the next. For example, Megyesi et al. (2005) illustrated that advanced decay is most prominently noted with the appearance of bone on less than half the area scored and/or mummified/desiccated tissue present on less than one-half of the area observed, with minor differences based on the body part examined. This liminal point differs from Gelderman et al. (2018) which lists gross skeletonization as the liminal morphological feature and eliminates the subjective 50% bone appearance.

Further confusing the task of assessing decomposition and estimating the PMI is the approach to scoring. Methods vary in how assessments of each body region, such as how limbs are examined, categorization of morphological features whether by points or stages, and how these scores are assessed in their totality (Gelderman et al. 2017; Megyesi et al., 2005; Vidoli, Kenyhercz, and Steadman 2020). Due to the variety of factors that affect decomposition and the variance in objectively scoring these stages by human observers, AI may provide a consistent methodology to evaluate human decomposition and eliminate potential errors. However, for an accurate AI method to be developed, a training sample and decomposition criteria must be given that are consistent and accurate. To strengthen AI's performance, an agreed-upon methodology with clearly defined morphological features by forensic experts would further enhance AI's ability to determine the state of human decomposition.

Out of respect for the donors, their families, and the community, Figures representing examples of the stages of human decomposition and method labeling can be requested from the author, committee chair, or director of the Anthropology Research Facility at the University of Tennessee, Knoxville. They will not be included in this thesis.

CHAPTER THREE

MATERIALS AND METHODS

This chapter discusses how more than twenty thousand photos were utilized to train AI deep learning algorithms to categorize human decomposition accurately. A total of 20,848 digital photos of human decomposition at The University of Tennessee, Knoxville, Anthropology Research Facility (ARF) from 2012 to 2023 were included in this study. The ARF at UTK is a human decomposition facility where human donors are placed outdoors to decompose naturally. The skeletons are recovered and accessioned into the UTK Donated Skeletal Collection. During the decomposition process, which can last weeks to more than a year, depending largely upon the season of placement, photos of morphological features of decomposition are captured daily. These daily photos are then saved to the facility's digital catalog and categorized by donor and date. Photos of donors capture the present morphological features of the donor in each body region, such as the head, torso, and limbs. Photos of each donor continue until the donor has skeletonized, recovered from the facility, processed, and assessed into the UTK Donated Skeletal Collection. Originally, all the photos were electronically stored in file folders by donor but were not searchable. The development of the web-based platform ICPUTRD, discussed above, has created searchable functions and new platforms to be developed, allowing this project to be possible (Nau et al. 2023).

Photo Sample for Training AI

Successful AI training depends upon including large data samples, which assists in accounting for human bias and is representative of the proposed population. As such, a large independent sample was used to capture morphological variety and account for potential bias in image selection (Bhushetty, 2023). Three groups of photos were randomly generated from the ARF's digital catalog. The three groups were placed in ICPUTRD for analysis. All photos represent donors who decomposed outside on the ground surface. No photos of human decomposition in indoor settings, burials, vehicles, or other settings were considered.

Donor demographics included males and females aged 18 and older and of any population affinity. Photos were excluded from the sample if the images were of poor quality, such as those that were blurred or too dark to discern body regions visually.

Human Decomposition Categorical Methods and Photo Labeling Process

Three human decomposition assessment methods were tested to determine if AI can categorize the stage of human decomposition achieved for the donor in each photo. These methods were Megyesi et al. (2005), Gelderman et al. (2018), and DittoNau. The Megyesi et al. (2005) method was modified for this study by assigning the present morphological features in human decomposition to the Stage of Decay (SOD) rather than utilizing the TBS. Gelderman et al. (2018) was modified such that the associated TDS score for the morphological feature assessed in the photo was not used as a score for calculation and rather as a proxy for SOD. Gelderman et al. (2018) was the only method selected that did not categorize morphological features into stages and would have required significant modifications to the method to do so. As such, instead of a SOD, Gelderman et al. (2018) were categorized into SCORES, which ranged from 1-6 based on the morphological features listed in the method. These modifications were necessary so each method's performance could be tested and compared after AI training to determine if there was any difference in AI success. No modifications were made to DittoNau as it was created for AI compatibility. Modifying these methods into similar categories allows AI to be trained and tested to determine the morphological features present in the photo and assign them to the appropriate category. For example, morphological features consistent with the fresh stage would be categorized into SOD1 for Megyesi et al. (2005), comparable to SCORE1 for Gelderman et al. (2018) and SOD1 for DittoNau. Similarly, morphological features consistent with the advanced or halted stage would be categorized into SOD3 for Megyesi et al. (2015) is comparable to SCORE4 for Gelderman et al. (2018) and SOD4 for DittoNau. The wide variation in decomposition categorization methods necessitated some modifications to evaluate the efficacy of which models best informed AI on the morphological features of human decomposition.

ICPUTRD is a web-based platform utilized to view and label the stage or score of human decomposition in the image via drop-down menus compiled by the computer and data science team. Each page displayed a photo array with multiple rows of photos of multiple donors in various states of decomposition. While labeling the stage or score of decomposition, only one method was tested at a time. Photo arrays were created for three specific body regions - the head, torso, and limbs. Images that contained the wrong body region were recategorized to the appropriate body region array. For example, when assessing and labeling human decomposition in the head region, if a torso photo was found in the array, the image would be labeled torso and recategorized into the appropriate body region for assessment.

Once a photo array was generated in ICPUTRD, I evaluated each photo for the present morphological features of decomposition. I labeled the evident human decomposition present based on the method being assessed, selecting the appropriate stage or score in the drop-down menu for each photo. This process was completed for each body region utilizing the three different methods. Labeling for the limbs was based on the most visible limb present in the photo.

The labeled photos from each body region based on each method in ICPUTRD were used to train AI machine algorithms. Three independent AI algorithms were developed by computer scientists using Xception Convolutional Neural Network to be trained based on labeled imaging data for each method. AI-model-specific algorithms then evaluated a photo population from the labeled data and predicted the stage of human decomposition. Figure 1 illustrates the research design for this thesis.

Megyesi et al. (2005)

Megyesi et al. (2005) was the first method used to assess decomposition in the photo sample. This method was used to categorize human decomposition into four stages: (a) Fresh (SOD1), (b) Early Decomposition (SOD2), (c) Advanced Decomposition (SOD3), and (d) Skeletonization (SOD4).

The photo sample generated to assess Megyesi et al. (2005) contained 8,351 images. Each photo was labeled with the SOD based on the most advanced visible morphological features corresponding to the applicable stage. For example, if evaluating a limb with gray to green discoloration and bone exposure less than one-half of the area scored, the image would be labeled as SOD3 (advanced decomposition). The morphological features categorized in each stage in Megyesi et al. (2005) and their corresponding SOD are presented in Tables 1.1, 1.2, and 1.3 by body region of TBS.

The Megyesi-based labeled photo sample was then utilized to train the AI algorithm (Figure 1). To evaluate the efficacy of the AI training algorithm for the Megyesi method, the first body region assessed was Head/Neck because this was hypothesized to be the body region easiest for AI to recognize features. A total of 4,220 head images were labeled. It was determined that body region photo sample populations would need approximately 2,000 images to adequately train the machine learning AI algorithm based on AI population training for Machine and Deep Learning models (Bhushetty, 2023; Shivaprasad, 2020). Each body region was assessed and labeled once the appropriate photo sample threshold was established. For AI training, 1,979 torso images and 2,152 limb images were labeled and utilized.

Gelderman et al. (2018)

Gelderman et al. (2018) was the second method assessed to train AI in categorizing human decomposition. Repeating the process used with Megyesi et al. (2005), a second randomly generated photo sample was created in ICPUTRD. A total of 6,055 photos of donors in various stages of human decomposition of the same donor pool were assessed and labeled. The sample size is notably lower than used to assess Megyesi. After evaluating the efficacy of the Megyesi algorithm, it was determined that each model required a data set of approximately 2,000 photos per body region to train and assess each AI-method-specific algorithm. Body regions were evaluated based on Gelderman et al. (2018). Sample sizes comprised 2,041 head photos, 1,982 toros photos, and 2,032 limb photos. These images were then utilized to train AI.

As with Megyesi, photographs were labeled based on the most advanced morphological feature present for the assessed body region. Based on the most advanced morphological feature present in each photo, the point value that would typically be assigned was used as an analog to SOD and recategorized as a SCORE. For example, as seen in Table 1.4, a point would be converted to a SCORE based on the morphological feature. For example, when evaluating the head region, if the morphological feature present was "no visible changes."

Instead of receiving one point for use in the TDS, the photo would be labeled SCORE1.

DittoNau Novel Method

The DittoNau method is a new approach to categorizing human decomposition based on repetitive object recognition, in this case, visible morphological features of thousands of individual photographs examined for the first two methods of this study. The DittoNau method incorporates two modifiers of human decomposition - the presence of entomological activity and scavenging - that are not considered in the other two methods. This method was constructed considering AI deep learning to emphasize the most visually recognizable characteristics, such as color changing and entomological activity.

The DittoNau method was developed to test AI's ability to learn decomposition by simplifying and categorizing visible morphological features. The approach utilizes the foundation of Megyesi et al. (2005) and Gelderman et al. (2018), incorporating five separate stages: (a) Fresh, (b) Initial Decomposition, (c) Active Decomposition, (d) Advanced or Halted Decomposition, and (e) Skeletonization. These stages represent the corresponding SOD. SOD1 (Fresh), SOD2 (Initial Decomposition), SOD3 (Active Decomposition), SOD4 (Advanced or Halted Decomposition), SOD5 (Skeletonization). The DittoNau simplifies the observation of skeletal exposure by removing specific presentations such as bone appearance percentage or whether joints are articulated or not. The morphological features are similar to Megyesi et al. (2005) and Gelderman et al.

(2018) and are grouped by the same body regions as the other methods: head, torso, and limbs. Additionally, DittoNau notes the presence of entomological activity and scavenging as an indicator of active decomposition. For example, if there appears to be no noticeable visible morphological changes, but insect activity is noted on the body, the stage of decomposition would still be categorized as active. The DittoNau method does not assign point values to the morphological features present, as this method has no PMI calculation. The stages and morphological features of the DittoNau method are presented in Table 1.5.

A third randomly generated photo sample from the donor pool was created and evaluated using the DittoNau method. A total of 6,442 photos were assessed and labeled according to the stage of human decomposition present based on the DittoNau method. This sample consisted of 2,076 head photos, 2,020 torso photos, and 2,346 limb photos and was used to train the AI-model-specific algorithm.

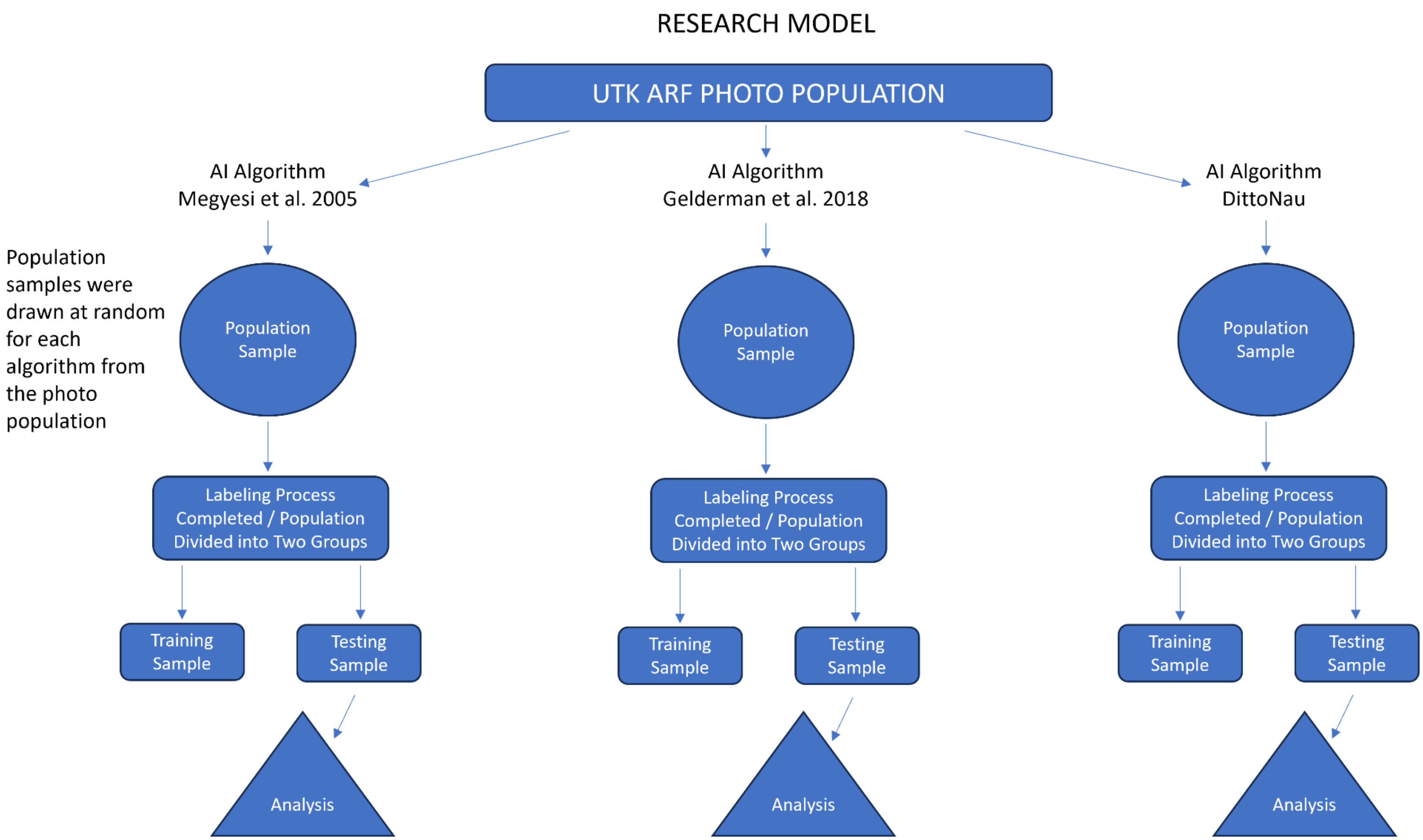


Figure 1. Research Design

Table 1. Categories and stages of decomposition for the head and neck. Megyesi et al. (2005). Fresh (SOD1), Early Decomposition (SOD2), Advanced Decomposition (SOD3), and Skeletonization (SOD4). Each photo was labeled with the SOD based on the visible morphological features corresponding to the applicable stage.

STAGE:	DESCRIPTION: HEAD and NECK
Fresh	
1 PT	Fresh, No discoloration.
Early Decomposition	
2 PTS	Pink-white appearance with skin slippage and some hair loss.
3 PTS	Gray to green discoloration: some flesh still relatively fresh.
4 PTS	Discoloration and/or brownish shades particularly at edges, drying of nose, ears, lips
5 PTS	Purging of decompositional fluids out of eyes, ears, nose, mouth, some bloating may be present.
6 PTS	Brown to black discoloration of flesh.
Advanced Decomposition	
7 PTS	Caving in of flesh and tissues of eyes and throat.
8 PTS	Moist Decomposition with bone exposure less than one half that of the area being scored.
9 PTS	Mummification with bone exposure less one half that of the area being scored.
Skeletonization	
10 PTS	Bone exposure of more than half the area scored with greasy substances and decomposed tissue.
11 PTS	Bone exposure of more than half the area scored with desiccated or mummified tissue.
12 PTS	Bones largely dry, but retaining some grease.
13 PTS	Dry Bone.

Table 2. Categories and stages of decomposition for the trunk. Megyesi et al. (2005). Fresh (SOD1), Early Decomposition (SOD2), Advanced Decomposition (SOD3), and Skeletonization (SOD4).

STAGE:	DESCRIPTION: TRUNK
Fresh	
1 PT	Fresh, No discoloration.
Early Decomposition	
2 PTS	Pink-white appearance with skin slippage and marbling.
3 PTS	Gray to green discoloration: some flesh relatively fresh.
4 PTS	Bloating with green discoloration and purging of decompositional fluids.
5 PTS	Postbloating following release of abdominal gasses, with discoloration changing from green to black.
Advanced Decomposition	
6 PTS	Decomposition of tissue producing sagging of flesh; caving in of the abdominal cavity.
7 PTS	Moist Decomposition with bone exposure less than one half that of the area being scored.
8 PTS	Mummification with bone exposure less one half that of the area being scored.
Skeletonization	
9 PTS	Bone with decomposed tissue, sometimes with body fluids and grease still present.
10 PTS	Bones with desiccated or mummified tissue covering less than one half of the area being scored.
11 PTS	Bones largely dry, but retaining some grease.
12 PTS	Dry Bone.

Table 3. Categories and stages of decomposition of the limbs. Megyesi et al. (2005). Fresh (SOD1), Early Decomposition (SOD2), Advanced Decomposition (SOD3), and Skeletonization (SOD4).

STAGE:	DESCRIPTION: Limbs
Fresh	
1 PT	Fresh, No discoloration.
Early Decomposition	
2 PTS	Pink-white appearance with skin slippage of hand and/or feet.
3 PTS	Gray to green discoloration; marbling; some flesh still relatively fresh.
4 PTS	Discoloration and/or brownish shades particularly at edges, drying of fingers, toes, and other projecting extremities.
5 PTS	Brown to black discoloration, skin having a leathery appearance.
Advanced Decomposition	
6 PTS	Moist decomposition with bone exposure less than one half that of the area being scored.
7 PTS	Mummification with bone exposure less one half that of the area being scored.
Skeletonization	
8 PTS	Bone exposure more than half the area scored, some decomposed tissue and body fluids remaining.
9 PTS	Bones largely dry, but retaining some grease.
10 PTS	Dry Bone.

Table 4. Developed decomposition scoring method. Gelderman et al. (2018). The morphological features listed by point are represented as a SCORE for each body region in this study. For example, the torso (BDS) "no visible changes" would be categorized as SCORE1.

FDS		
Points		
	1	1.1 No visible changes
	2	2.1 Livor mortis, rigor mortis and vibices
		2.2 Eyes: cloudy and/or tache noir
		2.3 Discoloration: brownish shades particularly at the edges. Drying of nose, ears, lips
	3	3.1 Grey to green discoloration
		3.2 Bloating of neck and face is present and/or skin blisters, skin slippage and/or marbling
		3.3 Purging of decompositional fluids out of ears, nose and mouth and/or brown to black discoloration
	4	4.1 Caving in of the flesh and tissues of eyes and throat. Skin having a leathery appearance
		4.2 Partial skeletonization, joints still together
	5	5.1 Gross skeletonization, some joints disarticulated
	6	6.1 Complete skeletonization
BDS		
Points		
	1	1.1 No visible changes
	2	2.1 Livor mortis, rigor mortis and vibices
	3	3.1 Grey to green discoloration
		3.2 Bloating with green discoloration and/or skin blisters, skin slippage and/or marbling
		3.3 Rectal purging of decompositional fluids
		3.4 Post-bloating: release of abdominal gases with discoloration changing from green to black
	4	4.1 Decomposition of tissue producing sagging of flesh. Caving in of the abdominal cavity
		4.2 Skin having a leathery appearance
		4.3 Partial skeletonization, joints still together
	5	5.1 Gross skeletonization, some joints disarticulated
	6	6.1 Complete skeletonization
LDS		
Points		
	1	1.1 No visible changes
	2	2.1 Livor mortis, rigor mortis and vibices
		2.2 Discoloration: brownish shades particularly at the edges. Drying of fingers and toes
	3	3.1 Skin blisters, skin slippage and/or marbling
		3.2 Grey to green discoloration
		3.3 Brown to black discoloration
	4	4.1 Skin having a leathery appearance
		4.2 Partial skeletonization, joints and tendons still together
	5	5.1 Gross skeletonization, some joints disarticulated
	6	6.1 Complete skeletonization

Table 5. The DittoNau novel method. Fresh (SOD1), Initial Decomposition (SOD2), Active Decomposition (SOD3), Advanced or Halted Decomposition (SOD4), Skeletonization (SOD5).

Stage:	Morphological Change:
Fresh	No appearance of decomposition
Initial Decomposition	Lividity Marbling
Active Decomposition	Discoloration: Grey-Green-Brown-Black Insect Activity Present Indications of scavenging (Torn flesh, missing anatomical features) Purging of fluids Bloat (primarily in Torso) Deflation (primarily in Torso) Hair loss Skin Slippage
Advanced or Halted Decomposition	Mummified tissue with or without bone appearance (Black, white, leatherlike, without moisture or active decomposition)
Skeletonization	Skeletonization, Articulated or Disarticulated

STAGE:	DESCRIPTION: HEAD and NECK
Fresh	
1 PT	Fresh, No discoloration.
Early Decomposition	
2 PTS	Pink-white appearance with skin slippage and some hair loss.
3 PTS	Gray to green discoloration: some flesh still relatively fresh.
4 PTS	Discoloration and/or brownish shades particularly at edges, drying of nose, ears, lips
5 PTS	Purging of decompositional fluids out of eyes, ears, nose, mouth, some bloating may be present.
6 PTS	Brown to black discoloration of flesh.
Advanced Decomposition	
7 PTS	Caving in of flesh and tissues of eyes and throat.
8 PTS	Moist Decomposition with bone exposure less than one half that of the area being scored.
9 PTS	Mummification with bone exposure less one half that of the area being scored.
Skeletonization	
10 PTS	Bone exposure of more than half the area scored with greasy substances and decomposed tissue.
11 PTS	Bone exposure of more than half the area scored with desiccated or mummified tissue.
12 PTS	Bones largely dry, but retaining some grease.
13 PTS	Dry Bone.

Megyesi et al. 2005

Stage of Decay (SOD) 1-4

FDS	
Points	
1	1.1 No visible changes
2	2.1 Livor mortis, rigor mortis and vibices
	2.2 Eyes: cloudy and/or tache noir
	2.3 Discoloration: brownish shades particularly at the edges. Drying of nose, ears, lips
3	3.1 Grey to green discoloration
	3.2 Bloating of neck and face is present and/or skin blisters, skin slippage and/or marbling
	3.3 Purging of decompositional fluids out of ears, nose and mouth and/or brown to black discoloration
4	4.1 Caving in of the flesh and tissues of eyes and throat. Skin having a leathery appearance
	4.2 Partial skeletonization, joints still together
5	5.1 Gross skeletonization, some joints disarticulated
6	6.1 Complete skeletonization

Gelderman et al. 2018

Score (SCORE) 1-6

Stage:	Morphological Change:
Fresh	No appearance of decomposition
Initial Decomposition	Lividity
	Marbling
Active Decomposition	Discoloration: Grey-Green-Brown-Black
	Insect Activity Present
	Indications of scavenging (Torn flesh, missing anatomical features)
	Purging of fluids
	Bloat (primarily in Torso)
	Deflation (primarily in Torso)
	Hair loss
	Skin Slippage
Advanced or Halted Decomposition	Mummified tissue with or without bone appearance (Black, white, leatherlike, without moisture or active decomposition)
Skeletonization	Skeletonization, Articulated or Disarticulated

DittoNau

Stage of Decay (SOD) 1-5

Figure 2. Stage of Decay Classifications. Groupings were developed to keep the integrity of the method intact for evaluating performance and to compare similar morphological features in each of the three methods.

Inter-observer Testing

An inter-observer study was conducted to evaluate reliability among observers to test the reliability in applying the three methods. Two additional evaluators were given access to ICPUTRD, in addition to the author, and instructions on labeling photos based on the criteria established from the three methods. The three observers in the inter-observer testing for each of the AI-method-specific algorithms were: a master's candidate in biological anthropology with training in osteology and forensics (the author), a Ph.D. candidate in biological anthropology and expertise in forensics, and a full-time forensics expert with a master's degree in biological anthropology.

For the inter-observer testing sample population, the computer science and data team selected 300 new torso images generated from the ARF's digital catalog that were not used previously and placed them into ICPUTRD to be used as the inter-observer testing sample. Each rater received the same 300 photos. The photographs were batched into 18 photo arrays, with six batches evaluated per method. Each batch contained 50 random images from the photo sample. No identifying information was present regarding the donor in the sample. This same 300 photo sample would be labeled a second time (by the author) to evaluate the intra-observer error.

To reduce bias in the interobserver testing, each photo batch was randomly assigned a method so that no method was repeated from array to array. As the rater observes, assesses, and labels each photo from the array, the next batch would require the rater to observe, assess, and label those photos using a method different from the one previously used. For example, Batch 1 may require the rater to utilize Megyesi et al. (2005) to evaluate each photo, Batch 2 then would require the rater to use Gelderman et al. (2018), and Batch 3 evaluated the DittoNau method. These batches were created so that each observer could concentrate on the specific method evaluation criteria required for the presented batch of photos. The photos, batches, and methods were identical for each rater for comparison.

The inter-observer study analysis was completed in R Studio using the "irr" package, computing Fleiss' Kappa score and p-value. Values were analyzed on a scale of 1 to 0, with 1 representing perfect agreement among observers. Intra-observer analysis was conducted in R Studio to analyze reliability using the intraclass correlation coefficient or ICC, with 1 representing strong reliability and 0 representing poor reliability. Substantial agreement is represented with a value of 0.8-0.61, moderate agreement with values between 0.6-.41, fair agreement with values between 0.4-0.21, and slight agreement with values between 0.2-0.01.

Intra-observer Testing

To evaluate intra-observer error, utilizing the same sample photo array from the inter-observer testing, I repeated the labeling process on these 300 photos a second time three weeks after the initial labeling was completed. The error was calculated using an interclass correlation coefficient. Values are assessed for reliability strength, with values less than 0.5 being poor, 0.5-0.75 moderate, 0.75-0.9 good, and values greater than 0.9 indicating excellent reliability (Koo and Li, 2016).

Statistical Analysis

A statistical analysis was conducted based on AI's performance with labeled photos not used in the training algorithm to examine the performance of AI-method-specific machine learning algorithms in learning each model's visual morphological features (See Figure 1). The statistical performance values are Precision, Recall, and the micro F1 score (MaF1).

Precision is the proportion of accurate predictions that are true and is calculated by dividing the number of true positives and true positives plus false positives. Models with greater than .80 or 80% are generally considered acceptable models. Recall is the proportion of true positive predictions correctly made by the examiner and is calculated by dividing true positives by true positives plus false negatives. Models with a recall of more than .50 or 50% recall are considered acceptable models (Shivaprasad, 2020). The MaF1 score, also known as the harmonic mean, is the value to evaluate a test's performance. It is calculated by multiplying by two, the value of precision multiplied by recall, divided by precision, and recall added together. A MaF1 score between 0 and 1 is calculated, with 1 representing high performance and 0 representing low performance. MaF1 scores greater than 0.9 are considered very good, 0.8-0.9 are good, 0.5-0.8 are marginal, and scores less than 0.5 are poor. The MaF1 score evaluates the balance of how precise a model is and or how sensitive it is. This score may indicate if the data is unbalanced or if there are potential errors with sampling (Bajaj, 2019).

Quality Control and Assurance

After each AI-model-specific algorithm was trained, the model's predictions of the stage of each photo were compiled into an Excel spreadsheet. AI predictions inconsistent with that photo's labeling data were evaluated for causes and errors. This process was implemented to evaluate whether inconsistencies were related to human error, or the AI algorithm incorrectly assessed the image. This process was generated and repeated for each body region for each method.

$$\frac{\text{True Positives}}{\text{True Positives + False Positives}} = \text{Precision}$$

$$\frac{\text{True Positives}}{\text{True Positives + False Negatives}} = \text{Recall}$$

$$2 * \frac{\dfrac{\text{True Positives}}{\text{True Positives + False Positives}} * \dfrac{\text{True Positives}}{\text{True Positives + False Positives}}}{\dfrac{\text{True Positives}}{\text{True Positives + False Positives}} + \dfrac{\text{True Positives}}{\text{True Positives + False Positives}}} = \text{MaF1}$$

Figure 3. Calculation formulas for Precision, Recall, and MaF1 Scores

CHAPTER FOUR

RESULTS

This chapter will present the results of this study. Statistical analysis was generated by the AI-method-specific algorithm testing data exhibiting the model's overall performance. These scores are the MaF1 score, the precision score illustrating the model's accuracy, and the sensitivity of the model's performance in the recall score. Overall results and comparison are presented first, followed by an analysis of the performance of each method, the inter-observer testing, and the intra-observer testing.

Overall Results

Figure 3 illustrates the precision scores for each body region for the method utilized. Precision evaluates the percentage of true positive predictions (Shivaprasad, 2020). The Megyesi et al. (2005) method had the following precision scores: head region - 0.89, torso region - 0.87, and limbs region - 0.79. Gelderman et al. (2018) had the following precision scores: head region - 0.90, torso region - 0.85, and limbs region - 0.89. The DittoNau method had the following precision scores: head region - 0.94, torso region - 0.88, and limbs region - 0.88. Precision performance was highest in the DittoNau method for the head region, and Megyesi et al. (2005) had the lowest precision in the limbs region. Torso, like MaF1 performance, was consistent throughout without significant differences.

Figure 4 illustrates the recall scores for each body region for the method utilized. Recall evaluates the percentage of positive predictions of the stage of decay that were correctly identified from the training data, also known as sensitivity (Shivaprasad, 2020). The Megyesi et al. (2005) method had the following recall scores: head region - 0.86, torso region - 0.88, and limbs region - 0.66. Gelderman et al. (2018) had the following recall scores: head region – 0.85, torso region - 0.91, and limbs region - 0.87. The DittoNau method had the following recall scores: head region - 0.93, torso region - 0.90, and limbs region - 0.86. Recall performance was highest with DittoNau in the head region, while Megyesi et al. (2005) had the lowest recall performance in the limbs region.

Figure 5 demonstrates the MaF1 scores of each body region for the methods utilized. MaF1 evaluates the machine learning model's accuracy by taking the harmonic mean of precision and recall (Kundu, 2022). The Megyesi et al. (2005) method had MaF1 scores of 0.87 for the head region, 0.88 for the torso region, and 0.70 for the limbs region. Gelderman et al. (2018) MaF1 scores were 0.88 for the head region, 0.87 for the torso region, and 0.87 for the limbs region. The DittoNau method MaF1 scores were 0.93 for the head region, 0.89 for the torso region, and 0.87 for the limbs region. Performance in the limbs region was notably lower for Megyesi et al. (2015), and the DittoNau method had the highest performance in the head region. The torso region was consistent among all three methods, with insignificant variation and overall good performance.

Megyesi et al. (2005)

Each method was evaluated based on AI-generated precision and recall scores for the applicable stage of decomposition of each body region. Figure 15 illustrates the precision scoring for Megyesi et al. (2005). SOD1 had the following precision scores: head region - 0.93, torso region – 1.0, limbs region – 0.75. SOD2 had the following precision scores: head region - 0.87, torso region – 0.92, and limbs region – 0.92. SOD3 had the following precision scores: head region - 0.88, torso region – 0.75, and limbs region – 0.6. SOD4 had the following precision scores: head region - 0.89, torso region – 0.82, limbs region – 0.92. Overall, the precision scoring for Megyesi et al. (2005) was consistently good, with a precision greater than 0.8, except for the limbs region scores in SOD 1 and SOD 2.

Figure 6 demonstrates the recall scoring for Megyesi et al. (2005). SOD1 had the following recall scores: head region - 0.84, torso region – 1.0, limbs region – 0.5. SOD2 had the following recall scores: head region - 0.96, torso region – 0.89, limbs region – 0.97. SOD3 had the following recall scores: head region - 0.72, torso region – 0.76, limbs region – 0.21. SOD4 had the following recall scores: head region - 0.95, torso region – 0.87, limbs region – 0.96. Recall scores for the head and torso regions were relatively consistent. However, the limbs region in both SOD1 and SOD3 scored notably lower than other regions.

Gelderman et al. (2018)

Figure 7 exhibits the precision scoring for Gelderman et al. (2018). SCORE1 had the following precision scores: head region – 1.0, torso region – 0.6, limbs region – 1.0. SCORE2 had the following precision scores: head region - 0.88, torso region – 0.96, limbs region – 0.85. SCORE3 had the following precision scores: head region - 0.94, torso region – 0.97, limbs region – 0.9. SCORE4 had the following precision scores: head region - 0.8, torso region – 0.79, limbs region – 0.75. SCORE5 had the following precision scores: head region - 0.81, torso region – 0.78, limbs region – 0.82. SCORE6 had the following precision scores: head region – 1.0, torso region – 0.94, limbs region- 1.0. Precision scoring for Gelderman et al. (2018) had overall higher yielding scores than Megyesi et al. (2005) and DittoNau, with the heads and limbs in SCORE1 and SCORE 6 achieving perfect precision. However, there was more variation of scores throughout, although none were significant. The lowest score was for the torso region in SCORE1 at 66%.

Figure 8 displays the recall scoring for Gelderman et al. (2018). These are the following recall scores for SCORE1: head region - 0.6, torso region – 1.0, limbs region – 1.0. SCORE2 had the following recall scores: head region – 1.0, torso region – 0.93, limbs region – 0.94. SCORE3 had the following recall scores: head region - 0.91, torso region – 0.86, limbs region – 0.86. SCORE4 had the following recall scores: head region - 0.82, torso region – 1.0, limbs region – 0.82. SCORE5 had the following recall scores: head region - 0.9, torso region –

0.88, limbs region – 0.90. SCORE6 had the following recall scores: head region - 0.88, torso region – .80, limbs region – 0.70. Recall performance for Gelderman et al. (2018) was consistently good, with four body regions achieving perfect recall. The lowest recall was SCORE1 in the torso region.

DittoNau Method

Figure 8 demonstrates the Precision scoring for DittoNau. SOD1 had the following precision scores: head region – 1.0, torso region – 0.81, limbs region – 0.94. SOD2 had the following precision scores: head region - 0.86, torso region – 0.87, and limbs region – 0.93. SOD3 had the following precision scores: head region - 0.92, torso region – 0.98, and limbs region – 0.90. SOD4 had the following precision scores: head region – 1.0, torso region – 0.81, limbs region – 0.77. SOD5 had the following precision scores: head region - 0.93, torso region – 0.95, limbs region – 0.84. Precision scoring for the DittoNau method was consistent and good, with overall better precision than Megyesi and Gelderman, with only SOD4 limbs scoring less than 80%, with a score of 77%.

Figure 9 shows the recall scoring for DittoNau. SOD1 had the following recall scores: head region – 1.0, torso region – 0.75, limbs region – 0.94. SOD2 had the following recall scores: head region – 1.0, torso region – 0.98, limbs region – 0.96. SOD3 had the following recall scores: head region - 0.98, torso region – 0.88, limbs region – 0.95. SOD4 had the following recall scores: head region - 0.73, torso region – 0.94, limbs region – 0.67. SOD5 represents the skeletonization stage and had the following recall scores: head region - 0.96, torso region – 0.95, limbs region – 0.77.

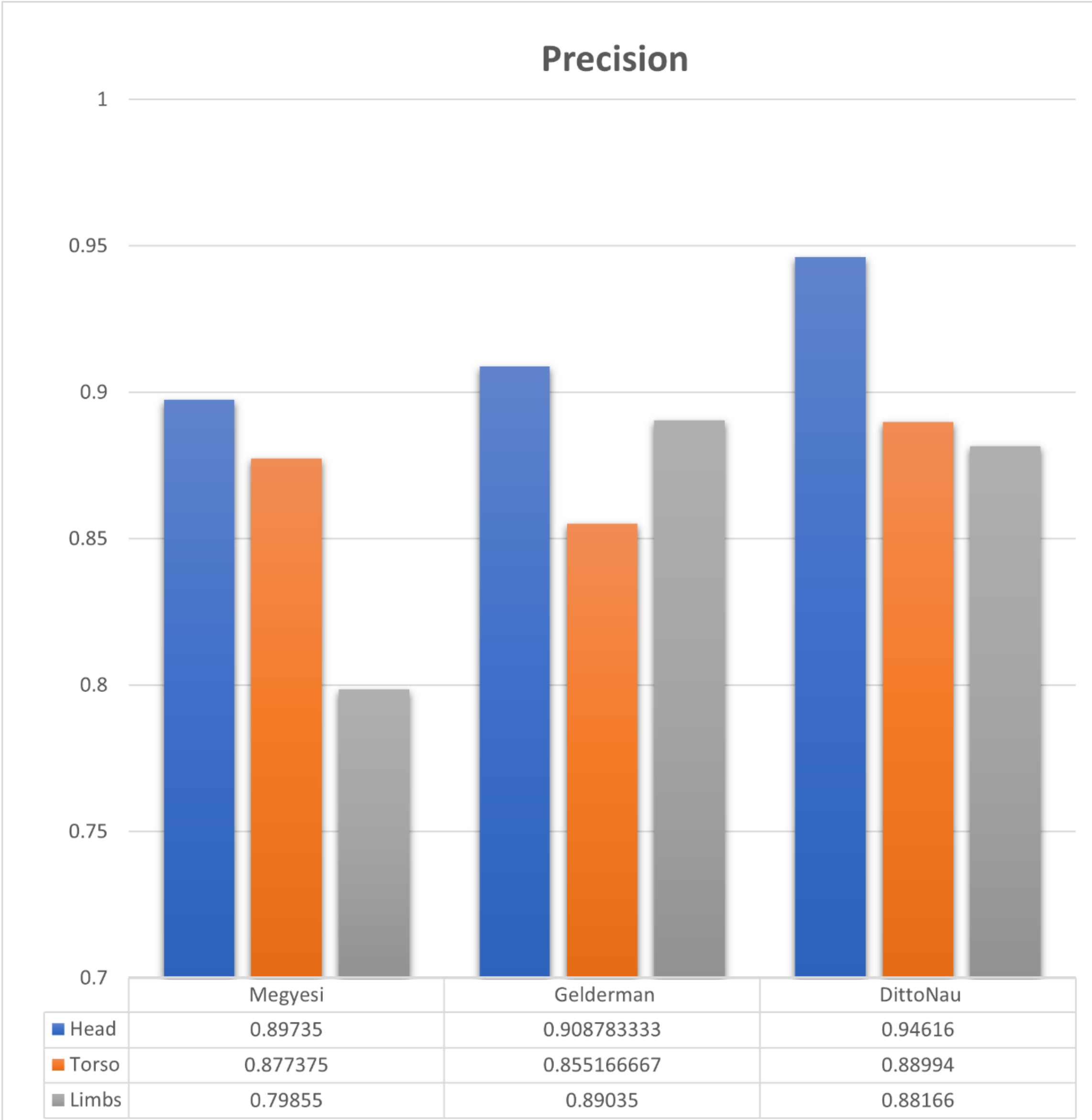
Inter-observer Testing

Inter-observer scores were evaluated utilizing a Fleiss' Kappa test for inter-rater reliability among two or more raters with ordinal data (Fleiss, Levin, and Paik, 2003). Intra-observer error (IOE) using Megyesi et al. (2005) resulted in a Kappa value of 0.67, indicating substantial agreement, and a p-value of 0.00. For the Gelderman et al. (2018) method, IOE resulted in a Kappa value of 0.59, indicating moderate agreement and a p-value of 0.00. Utilizing the DittoNau method, IOE resulted in a Kappa value of 0.57, indicating moderate agreement, with a p-value of 0.00, IOE results are presented in Figure 11.

Intra-observer Testing

The intra-observer error was evaluated using an intraclass correlation coefficient (ICC) for intra-rater reliability. Each method was evaluated for reliability.

Megyesi et al. (2005) resulted in an ICC value of 0.95, indicating excellent reliability and a p-value of < 0.00 with a 95% confidence interval of 0.94-0.96. Gelderman et al. (2018) resulted in an ICC value of 0.95, indicating excellent reliability and a p-value of < 0.00 with a 95% confidence interval of 0.94-0.96. DittoNau resulted in an ICC value of 0.90, indicating excellent reliability, and a p-value of < 0.00 with a 95% confidence interval of 0.88-0.93, indicating statistically good to excellent reliability (see Figure 12).



	Megyesi	Gelderman	DittoNau
Head	0.89735	0.908783333	0.94616
Torso	0.877375	0.855166667	0.88994
Limbs	0.79855	0.89035	0.88166

Figure 4. Precision scores exhibit each model's accuracy in determining the SOD based on body region.

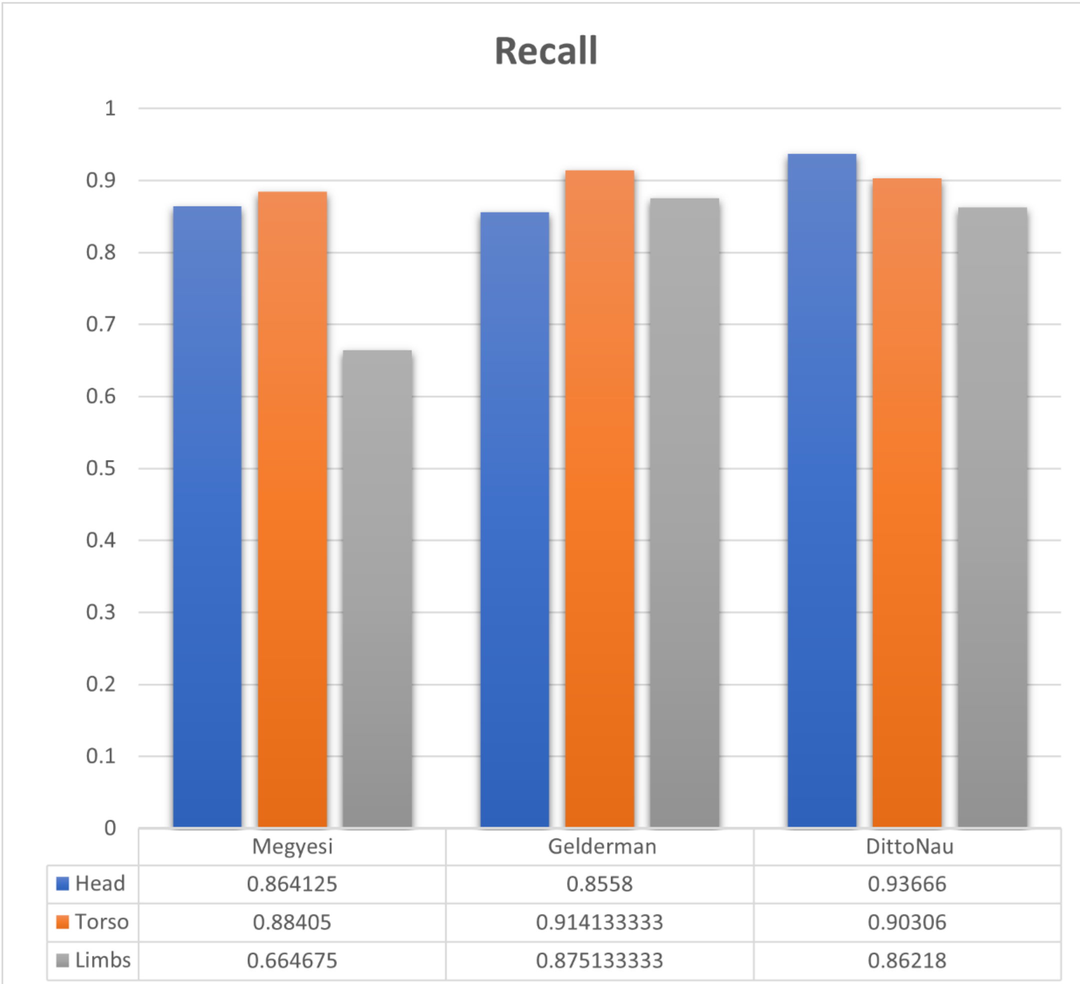


Figure 5. Recall scores exhibit each model's sensitivity in determining the SOD based on body region.

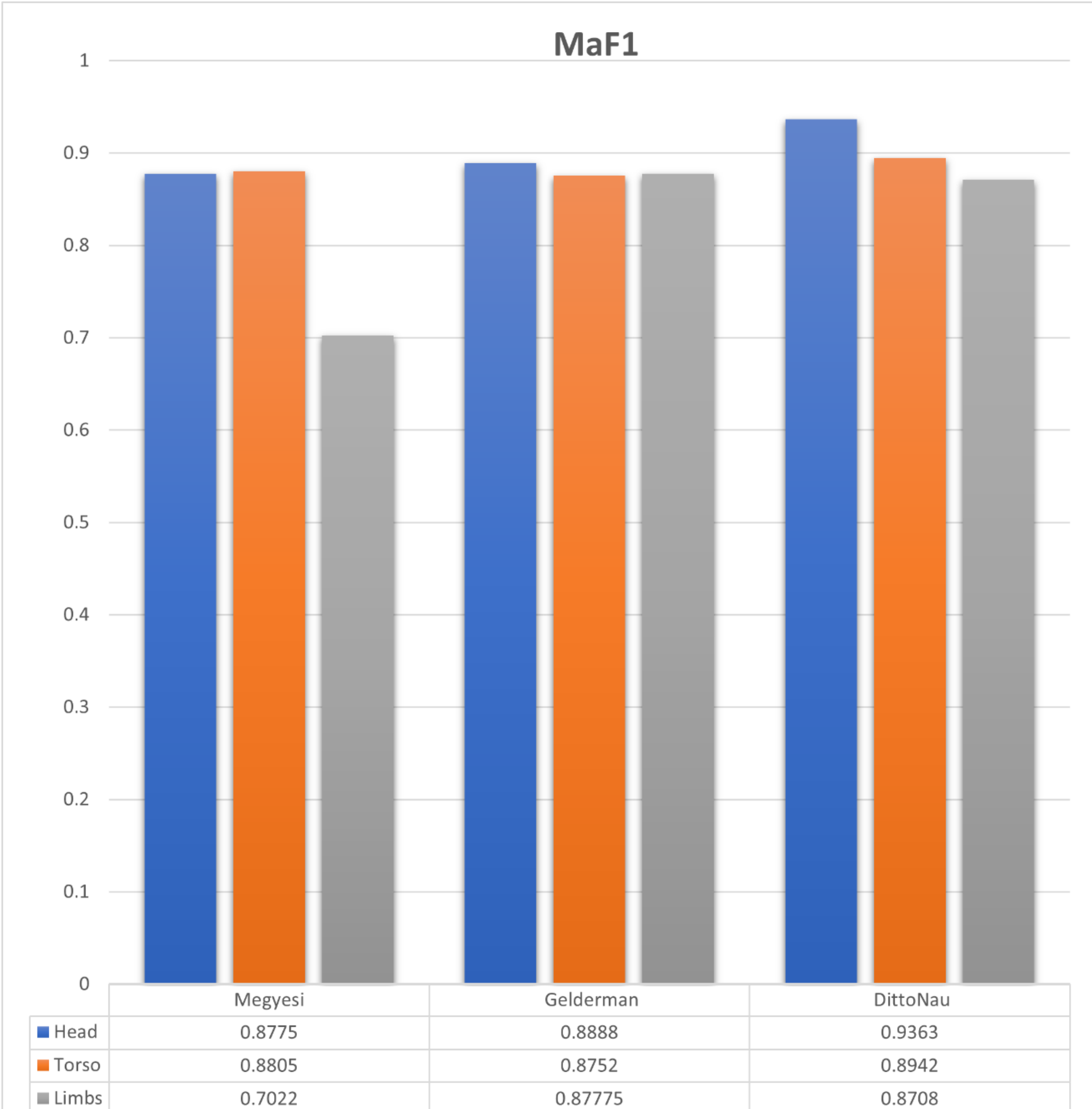
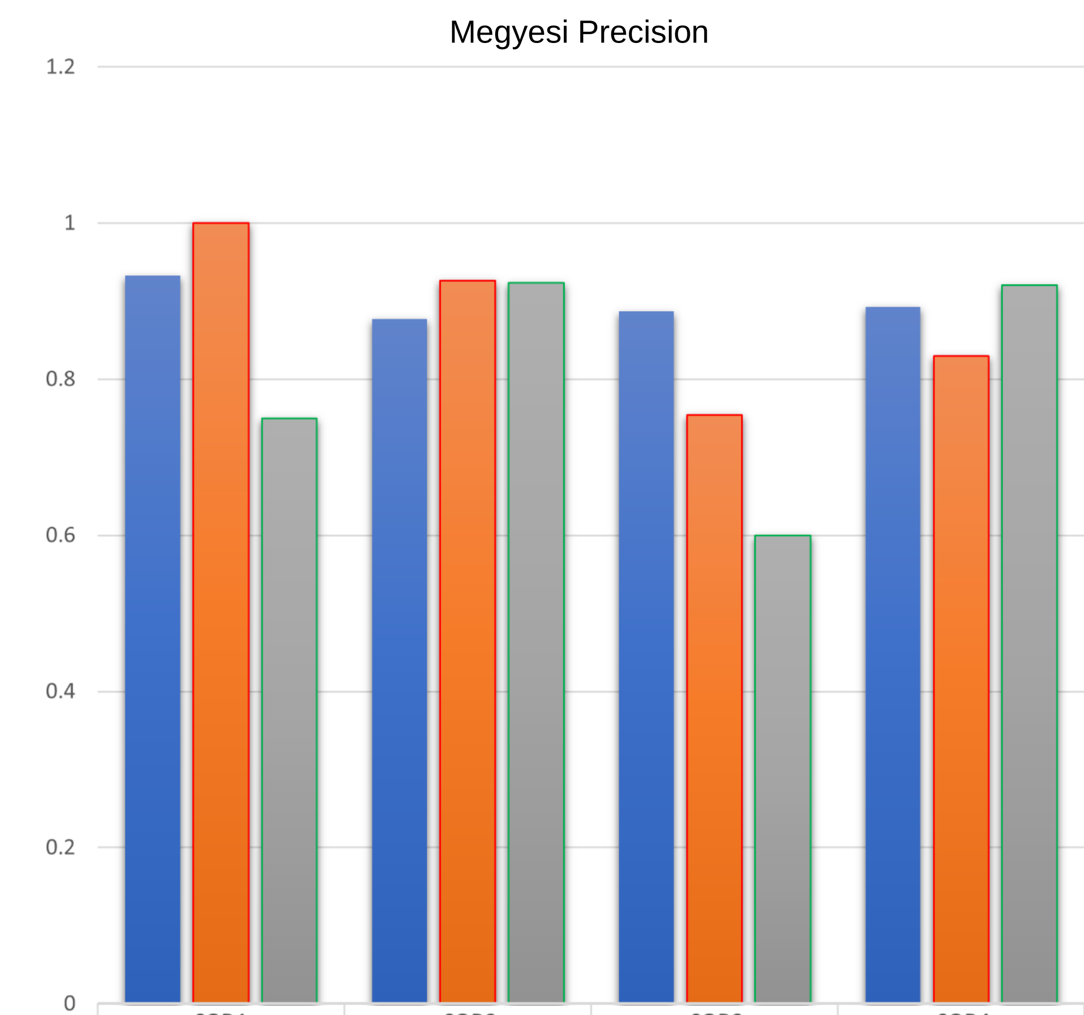
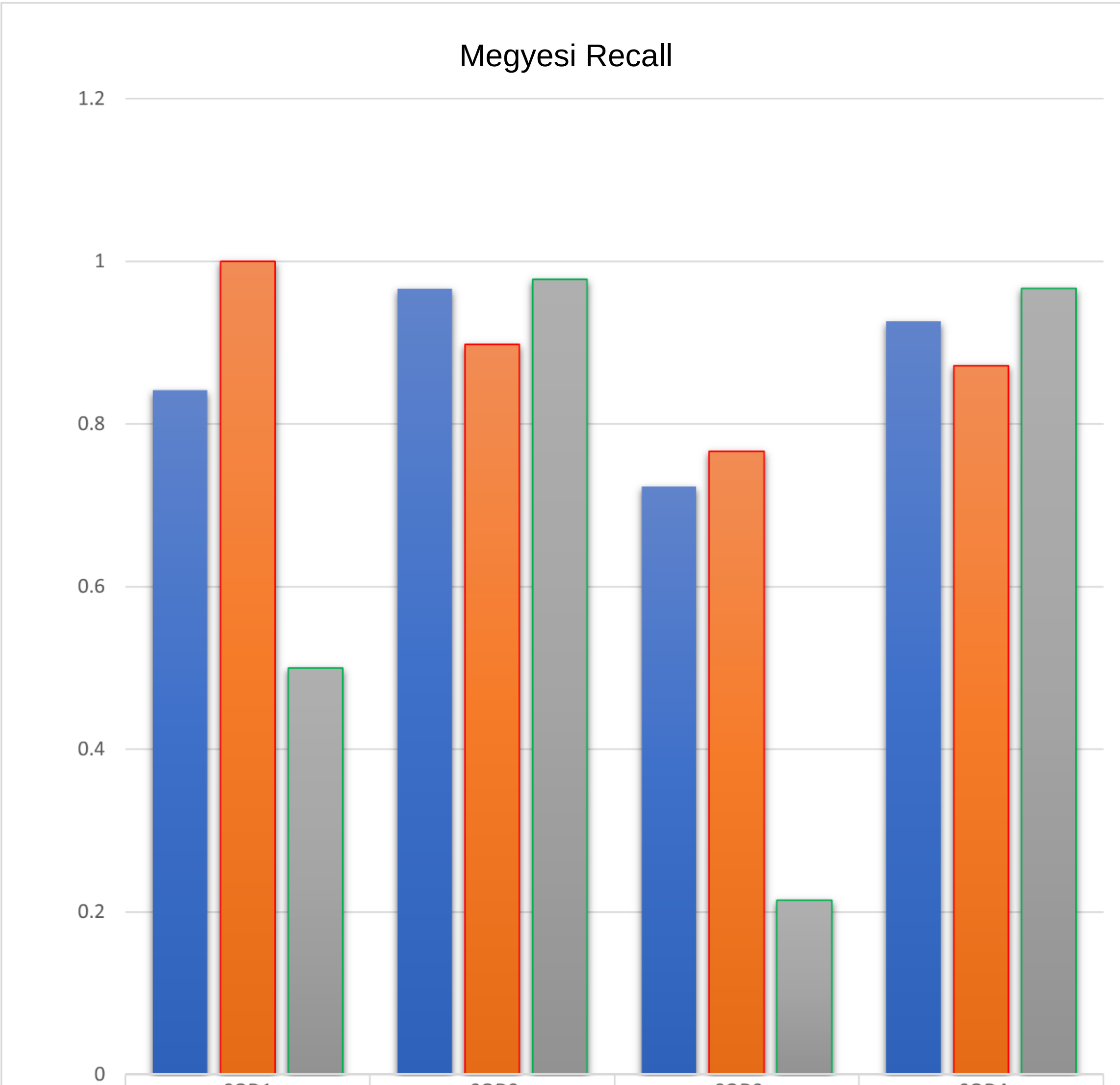


Figure 6. MaF1 scores for each model indicate model efficacy in determining the SOD by body region.



	SOD1	SOD2	SOD3	SOD4
Head	0.9324	0.8775	0.8867	0.8928
Torso	1.00	0.9263	0.754	0.8292
Limbs	0.75	0.9236	0.6	0.9206

Figure 5. Megyesi et al. (2005) Precision scores indicate the accuracy of AI models by SOD and body region. Fresh (SOD1), Early Decomposition (SOD2), Advanced Decomposition (SOD3), and Skeletonization (SOD4).



	SOD1	SOD2	SOD3	SOD4
Head	0.8414	0.9662	0.723	0.9259
Torso	1.00	0.8979	0.7666	0.8717
Limbs	0.5	0.9779	0.2142	0.9666

Figure 6. Megyesi et al. (2005) Recall scores indicate the sensitivity of AI models by SOD and body region.

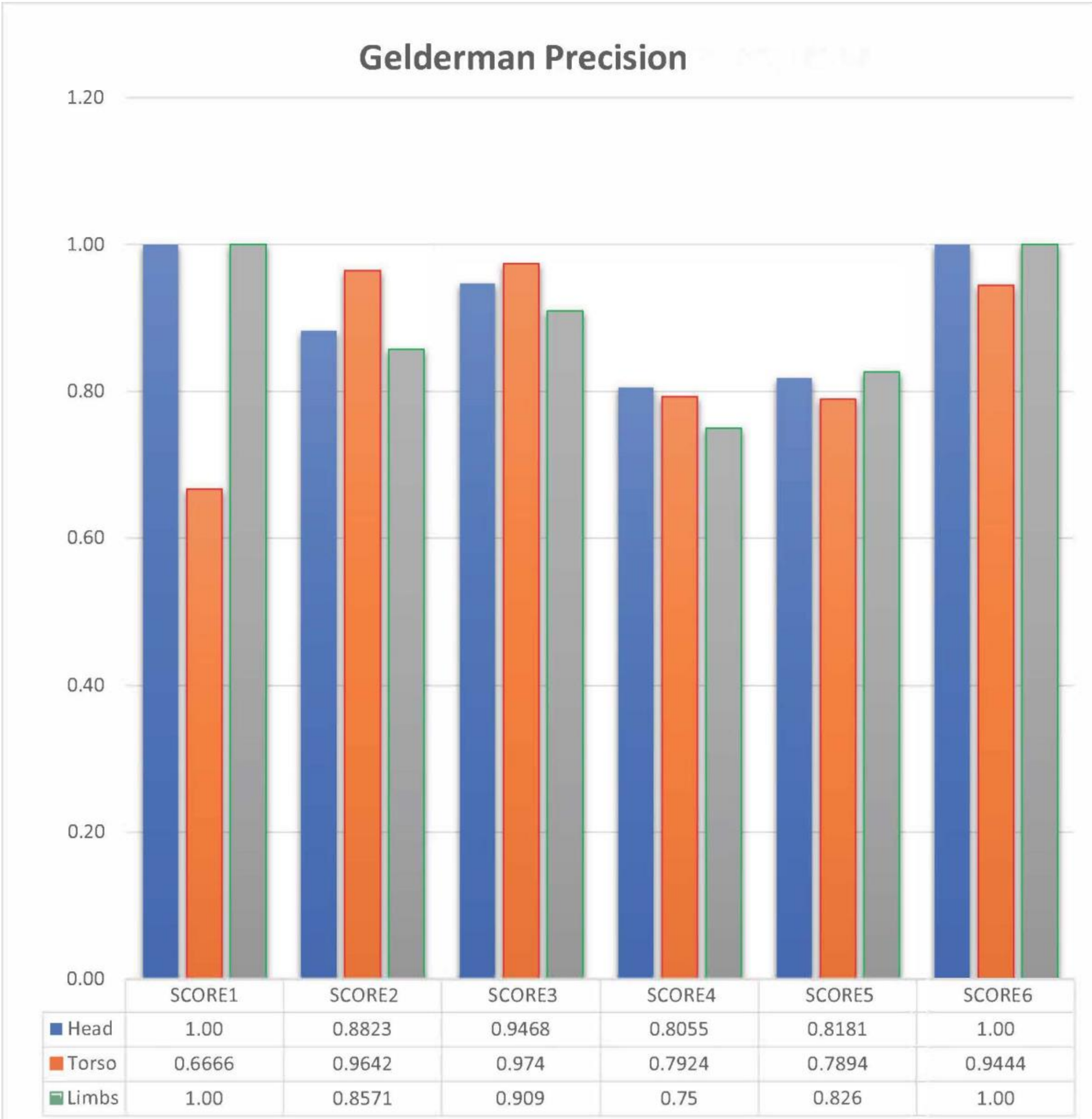
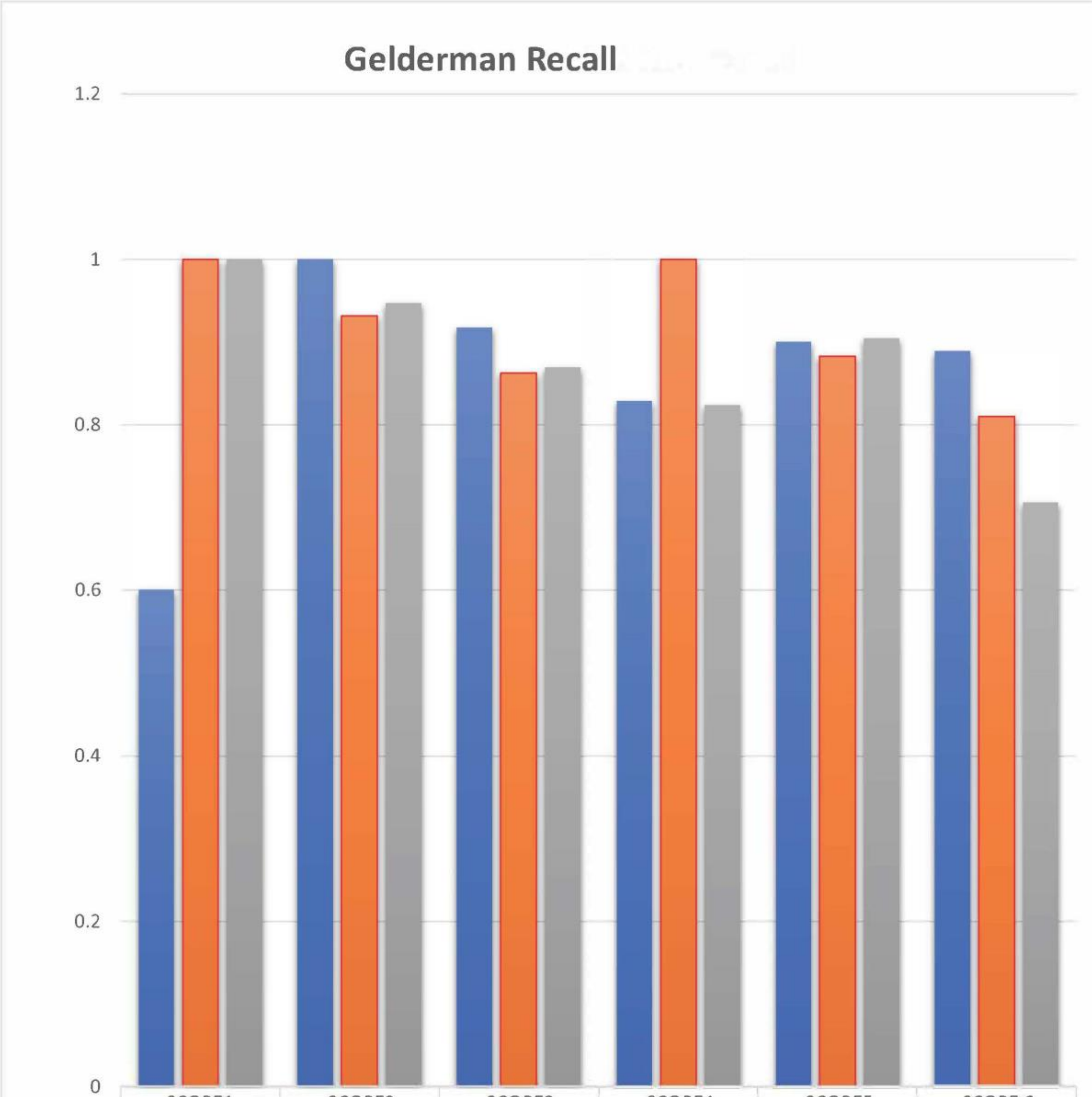
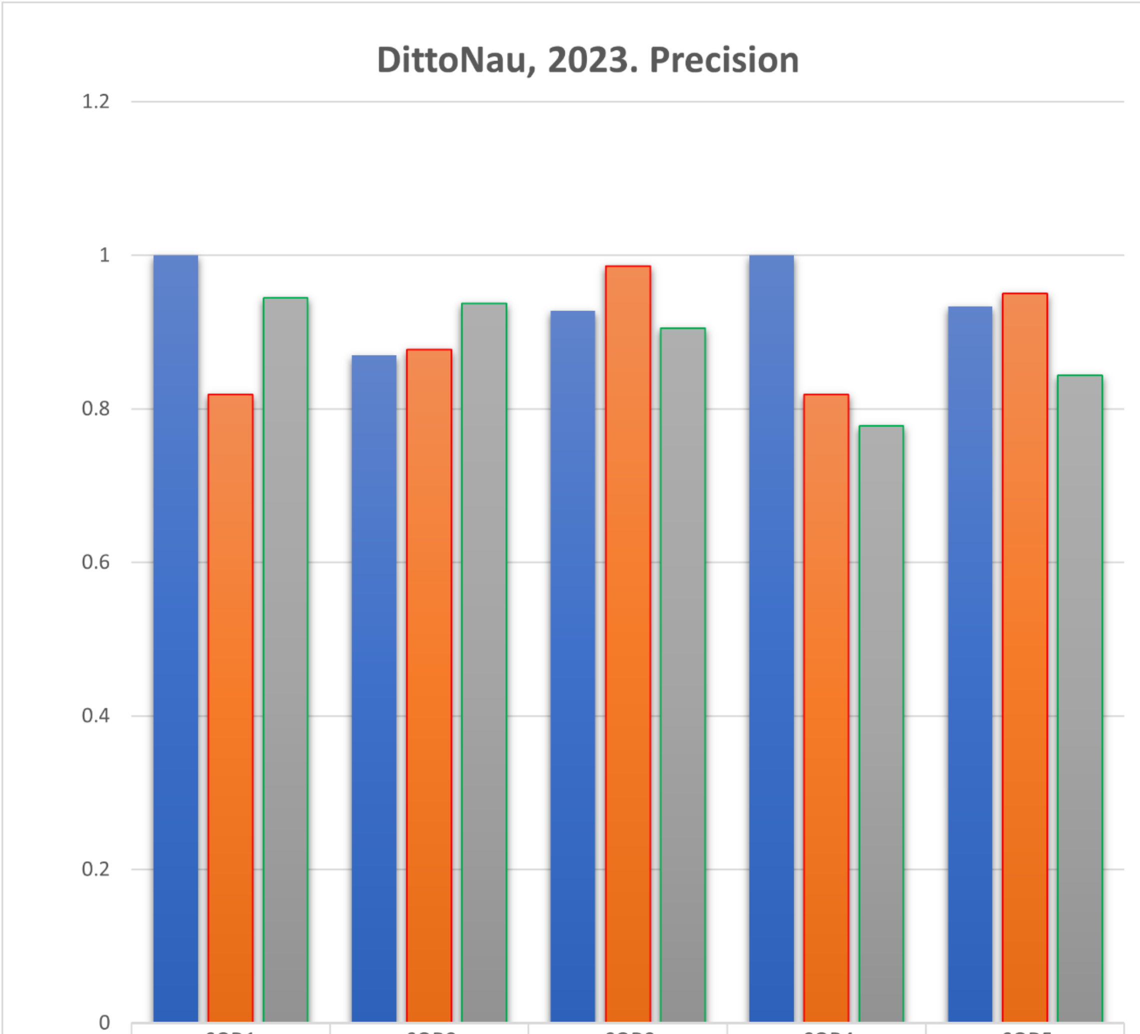


Figure 7. Gelderman et al. (2018) Precision scores indicate the accuracy of AI models by SCORE (similar to SOD) and body region.



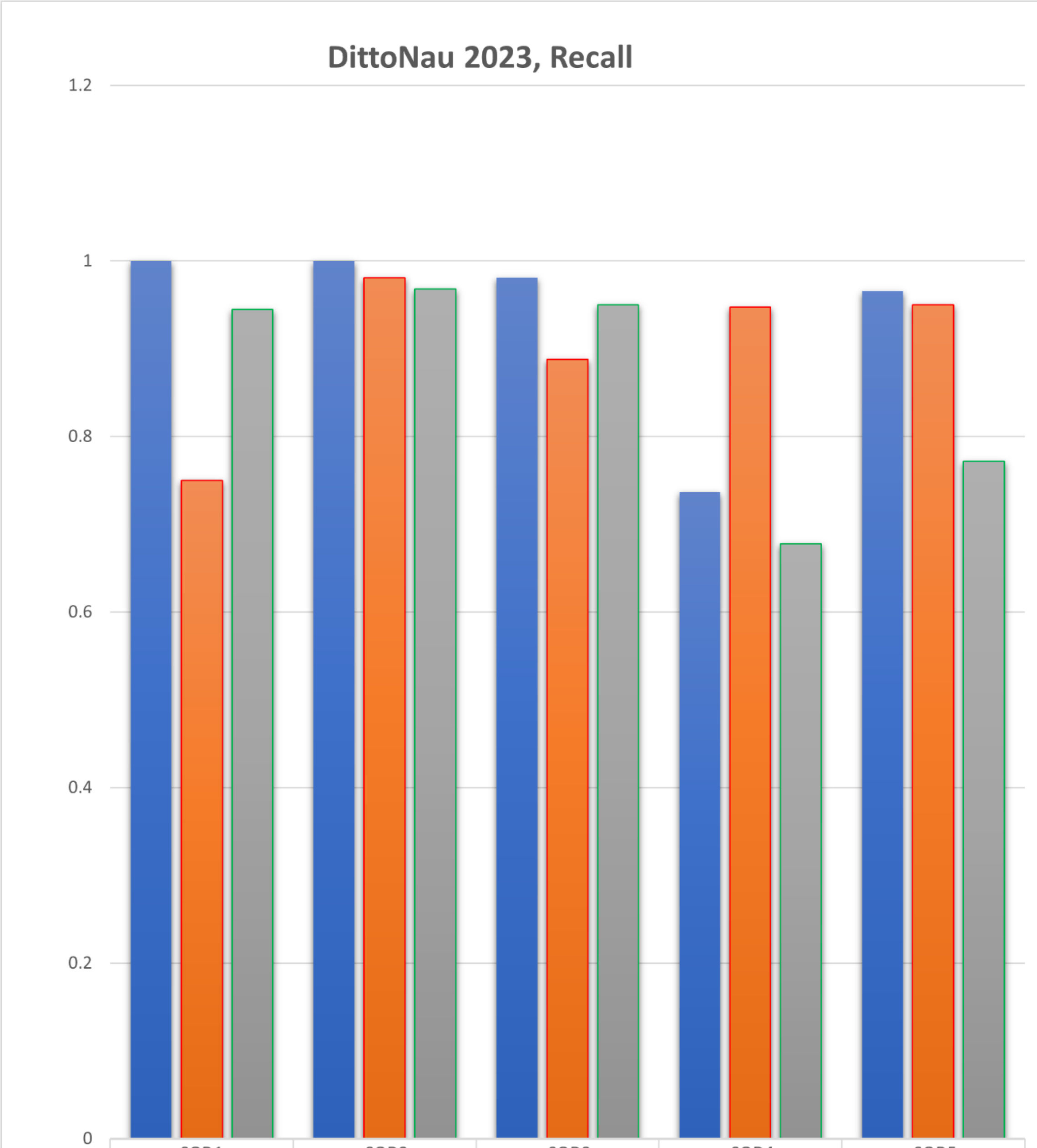
	SCORE1	SCORE2	SCORE3	SCORE4	SCORE5	SCORE 6
Head	0.6	1.00	0.9175	0.8285	0.9	0.8888
Torso	1.00	0.931	0.862	1.00	0.8823	0.8095
Limbs	1.00	0.9473	0.8695	0.8235	0.9047	0.7058

Figure 8. Gelderman et al. (2018) Recall scores indicate the sensitivity of AI models by SCORE (like SOD) and body region.



	SOD1	SOD2	SOD3	SOD4	SOD5
Head	1	0.8696	0.9279	1	0.9333
Torso	0.8182	0.8772	0.9861	0.8182	0.95
Limbs	0.9444	0.9375	0.9048	0.7778	0.8438

Figure 9. DittoNau Precision scores indicate the accuracy of AI models by SOD and body region. Fresh (SOD1), Initial Decomposition (SOD2), Active Decomposition (SOD3), Advanced or Halted Decomposition (SOD4), Skeletonization (SOD5).



	SOD1	SOD2	SOD3	SOD4	SOD5
Head	1	1	0.981	0.7368	0.9655
Torso	0.75	0.9804	0.8875	0.9474	0.95
Limbs	0.9444	0.9677	0.95	0.6774	0.7714

Figure 10. DittoNau Recall scores indicate the sensitivity of AI models by SOD and body region.

Inter-Observer Testing

Megyesi et al. 2005

```
##  Fleiss' Kappa for m Raters
##
##  Subjects = 300
##    Raters = 3
##     Kappa = 0.67
##
##         z = 30.9
##   p-value = 0
##
##    Kappa      z p.value
## 1  0.439 13.165   0.000
## 2  0.659 19.766   0.000
## 3  0.586 17.581   0.000
## 4  0.839 25.171   0.000
```

Gelderman et al. 2018

```
##  Fleiss' Kappa for m Raters
##
##  Subjects = 300
##    Raters = 3
##     Kappa = 0.593
##
##         z = 33.5
##   p-value = 0
##
##    Kappa      z p.value
## 1  0.465 13.937   0.000
## 2  0.716 21.471   0.000
## 3  0.654 19.620   0.000
## 4  0.602 18.061   0.000
## 5  0.417 12.523   0.000
## 6  0.462 13.852   0.000
```

DittoNau

```
##  Fleiss' Kappa for m Raters
##
##  Subjects = 300
##    Raters = 3
##     Kappa = 0.575
##
##         z = 30
##   p-value = 0
##
##    Kappa      z p.value
## 1  0.542 16.254   0.000
## 2  0.764 22.913   0.000
## 3  0.554 16.628   0.000
## 4  0.424 12.720   0.000
## 5  0.576 17.292   0.000
```

Figure 11. Inter-Observer Correlation (Fleiss' Kappa)

INTRA-OBSERVER
CORRELATION

MEGYESI

```
Single Score Intraclass Correlation

   Model: twoway
   Type : agreement

   Subjects = 300
     Raters = 2
   ICC(A,1) = 0.951

 F-Test, H0: r0 = 0 ; H1: r0 > 0
 F(299,299) = 39.7 , p = 1.11e-154

 95%-Confidence Interval for ICC Population Values:
  0.939 < ICC < 0.961
```

GELDERMAN

```
Single Score Intraclass Correlation

   Model: twoway
   Type : agreement

   Subjects = 300
     Raters = 2
   ICC(A,1) = 0.95

 F-Test, H0: r0 = 0 ; H1: r0 > 0
 F(299,300) = 39.1 , p = 4.28e-154

 95%-Confidence Interval for ICC Population Values:
  0.938 < ICC < 0.96
```

DITTONAU

```
Single Score Intraclass Correlation

   Model: twoway
   Type : agreement

   Subjects = 300
     Raters = 2
   ICC(A,1) = 0.908

 F-Test, H0: r0 = 0 ; H1: r0 > 0
 F(299,300) = 20.7 , p = 1.33e-115

 95%-Confidence Interval for ICC Population Values:
  0.886 < ICC < 0.926
```

Figure 12. Intra-Class Correlation (ICC)

CHAPTER FIVE

DISCUSSION

This chapter will discuss the results of this study and evaluate the performance of the three methods. The discussion will also focus on body regions, the performance of AI as it relates to the training of each method-specific model received, the statistical analysis, limitations, and future research.

Human decomposition is difficult to quantify and relies largely on subjective evaluation practices by human observers. Identifying specific morphological features is the central dogma to scoring decomposition, which can then be used to correlate with one of the four stages of decomposition. A decomposition score can also be used to calculate PMI. However, when the morphological features appear, the variables that may influence their appearance are not well studied.

Challenges remain regarding which morphological features to include in a decomposition method, especially in the later stages of human decomposition, such as when bone appears. Other challenges include when certain features appear and how different features should be weighted when calculating PMI. Many factors influence the rate and presentation of human decomposition, which compounds the evaluation process of PMI.

A consensus is needed in the forensic community to establish an accurate methodology that can be applied to train AI machine learning algorithms. In general, decomposition evaluation methods are not well outlined. For example, Megyesi et al. (2005) list 50% bone appearance as a criterion for advanced decomposition; however, how to calculate 50% is not clearly described. Basing AI training on unclear model criteria can lead to decreased interobserver reliability.

AI machine learning algorithms trained on visual morphological features from the photo population strongly assessed and accurately categorized stages of decomposition in all three methods with notable exceptions. Each of the three methods relied upon the photo population presented, particularly the photograph's quality and what was captured in the image, which may not account for all the morphological features present during the decomposition process when the image was taken. Only the morphological features present in the photograph can be evaluated by the observer.

The limbs body region consistently had the poorest performance among all three methods. The limbs body region was sub-optimal, with precision scores of less than 80% and recall scores of less than 50% when utilizing the Megyesi et al. (2005) method. This poor performance is likely based on the morphological features specified in Megyesi et al. (2005) regarding how 50% of bone appearance should be evaluated and scored. Limbs of decedents in outdoor environments are often subjected to external decomposition modifiers, such as entomological and scavenging activity, which may hasten or stall the decomposition process (Jeong, Jantz, and Smith, 2016; Simmons et al., 2010; Steadman et al., 2018).

These results may reflect AI's inability to accurately stage limb decomposition using the Megyesi et al. (2005) method. One potential cause for the poor performance is that the AI-method-specific requires more training data, especially for SOD3, where bone appearance scoring is more subjective. Another cause could be that the Megyesi et al. (2005) method is a poor candidate for AI deep learning as the limbs criteria remain too subjective without specific parameters that AI can understand, such as anatomical markers to measure the percentage of bone appearance. It is likely the latter, as AI performance in limbs was significantly better in the other two methods that simplified morphology in the later stages of decomposition.

Exploring Megyesi et al. (2005) further, multiple rationales are possible when evaluating the sub-optimal scoring in the method. In SOD1, variation in precision and recall in different body regions is likely caused by poor sampling distribution, as populations with fresh stage morphological features are significantly fewer based on the donors' time of death and placement in the facility and when the photograph was captured. The body region variation findings in Megyesi et al. (2005) are consistent in both Gelderman et al. (2018) and DittoNau in the first stage of decomposition. However, DittoNau had strong precision but a recall of 75% in the torso region. These inconsistencies in precision and recall in different body regions within the fresh or first stage of decay are consistent with poor sampling distribution for this stage. Another possibility is that DittoNau outperformed the other two methods in the torso region, as a recall of 75% is still significant and could merely represent slightly unbalanced data in machine learning (Bajaj, 2019).

Sub-optimal performances noted throughout a region indicate potential issues regarding the methodology, the applicability of the method to train AI, or a training sample size that is too small. Poor performance is observed in all body regions in SOD3 in Megyesi et al. (2005), indicating difficulty in determining the correct stage of decay in the Megyesi et al. (2005) method when the morphological feature of bone appearance and the 50% quantifier were applied. This analysis is strengthened when sensitivity is applied in this stage, with notably poor sensitivity. These performance patterns are not present in the other two methods, suggesting that the morphological criteria outlined in the later decomposition stages of Megyesi et al. (2005) do not inform AI well.

Statistical Analysis

Gelderman et al. (2018) and DittoNau scored consistently well with few variations and fewer sub-optimal performances. Additionally, DittoNau had the highest performing MaF1 scores, indicating strong performance, and was only marginally higher than Gelderman et al. (2005). Megyesi et al. (2005) had comparable performance metrics with notable limb region deviation. This deviation may be attributed to the Megyesi et al. (2005) scoring technique regarding bone appearance, with a 50% modifier in which a limb may be classified as advanced decomposition or skeletonization. Megyesi et al. (2005) do not specify how to assess bone appearance during examination.

The strong performance of the DittoNau method is most likely attributed to repetitive object recognition of thousands of randomized photos of human decomposition, simplified for scoring by eliminating specific conditions, such as the presence of bone. It focuses on whether decomposition activity is present (Contini, Wardle, and Carlson, 2017). These two factors produced a model that was more optimized for AI.

The two published methods differed notably, as Megyesi et al. (2005) primarily consisted of outdoor cases, and Gelderman et al. (2018) consisted mostly of indoor cases. Performance analysis from AI showed a compelling performance of Gelderman et al. (2018) for outdoor cases. This finding is significant due to the notable increase in decomposition variables and morphological variation in outdoor scenarios (Janaway, Percival, and Wilson, 2009; Mann, Bass, and Meadows, 1990; Marks, Love, and Dadour, 2016).

Inter and Intra-Observer Testing

Fleiss' Kappa analysis of the inter-rater reliability examination showed that overall, there was moderate agreement among all three methods when comparing the three human raters. Notably, Megyesi et al. (2005) is well known to each of the raters and is the foundation for the other two methods. DittoNau is a novel method and was least known among the raters and was only marginally lower than Gelderman, likely for the same reason.

Intra-class reliability was performed on the rater training the AI in all three methods, which showed exceptional reliability. The rater reliability indicates strong agreement with the first and second image labeling performance by the rater.

Limitations, Future Research, and Disclosures

Future research should evaluate morphological features in decomposition in multiple geographic regions with varied climates, identify the most important features, and establish a consensus on a method incorporating variables that may influence morphological presentation, such as entomological activity and scavenging.

AI machine learning would benefit from an increased variety of samples (e.g., indoor, outdoor, and other geographic locations), testing multiple available or new methods, and an increased and varied trainer cohort to decrease potential training bias. Training bias is likely present based on repetitive image viewing and model construction within ICPUTRD. While ICPUTRD is efficient and large populations can be assessed quickly, the trade-off is that training data can be exposed to bias from photo sample exposure and influencing subsequent photos' labeling.

Bias is also likely present in the inter-observer testing due to the photo array experimental design potentially influencing behavioral aspects of human raters and their labeling of human decomposition. For example, when a rater observes more than one photo of decomposition at a time, the exposure of these other morphological features may influence the observation and assessment of the intended photo. For example, when utilizing a method, the rater observes a limb with bone exposures and sees a similar photo in the array with bone exposure. These observations can influence the application of the method, and the rater may choose a category for each photo based on exposure to the others.

Future studies can address these potential issues by having multiple AI trainers processing one photo at a time, utilizing a method or multiple methods that can account for potential bias while still providing a large enough sample population to train AI adequately.

Photographs are a useful and affordable tool to document human decomposition. However, evidence suggests that two-dimensional images may influence the performance of models, such as Megyesi et al. (2005), in those evaluators, when assessing remains in person, performed better than when assessing decomposition through photographs (Ribéreau-Gayon et al., 2018). Additionally, photographic examination is limited to what is captured in the image, which may influence the determination of all the morphological features present. Further research on the applicability and possible effect of method performance from photographs is needed. An experimental design that can include an increased and varied population, trained by multiple experts on the agreed-upon method(s), and tested for reliability as well as replicability will further enhance our understanding of AI's ability to evaluate human decomposition and better understand the bias that occurs from raters and influence from photos.

Funding for this research was made possible by a Department of Justice, National Institute for Justice grant (15PNIJ-21-GG-04161-SLFO to Mockus and

Steadman). The opinions, findings, and conclusions or recommendations expressed in this thesis are those of the author and do not necessarily reflect those of the US Department of Justice. There are no known conflicts of interest.

CHAPTER SIX

CONCLUSION

This chapter will summarize the thesis research study and present the model that performed best to inform AI deep learning algorithms based on visual morphological features of human decomposition from photos.

Over 20,000 individual images of human decomposition were assessed for decomposition characteristics, labeled according to three decomposition categorical methods - two published and one novel method - Megyesi et al. (2005), Gelderman et al. (2018), and DittoNau, respectively. Once labeled for their evident state of decomposition in ICPUTRD, these images were then utilized by the computer science and data team to train AI-method-specific machine learning algorithms. Trained AI models made predictions on the stage of decay based on the AI-trained method utilized and produced statistical performance analyses.

Regarding the research questions of this thesis study, while all three methods performed well and showed that AI could be trained to categorize human decomposition, the DittoNau method had the best overall performance, precision, sensitivity, and best-informed AI, though only marginally better than Gelderman et al. (2018). The three methods each approach categorizing human decomposition differently, with some overlapping, however, in the morphological features they possess. For example, Megyesi et al. (2005) and DittoNau both categorize morphological features into stages, while Gelderman et al. (2018) do not. DittoNau focuses on morphological features and activity when categorizing decomposition and this may have led to slightly better results. Certain morphological features and how they are categorized can lead to poor performance and potential confusion, such as Megyesi et al. (2005) 50% bone appearance feature. Additionally, how limbs undergo decomposition presents challenges to current methods in how and when morphological features appear, as they are susceptible to scavenging and increased entomological activity (Jeong, Jantz, and Smith, 2016; Simmons et al. 2010; Steadman et al., 2018). These factors should be considered when designing a categorical model of human decomposition.

This study also indicates that AI is capable of accurately and repetitively categorizing human decomposition with adequate training and programming, potentially opening opportunities for examinations by agencies lacking forensic experts and those with disabilities. However, a more varied sample of photos and expert trainers are needed to increase AI's reliability and training.

LIST OF REFERENCES

Bajaj, A. (2019). "What does your classification metric tell about your data? Study about evaluation strategies of classification metrics and how to evaluate the result." *Towards Data Science*. https://towardsdatascience.com/what-does-your-classification-metric-tell-about-your-data.

Bhushetty, J. (2023). "A Complete Guide to Testing AI and Machine Learning Applications." *QED42*. Retrieved September from https://www.qed42.com/insights/perspectives/biztech/complete-guide-testing-ai-and-ml-applications

Bugelli, V., et al. (2018). "Decomposition pattern and insect colonization in two cases of suicide by hanging." *Forensic Sci Res*, 3(1), 94-102. https://doi.org/10.1080/20961790.2017.1418622

Cheng-Tek Tai, M. 2020. "The impact of artificial intelligence on human society and bioethics." *Tzu Chi Medical Journal*. 32(4): 339-343.

Cockle, D. L., Bell, L. S. (2015). "Human decomposition and the reliability of a 'Universal' model for post mortem interval estimations." *Forensic Sci Int*, 253. https://doi.org/10.1016/j.forsciint.2015.05.018

Contini, E. W., Wardle, S. G., Carlson, T. A. (2017). "Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions." *Neuropsychologia*, 105, 165-176. https://doi.org/10.1016/j.neuropsychologia.2017.02.013

Cross, P., Simmons, T. (2010). The influence of penetrative trauma on the rate of decomposition. *J Forensic Sci*, *55*(2), 295-301. https://doi.org/10.1111/j.1556-4029.2009.01277.x

Dautartas, A. M. (2009). "The effect of various coverings on the rate of human decomposition." *University of Tennessee*. http://trace.tennessee.edu/utk_gradthes/69

Dautartas, A. M. (2018). "A Comparison of Human and Domestic Pig Decomposition Using Multivariate Methods." *University of Tennessee*.

Fakorede, S. T., et al. (2022). "Prevalence and population genetic data of colour vision deficiency among students from selected tertiary institutions in Lagos State, Nigeria." *Egyptian Journal of Medical Human Genetics*, 23(1), 73. https://doi.org/10.1186/s43042-022-00287-9

Fleiss, J. L., Levin, B., Paik, M. C. (2003). "The Measurement of Interrater Agreement." *Statistical Methods for Rates and Proportions*, 598-626. https://doi.org/https://doi.org/10.1002/0471445428.ch18

Franceschetti, L., et al. (2021). "Comparison of Accumulated Degree-Days and Entomological Approaches in Post Mortem Interval Estimation." *Insects*, 12(3). https://doi.org/10.3390/insects12030264

Galloway, A., et al. (1989). "Decay rates of human remains in an arid environment." *J Forensic Sci*, 34(3), 607-616. https://www.ncbi.nlm.nih.gov/pubmed/2738563

Gelderman, H. T., et al. (2018). "The development of a post-mortem interval estimation for human remains found on land in the Netherlands." *Int J Legal Med*, 132(3), 863-873. https://doi.org/10.1007/s00414-017-1700-9
Gelderman, H. T., et al. (2019). "Estimation of the post-mortem interval based on the human decomposition process." *J Forensic Leg Med*, 61, 122-127. https://doi.org/10.1016/j.jflm.2018.12.004
Gouse, S., et al. (2018). "Forensic photography: Prospect through the lens". *J Forensic Dent Sci*, 10(1), 2-4. https://doi.org/10.4103/jfo.jfds_2_16
Janaway, R. C., Percival, S. L., Wilson, A. S. (2009). "Decomposition of Human Remains." S. L. Percival (Ed.), *Microbiology and Aging: Clinical Manifestations*, 313-334. Humana Press. https://doi.org/10.1007/978-1-59745-327-1_14
Javan, G. T., et al. (2019). "An interdisciplinary review of the thanatomicrobiome in human decomposition." *Forensic Sci Med Pathol*, 15(1), 75-83. https://doi.org/10.1007/s12024-018-0061-0
Jeong, Y., Jantz, L. M., Smith, J. (2016). "Investigation into Seasonal Scavenging Patterns of Raccoons on Human Decomposition." *J Forensic Sci*, 61(2), 467-471. https://doi.org/10.1111/1556-4029.12992
Keough, N., Myburgh, J., Steyn, M. (2017). "Scoring of Decomposition: A Proposed Amendment to the Method When Using a Pig Model for Human Studies." *J Forensic Sci*, 62(4), 986-993. https://doi.org/10.1111/1556-4029.13390
Koo, T. K., Li, M. Y. (2016). "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *J Chiropr Med*, 15(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012
Kundu, R. (2022). "F1 Score in Machine Learning: Intro and Calculation." *Machine Learning*. https://www.v7labs.com/blog/f1-score-guide
Lennartz, A., Hamilton, M., Weaver, R. (2020). "Moisture Content in Decomposing, Desiccated, and Mummified Human Tissue." *Forensic Anthropology*, 3, 1-16. https://doi.org/10.5744/fa.2020.1001
Mann, R. W., Bass, W. M., Meadows, L. (1990). "Time since death and decomposition of the human body: variables and observations in case and experimental field studies." *Journal of Forensic Sciences*, 35(1), 103-111.
Marks, M. K., Love, J. C., Dadour, I. R. (2016). "Taphonomy and Time." D. W. Steadman (Ed.), *Hard Evidence: Case Studies in Forensic Anthropology*,165-178. Routledge.
Megyesi, M. S., et al. (2005). "Using accumulated degree-days to estimate the post-mortem interval from decomposed human remains." *J Forensic Sci*, 50(3), 618-626.
Mousavi S., et al. 2019. "Machine-assisted annotation of forensic imagery." *IEEE International Conference on Image Processing (ICIP 2019)*, pp. 1595-1599. DOI: 10.1109/ICIP.2019.8803068 (2019)
Nau, A. M., Mockus, A., Steadman, DW. (2023). "Stage of Decay Estimation Exploiting Exogenous and Endogenous Image Attributes to Minimize

Manual Labeling Efforts and Maximize Classification Performance." *2023 IEEE International Conference on Image Processing (ICIP)*.
Pittner, S., et al. (2020). "The applicability of forensic time since death estimation methods for buried bodies in advanced decomposition stages." *PLoS One*, 15(12), e0243395. https://doi.org/10.1371/journal.pone.0243395
Sauerwein, K. A. (2018). "Perceptions and Cognitive Bias in Decomposition Scoring Methods in Forensic Anthropology." *University of Tennessee*.
Shivaprasad, P. (2020). "Understanding Confusion Matrix, Precision-Recall, and F1-Score: Why accuracy shouldn't be the only performance metric you care about while evaluating a Machine Learning model." *Towards Data Science*. Retrieved September from https://towardsdatascience.com/understanding-confusion-matrix-precision-recall-and-f1-score-8061c9270011
Shrestha, R., Kanchan, T., Kirshan, K. (2022). "Methods of Estimation of Time Since Death." *National Institute of Health*. Retrieved March from https://www.ncbi.nlm.nih.gov/books/NBK549867/
Simmons, T., et al. (2010). "The influence of insects on decomposition rate in buried and surface remains." *J Forensic Sci*, 55(4), 889-892. https://doi.org/10.1111/j.1556-4029.2010.01402.x
Steadman, DW., et al. (2018). "Differential Scavenging Among Pig, Rabbit, and Human Subjects." *J Forensic Sci*, 63(6), 1684-1691. https://doi.org/10.1111/1556-4029.13786
Tai, M. C. (2020). The impact of artificial intelligence on human society and bioethics. *Tzu Chi Med J*, *32*(4), 339-343. https://doi.org/10.4103/tcmj.tcmj_71_20
Vass, A. A. (2001). "Beyond the grave-understanding human decomposition." *Microbiology Today*, 28, 190-193.
Vidoli, G. M., Kenyhercz, M., Steadman, DW. (2020). "Expanding on Total Body Score with Use of Geographic Information Systems (GIS)." *US Department of Justice*. 1-14.
Wells, J. D. (2019). "A Forensic Entomological Analysis Can Yield an Estimate of Post-mortem Interval, and Not Just a Minimum Post-mortem Interval: An Explanation and Illustration Using a Case." *Journal of Forensic Sciences*, 64(2), 634-637. https://doi.org/https://doi.org/10.1111/1556-4029.13912
Wescott, D. J. (2018). "Recent advances in forensic anthropology: decomposition research." *Forensic Sci Res*, 3(4), 327-342. https://doi.org/10.1080/20961790.2018.1488571
Zapico, S. C., Adserias-Garriga, J. (2022). "Post-mortem Interval Estimation: New Approaches by the Analysis of Human Tissues and Microbial Communitiesrsquo; Changes." *Forensic Sciences*, 2(1), 163-174. https://www.mdpi.com/2673-6756/2/1/13

VITA

Phillip D. Ditto was born in Nuremberg, Germany, in January 1984. Phil has traveled and lived all over the world. Before attending undergraduate school, he served in the United States Army as an intensive care nurse and flight medic, serving worldwide. Phil has been a first responder and paramedic for over 20 years. He attended Texas Christian University and graduated with a Bachelor of Arts in Anthropology and a minor in Criminal Justice in May 2020. Phil is currently pursuing a Master of Public Health degree at Johns Hopkins.