

To the Graduate Council:

I am submitting herewith a thesis written by Jianxin Xie entitled “Physics-Augmented Modeling and Optimization of Complex Systems: Healthcare Applications.” I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial Engineering.

Bing Yao, Major Professor

We have read this thesis
and recommend its acceptance:

Anahita Khojandi

Zhongshun Shi

Xiaopeng Zhao

Accepted for the Council:

Dixie Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Physics-Augmented Modeling and Optimization of Complex Systems: Healthcare Applications

A Dissertation Presented for the
Doctor of Philosophy
Degree

The University of Tennessee, Knoxville

Jianxin Xie

August 2023

© by Jianxin Xie, 2023
All Rights Reserved.

Acknowledgements

I would like to express my deepest appreciation to my advisor, Dr. Bing Yao, whose patience, resourcefulness, and guidance have been a beacon of light throughout my research journey. Her unwavering belief in me and acknowledgment of my efforts have been instrumental to my academic growth and self-worth.

My profound thanks go to Dr. Farzad Yousefian for his persistent support during my time at both Oklahoma State University and University of Tennessee at Knoxville. His endorsements, conveyed through the recommendation letters, have played a pivotal role in my achievements. The knowledge and inspiration I gained from his optimization courses have profoundly influenced my research and will continue to guide me in my future work. His consistent guidance and mentorship have been pillars in my academic journey.

I am immensely grateful to Dr. Zheyu Jiang, our esteemed research collaborator. His invaluable research suggestions have enriched my work and broadened my perspectives. His strong recommendation letters have been a crucial support in my academic pursuits, reflecting his trust and confidence in my abilities. His mentorship has been truly inspiring.

My sincere thanks to Dr. Anahita Khojandi, Dr. Zhongshun Shi, and Dr. Xiaopeng Zhao for their willingness to serve as my committee members. Their insights and feedback have enriched my dissertation and contributed significantly to my academic progress.

Lastly, but most importantly, I wish to acknowledge my mother, whose belief in me has been the fuel to my thirst for academic success. Her pride in my accomplishments has been my greatest motivation. I am forever grateful for her unconditional love and support, which have been my constant source of strength and motivation through all the challenges and triumphs of my academic journey.

To each one of you, thank you for your unwavering faith, continuous support, and invaluable contributions to my academic journey. Your confidence and guidance have made this achievement possible, and I am forever indebted to you all.

Abstract

The rapid advances in sensing technology have created a data-rich environment that tremendously benefits predictive modeling and decision-making for complex systems. Harnessing the full potential of this complexly-structured sensing data requires the development of novel and reliable analytical models and tools for system informatics. Such advancements in sensing present unprecedented opportunities to investigate system dynamics and optimize decision-making processes for smart health. Nevertheless, sensing data is typically characterized by high dimensionality and intricate structures. To fully unlock the potential of this data, we significantly rely on innovative analytical methods and tools that can effectively process information.

The objective of this dissertation is to develop innovative physics-augmented methodologies for modeling, monitoring, and optimizing complex systems. Specifically, the research focuses on the development of physics-augmented machine learning models that facilitate optimal decision-making in complex healthcare systems. This research will enable 1) robust predictive modeling of spatiotemporal systems; 2) extracting important feature information about system dynamics; 3) optimizing decision-making under uncertainty and sparse sensor observations. My research accomplishments include:

- *Physics-constrained deep learning for robust inverse modeling of spatiotemporal systems:* In Chapter 2, we developed a physics-constrained machine learning framework to address the high-dimensional inverse problem. This method integrates physics-based principles with the advanced deep learning infrastructure to predict the spatiotemporal system dynamics based on indirect and noisy sensor observations. This methodology is implemented in inverse electrocardiography (ECG) modeling.

- *Physics-constrained active learning and sequential design in 3D spatiotemporal systems:* In Chapter 3, a novel physics-constrained strategy for optimal sensor placement is proposed to actively explore and model the dynamics of 3D complex-structured systems. This active learning scheme combines uncertainty estimation in deep learning and space-filling design over the complex geometry, enabling effective learning of system dynamics from limited sensor exploration.
- *Defects identification in 3D heterogeneous complex systems:* In Chapter 4, we designed a new framework to characterize the location and extent of diseased cardiac tissue from indirect and noisy body sensor observations. Specifically, a hierarchical Gaussian Process-based active learning framework was developed to reliably identify the defect location and further extract spatial-varying properties in a complex 3D system.

Table of Contents

1	Introduction	1
1.1	Motivation	1
1.2	Research Background	2
1.3	Research Objectives	4
1.4	Organization of the Dissertation	5
2	Physics-constrained Deep Learning for Robust Inverse ECG Modeling	7
2.1	Introduction	9
2.2	Research Background	10
2.2.1	Inverse ECG Modeling and Regularization	10
2.2.2	Deep Neural Networks	13
2.3	Research Methodology	14
2.3.1	Cardiac Electrophysiological Model	14
2.3.2	Artificial Neural Networks	15
2.3.3	P-DL approach	16
2.3.4	Bayesian Optimization for Regularization Parameter	18
2.3.5	Parallel Computing to Increase Computation Efficiency	21
2.4	Experimental Design and Results	22
2.4.1	Impact of Neural Network Structure on the Prediction Performance	24
2.4.2	Impact of Collocation Points on the Prediction Performance	25
2.4.3	Comparison Study with Traditional Regularization Methods	25
2.4.4	Comparison Study with the Physics-constrained Statistical Method	27
2.4.5	Impact of Electrophysiological Model Parameters on the Prediction Performance	29

2.5	Discussion	30
2.5.1	Advantages of the Proposed P-DL Method in Inverse ECG Modeling	30
2.5.2	Training the P-DL Model	31
2.5.3	Define New Network Structures with Physics-based Models	32
2.6	Conclusions	33
3	Physics-constrained Deep Active Learning for Spatiotemporal modeling of Cardiac Electrodynamics	34
3.1	Introduction	36
3.2	Research Methodology	41
3.2.1	<i>Spatiotemporal Gaussian Process (GP) Modeling</i>	41
3.2.2	<i>Physics-constrained Deep Active Learning (P-DAL)</i>	43
3.3	Model Validation	49
3.3.1	<i>Healthy heart system</i>	50
3.3.2	<i>Diseased heart system</i>	53
3.3.3	Training time	55
3.4	Discussion	55
3.4.1	The main differences in methodological contributions between the proposed P-DAL framework and the P-DL developed in Chapter 2	55
3.4.2	Advantages of the P-DAL approach in modeling the spatiotemporal cardiac systems	56
3.5	Conclusions	58
4	Hierarchical Active Learning for Defect Localization in 3D Systems	60
4.1	Introduction	61
4.2	Related Works	64
4.3	Problem Formulation	66
4.3.1	Problem Setting	66
4.3.2	Gaussian Process Meta-modeling for Model Calibration	67
4.3.3	Gaussian Process Optimization	68
4.4	Research Methodology	69

4.4.1	Gaussian Process Regression on the 2D Embedding of a 3D Surface	70
4.4.2	Hierarchical GP-based Active Learning	72
4.5	Experimental Design and Results	77
4.5.1	Experimental Design	78
4.5.2	Experimental Results	79
4.6	Conclusions	81
5	Conclusions	83
	Bibliography	87
	Appendices	109
	Appendix A Supplement Materials of Chapter 2	109
A.1	Automatic Differentiation	109
A.2	The optimal regularization parameter w_{opt}	109
A.3	Physiological-Model-constrained Kalman filter (P-KF)	110
A.4	Statistical test for the comparison of RE generated from P-DL and P-KF	112
	Appendix B Supplement Materials of Chapter 3	113
B.1	Hyperparameter Section	113
B.1.1	Hyperparameter selection for neural network structures and regularization parameter w	113
B.1.2	The influence of parameter λ on the active learning performance	115
B.2	Statistical test to compare the results from P-DAL with P-DSL and P-DRL	116
	Appendix C Supplement Materials of Chapter 4	118
C.1	Refinement Procedure in HGPAL	118
C.2	Benchmark methods	119
C.2.1	Multi-resolution method	119
C.2.2	Multi-task Gaussian Process Upper Confidence Bound (MT-GPUCB)	120
	Appendix D Figures	124
	Appendix E Tables	140

List of Tables

E.1	Structure details of different neural networks	140
E.2	The <i>RE</i> , <i>CC</i> , and <i>MSE</i> from the P-KF algorithm with different initial conditions. Here, $\hat{\mathbf{u}}_0(\mathbf{s})$ and $\mathbf{u}_0(\mathbf{s})$ denote the specified and true initial condition, respectively; the initialization noise follows a normal distribution, i.e., $\boldsymbol{\epsilon}_{\mathbf{u}_0}(\mathbf{s}) \sim N(0, 0.0025)$	140
E.3	The computational time for different methods.	141
E.4	The influence of variations in AP model parameters ($\pm 10\%$) on the <i>RE</i> , <i>CC</i> , and <i>MSE</i> provided by the P-DL.	141
E.5	Training time for P-DL modeling	141
E.6	The comparison of <i>MSE</i> , <i>SDP</i> , and <i>OVP</i> generated by the HGPAL-REF approach, MT-GPUCB, and the multi-resolution method.	142
E.7	The estimated parameter sets for both healthy and diseased domains on the heart surface by the proposed HGPAL-REF approach and the MT-GPUCB method.	143
E.8	The computation of the forward and backward pass of the automatic differentiation example as shown in Fig. D.24.	144
E.9	The optimal regularization parameter w_{opt} and the iteration number in GP-UCB for each NN structure.	145
E.10	The structure details of different neural network.	145
E.11	The variation of <i>RE</i> by P-DAL that is trained based on sensor observations from 50 spatial locations with different λ 's in the healthy and diseased heart systems. The experiments are repeated 6 times to generate the standard deviation of <i>RE</i>	146

E.12	The RE 's generated by P-DAL, P-DSL, and P-DRL. The experiments are repeated 6 times to generate the standard deviation of RE	146
E.13	p -value for the two-sample statistical test	147

List of Figures

D.1	Illustration of the proposed P-DL framework for HSP prediction. Note that $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are the predicted HSP and recovery current, respectively. The physical constraint of AP model and boundary condition are converted into residuals as physical loss $\mathcal{L}_{ph}$ and added to the total loss function of the DNN for respecting the underlying physics principle.	124
D.2	(a) Illustration of the torso-heart geometry; (b) BSPM at two different time points t_1 and t_2	125
D.3	Scheme for parallel computing.	125
D.4	Experimental design for evaluating the performance of the proposed P-DL approach.	126
D.5	The variation of (a) RE , (b) CC , and (c) MSE with respect to neural network structure.	126
D.6	The variation of (a) RE , (b) CC , and (c) MSE with respect to the number of collocation points.	127
D.7	The estimated HSP evolvment $\hat{\mathbf{u}}(\mathbf{s}, t)$ from an arbitrary spatial location in the heart generated from pure data-driven DL and P-DL with 50000 collocation points compared with the true evolvment.	127
D.8	The comparison of (a) RE , (b) CC , and (c) MSE between the proposed P-DL method and other regularization models (i.e., Tikh_0th, Tikh_1st, and STRE) when different noise levels $\sigma_\epsilon = 0.01, 0.05, 0.1$ are added to the BSPM $\mathbf{y}(\mathbf{s}, t)$	128

D.9	(a) Reference HSP mapping $\mathbf{u}(\mathbf{s}, t)$ at $t = 300$; Estimated mappings $\mathbf{u}(\mathbf{s}, t)$ at $t = 300$ by P-DL and traditional methods (i.e., Tikh_0th, Tikh_1st, and STRE) with noise level of (b) $\sigma_\epsilon = 0.01$ (c) $\sigma_\epsilon = 0.05$ and (d) $\sigma_\epsilon = 0.1$ in BSPM $\mathbf{y}(\mathbf{s}, t)$	129
D.10	The comparison of HSP evolvment $\mathbf{u}(\mathbf{s}, t)$ over time for one specific spatial location, between the true HSP data and predictions by Tikh_0th, Tikh_1st, STRE, and P-DL under the noise level of (a) $\sigma_\epsilon = 0.01$, (b) $\sigma_\epsilon = 0.05$, (c) $\sigma_\epsilon = 0.1$	129
D.11	The flowchart of the proposed P-DAL framework. Note that $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are the predicted electrical potential and recovery current on the heart geometry, respectively. $\mathbf{u}_m$ and $\hat{\mathbf{u}}_m$ denotes the measured data at specific spatial locations and the corresponding predictions, respectively. The physical constraint of AP model and boundary condition are transformed into the form of residuals as the physical loss $\mathcal{L}_{ph}$ and incorporated into the total loss function $\mathcal{L}$ of the DNN to respect the underlying physics-based principle.	130
D.12	The illustration of the geodesic distance between points A and B, which is the length of the shortest path on the 3D geometry between the two points. . . .	131
D.13	The comparison of RE generated from STGP-E, STGP-M, and P-DL models in the healthy heart system with different noise levels of $\sigma_\epsilon = 0.01$, 0.05, and 0.1.	131
D.14	(a) Reference potential mapping $\mathbf{u}(\mathbf{s}, t)$ on the healthy heart surface at $t = 30$; Predictive potential mappings $\hat{\mathbf{u}}(\mathbf{s}, t)$ at $t = 30$ by STGP-E, STGP-M, and P-DL approaches with noise level of (b) $\sigma_\epsilon = 0.01$ (c) $\sigma_\epsilon = 0.05$ and (d) $\sigma_\epsilon = 0.1$	132
D.15	(a) The RE estimated by P-DL model based on training data collected from random-sampling (P-DRL), space-filling (P-DSL), and active learning (P-DAL); (b) The estimated evolvment of cardiac electrodynamics over time for the healthy heart system.	132
D.16	The comparison of RE generated from STGP-E, STGP-M, and P-DL models in diseased heart system with different noise levels of $\sigma_\epsilon = 0.01$, 0.05, and 0.1. . . .	133

D.17 (a) Reference potential mapping $\mathbf{u}(\mathbf{s}, t)$ on the diseased heart surface at $t = 33$; Predictive potential mappings $\hat{\mathbf{u}}(\mathbf{s}, t)$ at $t = 33$ by STGP-E, STGP-M, and P-DL with noise level of (b) $\sigma_\epsilon = 0.01$ (c) $\sigma_\epsilon = 0.05$ and (d) $\sigma_\epsilon = 0.1$ for the diseased heart system.	133
D.18 (a) The RE estimated by P-DL based on training data collected from random-sampling (i.e., P-DRL), space-filling (i.e., P-DSL), and active learning (i.e., P-DAL); (b) The estimated evolvement of cardiac electrodynamics over time in the diseased heart system.	134
D.19 The RE estimated by P-DL model with different neural network structures for (a) healthy and (b) diseased heart systems. The experiments are repeated 6 times to generate the standard deviation of RE_v (i.e., the error bar).	134
D.20 The flowchart of the proposed research methodology.	135
D.21 Illustration of the experimental design for evaluating the performance of our HGPAL-REF method.	135
D.22 The different diseased regions (marked as red): (a) small MI at Loc_1, (b) large MI at Loc_1, (c) small MI at Loc_2, (d) large MI at Loc_2, (e) small MI at Loc_3, (f) large MI at Loc_3 on the heart surfaces estimated by the multi-resolution approach, MT-GPUCB, and the proposed HGPAL-REF method. Note that parameter set θ_2 is used to simulate the true diseased locations.	136
D.23 Evolvement curves of estimated BSPs produced by different methods with (a) small MI at Loc_1, (b) large MI at Loc_1, (c) small MI at Loc_2, (d) large MI at Loc_2, (e) small MI at Loc_3, (f) large MI at Loc_3, at one specific spatial point on the body surface.	137
D.24 A simple network example to illustrate automatic differentiation.	137
D.25 The illustration of the convergence of GP-UCB.	138
D.26 The RE estimated by P-DL model with different neural network structures for (a) healthy and (b) diseased heart systems. The experiments are repeated 6 times to generate the standard deviation of RE_v (i.e., the error bar).	138
D.27 The illustration of GP-UCB convergence for (a) healthy and (b) diseased heart systems.	139

D.28 The RE 's estimated by P-DAL model with different λ 's for (a) healthy and (b) diseased heart systems as the number of training points increases.	139
---	-----

Chapter 1

Introduction

1.1 Motivation

As we navigate through the increasing intricacies of modern industrial processes, advanced sensing technology has emerged as a potent and indispensable tool to monitor, model, and control such complexities. This is especially apparent in the field of healthcare. For example, the usage of distributed electrocardiography (ECG) sensing systems (also known as body sensor network) allows us to capture multi-dimensional information from the electrical readings on the body surface as a reflection of the heart's electrical activities. Additionally, cardiac catheterization, an invasive medical procedure, provides direct electrical readings from within the heart, offering valuable insights into its electrical behavior and aiding in diagnosis and treatment planning. These obtained signals can facilitate an in-depth exploration of cardiac electrodynamics, which could lead to the early detection and diagnosis of cardiac irregularities and improving the patient outcomes.

While rapid advances in sensing and imaging have resulted in a data-rich environment, leveraging this abundance of data for disease diagnosis and optimal treatment planning heavily relies on the development of innovative analytical models. There is an urgent need for novel methodologies and tools that can effectively process the high-dimensionality and complex structures of sensor data, to fully realize its potential in informing and optimizing decision-making in health systems. My research aims to address the following challenges:

- Distributed sensing brings an abundance of spatiotemporal data. It is of great importance to effectively address the spatial and temporal correlations for high-dimensional predictive modeling.
- Advanced sensing leads to a large amount of data with nonlinear properties and heterogeneous patterns, which calls for effective analytical models to comprehend the underlying features for reliable healthcare process control and decision-making.
- The presence of extraneous noise, sparse sensor placements, and other uncertainty factors escalate the difficulties in pure data-driven modeling. It is of an urgent need to exploit the underlying system physics and develop novel machine-learning approaches that integrate physics-based prior knowledge with real sensing data for robust modeling and monitoring of system dynamics.

Our goal is to develop innovative physics-augmented methodologies to gain deeper understanding of complexly structured sensor data for effective modeling and decision making in healthcare.

1.2 Research Background

Heart disease is among the leading causes of death in the US, responsible for about 21% of the total death in the year 2020 ([Ahmad and Anderson \(2021\)](#)). The average yearly medical expenses associated with heart failure care were approximated at \$24,383 per patient, with the majority of these costs being attributed to hospitalizations specifically related to heart failure, which had a median cost of \$15,879 per patient ([Urbich et al., 2020](#)). The significant mortality rates and escalating medical costs underscore the urgent need for effective and timely interventions for patients with heart conditions. Such progress depends on an in-depth comprehension of cardiac dynamic mechanism and tissue properties. By leveraging sophisticated sensor data and developing advanced analysis methodologies, we aim to provide valuable insights in cardiac problems and potentially contribute to a decrease in heart disease mortality.

In clinic, electrocardiography (ECG) is a commonly employed diagnostic method for identifying and studying abnormal heart conditions. It functions by tracking the electrical signals generated by the depolarization and repolarization of the heart’s chambers, which are detected on the surface of the torso. This process creates a spatiotemporal map of the heart’s electrical activity, providing critical insights into cardiac function (Oster et al., 1998). Building on the fundamentals of electrocardiography (ECG), the twelve-lead ECG system has been designed for enhancing the depth cardiac examination by using 12 leads to capture the heart’s electrical activity, which provides a comprehensive view of the heart’s function (Yang et al., 2013). The twelve leads are achieved through ten physical electrodes placed strategically on the patient’s body, creating various electrical pathways that represent the heart’s activity from different angles. On the other hand, the EASI ECG system, an alternative to the twelve-lead ECG, provides a simplified yet effective method for capturing cardiac electrical activity (Dower et al., 1988). This system employs just five electrodes placed in specific positions on the torso, offering a less complex setup and allowing for a more comfortable experience for the patient.

The precise diagnosis of heart conditions like acute ischemia and myocardial infarction necessitate the use of high-resolution ECG mapping (Herring and Paterson, 2006), which provides a comprehensive 3D indication of the heart’s electrical signals, and further enhances our understanding of the cardiac electrodynamics. This is where body surface potential mapping (BSPM) comes into play. BSPM employs a large number of ECG electrodes to record high-resolution spatiotemporal electrical patterns on the body surface as a manifestation of heart electrical activity. This detailed mapping can be invaluable in determining the necessity of interventions such as the implementation of a pacemaker (Pastore et al., 2007). By providing a detailed view of the heart’s electrical activity, BSPM can assist physicians in making critical treatment decisions, thus underlining its immense importance in cardiac care.

Another technique that directly captures the electrical signals from the heart is cardiac catheterization. This invasive procedure offers a more direct view at the heart’s electrical activity, compared to non-invasive body surface ECG. Cardiac catheterization involves threading a thin tube, or catheter, through a blood vessel to the heart. Multiple electrodes

can be installed at the distal end of the catheter. The sensors are in direct contact with the surface of the myocardium to record electrical signals, providing useful information for diagnosing and treating a variety of heart conditions (Jenkins et al., 1993).

Nonetheless, both body ECG and cardiac catheterization comes with the limitation. For ECG, the electrical signals, originating from the heart, tend to weaken and become distorted as they propagate towards the torso. This presents a challenge in precisely reconstructing the cardiac electrodynamics based on body-surface ECG signals, which necessitates us to address an inverse ECG problem. However, the inverse ECG problem is generally recognized as a challenging issue due to its ill-conditioned nature. The solutions are susceptible to inaccuracies and noise in body surface measurements, which can significantly variate the prediction outcomes. This often results in a less than optimal performance when it comes to reliably estimating the heart's dynamic electrical behavior.

Conversely, while cardiac catheterization offers a more direct and reliable source of cardiac electrical signals, its invasive nature limits the number of sensors that can be safely placed onto the heart surface. This limitation, unfortunately, results in sparse sensor data. Such limited data coverage is insufficient for generating a high-resolution mapping of cardiac electrodynamics for comprehensive cardiac assessment and precise diagnosis. Given the challenge, it's imperative to develop a reliable machine learning model that could maximize the value of the sparse sensor data and extract meaningful insights. By employing advanced algorithms that can interpret and analyze sparse data effectively, we could reconstruct a more comprehensive and high-resolution representation of the heart's electrical dynamics.

1.3 Research Objectives

My research goal is to develop innovative physics-augmented methodologies for modeling, monitoring, and optimizing complex systems with the applicaiton in healthcare. Specifically, the research objectives of this dissertation include:

- Developing a physics-constrained machine learning framework to address the high-dimensional inverse problem based on indirect and noisy sensor observations.

- Developing a physics-augmented strategy for optimal sensor placement to actively explore and model the dynamics of 3D complex-structured systems based on sparse observations.
- Proposing a physics-guided statistical framework to characterize the properties of the inner profile and locate the defect region based on indirect and noisy external sensor observations.

The proposed research aims to achieve several key objectives that include: 1) managing and processing substantial volume of spatiotemporal data to facilitate detailed high-dimensional modeling of systems with complex 3D geometry; 2) distilling crucial informations about system dynamics from complex sensor data; 3) enhancing decision-making strategies in the face of extraneous noises and factors of uncertainty with direct or indirect sensor observations.

1.4 Organization of the Dissertation

This dissertation is structured around several manuscripts, with each of Chapters 2 through Chapter 4 composed as a research paper, either published or currently under review. The rest of the dissertation is arranged in the following manner:

Chapter 2 present presents a physics-constrained deep learning (P-DL) framework for robust inverse ECG modeling. We first integrate the physics prior knowledge of cardiac electrodynamics with advanced deep learning (DL) infrastructures to solve the inverse ECG problem. Secondly, we investigate the effect of the physics-driven regularization on the inverse solution, which is controlled by an introduced regularization parameter. Finally, we validate and evaluate the performance of the proposed framework to reconstruct the cardiac electrodynamics from BSPM data in a 3D torso-heart geometry. And compare our approach with existing regularization methods, i.e., Tikhonov zero-order, Tikhonov first-order, spatiotemporal regularization, and the physiological-model-constrained Kalman Filter method.

In Chapter 3, we presents a novel physics-constrained deep active learning (P-DAL) framework to model the spatiotemporal cardiac electrodynamics from the limited intra-cardiac sensor observations. We first adapt the P-DL framework presented in Chapter 2, embedding physics-based knowledge of cardiac dynamic propagation into neural network to generate the prediciton of electrodynamics. Next, we propose an innovative active learning approach aimed at identifying informative spatial locations on the heart’s surface. These locations hold potential to deliver a more thorough understanding of the heart’s electrical activity, thereby enhancing the predictive capability of the P-DL method. The experimental results verifies the efficacy of P-DAL model in reconstructing the comprehensive spatiotemporal cardiac electrodynamics using sparse intra-cardiac observations.

In Chapter 4, a Hierarchical Gaussian Process (GP)-based Active Learning (HGPAL) framework is proposed for relizble defect localization in complex 3D systems. We first develop a GP regression model on the 2D embedding of a 3D geometry to capture the geometric information. Second, we propose a Hierarchical GP-based active learning approach composed of master GP and sub-GP modeling to account for the system heterogeneity and further estimate the spatial-varying model parameter for defect localization. We evaluate the performance of the proposed HGPAL framework in a 3D body-heart system to identify and localize infarct in the heart.

Chapter 5 encapsulates the key contributions of my research and outlines potential directions for future research.

Chapter 2

Physics-constrained Deep Learning for Robust Inverse ECG Modeling

A version of Chapter 2 was published in *IEEE Transactions on Automation Science and Engineering* on Jan 28th, 2022:

Citation:

J. Xie and B. Yao, “Physics-constrained Deep Learning for Robust Inverse ECG Modeling”, *IEEE Transactions on Automation Science and Engineering*, vol. 20, pp. 151 - 166, 2022. (Xie and Yao, 2022b)

Authors:

Jianxin Xie¹, Bing Yao¹

¹Department of Industrial and Systems Engineering, University of Tennessee at Knoxville

2.1 Introduction

The development of cutting-edge medical sensing and imaging technique greatly benefits the modeling of complex dynamic systems and further facilitates effective system monitoring and optimization for smart health. For instance, electrocardiography (ECG) is a widely-used diagnostic tool to investigate and detect abnormal heart conditions. It monitors electrical signals on the torso surface initiated by the electrical depolarization and repolarization of the heart chambers, establishing a spatiotemporal pattern of the cardiac electrical activity (Oster et al., 1998). Effective ECG monitoring systems enable medical scientists to non-invasively access the electrical behavior of the heart by simply measuring the electrical potentials on the body surface. Twelve-lead ECG, one of the most widely-used ECG systems, is implemented by strategically distributing four electrodes on limbs and six on the chest to harvest multi-angle views of the spatiotemporal cardiac electrodynamics (Yang et al., 2013). However, traditional ECG systems (e.g., 12-lead ECG, 3-lead VCG) only employ a limited number of ECG sensors on the thorax and thus the resulted cardiac electrical distribution is with low spatial resolution. Accurate diagnosis of heart diseases such as acute ischemia calls upon the high-resolution ECG mapping (Herring and Paterson, 2006) that is capable of capturing a comprehensive 3D view of the cardiac electrical signals. Thus, body surface potential mapping (BSPM) has been developed by deploying an increased number (e.g., 120) of ECG electrodes to record the high-resolution spatiotemporal cardiac electrodynamics projected on the body surface (Bond et al., 2009; Lacroix et al., 1991; Rudy and Burnes, 1999; Chen and Yang, 2015; Zhu et al., 2018).

Rapid advances in sensing and imaging bring abundant data and provide an unprecedented opportunity for data-driven disease diagnosis and optimal treatment planning (Piri et al., 2017; Lin et al., 2018; Si et al., 2017; Yao et al., 2017; Iquebal and Bukkapatnam, 2020; Yao et al., 2021). Realizing the full potential of the sensing and imaging data depends greatly on the development of innovative analytical models. In the human heart system, the ECG signal (or BSPM) describes the spatiotemporal distribution of electric potentials on the body surface, which reflects electrical behaviors of the source, i.e., epicardial potentials. When projecting from the heart to the body surface, the spatiotemporal pattern of the cardiac

electrodynamics becomes diminished and blurred, which defies the accurate characterization of local abnormal regions in the heart. Precision cardiology (Corral-Acero et al., 2020; Trayanova, 2019; Yao, 2021) calls for advanced analytical models that can effectively handle the spatiotemporal data structure for robust prediction of heart-surface electrical signals (i.e., heart-surface potentials (HSP)) from body-surface sensor measurements (i.e., BSPMs). This is also called the inverse ECG problem (Rudy and Burnes, 1999; Oster et al., 1997; Gulrajani, 1998; Yao et al., 2016; Yao and Yang, 2020).

Inverse ECG problem is commonly recognized as ill-conditioned, because a small input noise may cause highly unstable prediction. Great efforts have been made to achieve robust inverse solutions by adding proper spatial or/and temporal regularization (Brooks et al., 1999; Greensite, 2003; Wang et al., 2009; Yu et al., 2015; Karoui et al., 2018). However, most existing regularization methods (e.g., Tikhonov, L1-norm, or spatiotemporal regularization) focus on exploiting the data structure (i.e., data-driven regularization). Little has been done to integrate the well-established physics model of cardiac electrodynamics (i.e., physics-driven regularization) with advanced machine learning approaches for robust inverse ECG modeling. In order to further increase the estimation accuracy and robustness, it is imperative to incorporate the physics prior knowledge that theoretically describes the cardiac electrical wave propagation into the transformation from BSPM data to heart-surface signals. This paper proposes a novel physics-constrained deep learning (P-DL) framework to solve the inverse ECG problem and achieve a robust estimation of HSP from noisy BSPM data.

2.2 Research Background

2.2.1 Inverse ECG Modeling and Regularization

The goal of inverse ECG modeling is to enable non-invasive reconstruction of cardiac electrical activity from body-surface sensor measurements for optimal diagnosis and characterization of heart disease. It has been widely recognized that inverse ECG problem is ill-posed, the solution of which is sensitive to model uncertainty and measurement noise. Specifically, the relationship between BSPM $\mathbf{y}(\mathbf{s}, t)$ with HSP $\mathbf{u}(\mathbf{s}, t)$ can be expressed as

$\mathbf{y}(\mathbf{s}, t) = \mathbf{R}\mathbf{u}(\mathbf{s}, t) + \epsilon$, where $\mathbf{s}$ and t denotes the spatial and temporal coordinates respectively, and $\mathbf{R}$ is a transfer matrix solved by Divergence Theorem and Green's Second Identity (Yao and Yang, 2020; Barr et al., 1977). However, the rank-deficiency of transfer matrix $\mathbf{R}$ (i.e., $\text{rank}(\mathbf{R}) < \min\{\dim(\mathbf{u}), \dim(\mathbf{y})\}$) and its large condition number (i.e., $\text{cond}(\mathbf{R}) = \|\mathbf{R}\| \|\mathbf{R}^{-1}\|$) make the estimated HSP $\hat{\mathbf{u}}(\mathbf{s}, t)$ very sensitive to the BSPM data $\mathbf{y}(\mathbf{s}, t)$. Small fluctuation in BSPMs, i.e., $\Delta\mathbf{y}$, will cause a significant variation $\Delta\hat{\mathbf{u}}$ in $\hat{\mathbf{u}}$ (i.e., $\Delta\hat{\mathbf{u}}/\hat{\mathbf{u}} \simeq \text{cond}(\mathbf{R})\Delta\mathbf{y}/\mathbf{y}$). Therefore, regularizing the inverse ECG model for robust solution is crucially important for understanding the mechanism of the cardiac physiology and pathology, further facilitating the accurate disease diagnosis and optimal therapeutic design.

In existing literature, remarkable progress has been made in developing statistical regularization algorithms by exploiting the spatiotemporal data structure to improve the robustness of inverse ECG solution. Among them, Tikhonov method (including zero-order and first-order) is one of the most popular regularization approaches. It minimizes the mean squared error (i.e., residual) from body-heart transformation and penalizes the L_2 norm of inverse HSP solution (Rudy and Burnes, 1999; Dawoud et al., 2008; Wang et al., 2011; Throne and Olson, 2000) to suppress the unreliable component and improve the overall smoothness of the solution. The objective function for Tikhonov regularization is formulated in Eq.(2.1)

$$\min_{\mathbf{u}(\mathbf{s}, t)} \{ \|\mathbf{y}(\mathbf{s}, t) - \mathbf{R}\mathbf{u}(\mathbf{s}, t)\|_2^2 + \lambda_{Tikh}^2 \|\Gamma\mathbf{u}(\mathbf{s}, t)\|_2^2 \} \quad (2.1)$$

where $\|\cdot\|_2$ indicates the L_2 norm, λ_{Tikh} is a regularization coefficient, and Γ denotes the operator constraining HSP solution $\mathbf{u}(\mathbf{s}, t)$, which is an identity matrix for zero-order regularization and a spatial gradient operator for first-order regularization to increase the spatial smoothness of the inverse solution. Another widely-used regularization technique is the L_1 norm-based method including zero-order (Wang et al., 2011) and first-order (Ghosh and Rudy, 2009; Shou et al., 2010), which penalizes L_1 norm of inverse HSP solution to encourage the model sparsity and regularity. However, both Tikhonov and L_1 regularization methods only consider the spatial data structures in inverse ECG modeling, but do not effectively incorporate the temporal evolving patterns of the cardiac electrodynamics.

A variety of spatiotemporal regularization methods have been developed to fuse the space-time correlations into the inverse ECG problem for robust estimation of HSP (Brooks et al., 1999; Greensite, 2003; Wang et al., 2009; Yu et al., 2015; Yao and Yang, 2020). For example, Messnarz *et al.* Messnarz et al. (2004) reconstructed the potential distribution on the heart surface by defining a symmetric matrix of the spatial gradient to account for the spatial correlation and assuming that cardiac electric potentials are non-decreasing during the depolarization phase to address the temporal correlation. Yao and Yang (2016) proposed a spatiotemporal regularization algorithm by regularizing both the spatial smoothness and increasing temporal robustness in inverse ECG modeling, whose objective function is given by Eq. (2.2).

$$\min_{\mathbf{u}(s,t)} \sum_t \left\{ \|\mathbf{y}(s,t) - \mathbf{R}\mathbf{u}(s,t)\|_2^2 + \lambda_s^2 \|\Gamma \mathbf{u}(s,t)\|_2^2 + \lambda_t^2 \sum_{\tau=t-\omega/2}^{\tau=t+\omega/2} \|\mathbf{u}(s,t) - \mathbf{u}(s,\tau)\|_2^2 \right\} \quad (2.2)$$

where λ_s and λ_t denote the spatial and temporal regularization parameters respectively, and ω is a time window selected to incorporate the temporal correlation. However, most existing spatiotemporal regularization methods mainly focus on exploiting the data structure to capture the statistical correlations between adjacent spatial locations and neighboring time points, and do not incorporate the physics principle that governs the HSP evolving dynamics into regularization for robust inverse ECG modeling.

Note that Wang et al. (2009) proposed a physiological-model-constrained Kalman filter (P-KF) approach to solve the inverse ECG problem. They formulated the cardiac system into a high-dimensional stochastic state-space model and estimate the cardiac electric potential with the Kalman filter method. However, this method involves large matrices operation and may suffer from the initialization problem (Wang et al., 2009; Zhao and Huang, 2020). Little has been done to develop innovative machine learning approaches that can incorporate the underlying physics law for robust and effective inverse ECG modeling.

2.2.2 Deep Neural Networks

Deep Neural Network (DNN) or Deep Learning (DL) has been proved to be a powerful tool as a universal function approximator that is capable of handling a variety of problems with strong nonlinearity and high dimensionality (LeCun et al., 2015). Different advanced DL structures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been developed to extract efficient features from high-dimensional data for predictive modeling (Yu et al., 2020; Wang et al., 2019b; Park et al., 2016; Zheng et al., 2020; Wang et al., 2021a; Zheng et al., 2019). However, little has been done to utilize the DNN to solve the inverse ECG problem and predict the HSP. This is due to the fact that traditional DNN algorithms require an extensive amount of data for modeling complex physical or biological systems. Such data are generally sparse and costly to acquire, which makes traditional DNNs lack robustness and fail to converge. For example, the intracardiac electrical signals are often measured by placing electrodes within the heart via cardiac catheters, which are of low spatial resolution, expensive to acquire, and incur discomforts to patients (Schulman et al., 1999).

To cope with the problem of limited data in complex physical and biological systems, recent techniques have been developed to incorporate the underlying system physics into data-driven predictive modeling. For example, Wang et al. (2021b) proposed to constrain the GP modeling with the mechanistic model of proliferation invasion describing the dynamics of tumor growth for accurate prediction of the spatial distribution of tumor cell density. Raissi (2018a) innovatively developed a physics-informed neural network (PINN) framework that embeds the underlying physical law of nonlinear high-dimensional systems in the form of partial differential equations (PDE) into the DNN. The PINN algorithm has been further adapted to solve various problems including simulating incompressible fluid flow by Navier–Stokes equation (Jin et al., 2021), elastodynamic problem (Rao et al., 2021a), heat source layout optimization (Chen et al., 2020b), material designing (Liu and Wang, 2019), physical law discovery from scarce and noisy data (Berg and Nyström, 2019; Both et al., 2020), cardiac activation by solving Eikonal equation (Sahli Costabal et al., 2020), and cardiovascular flows modeling (Kissas et al., 2020).

As opposed to the traditional DNN modeling, where the mechanism underlying the physics system is generally unknown and has to be learned from the observed data, there exists a vast amount of prior knowledge about the cardiac dynamic system which is characterized by the physics-based PDEs. However, very little has been done to leverage the advanced deep learning infrastructure and incorporate the physics-based cardiac model for robust inverse ECG modeling. In this work, we propose a P-DL framework to encode physics-based principles into the DNN to robustly reconstruct the dynamic potential mapping on the heart surface from BSPMs. Additionally, we propose to efficiently search for the optimal balance between the physics-based loss and the data-driven loss using GP-UCB. The proposed P-DL framework not only effectively incorporates the body-surface sensing data into ECG modeling but also respects the underlying physics law, thereby significantly increasing the robustness and accuracy of the inverse ECG solution (Xie and Yao, 2022b).

2.3 Research Methodology

2.3.1 Cardiac Electrophysiological Model

Different phenomenological models have been developed to study cardiac electrophysiology and delineate the electrical wave propagation in the heart (Sermesant et al., 2006; Camara et al., 2011). Here, without loss of generality, we utilize Aliev–Panfilov (AP) model (Aliev and Panfilov, 1996b) that is widely used in various cardiac research to describe the spatiotemporal evolution of cardiac electrodynamics:

$$\frac{\partial u}{\partial t} = \nabla \cdot (D \nabla u) + k_r u(u - a)(1 - u) - uv \quad (2.3)$$

$$\frac{\partial v}{\partial t} = \xi(u, v)(-v - k_r u(u - a - 1)) \quad (2.4)$$

$$\mathbf{n} \cdot \nabla u|_{\Gamma_H} = 0 \quad (2.5)$$

where u denotes the normalized HSP, v represents the recovery current that controls the local depolarization behavior of the cardiac electric potential, and ∇ denotes a spatial gradient operator on the heart geometry. Parameters $\xi(u, v) = e_0 + \mu_1 v / (u + \mu_2)$ describes the coupling

between u and v , D is the diffusion conductivity, k_r is repolarization constant, and a controls the tissue excitability. The value of each parameter for the remainder of this research are set as $a = 0.1$, $D = 10$, $k_r = 8$, $e_0 = 0.002$, $\mu_1 = \mu_2 = 0.3$ from the documented literature (Aliev and Panfilov, 1996b). Note that Eq. (2.5) represents the boundary condition to guarantee there is no electric current flowing out from the heart, where Γ_H is the heart boundary and $\mathbf{n}$ denotes the surface normal of Γ_H . The electrophysiological model in Eq. (2.3)-(2.5) will further be used to constrain the DNN model to achieve a robust and accurate inverse ECG solution.

2.3.2 Artificial Neural Networks

Deep learning has increasingly drawn significant attention due to its successful applications in a variety of fields, such as image and character recognition, natural language processing, advanced manufacturing, and computer vision (Xiao et al., 2020; Wang and Yao, 2021; Wang et al., 2018). Feed-forward fully-connected neural network is the most fundamental architecture, which consists of an input layer, multiple hidden layers, and one output layer. The information passes layer-by-layer in one direction. The value of each neuron in a layer is calculated by the sum of the products of weights and outputs from the previous layer, which is then activated by a specific activation function. The relationship between two neighbor layers, layer- $(n - 1)$ and layer- (n) , is generally described by the following equation:

$$\mathbf{x}_n = \sigma(\mathbf{b}_n + \mathbf{W}_n \mathbf{x}_{n-1}), 1 < n \leq K \quad (2.6)$$

where $\sigma(\cdot)$ represents the activation function, K is the total number of layers, $\mathbf{W}_n$ and $\mathbf{b}_n$ are the weight matrix and bias vector for layer- (n) respectively, and $\mathbf{x}_n$ and $\mathbf{x}_{n-1}$ denote the outputs of the $(n - 1)$ -th and (n) -th layers, respectively.

Our objective is to solve the inverse ECG problem for the spatiotemporal HSP $\mathbf{u}(\mathbf{s}, t)$ through the DNN modeling. The selection of optimal neural network structure will be introduced in Section 2.4.1. The hyperbolic tangent function is chosen as an activation function due to its great training performance in multi-layer neural networks. Here, we implement the basic neural network structure with a tailored loss function that is constrained

by the physics law (see Eq. (2.3)-(2.5)) to solve the inverse ECG problem, and eventually reconstruct the electrical signals on the heart surface. The detailed construction of the P-DL framework will be presented in the following subsection.

2.3.3 P-DL approach

Fig. D.1 shows the proposed P-DL framework. The solution of the inverse ECG problem is parameterized by a DNN that is trained to not only satisfy the physics law but also meet the data-driven constraint incurred by the body-surface sensor measurements (i.e., BSPM). Specifically, we estimate the spatiotemporal mapping of HSP as

$$[\mathbf{s}, t] \xrightarrow{\mathcal{N}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})} [u(\mathbf{s}, t), v(\mathbf{s}, t)]$$

where $\mathcal{N}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$ defines the DNN model, $\mathbf{s}$ and t denotes the spatial and temporal coordinates respectively, and $\boldsymbol{\theta}_{NN}$ represents the DNN hyper-parameters. Note that the DNN, i.e., $\mathcal{N}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$, is constructed with an input layer composed of the spatiotemporal coordinates $[\mathbf{s}, t]$, the hidden layers to approximate the complex functional relationship, and the output layer with the estimation of $\hat{u}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$ and $\hat{v}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$. We further encode the physics law into the DNN by defining a unique loss function as

$$\mathcal{L}(\boldsymbol{\theta}_{NN}) = \mathcal{L}_{hb} + w \cdot \mathcal{L}_{ph} \quad (2.7)$$

where w is introduced to control the regularization effect imposed by the physics principles. The total loss $\mathcal{L}(\boldsymbol{\theta}_{NN})$ consists of the following two key elements:

(1) **Data-driven loss $\mathcal{L}_{hb}$:** BSPMs provide the evolving dynamics of electrical potential distribution $\mathbf{y}(\mathbf{s}, t)$ on the torso surface, which can be leveraged to estimate the heart-surface electrical signals. Specifically, the HSP $\mathbf{u}(\mathbf{s}, t)$ and body surface potential (BSP) $\mathbf{y}(\mathbf{s}, t)$ can be related by the transfer matrix $\mathbf{R}$ (i.e., $\mathbf{y}(\mathbf{s}, t) = \mathbf{R}\mathbf{u}(\mathbf{s}, t)$), which is derived according to the divergence theorem and Green's Second Identity (see more details at Yao and Yang (2020); Barr et al. (1977)). The data-driven loss generated from the transformation between BSP

and HSP, denoted as $\mathcal{L}_{hb}$, is thus defined as:

$$\mathcal{L}_{hb} = \frac{1}{N_{st}} \sum_t \sum_s \|\mathbf{y}(\mathbf{s}, t) - \mathbf{R}\hat{\mathbf{u}}(\mathbf{s}, t)\|^2 \quad (2.8)$$

where N_{st} stands for the total number of the descrtized spatiotemporal instances.

As stated before, due to the big condition number of the matrix $\mathbf{R}$, the prediction of HSP lacks the accuracy and robustness if only minimizing the mean squared error (i.e., training the DNN with only $\mathcal{L}_{hb}$) without any regularization. Additional constraint needs to be imposed to regularize the DNN-based estimation for robust inverse ECG modeling.

(2) **Physics-based loss $\mathcal{L}_{ph}$:** In order to further increase the estimation accuracy and robustness, physics constraints are imposed over the spatiotemporal collocation points, $[\mathbf{s}_i, t_i]$'s, that are randomly selected within the spatial domain defined by the heart geometry and the temporal domain defined by the cardiac cycle. The physics constraints include two parts: the boundary condition and cardiac electrophysiological model (see Eq. (2.3)-(2.5)). Specifically, we define the boundary condition-based residual as:

$$r_{bc}(\mathbf{s}, t; \boldsymbol{\theta}_{NN}) := \mathbf{n} \cdot \nabla \hat{\mathbf{u}}(\mathbf{s}, t; \boldsymbol{\theta}_{NN}), \quad \mathbf{s} \in \Gamma_H \quad (2.9)$$

The boundary condition will then be achieved by encouraging $r_{bc}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$ to be close to zero. As such, we define the boundary condition-based loss as

$$\mathcal{L}_{bc} = \frac{1}{N_{bc}} \sum_{i=1}^{N_{bc}} (r_{bc}(\mathbf{s}_i, t_i; \boldsymbol{\theta}_{NN}))^2, \quad \mathbf{s}_i \in \Gamma_H \quad (2.10)$$

where N_{bc} denotes the number of collocation points selected on the boundary Γ_H . Additionally, the electrophysiological model-based residuals are defined as

$$\begin{aligned} r_u(\mathbf{s}, t; \boldsymbol{\theta}_{NN}) &:= \frac{\partial \hat{u}}{\partial t} - \nabla \cdot (D \nabla \hat{u}) - k \hat{u}(\hat{u} - a)(1 - \hat{u}) + \hat{u} \hat{v} \\ r_v(\mathbf{s}, t; \boldsymbol{\theta}_{NN}) &:= \frac{\partial \hat{v}}{\partial t} - \xi(\hat{u}, \hat{v})(-\hat{v} - k \hat{u}(\hat{u} - a - 1)) \end{aligned} \quad (2.11)$$

The physiological-based constraint will be achieved by minimizing the magnitude of both $r_u(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$ and $r_v(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$. Thus, the electrophysiological model-based loss is defined as

$$\mathcal{L}_f = \frac{1}{N_f} \sum_{i=1}^{N_f} ((r_u(\mathbf{s}_i, t_i; \boldsymbol{\theta}_{NN}))^2 + (r_v(\mathbf{s}_i, t_i; \boldsymbol{\theta}_{NN}))^2) \quad (2.12)$$

where N_f denotes the total number of selected spatiotemporal collocation points to incorporate the AP model. The physics-based loss combines the effect from both the boundary condition and the electrophysiological model, i.e., $\mathcal{L}_{ph} = \mathcal{L}_{bc} + \mathcal{L}_f$, which is encoded to the DNN loss function to respect the underlying physics laws in inverse ECG modeling.

In order to compute the loss function during the DNN training, it is necessary to acquire the value of partial derivatives of $u(\mathbf{s}, t)$ and $v(\mathbf{s}, t)$ with respect to spatial or temporal variables. This can be easily achieved through automatic differentiation, which is able to yield the exact result of derivatives without approximation error (except for the round-off error) (Baydin et al., 2018). Automatic differentiation currently is the most popular method to train the deep neural networks by calculating the derivatives *via* breaking down the differentials in multiple paths using the chain rule (See more details in Appendix A). In this study, automatic differentiation is engaged to obtain the derivatives with respect to the spatiotemporal coordinates $[\mathbf{s}, t]$ in order to encode the physical laws in the form of a partial differential equation into the DNN.

2.3.4 Bayesian Optimization for Regularization Parameter

The performance of the proposed P-DL framework also depends on the value of the regularization parameter w , as shown in Eq. (3.8). In traditional regularization methods, the regularization parameter is generally selected according to the L-curve method, which is ambiguous, insecure, and requires solving the inverse ECG problem repeatedly given different parameter values. This may incur a huge computation burden for large-scale inverse ECG modeling. Here, we propose to actively search for the optimal regularization parameter through the Gaussian Process (GP)-based surrogate modeling.

GP has been widely used in machine learning as a method for solving complex black-box optimization problems (Seeger, 2004). Here, we propose to search for the optimal

regularization parameter by GP-based Bayesian optimization. Before establishing the Bayesian optimization procedure, we first define a balance metric as the criterion to evaluate the prediction performance given the parameter w as

$$m(w) = \log \left[\left(\frac{\mathcal{L}_{hb}}{\mathcal{L}_{ph}} + \frac{\mathcal{L}_{ph}}{\mathcal{L}_{hb}} \right) \cdot \mathcal{L}(\theta_{NN}) \right] \quad (2.13)$$

The term of $\left(\frac{\mathcal{L}_{hb}}{\mathcal{L}_{ph}} + \frac{\mathcal{L}_{ph}}{\mathcal{L}_{hb}} \right)$ in Eq. (2.13) controls the balance degree between the data-driven loss $\mathcal{L}_{hb}$ and physics-based loss $\mathcal{L}_{ph}$, which will achieve its minimum value of 2 if $\mathcal{L}_{hb} = \mathcal{L}_{ph}$ (i.e., the training result is perfectly balanced between the data-driven term and physics-regularization term). The term of $\mathcal{L}(\theta_{NN})$ controls the overall magnitude of the total loss, minimizing which will help to reduce either the data-driven or physics-based loss or both. A combination of the two terms will serve as an effective metric to measure the prediction performance of the P-DL given different values of the regularization parameter. In other words, minimizing Eq. (2.13) will enable to not only reduce the total loss $\mathcal{L}(\theta_{NN})$ but also help to balance between the data-driven loss $\mathcal{L}_{hb}$ and physics-based loss $\mathcal{L}_{ph}$, thereby improving the performance of the P-DL for robust inverse ECG modeling.

The input of the GP modeling is the regularization parameter, i.e., w 's, and the output is the corresponding balance metric, i.e., $m(w)$'s. As such, we assume

$$m(w) = \hat{m}(w) + \delta_m, \quad \delta_m \sim \mathcal{N}(0, \sigma_{\delta_m}^2) \quad (2.14)$$

where δ_m is Gaussian white noise added for numerical stability, and $\hat{m}(w)$ is further modeled as a GP, i.e., $\hat{m}(w) \sim \mathcal{GP}(0, \mathcal{K})$, where $\mathcal{K}$ denotes the kernel function which is selected as the widely used square-exponential kernel:

$$\mathcal{K}(w_i, w_j) = \sigma_m^2 \cdot \exp \left(-\frac{\|w_i - w_j\|^2}{2l_m^2} \right) \quad (2.15)$$

where σ_m and l_m denote the GP hyperparameters, which can be determined by maximizing the log marginal likelihood for GP regression (Rasmussen, 2003). Given the collected observations $\{\mathbf{m}_{1:n}, \mathbf{w}_{1:n}\}$, where $\mathbf{m}_{1:n} = [m_1 \dots m_n]^T$ and $\mathbf{w}_{1:n} = [w_1 \dots w_n]^T$, the predictive

distribution of $m(w)$ given an arbitrary unevaluated regularization parameter w_* is

$$P(m(w_*)|\mathbf{m}_{1:n}, \mathbf{w}_{1:n}, w_*) = \mathcal{N}(\mu_n(w_*), \sigma_n^2(w_*)) \quad (2.16)$$

The predictive mean $\mu_n(w_*)$ and variance $\sigma_n^2(w_*)$ is given as

$$\begin{aligned} \mu_n(w_*) &= \mathcal{K}(w_*, \mathbf{w}_{1:n})(\mathcal{K}(\mathbf{w}_{1:n}, \mathbf{w}_{1:n}) + \sigma_{\delta_m}^2 \mathcal{I})^{-1} \mathbf{m}_{1:n} \\ \sigma_n^2(w_*) &= \mathcal{K}(w_*, w_*) - \mathcal{K}(w_*, \mathbf{w}_{1:n})(\mathcal{K}(\mathbf{w}_{1:n}, \mathbf{w}_{1:n}) + \sigma_{\delta_m}^2 \mathcal{I})^{-1} \mathcal{K}(\mathbf{w}_{1:n}, w_*) \end{aligned} \quad (2.17)$$

where $\mathcal{I}$ represents an identity matrix with the same dimensionality of $\mathcal{K}(\mathbf{w}_{1:n}, \mathbf{w}_{1:n})$.

With the GP-surrogate modeling, we can evaluate the prediction performance of P-DL given an arbitrary value of w without actually solving the inverse ECG problem. As such, we can intuitively select the optimal parameter by minimizing the GP predictive mean, i.e., $w_{opt} = \arg \min_w \{\mu_n(w)\}$. However, the accuracy of the GP surrogate depends on the quality of the training data $\{\mathbf{m}_{1:n}, \mathbf{w}_{1:n}\}$, which requires an effective strategy to collect the most informative data points. To find a best next query point w_{n+1} in the parameter space for updating GP, a good acquisition function should balance the trade-off between exploitation of current knowledge and exploration of uncertain parameter space, which is constructed based on GP upper-confidence-bound (GP-UCB) (Srinivas et al., 2012) as

$$w_{n+1} = \arg \max_w \{-\mu_n(w) + \beta^{1/2} \sigma_n(w)\} \quad (2.18)$$

where β is a positive constant that balances between the global search and local optimization. The optimal value of β depends on the context of the problem. The theoretical rule to select an appropriate value for β is provided in Srinivas et al. (2009). Specifically, a good regret bound can be achieved with high probability, if $\beta_i = 2 \log \left(\frac{|\mathcal{A}|^2 \pi^2}{6\delta} \right)$, where i denotes the iteration index, $\mathcal{A}$ is the input space, and δ is a parameter between 0 and 1. However, this rule does not restrict other choices for β . The authors in (Srinivas et al., 2009) mentioned that the GP-UCB algorithm can be improved by scaling the theoretical value for β_i down by a factor of 5. In our work, we simplify the process for β selection and empirically select $\beta = 9$.

After the new point w_{n+1} is selected, the P-DL framework will be run to solve the inverse ECG problem and calculate the corresponding balance metric $m(w_{n+1})$ with the new regularization parameter w_{n+1} . When the new input-output pair $\{w_{n+1}, m_{n+1}\}$ becomes available, the GP surrogate will be updated according to:

$$\begin{bmatrix} \mathbf{m}_{1:n} \\ m_{n+1} \end{bmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} \mathcal{K}(\mathbf{w}_{1:n}, \mathbf{w}_{1:n}) + \sigma_{\delta_m}^2 \mathbf{I} & \mathcal{K}(\mathbf{w}_{1:n}, w_{n+1}) \\ \mathcal{K}(w_{n+1}, \mathbf{w}_{1:n}) & \mathcal{K}(w_{n+1}, w_{n+1}) \end{bmatrix} \right) \quad (2.19)$$

The GP-UCB process is iterated until the training points (i.e., values of the regularization parameter w), collected by Eq. (2.18) converge, where the convergence criterion is defined as $\|w_{n+1} - w_n\| < 10^{-2}$. The converged value of w will be selected as the optimal regularization parameter.

2.3.5 Parallel Computing to Increase Computation Efficiency

In the proposed P-DL method, the DNN parameters (i.e., θ_{NN}) can be estimated by back-propagation algorithms incorporated with optimization methods such as stochastic gradient descent (SGD), adaptive moment estimation (Adam), and other optimization methods. In our work, we engage Adam as the optimizer due to its improved generalization performance (Keskar and Socher, 2017; Kingma and Ba, 2014). This process involves evaluating the gradients of the loss function at each training point (or collocation point) and updating θ_{NN} iteratively. A large number of collocation points will increase the computational intensity of the training process. Fortunately, an appealing feature of the P-DL method is that computing tasks are much easier to be operated in parallel than traditional inverse ECG modeling. Here, we propose to scale up the P-DL algorithm by effectively harnessing the computing power of multiple processors collaboratively.

Fig. D.3 shows the proposed strategy for parallel computing. To increase the computational parallelism and escape from the local minima (Bottou, 2010; Ge et al., 2015), the whole collocation dataset is divided into mini-batch data (denoted as $\mathcal{D}$), which are further split for the Adam algorithm into n subsets during network training, i.e., $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup \dots \cup \mathcal{D}_n$. Thus, the gradient of the loss function with data $\mathcal{D}$ will be

calculated as

$$\nabla \mathcal{L}(\theta_{NN}) = \frac{1}{|\mathcal{D}|} [\nabla \mathcal{L}_1(\theta_{NN}) + \cdots + \nabla \mathcal{L}_{n_D}(\theta_{NN})] \quad (2.20)$$

where $\nabla \mathcal{L}_i(\theta_{NN}) = \sum_{j \in \mathcal{D}_i} \nabla \mathcal{L}_j(\theta_{NN})$, denoting the sum of gradients on subset $\mathcal{D}_i$. As such, parallel computing can be readily implemented to update network hyperparameters by assigning each subset $\mathcal{D}_i$ to an individual processor and combining results from many computing units.

2.4 Experimental Design and Results

We validate and evaluate the performance of the proposed P-DL framework in a 3D torso-heart geometry to estimate the HSP from BSPM measurements. As shown in Fig. D.2(a), the heart geometry is formed by 1094 nodes and 2184 mesh elements, and the torso surface consists of 352 nodes and 677 mesh elements, which is obtained from the 2007 PhysioNet Computing in Cardiology Challenge (Dawoud et al., 2008; Goldberger et al., 2000). This torso-heart geometry will result in a transfer matrix $\mathbf{R}$ with a dimensionality of 352×1094 . The spatiotemporal HSP and BSPM data are developed from the two-variable cardiac reaction-diffusion system through the procedure detailed in (Wang et al., 2009). The reference HSP data $\mathbf{u}(\mathbf{s}, t)$ consists of 1094×661 data points, which are used to evaluate the quality of inverse ECG solution $\hat{\mathbf{u}}(\mathbf{s}, t)$. In other words, there are 1094 reference time-series data from the heart geometry. Note that the cardinality of the discretized time steps is 661 in the present investigation. The BSPM data $\mathbf{y}(\mathbf{s}, t)$ has a dimension of 352×661 , which goes into the data-driven loss (i.e., $\mathcal{L}_{hb}$) to train the P-DL and inversely reconstruct the HSP signal. Fig. D.2(b) illustrates the spatiotemporal BSPMs that will be used to predict the HSP.

We investigate the impact of neural network structure, number of collocation points, noise level, and physics model parameters on the prediction performance as shown in Fig. D.4. The proposed P-DL approach will be benchmarked with traditional regularization methods including Tikhonov zero-order (Tikh_0th), Tikhonov first order (Tikh_1st), the

spatiotemporal regularization (STRE) (see Eq. (2.1)-(2.2) for the objective functions), and the physiological-model-constrained Kalman filter (P-KF) method. The model performance will be evaluated according to three metrics: Relative Error (RE), Correlation Coefficient (CC), and Mean Squared Error (MSE), which are defined as

$$RE = \frac{\sqrt{\sum_{s,t} \|\hat{\mathbf{u}}(s, t) - \mathbf{u}(s, t)\|^2}}{\sqrt{\sum_{s,t} \|\mathbf{u}(s, t)\|^2}} \quad (2.21)$$

$$CC = \frac{\sum_s (\hat{\mathbf{u}}(s, \cdot) - \bar{\hat{\mathbf{u}}}(s, \cdot))^T \cdot (\mathbf{u}(s, \cdot) - \bar{\mathbf{u}}(s, \cdot))}{\sum_s \|(\hat{\mathbf{u}}(s, \cdot) - \bar{\hat{\mathbf{u}}}(s, \cdot))\| \cdot \|(\mathbf{u}(s, \cdot) - \bar{\mathbf{u}}(s, \cdot))\|} \quad (2.22)$$

$$MSE = \frac{1}{N_{st}} \sum_{s,t} \|\hat{\mathbf{u}}(s, t) - \mathbf{u}(s, t)\|^2 \quad (2.23)$$

where N_{st} stands for the total number of spatiotemporal instances, $\mathbf{u}(s, t)$ and $\hat{\mathbf{u}}(s, t)$ represent the reference and estimated HSPs, $\hat{\mathbf{u}}(s, \cdot)$ and $\mathbf{u}(s, \cdot)$ denote the estimated and reference time series vectors at spatial location s , and $\bar{\hat{\mathbf{u}}}(s, \cdot)$ and $\bar{\mathbf{u}}(s, \cdot)$ are the mean of $\hat{\mathbf{u}}(s, \cdot)$ and $\mathbf{u}(s, \cdot)$, respectively. The metric RE quantifies how large the discrepancy between the estimation $\hat{\mathbf{u}}(s, t)$ and the truth $\mathbf{u}(s, t)$ compared to the magnitude of the true signals. A smaller value of RE indicates a better model performance. CC characterizes the degree of pattern similarity between $\mathbf{u}(s, t)$ and $\hat{\mathbf{u}}(s, t)$ with $CC = 1$ for perfect pattern match and $CC = 0$ for zero correlation (i.e., $\mathbf{u}(s, t)$ and $\hat{\mathbf{u}}(s, t)$ are not similar at all). MSE measures the average squared difference between $\mathbf{u}(s, t)$ and $\hat{\mathbf{u}}(s, t)$.

The proposed P-DL algorithm is successfully implemented on TensorFlow, which is a popular platform for DL. TensorFlow can be operated on the heterogeneous environment including a multi-core Central Processing Unit (CPU), Graphics Processing Unit (GPU), and Tensor Processing Unit (TPU) (Abadi et al., 2016). Such cutting-edge hardware developments make parallelization applicable, which significantly increases the computation efficiency in DL training (Shi et al., 2016). In our work, we process the P-DL algorithm on the parallel computing platform NVIDIA CUDA, which is a software framework for programs running across different environments such as CPU and GPU. The neural network is carried on TensorFlow-GPU with Python application programming interface (API), and we found that the computation speed is around 6.5 times faster than running on CPU only. Note that

the CPU used for the computation is Intel(R) Xeon(R) W-2223CPU @ 3.6GHz. The GPU is NVIDIA Quadro P2200.

2.4.1 Impact of Neural Network Structure on the Prediction Performance

In the present investigation, we study the impact of network structure on the prediction performance of the proposed P-DL method. Specifically, five feedforward fully-connected neural network structures are evaluated, whose structure details are listed in Table E.1. Each network structure is evaluated by executing P-DL with the same data input and hyperparameters. Note that a Gaussian noise $\epsilon(\mathbf{s}, t) \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with standard deviation $\sigma_\epsilon = 0.01$ is added to BSPM data $\mathbf{y}(\mathbf{s}, t)$. GP-UCB algorithm is first implemented to find the best regularization parameter to balance between $\mathcal{L}_{hb}$ and $\mathcal{L}_{ph}$ for each neural network structure. The optimal value w_{opt} for each network structure is shown in Appendix B.

Fig. D.5(a-c) compares the *RE*, *CC* and *MSE*, for different network structures. Note that the experiments are repeated 10 times, and therefore every obtained *RE*, *CC* and *MSE* comes with a corresponding error bar (i.e., one standard deviation of *RE*, *CC* and *MSE*). The network being too thin (e.g., NN1 with 5 neurons for each layer) or too shallow (e.g., NN2 with only 3 layers) may cause relatively high *RE* with a value of (0.2084 ± 0.0188) or (0.1839 ± 0.0052) , big *MSE* value of (0.0139 ± 0.0025) or (0.0107 ± 0.0006) , and low *CC* value of (0.9641 ± 0.0064) or (0.9722 ± 0.0016) . Thus, a more complex structure is preferred in terms of prediction performance. If we increase the number of neurons in each layer (e.g., NN3 and NN5), or add more layers to the network (e.g., NN4), the resulted *RE* and *MSE* are significantly reduced. According to Fig.D.5 (a-c), there is no significant difference among the *RE*, *CC*, and *MSE*'s obtained from P-DL with NN3, NN4, and NN5, indicating further increasing the network complexity may not be necessary. As such, we choose NN3 as the network structure for later experiments.

2.4.2 Impact of Collocation Points on the Prediction Performance

To verify the necessity and efficacy of the physics-driven regularization, we further compare the performance of our proposed P-DL model with the data-driven deep learning (DL) method. The loss function of the data-driven DL is the pure data-driven loss defined in Eq. (9). In other words, zero collocation points will be selected to encode the physics-based principle. We also randomly select different numbers of spatiotemporal collocation points to enforce the physics laws imposed by the AP model. Specifically, we choose five numbers of collocation points: $N_{ph} = 10000, 25000, 50000, 75000, 100000$ to quantify its impact on the predictive accuracy of the proposed P-DL framework. Similarly, we engage the feedforward fully connected DNN with 5 layers and 10 neurons in each layer as the model infrastructure.

As shown in Fig. D.6(a-c), the *RE* and *MSE* decrease, and *CC* increases as we raise the number of collocation points. Specifically, with $N_{ph} = 0$, i.e., the pure data-driven DL has the worst estimation performance, which is due to the fact that the physics principle is absent from the model construction. By increasing N_{ph} to 50000, *RE* and *MSE* are reduced to (0.1490 ± 0.0123) and (0.0071 ± 0.0012) , while *CC* increases to (0.9817 ± 0.0031) . If we continue increasing the number of collocation points, the estimated *RE*, *MSE*, and *CC* do not improve significantly, besides it may potentially increase the computational burden. Therefore, we select 50000 collocation points for later numerical experiments. Fig. D.7 depicts the HSP evolvment at one arbitrary location on the heart geometry yielded from the data-driven DL and the P-DL with 50000 collocation points, compared with true evolvment. The predicted evolvment by the data-driven DL shows a more significant discrepancy from the true evolvment pattern compared with the one estimated by the P-DL model.

2.4.3 Comparison Study with Traditional Regularization Methods

We further benchmark the proposed P-DL framework with traditional regularization methods that are commonly used in current practice (i.e., Tikh_0th, Tikh_1st, and STRE). Fig. D.8 (a-c) shows the comparison of *RE*, *CC*, and *MSE* obtained from the four methods at different noise levels of $\sigma_\epsilon = 0.01, 0.05$, and 0.1 . As the noise level increases from $\sigma_\epsilon = 0.01$ to $\sigma_\epsilon = 0.1$, the *RE* and *MSE* increase monotonically for all the methods. Specifically, *RE* and

MSE increase from (0.4070 ± 0.0004) , $(0.0524 \pm 7E-5)$ to (0.7060 ± 0.0025) , (0.1583 ± 0.0007) for Tikh_0th, from (0.2666 ± 0.0008) , (0.0225 ± 0.0002) to (0.4625 ± 0.0013) , (0.0681 ± 0.0006) for Tikh_1st, from (0.2185 ± 0.0005) , (0.0151 ± 0.0002) to (0.2481 ± 0.0013) , (0.0195 ± 0.0002) for STRE, and from (0.1490 ± 0.0123) , (0.00071 ± 0.0012) to (0.1544 ± 0.0086) , (0.0076 ± 0.0008) for the P-DL model. Correspondingly, CC decreases for each method as the noise level is raised up to $\sigma_\epsilon = 0.1$. Overall, the proposed P-DL model generates the smallest RE and MSE , and highest CC compared to other regularization methods regardless of the noise level.

As illustrated in Fig. D.8, it is also worth noting that the estimated RE and MSE obtained from the commonly used Tikh_0th and Tikh_1st methods increases dramatically as the bigger noise being added to BSPM data $\mathbf{y}(\mathbf{s}, t)$, where the RE reaches a 73.4% and 73.5% growth, and MSE increases by 202.0% and 202.7% as the noise level increases from $\sigma_\epsilon = 0.01$ to $\sigma_\epsilon = 0.1$, respectively. The STRE model reduces the effect of noise level but still yields an increase level of 13.5% in RE and 29.1% in MSE . On the other hand, the proposed P-DL method achieves only a 3.6% increase in RE and 7.0% in MSE as the noise level reaches $\sigma_\epsilon = 0.1$. Similarly, CC drops 26.7%, 10.7%, and 1.2% for Tikh_0th, Tikh_1st, and STRE model, respectively, whereas for P-DL model, CC only declines by 0.12% as the noise level grows from $\sigma_\epsilon = 0.01$ to $\sigma_\epsilon = 0.1$, exhibiting extraordinary anti-noise performance. Thus, P-DL model not only accomplishes superior prediction accuracy, but also demonstrates the predictive robustness when encountering a high level of noise in BSPM.

Fig. D.9(a) delineates the true potential mappings on the heart surface. Because the potential distribution is changing over time dynamically, Fig. D.9 only depicts the mappings at one specific time step, i.e., $t = 300$. Note that the HSP $\mathbf{u}(\mathbf{s}, t)$ is normalized, and both $\mathbf{u}$ and t are unitless. Fig. D.9(b-d) show the estimated HSP distribution generated by the four different methods, with noise level of $\sigma_\epsilon = 0.01$, 0.05, and 0.1 added to BSPMs $\mathbf{y}(\mathbf{s}, t)$, respectively. The estimated potential mappings by Tikh_0th, Tikh_1st, and STRE under each noise level present significantly different color patterns compared with the reference mapping in Fig. D.9(a). Typically, under the noise level as large as $\sigma_\epsilon = 0.1$, the proposed P-DL yields the RE and MSE of 0.1469 and 0.0068, which are remarkably lower than the RE and MSE estimated by other traditional regularization methods, i.e., 0.7067 and 0.1594 by

Tikh_0th, 0.4641 and 0.0683 by Tikh_1st, 0.2479 and 0.0195 by STRE, respectively, leading to a similar HSP pattern with the ground truth shown in Fig. D.9(a). Similarly, the P-DL model generates the highest CC of 0.9824, compared to CC value of 0.6629, 0.8402, and 0.9481 estimated by Tikh_0th, Tikh_1st, and STRE, respectively, indicating P-DL model yields the smallest pattern difference with the real potential distribution. This is due to the fact that the P-DL model successfully incorporates the bio-electrical propagation (i.e., AP model) and boundary condition, yielding a robust inverse ECG solution.

Furthermore, Fig. D.10 shows the comparison of estimated HSPs $\hat{\mathbf{u}}(\mathbf{s}, t)$ evolving over time for one specific spatial location in the heart by the four methods under different noise levels. Note that both Tikh_0th and Tikh_1st generate noisy estimation whose noisy pattern becomes more significant when σ_ϵ increases from 0.01 to 0.1. The STRE method smoothens out the noise over time in the estimated signal due to the imposed extra temporal regularization, but the temporal regularization does not reflect the physics-based prior knowledge of cardiac electrodynamics, leading to a morphology deviation from the true HSP signal. In contrast, by incorporating the physics-based principles, our proposed P-DL model yields a smooth and robust estimation that matches the morphologic pattern of the true HSP signal well under all the three noise levels as illustrated in Fig. D.10(a-c).

2.4.4 Comparison Study with the Physics-constrained Statistical Method

Wang *et al.* proposed to solve the inverse ECG model by fusing the physics-based cardiac model (i.e., the AP model) into the Kalman filter (KF) algorithm (i.e., P-KF) (Wang et al., 2009). The detailed algorithm is shown in Appendix C. Note that the prediction accuracy of the P-KF algorithm depends heavily on the assumptions about the initial conditions (i.e., the initial distribution of heart electrical potential $\mathbf{u}_0(\mathbf{s})$), which is a common problem in KF-based methods (Zhao and Huang, 2020). However, the initial potential distribution in the heart, i.e., $\mathbf{u}_0(\mathbf{s})$, at the beginning of the cardiac cycle, depends on the anatomical location of the earliest excitation in the heart, which is difficult to accurately obtain in clinical practice. An inaccurately specified initial condition will significantly reduce the

prediction accuracy of the inverse ECG solution. Table E.2 summarizes the performance of P-KF under different initial conditions with a noise level of 0.01 added to the BSPM. Note that the hyperparameters in this algorithm are adopted from (Wang et al., 2009). The experiments are repeated 10 times for each initial condition.

According to Table E.2, when the initial potential mapping is assigned with the true condition, i.e., $\hat{\mathbf{u}}_0(\mathbf{s}) = \mathbf{u}_0(\mathbf{s})$, the P-KF algorithm generates the smallest RE ($0.1426 \pm 1e-5$) and MSE ($0.0064 \pm 1e-6$), and the highest CC ($0.9848 \pm 3e-6$), which is comparable with our P-DL method with RE , MSE , and CC of (0.1490 ± 0.0123), (0.0071 ± 0.0012), and (0.9817 ± 0.0031), respectively. It is also worth mentioning that statistical test shows that there is no significant difference of the RE generated from the P-KF and P-DL (see Appendix D for more detail). However, if the P-KF algorithm is initialized with an inaccurate initial condition, the estimation accuracy will be significantly reduced. For example, if an initialization noise level of $\sigma_{\epsilon_{u_0}} = 0.05$, i.e., $\epsilon_{u_0}(\mathbf{s}) \sim N(0, 0.0025)$, is added to the overall potential mapping at $t = 0$ (i.e., $\hat{\mathbf{u}}_0(\mathbf{s}) = \mathbf{u}_0(\mathbf{s}) + \epsilon_{u_0}(\mathbf{s})$), the RE and MSE rise to (0.1626 ± 0.0027) and (0.0084 ± 0.0003), and CC drops to (0.9803 ± 0.0006). A random initialization (i.e., randomly picking the earliest excitation on the heart geometry) can drastically increase the RE to (0.8940 ± 0.0458) and MSE to (0.2539 ± 0.0260), and reduce the CC to (0.3247 ± 0.0712). Similarly, under the zero initialization (i.e., $\hat{\mathbf{u}}_0(\mathbf{s}) = 0$), the resulted RE and MSE are as high as ($0.9960 \pm 7e-006$) and (0.3143 ± 0.0028) respectively, and CC decreases to (0.1563 ± 0.0007).

As such, the performance of P-KF is comparable to our P-DL model only when the true initialization condition is provided (i.e., $\hat{\mathbf{u}}_0(\mathbf{s}) = \mathbf{u}_0(\mathbf{s})$). Incorrect or inaccurate initial conditions will significantly deteriorate the model performance. However, in clinical reality, the knowledge of the true initial condition of HSP mapping is difficult to track accurately, which makes P-KF less applicable in clinical practice. In contrast, our P-DL model does not require any prior information of initial HSP mapping at $t = 0$.

Moreover, the KF-based model is more computational-intensive because the algorithm is iterated along the discretized time domain and involves repeated numerical evaluations of the AP model and large matrix manipulation at each time instance. Specifically, for each time instance, the discretized AP model is evaluated $2N+1$ times to generate the *priori* prediction of HSP, where N is the total number of the discretized nodes on the 3D heart geometry (see

more details at Appendix C). Table E.3 further summarizes the program running time for each method. It takes about 19 minutes for the proposed P-DL to converge and generate the inverse ECG solution, which is much more computation-efficient compared with the P-KF method (i.e., 3 hours and 10 minutes) and the STRE method (i.e., 2 hours and 26 minutes). Note that Tikh_0th and Tikh_1st can generate the inverse results within a second. This is due to the fact that both Tikh-based methods have closed-form solutions and only focus on regularizing the spatial structure of the inverse ECG solution but ignore the patterns in the temporal domain.

2.4.5 Impact of Electrophysiological Model Parameters on the Prediction Performance

In the case that AP model could not precisely describe the underlying cardiac electrodynamics, it is necessary to investigate how the variations in the model parameters will influence the predictive accuracy of P-DL model. Table E.4 lists the *RE* and *CC* when the individual hyperparameter (i.e., a , k_r , D , $e0$, μ_1 , μ_2) deviates from the reference value by $\pm 10\%$, given that a noise level of $\sigma_\epsilon = 0.01$ is added to BSPM $\mathbf{y}(\mathbf{s}, t)$. Note that *RE*, *MSE*, and *CC* evaluated with the reference parameter setting (i.e., $a = 0.1$, $D = 10$, $k_r = 8$, $e0 = 0.002$, $\mu_1 = \mu_2 = 0.3$) are (0.1490 ± 0.0123) , (0.0071 ± 0.0012) , and (0.9817 ± 0.0031) as shown in Fig. D.8.

After designated $\pm 10\%$ variations to the AP parameters, the overall prediction performance deteriorates with an increase in *RE* and *MSE* and a decrease in *CC*. Nevertheless, as shown in Table E.4, *RE* and *MSE* predicted by the proposed P-DL model with $\pm 10\%$ variations in the AP hyperparameters still rank the smallest compared with traditional regularization methods, i.e., (0.4070 ± 0.0004) and $(0.0524 \pm 7e - 5)$ by Tikh_0th, (0.2666 ± 0.0008) and (0.0225 ± 0.0002) by Tikh_1st, and (0.2185 ± 0.0005) and $(0.0151 \pm 6e - 5)$ by STRE, respectively. Similarly, the *CC* estimated by the P-DL model still scores the highest value among the four methods, i.e., (0.8609 ± 0.0002) by Tikh_0th, (0.9413 ± 0.0005) by Tikh_1st, and (0.9606 ± 0.0002) by STRE even when the model is provided with an inaccurate AP parameters. Although the AP model with imperfect parameter setting is encoded as the

physics constraint into the P-DL model, the body-heart transformation loss $\mathcal{L}_{hb}$ also plays an important role in optimizing the neural network. These two parts interact with each other, making the P-DL model robust against the measurement noise and parameter uncertainties.

2.5 Discussion

2.5.1 Advantages of the Proposed P-DL Method in Inverse ECG Modeling

The advantages of the P-DL method in solving the inverse ECG problem over the statistical models are listed as follows:

(1) **The physics prior knowledge can be easily incorporated into DNN:** The spatiotemporal cardiac electrodynamics is generally described by the physics-based principles in terms of partial differential equations (PDEs). Due to the rapid advancements in deep learning infrastructures, the physics-based PDEs (i.e., the physics-based cardiac model) can be easily integrated with data-driven loss in the DNN through automatic differentiation. The physics-based cardiac model in the P-DL will serve as a spatiotemporal regularization to respect the evolving dynamics of cardiac electrical signals. On the other hand, it is difficult to encode the physics law into traditional statistical modeling. In fact, in order to achieve high-quality prediction with statistical methods, researchers have to regularize the inverse ECG solution by delicately designing a proper constraint, usually by mathematically developing a spatial operator to characterize the correlation in the space domain and/or adding temporal regularization to increase the prediction robustness in the time domain. In other words, traditional regularization methods in inverse ECG modeling focus on exploiting the data structure (i.e., data-driven regularization), but ignore existing physics prior knowledge for regularization. As such, the proposed P-DL-based inverse ECG method achieves a better prediction performance compared with traditional regularization methods as shown by the experimental results in Fig. [D.8-D.10](#).

(2) **Computation Efficiency:** Another key attractive feature of deep learning-based modeling is that computing tasks are much easier to be executed in parallel, which will

increase the computation efficiency by harnessing the computing power of multiple processors collaboratively. However, solving the statistical models with spatiotemporal regularization (e.g., the STRE or P-KF method) requires either large matrix manipulation, which can be computationally prohibitive for large-scale inverse ECG modeling, or the development of complex optimization algorithms, which can be exceedingly mathematically heavy, as shown in Table E.3.

2.5.2 Training the P-DL Model

The proposed P-DL network is trained by solving the physics-encoded loss function in Eq. (3.8). Specifically, the training data comes from the body surface measurements (i.e., BSPM $y(\mathbf{s}, t)$), which is related to the HSP through a transfer matrix $\mathbf{R}$, and the spatiotemporal collocation points to encode the physics-based principles (i.e., the AP model). Note that we did not use the traditional training-validation-testing split to train and evaluate the P-DL model. This is due to the fact that the deep learning (DL) implementation of the inverse ECG modeling is different from traditional DL predictive models. The ultimate goal of inverse ECG modeling is to estimate the HSP from BSPM. In other words, the data directly from the object of interest (i.e., the heart surface data $\mathbf{u}(\mathbf{s}, t)$) does not go into the training set or serve as input of the network. As such, it will not introduce the common problem of overfitting as in traditional machine learning. Additionally, because the electrical signals from each heart node affect the electrical signals on the entire body surface, the electrical potentials either on the heart or body surface cannot be analyzed locally. In other words, the HSP cannot be reliably estimated by a local portion of the BSPM. As such, training-validation-testing split is not generally applicable in inverse ECG modeling Wang et al. (2009); Messnarz et al. (2004); Ghosh and Rudy (2009); Shou et al. (2010). The quality of inverse ECG solution is commonly evaluated by comparing the estimations $\hat{\mathbf{u}}(\mathbf{s}, t)$ with the reference signals $\mathbf{u}(\mathbf{s}, t)$ on the heart surface as shown in existing literature Wang et al. (2009); Messnarz et al. (2004); Ghosh and Rudy (2009); Shou et al. (2010).

2.5.3 Define New Network Structures with Physics-based Models

In the present investigation, we engage a fully-connected feed-forward neural network (NN) structure to solve the inverse ECG problem by defining a unique loss function that fuses the data-driven transformation loss and physics-based loss. In the existing literature, several approaches have been developed to leverage the physics-based knowledge to construct more complicated NN architectures [Willard et al. \(2020\)](#). For example, the NN structures can be adapted by defining intermediate physics variables to produce physics-consistent predictions. Daw *et al.* [Daw et al. \(2020\)](#) designed a special long-short-term-memory (LSTM) NN by introducing a monotonicity-preserving structure into the LSTM cell, which can help to predict physically consistent densities in lake temperature modeling. Additionally, invariances and symmetries of the systems can be leveraged to design a physics-based NN structure. For instance, Wang *et al.* [Wang et al. \(2020\)](#) incorporated the system symmetries into the convolutional NN (CNN) to improve the accuracy and generalizability of DNN-based predictive models for learning physical dynamics. The NN architecture can also be modified by encoding physics-based domain-specific knowledge. For example, Sadoughi *et al.* [Sadoughi and Hu \(2019\)](#) designed a CNN by incorporating domain-specific knowledge of bearings and their fault characteristics for the defect detection of rolling element bearings.

Designing new NNs with the physics-based cardiac model can be challenging for inverse ECG modeling. It is difficult to find a meaningful intermediate physics variable that is able to improve the overall predictive performance. Additionally, the AP model that describes the cardiac electrodynamics is governed by non-linear PDEs over the complexly structured heart geometry, which does not involve simple invariance or symmetric properties to be used to design effective structures. Thus, in this paper, we focus on incorporating the physics prior knowledge (i.e., the AP model) into the DNN by defining a loss function that combines both the data-driven and physics-based loss. Specifically, the AP model serves as a physics-based spatiotemporal regularization to respect the spatiotemporal evolving dynamics of cardiac signals to improve the predictive performance in inverse ECG modeling. Experimental results demonstrate the superior performance of the proposed P-DL framework compared with existing approaches in solving the inverse ECG problem. In our future work, we will

focus on leveraging the physics-based cardiac model to design new network structures to further improve the predictive performance of the P-DL for inverse ECG modeling.

2.6 Conclusions

In this paper, we propose a physics-constrained deep learning (P-DL) strategy to predict the heart-surface electrical signals from body-surface sensor measurements. First, we encode the underlying physics law of cardiac electrodynamics into a DNN to solve the inverse ECG model. Second, in order to balance between the body-heart transformation (i.e., data-driven loss) and physics-based loss, we defined a balance metric as an indicator to efficiently evaluate the predictive performance given different regularization parameters. Third, we propose to solve for the optimal regularization parameter using the GP-UCB algorithm. Finally, we propose to implement the P-DL with TensorFlow-GPU to increase the computation efficiency. We validate and evaluate the performance of the P-DL framework in a 3D torso-heart system to estimate HSP from BSPM data. Numerical experiments demonstrate that the proposed P-DL model is robust against measurement noise and other uncertainty factors. The inverse ECG solution by the P-DL method significantly outperforms existing regularization methods, i.e., Tikhonov zero-order, Tikhonov first-order, STRE model, and physiological-model-constrained Kalman filter. The presented P-DL framework can be broadly implemented to investigate other spatiotemporal high-dimensional systems such as thermodynamics and tumor-growth systems.

Chapter 3

Physics-constrained Deep Active Learning for Spatiotemporal modeling of Cardiac Electrodynamics

A version of Chapter 3 was published in *Computers in Biology and Medicine* on Jul 1st, 2022:

Citation:

J. Xie and B. Yao, “Physics-constrained Deep Active Learning for Spatiotemporal modeling of Cardiac Electrodynamics”, *Computers in Biology and Medicine*, vol. 146, p. 105586, 2022. (Xie and Yao, 2022a)

Authors:

Jianxin Xie¹, Bing Yao¹

¹Department of Industrial and Systems Engineering, University of Tennessee at Knoxville

3.1 Introduction

Heart disease is among the leading causes of death in the US, responsible for about 21% of the total death in the year 2020 ([Ahmad and Anderson \(2021\)](#)). Scientists have developed effective medical procedures to directly or indirectly collect clinical data and study the electrical activity of the heart. For example, electrocardiography (ECG) is a widely-used multi-sensor technique to non-invasively infer the heart electrical signals from body-surface potential measurements by strategically placing electrode sensors at various positions on the skin surface of the thorax and limbs ([Yao et al. \(2021\)](#); [Zhu et al. \(2018\)](#)). The electrical activity (i.e., electric potential) of the heart can be further estimated given the body-surface ECG measurements, which is called the inverse ECG problem ([Gulrajani \(1998\)](#)). However, as the electrical signals originate from the heart, they become diminished and blurred when propagating toward the torso, which makes it difficult to accurately reconstruct the cardiac electrodynamics from body-surface ECG signals. Moreover, the inverse ECG problem is widely acknowledged as ill-conditioned because the predictive result is sensitive to the uncertainty and noise inherited from the body surface measurements, leading to unsatisfactory performance in reliably estimating the dynamic electrical behavior of the heart.

Cardiac catheterization, a standard medical procedure to directly access the cardiac electrical signals, has been developed to help physicians to investigate and diagnose pathological heart conditions ([Bourassa \(2005\)](#)). Specifically, a thin and flexible catheter is inserted into an artery or vein from the groin, neck, or arm, and threaded through a blood vessel into the heart chamber to track the dynamic evolvement of the cardiac electric potential ([Jenkins et al. \(1993\)](#)). Multiple electrodes can be installed at the distal end of the catheter, which directly contact the surface of the myocardium to collect the time series signals (i.e., the time course of the electric potential). As such, cardiac catheterization provides a more reliable data source to investigate the heart electrical activity. Although the catheterization-based measurement straightforwardly provides the cardiac electrical signals, such data are sparse and limited for traditional data-driven machine learning or statistical

modeling to construct a complete mapping of the cardiac electrodynamics due to the invasive nature of the catheterization.

Cardiac modeling and simulation provide unprecedented opportunities for the quantitative understanding of heart disease mechanisms and the exploration of optimal therapeutic designs (Niederer et al. (2019); Yao (2021); Xie and Yao (2022a)). A variety of electrophysiological models have been proposed to quantitatively describe the electrical wave propagation in the heart and further facilitate the investigation of cardiac disease mechanisms. The computational models typically describe the multi-scale cardiac electrophysiology at the cellular or tissue level in the form of PDEs (Mao (2015)). The cellular-level model is abstracted from experimental findings of the electrophysiology of heart cells to simulate the electrical behavior induced by the dynamics of ion concentration across cardiac cells. The tissue-level model homogenizes the electrophysiological property of the heart tissue to macroscopically describe the cardiac electrical activity, which reduces the model complexity with fewer variables and equations. In order to quantitatively elucidate the cardiac electrodynamics, the finite element method (FEM) is widely-used to numerically solve the electrophysiological models (McInerney and Terzopoulos (1995); Sundnes et al. (2006)). However, FEM involves the discretization of both the spatial domain defined by the heart geometry and the temporal domain of the cardiac cycle. The computation complexity to solve the PDEs that govern the cardiac electrical behavior at each time step is proportional to the number of discretized nodes on the heart geometry, resulting in expensive computational costs for detailed cardiac modeling. More importantly, real-world clinical measurements cannot be readily integrated with the FEM numerical procedure, leading to low reliability of the model results and inferior applicability of FEM-based simulation in clinical practice.

Gaussian process (GP) modeling has been a popular method to solve the spatiotemporal dynamics of various complex systems (Wang et al. (2019a); Ezzat et al. (2019); Yao and Yang (2018)). GP simulates a stochastic process where any finite collection of spatial and temporal random variables follows a multivariate normal distribution (Xia et al. (2008); Ba and Joseph (2012)). Various cardiac-related studies have involved GP to model the biophysical process due to its relatively good predictive ability. For instance, Plumlee et al. (2016) developed a Bayesian GP to estimate functional parameters in the cardiac ion channel

model. [Dhamala et al. \(2018\)](#) used GP as a surrogate to improve the sampling process and further estimate the local tissue excitability in a 3D cardiac model. [Feng et al. \(2017\)](#) employed GP-based active learning to find the next best location for voltage mapping from the right ventricle to guide the procedure of catheter ablation. Previously, researchers have also engaged GP modeling to investigate the spatiotemporal propagation of electrical waves in the heart ([Hu et al. \(2020\)](#)). GP-based cardiac modeling involves manipulating a large kernel matrix with the dimensionality of $(P * Q) \times (P * Q)$, where P and Q denote the cardinality of the discretized spatial and temporal coordinates. This can be impractically computation-intensive for detailed modeling of the spatiotemporal cardiac electrodynamics with big P and Q . Moreover, the cardiac simulation from GP models heavily depends on the input data *priori* which is often hard to collect clinically.

With the recent fast proliferation of available data and advanced computational tools, deep learning or deep neural network (DNN) has drawn drastic attention due to its broad applications in various scientific regions, including natural language processing ([Goldberg \(2016\)](#)), image recognition ([He et al. \(2016\)](#)), healthcare management ([Chen et al. \(2020c\)](#); [Wang and Yao \(2021\)](#)), etc. Deep learning has been widely implemented to solve nonlinear PDEs of complex physical systems due to its strong predictive power. For example, [Rao and Liu \(2020\)](#) developed a 3D deep convolutional neural network (3D-CNN) to model the microstructure using representative volume elements and predict the anisotropic material properties of composites. [Han et al. \(2019\)](#) proposed to leverage the CNN to model the multiphysics process of evolving topology in the electrosurgical dissection of soft hydrated tissues. [Zhang et al. \(2019b\)](#) engaged a deep long short-term memory (LSTM) network to replace FEM-based modeling and predict the nonlinear structural seismic response in a data-driven fashion. However, traditional data-driven DNNs ignore the underlying physics principles governing the system behavior, and thus require a large quantity of experimental or computation-yielded data to train the network.

It is worth noting that there exists a vast amount of prior knowledge to describe general physical and biological systems, which can be embedded into DNN as model constraints to simulate the system dynamics ([Raissi \(2018b\)](#); [Raissi et al. \(2019\)](#)). The physics-based constraints can suppress the demand of large datasets and further enhance the

robustness of deep learning for accurate predictive modeling by respecting the physics-based ground truth. [Raissi et al. \(2019\)](#) innovatively leveraged the strong function approximation ability of DNN with the well-established physics laws to build a general physics-informed neural network (PINN) framework, which is able to effectively solve non-linear PDEs with limited training data. The PINN technique arouses tremendous innovations in modeling the physical and biophysical systems, including the liquid flow simulation ([Raissi et al. \(2020\)](#)), elastodynamic problems ([Rao et al. \(2021b\)](#)), non-linear structural systems ([Zhang et al. \(2020\)](#)), cardiovascular flow modeling ([Kissas et al. \(2020\)](#)), and cardiac activation mapping ([Sahli Costabal et al. \(2020\)](#)). [Xie and Yao \(2022b\)](#) developed a physics-constrained deep learning (P-DL) approach for robust inverse ECG modeling, which regularizes the inverse ECG solution with the physics-based cardiac model. However, little has been done to incorporate the cardiac electrophysiological model as the physics constraint into DNNs to model the spatiotemporal cardiac electrodynamics from sparse intracardiac sensor observations.

Moreover, the performance of predictive modeling depends to a great extent on the quantity and quality of training data. However, intracardiac signals are difficult to collect in clinical reality because the procedure requires invasively inserting a thin catheter, which carries a limited number of sensors, into the heart chamber. To further increase the predictive power of DNN-based methods, an effective active learning strategy is urgently needed to seek the spatial locations in the heart that can potentially provide comprehensive information of the cardiac electrodynamics. One possible approach is to quantify the predictive uncertainty of the DNN modeling and select the spatial locations with large uncertainty for subsequent data collection. For example, the Bayesian-based approach has been widely implemented for uncertainty quantification in DNN. The Bayesian DNN assigns a prior probability distribution to the network parameters and Markov chain Monte Carlo (MCMC) is commonly employed to estimate the posterior distribution and quantify the predictive uncertainty ([Maddox et al. \(2019\)](#); [Neal \(2012\)](#); [MacKay \(1992\)](#)). However, the traditional Hamiltonian Monte Carlo requires the information of full gradient of the log-likelihood function, which is difficult to obtain for modern DNNs ([Maddox et al. \(2019\)](#)). Moreover, the MCMC method

can be prohibitively expensive when evaluations of the posterior distribution of the network parameters invoke large-scale DNN training and modeling.

[Gal and Ghahramani \(2016\)](#) has theoretically proved that the DNN with dropout is equivalent to an approximation of a probabilistic deep GP. Dropout, a deep learning technique, now can be viewed as an alternative for convenient and computation-effective uncertainty quantification. This appealing approach has been successfully implemented in several physics-informed DNN models ([Zhang et al. \(2019a\)](#); [Yang et al. \(2021\)](#); [Daw et al. \(2020\)](#)). Here, we propose to engage dropout as the uncertainty estimation technique since it only collects the stochastic predictions at the test time and does not require repeated DNN training as opposed to traditional Bayesian models. Furthermore, uncertainty quantification serves as the sole criterion in traditional active learning strategies ([Sahli Costabal et al. \(2020\)](#)), which may not work well if the high uncertainty locations tend to aggregate together at local regions. Thus, there is a need to develop new active learning criteria that can not only identify highly uncertain regions but also guarantee the predictive model can estimate the spatiotemporal dynamics globally well.

This work presents a novel physics-constrained deep active learning (P-DAL) framework to model the spatiotemporal cardiac electrodynamics from the limited sensor observations. We first adapt the physics-constrained deep learning (P-DL) framework developed in [Xie and Yao \(2022b\)](#) to incorporate the physics prior knowledge that describes the electrical wave propagation of the heart into deep learning to infer the spatiotemporal cardiac electrodynamics from sparse sensor measurements. Second, we develop a novel active learning strategy to seek the informative spatial locations on the heart surface that can potentially provide more comprehensive information about the heart electrical activity to increase the predictive power of the P-DL method. The proposed active learning criterion combines both the prediction uncertainty of the P-DL and the space-filling design over the heart domain. We evaluate and validate the performance of the proposed framework to model the cardiac electrodynamics in both healthy and diseased heart systems. Experimental results show that the P-DL approach significantly outperforms traditional GP-based methods. We also demonstrate the efficacy of the proposed active learning procedure

by benchmarking with traditional learning strategies in both the healthy and diseased heart systems.

3.2 Research Methodology

3.2.1 *Spatiotemporal Gaussian Process (GP) Modeling*

In this subsection, we introduce a spatiotemporal GP (STGP) to model the cardiac electrodynamics. We denote the electrical potential as $u(\mathbf{s}, t)$, where $\mathbf{s}$ and t refers to the 3D spatial locations in the heart and temporal coordinates over the cardiac cycle, respectively. In GP modeling, we assume $u(\mathbf{s}, t) = \hat{u}(\mathbf{s}, t) + \delta$, $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$, where $\{(\mathbf{s}, t), u\}$ is a set of input-output pairs and δ is a Gaussian random variable introduced to model the observation noise. We further assume $\hat{\mathbf{u}} \sim \mathcal{GP}(\mathbf{0}, \mathcal{K})$, which defines the prior distribution over the latent function $\hat{u}(\mathbf{s}, t)$. $\mathcal{K}$ represents the kernel function to manifest important model properties, which needs to properly account for the spatiotemporal structure of the cardiac electrodynamics. It should also be noted that GP modeling involves evaluating the inverse and determinant of the covariance matrix $\mathcal{K}$, which can be highly computationally intensive. Specifically, in the STGP modeling, $\mathcal{K}$ has a dimensionality of $(P * Q) \times (P * Q)$, where P and Q denote the cardinality of observed spatial and temporal points, respectively. To address the computational issue, we take the advantage of the Kronecker property and write the kernel function as the Kronecker product of two sub-covariance matrices as:

$$\mathcal{K}[(s, t), (s', t')] = \mathcal{K}_s(s, s') \mathcal{K}_t(t, t') \implies \mathcal{K} = \mathcal{K}_s \otimes \mathcal{K}_t \quad (3.1)$$

where $\mathcal{K}_s$ and $\mathcal{K}_t$ represent the spatial and temporal kernel function with the dimension of $P \times P$ and $Q \times Q$, respectively. Here, we engage two different kernel functions, i.e., exponential and a Matérn 3/2 kernel, to capture the spatiotemporal correlations. The specific

kernel definition is shown below:

$$\begin{aligned}
\textbf{Exponential Kernel:} \quad \mathcal{K}_{se}(s, s') &= \sigma_e \cdot \exp\left(-\frac{d(s, s')}{2l_{se}}\right) \\
\mathcal{K}_{te}(t, t') &= \sigma_e \cdot \exp\left(-\frac{|t - t'|}{2l_{te}}\right)
\end{aligned} \tag{3.2}$$

$$\mathcal{K}_e[(s, t), (s', t')] = \mathcal{K}_{se}(s, s') \otimes \mathcal{K}_{te}(t, t')$$

$$\begin{aligned}
\textbf{Matérn 3/2 kernel:} \quad \mathcal{K}_{sm}(s, s') &= \sigma_m \cdot \left(1 + \frac{\sqrt{3}d(s, s')}{l_{sm}}\right) \exp\left(-\frac{\sqrt{3}d(s, s')}{2l_{sm}}\right) \\
\mathcal{K}_{tm}(t, t') &= \sigma_m \cdot \left(1 + \frac{\sqrt{3}|t - t'|}{l_{tm}}\right) \exp\left(-\frac{\sqrt{3}|t - t'|}{2l_{tm}}\right)
\end{aligned} \tag{3.3}$$

$$\mathcal{K}_m[(s, t), (s', t')] = \mathcal{K}_{sm}(s, s') \otimes \mathcal{K}_{tm}(t, t')$$

where $\mathcal{K}_e$ and $\mathcal{K}_m$ stand for the spatiotemporal correlations engaging exponential and Matérn 3/2 kernel with hyper-parameters $\{\sigma_e, l_{se}, l_{te}\}$ and $\{\sigma_m, l_{sm}, l_{tm}\}$, respectively. The hyper-parameters can be estimated by the maximum likelihood method (Rasmussen (2003)). $d(s, s')$ denotes the distance between two locations to capture the spatial correlation.

We further define $\mathbf{S} = [s_1 \dots s_P]^T$, $\mathbf{t} = [t_1 \dots t_Q]^T$, and $\mathbf{u}_{1:Q \times P} = [u(s_i, t)]|_{s_i \in \mathbf{S}, t \in \mathbf{t}}$, which is the observed electric potentials at space-time location set $\{\mathbf{S}, \mathbf{t}\}$. Given the observations $\{(\mathbf{S}, \mathbf{t}), \mathbf{u}_{1:Q \times P}\}$, the predictive posterior mean of $u(s_*, t_*)$ at an arbitrary location (s_*, t_*) can be derived as $\mu_u(s_*, t_*) = \mathcal{K}_u^* \Sigma^{-1} \mathbf{u}_{1:Q \times P}$, where $\Sigma = \mathcal{K}[(\mathbf{S}, \mathbf{t}), (\mathbf{S}, \mathbf{t})] + \sigma_\delta^2 \mathcal{I}_{QP}$, $\mathcal{K}_u^* = \mathcal{K}[(s_*, t_*), (\mathbf{S}, \mathbf{t})] = \mathcal{K}_s^* \otimes \mathcal{K}_t^*$, and $\mathcal{I}_{QP}$ is an identity matrix with dimension of $(P * Q) \times (P * Q)$.

In order to increase the computational efficiency in calculating the precision matrix Σ^{-1} , we first compute the singular value decomposition (SVD) of the spatial and temporal covariance matrices as $\mathcal{K}_s = U_s \Lambda_s U_s^T$ and $\mathcal{K}_t = U_t \Lambda_t U_t^T$. Then, the SVD of Σ is given as

$$\Sigma = (U_s \otimes U_t)(\Lambda_s \otimes \Lambda_t + \sigma_\delta^2 \mathcal{I}_{QP})(U_s \otimes U_t)^T \equiv (U_s \otimes U_t)\Lambda(U_s \otimes U_t)^T \tag{3.4}$$

where $\Lambda \equiv \Lambda_s \otimes \Lambda_t + \sigma_\delta^2 \mathcal{I}_{QP}$ is a diagonal matrix. The predicted posterior mean $\mu_u(\mathbf{s}_*, t_*)$ can be reformulated in the form of Kronecker product:

$$\begin{aligned} \mu_u(\mathbf{s}_*, t_*) &= (\mathcal{K}_s^* \otimes \mathcal{K}_t^*) \Sigma^{-1} \mathbf{u}_{1:Q \times P} = (\mathcal{K}_s^* \otimes \mathcal{K}_t^*) (U_s \otimes U_t) \Lambda^{-1} (U_s \otimes U_t)^T \mathbf{u}_{1:Q \times P} \\ &= (\mathcal{K}_s^* U_s \otimes \mathcal{K}_t^* U_t) \Lambda^{-1} \text{vec} \left(U_t^T \mathbf{u}_{Q \times P} U_s \right) \equiv (\mathcal{K}_s^* U_s \otimes \mathcal{K}_t^* U_t) \mathbf{u}'_{1:Q \times P} \end{aligned} \quad (3.5)$$

where $\mathbf{u}_{1:Q \times P} = \text{vec}(\mathbf{u}_{Q \times P})$, and $\mathbf{u}'_{1:Q \times P} = \Lambda^{-1} (U_s \otimes U_t)^T \mathbf{u}_{1:Q \times P}$, which is a vector with the same length as $\mathbf{u}_{1:Q \times P}$. Note that $\text{vec}(\mathbf{A})$ is the vectorization of the matrix $\mathbf{A}$ by stacking the columns of $\mathbf{A}$ into a single column vector. Additionally, we define $\mathbf{u}'_{Q \times P} = \text{vec}^{-1}(\mathbf{u}'_{1:Q \times P})$, where $\text{vec}^{-1}(\cdot)$ denotes the inverse operation of $\text{vec}(\cdot)$. Then, the posterior mean at $(\mathbf{s}_*, t_*)$ from Eq. (3.5) is given as

$$\mu_u(\mathbf{s}_*, t_*) = \text{vec} \left(\mathcal{K}_t^* U_t \mathbf{u}'_{Q \times P} U_s^T \mathcal{K}_s^{*T} \right) \quad (3.6)$$

Taking advantage of the Kronecker property and exploiting the structure of the kernel function, we can increase the computational efficiency of the STGP modeling by manipulating much smaller matrices with dimensions of $P \times P$ and $Q \times Q$. However, the prediction performance of STGP modeling relies heavily on the amount of clinically-measured data, which is usually sparse due to the invasive nature of accessing intracardiac signals. Moreover, the traditional GP method only captures the data correlation while ignoring the underlying physics-based principles, leading to unsatisfactory prediction performance. Therefore, a more comprehensive method needs to be established by both exploiting the data structure and respecting the underlying physics laws simultaneously.

3.2.2 *Physics-constrained Deep Active Learning (P-DAL)*

Fig. D.11 presents the proposed P-DAL modeling framework. We first adapt the P-DL framework proposed in (Xie and Yao (2022b)) to model the spatiotemporal cardiac electrodynamics from sparse sensor signals. Note that the cardiac electrophysiological model is detailed in Chapter 2.3.1. Second, we develop a novel active learning strategy that combines the prediction uncertainty of the P-DL and space-filling design over the heart

geometry to seek the informative spatial locations for subsequent sensor measurements and increase the predictive power of the P-DL predictive modeling (Xie and Yao, 2022a).

3.2.2.1 Physics-constrained Deep Learning (P-DL)

The P-DL framework is adapted from (Xie and Yao (2022b)). Specifically, the spatiotemporal cardiac electrodynamics described by $u(\mathbf{s}, t)$ and $v(\mathbf{s}, t)$ is modeled using a DNN that is trained to not only satisfactorily match the sensor measurements but also respect the underlying physics law. We model the spatiotemporal cardiac potential distribution as:

$$[\mathbf{s}, t] \xrightarrow{\mathcal{N}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})} [u(\mathbf{s}, t), v(\mathbf{s}, t)] \quad (3.7)$$

where $\mathcal{N}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$ represents the DNN model, $\boldsymbol{\theta}_{NN}$ stands for the DNN hyperparameters. The DNN is constructed with an input layer consisting of spatiotemporal coordinates $[\mathbf{s}, t]$, multiple hidden layers to perform nonlinear transformations, and one output layer for the prediction of $\hat{u}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$ and $\hat{v}(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$. The physics principles are further encoded into the DNN along with the data constraint through a unique loss function defined as

$$\mathcal{L}(\boldsymbol{\theta}_{NN}) = \mathcal{L}_D + w \cdot \mathcal{L}_{ph} \quad (3.8)$$

where w is a parameter to control the extent of constraint effect imposed by the physics principles, which can be selected according to GP upper-confidence bound algorithm. The total loss $\mathcal{L}(\boldsymbol{\theta}_{NN})$ is composed of the following two critical components:

(1) **Data-driven loss $\mathcal{L}_D$:** Electrical signals in the heart can be collected by recording the electric potential over time using the electrode distals at the end of the catheter that directly contact the epicardium or endocardium (Markowitz and Lerman (2006)). As such, each electrode fetches a time series signal of electric potentials at a specific location in the heart, denoted as $\mathbf{u}_m(\mathbf{s}, t)$. The predictive model needs to be trained to generate predictions that match the sensor measurements closely. Hence, the data-driven loss $\mathcal{L}_D$, incurred from

the discrepancy between the measured and predicted signals, is defined as:

$$\mathcal{L}_D = \frac{1}{N_m} \sum_{i=1}^{N_m} (\mathbf{u}_m(\mathbf{s}_i, t_i) - \hat{\mathbf{u}}_m(\mathbf{s}_i, t_i))^2 \quad (3.9)$$

where N_m denotes the number of spatiotemporal measurements.

(2) **Physics-based loss $\mathcal{L}_{ph}$** : To further enhance the model robustness, physics-based regularization is imposed over the spatiotemporal collocation points, $[\mathbf{s}_i, t_i]$'s, which are randomly sampled from the spatiotemporal domain. The physics constraints consist of two components: the boundary condition and the cardiac electrophysiological model (see Eq. (2.3)-(2.5)), both of which are enforced by encouraging the physics-based residuals to be close to zero. Specifically, the boundary condition-based residual is defined as:

$$rs_{bc}(\mathbf{s}, t; \boldsymbol{\theta}_{NN}) := \mathbf{n} \cdot \nabla \hat{\mathbf{u}}(\mathbf{s}, t; \boldsymbol{\theta}_{NN}), \quad \mathbf{s} \in \Gamma_H \quad (3.10)$$

The boundary condition-based loss is then defined as:

$$\mathcal{L}_{bc} = \frac{1}{N_{bc}} \sum_{i=1}^{N_{bc}} (rs_{bc}(\mathbf{s}_i, t_i; \boldsymbol{\theta}_{NN}))^2, \quad \mathbf{s}_i \in \Gamma_H \quad (3.11)$$

where N_{bc} is the total number of collocation points chosen from Γ_H to enforce the boundary condition. Similarly, the electrophysiological model-based residuals are defined as:

$$\begin{aligned} rs_u(\mathbf{s}, t; \boldsymbol{\theta}_{NN}) &:= \frac{\partial \hat{u}}{\partial t} - \nabla \cdot (D \nabla \hat{u}) - k \hat{u}(\hat{u} - a)(1 - \hat{u}) + \hat{u} \hat{v} \\ rs_v(\mathbf{s}, t; \boldsymbol{\theta}_{NN}) &:= \frac{\partial \hat{v}}{\partial t} - \xi(\hat{u}, \hat{v})(-\hat{v} - k \hat{u}(\hat{u} - a - 1)) \end{aligned} \quad (3.12)$$

The physiological-based constraint will then be realized by encouraging the values of both $rs_u(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$ and $rs_v(\mathbf{s}, t; \boldsymbol{\theta}_{NN})$ to be close to zero. Thus, we define the electrophysiological model-based loss as:

$$\mathcal{L}_f = \frac{1}{N_f} \sum_{i=1}^{N_f} \left[(rs_u(\mathbf{s}_i, t_i; \boldsymbol{\theta}_{NN}))^2 + (rs_v(\mathbf{s}_i, t_i; \boldsymbol{\theta}_{NN}))^2 \right] \quad (3.13)$$

where N_f represents the total number of selected spatiotemporal collocation points to imprint the AP model into the DNN. Note that the partial derivatives of $u(\mathbf{s}, t)$ and $v(\mathbf{s}, t)$ with respect to the spatiotemporal variables $\mathbf{s}$ and t in $\mathcal{L}_f$ can be easily computed using automatic differentiation, which is a widely used technique in deep learning and is more efficient, accurate, and reliable than numerical differentiation used in FEM (Paszke et al. (2017)). The physics-based loss is then defined to integrate the effect from both the boundary condition and the AP model, i.e., $\mathcal{L}_{ph} = \mathcal{L}_{bc} + \mathcal{L}_f$, which is further embedded into the overall loss function in Eq. (3.8) to respect both the data-driven loss and the underlying system physics for the reliable modeling of the spatiotemporal cardiac electrodynamics.

3.2.2.2 *Active Learning in P-DL Predictive Modeling*

The predictive power of the P-DL approach depends heavily on the quality and volume of the collected training data. During cardiac catheterization, only a limited number of electrodes can be transported to the heart surface to collect the intracardiac signals from the corresponding limited locations. The whole mapping of cardiac electrodynamics is then estimated based on the collected time-series data from the sparse spatial locations. In order to further reduce the predictive error, we develop a novel active learning strategy by integrating the neural network dropout with the space-filling design to collect the most informative data for model training.

(1) **Dropout for uncertainty estimation in deep learning:** Although deep learning has been widely utilized in approximating any complex functional relationship, standard DNNs are limited in the ability to capture the prediction uncertainty. Traditional methods for uncertainty quantification such as Markov Chain Monte Carlo will be computationally expensive for large-scale P-DL modeling due to the vast computational efforts in DNN training. Fortunately, the dropout technique can serve as a convenient and effective way to quantify uncertainty in deep learning. DNN with dropout has been proved to be mathematically equivalent to the Bayesian deep GP (Gal and Ghahramani (2016)), which has wide applications for uncertainty quantification in various fields (Kim and Liu (2020); Miyato et al. (2018); Althoff et al. (2021); Abdar et al. (2021)). Here, we propose to engage the dropout technique to quantify the prediction uncertainty in our P-DL model.

The critical idea of dropout is to drop a subset of DNN parameters (i.e., set the value of those parameters to zero) randomly with a pre-defined probability $p \in (0, 1)$. The dropout loss function $\mathcal{L}_{\text{dropout}}$ is defined as the summation of original total loss of the P-DL model and a L_2 regularization on the DNN hyperparameters (Abdar et al. (2021)):

$$\mathcal{L}_{\text{dropout}} = \mathcal{L}(\theta_{NN}) + \frac{1-p}{2N} \sum_{i=1}^l \theta_{NN_i}^2 \quad (3.14)$$

where N denotes the total number of training samples, l is the number of neural network hidden layers, and θ_{NN_i} is the weights and bias for the i th layer. The units from each layer of the P-DL (excluding the input and output layers) are dropped independently and randomly with a fixed dropout rate p during the network training.

Once the training is completed, the dropout technique further provides a stochastic prediction for each testing spatial location by randomly dropping units of the P-DL network during the inference phase. As such, we can compute the standard deviation of each spatial point on the heart surface by simply running the stochastic forward propagation of the P-DL for M times, and further empirically estimate the prediction uncertainty $\hat{\sigma}(\mathbf{s})$ at an arbitrary spatial location $\mathbf{s}$ as:

$$\hat{\sigma}(\mathbf{s}) = \frac{1}{t_N} \sum_{j=1}^{t_N} \sqrt{\frac{\sum_{i=1}^M (\hat{u}_i(\mathbf{s}, t_j) - \bar{\hat{u}}(\mathbf{s}, t_j))^2}{M-1}} \quad (3.15)$$

where $\hat{u}_i(\mathbf{s}, t_j)$ denotes the estimation for the i th prediction at time t_j , $\bar{\hat{u}}(\mathbf{s}, t_j) = \frac{\sum_{i=1}^M \hat{u}_i(\mathbf{s}, t_j)}{M}$ is the mean estimation, and t_N denotes the total number of discretized temporal points (i.e., time stamps) at the spatial location $\mathbf{s}$.

(2) **Geodesic distance-based space-filling design:** The DNN dropout enables the effective evaluation of the prediction uncertainty at each spatial location, which can be utilized for active learning. For example, the next sensor observation point can be selected as the spatial location with the maximum prediction uncertainty. However, uncertainty quantification alone may not be able to yield reliable next query points if the spatial locations with high estimated uncertainty tend to aggregate together at local regions on

the heart surface. This will potentially result in an inferior predictive performance. In order to guarantee that the P-DL can predict the cardiac electrodynamics globally well, we further involve the maximin-distance design, a widely used space-filling scheme for the optimal design of computer experiments, into the active learning framework.

In the traditional sequential learning procedure based on pure space-filling design, the next query point $\mathbf{s}_{n+1}$ is selected according to the max-min criterion

$$\mathbf{s}_{n+1} = \arg \max_{\mathbf{s}} \min_{i \in \{1, 2, \dots, n\}} \text{dist}(\mathbf{s}, \mathbf{s}_i) \quad (3.16)$$

which produces the next point $\mathbf{s}_{n+1}$ that maximizes the minimum distance between $\mathbf{s}_{n+1}$ and existing observation locations $\mathbf{s}_i$'s, $i \in \{1, \dots, n\}$. The distance function $\text{dist}(\cdot)$ in Eq. (3.16) is selected as the geodesic metric d_g to preserve geometric information of the heart for reliable cardiac modeling, i.e., $\mathbf{s}_{n+1} = \arg \max_{\mathbf{s}} \min_{i \in \{1, 2, \dots, n\}} d_g(\mathbf{s}, \mathbf{s}_i)$, which will then be integrated with the prediction uncertainty of P-DL to form the active learning criterion.

It should be noted that the geodesic distance d_g is able to effectively measure the distance between two points on the heart geometry, which cannot be achieved by Euclidean distance. This is due to the fact that the Euclidean metric measures the distance of the straight line connecting two nodes and ignores the topology of the geometry (Chen and Yang (2016)). Instead, we choose the geodesic distance metric, which is defined as the length of the shortest path connecting two nodes on the geometry, as illustrated by Fig. D.12. In order to calculate the geodesic distance matrix, a k-nearest neighbor graph is first constructed by connecting each location $\mathbf{s}_i$ to its k_G nearest neighbors $\mathbf{s}_{j, i \leftrightarrow j}$ (where $i \leftrightarrow j$ means node j is one of the k_G nearest neighbors for node i) on the heart geometry. Second, the distance between each pair of nodes on the graph is initialized as: $d_G(i, j) = \|\mathbf{s}_i - \mathbf{s}_j\|$ if i and j are connected; $d_G(i, j) = \infty$, otherwise. Third, Dijkstra's algorithm (Dijkstra et al. (1959)) is implemented to calculate the shortest path between each pair of the nodes in the k-nearest neighbor graph, which provides the pairwise geodesic distance $d_g(\mathbf{s}_i, \mathbf{s}_j)$ across the heart geometry.

(3) **Active learning criterion:** An effective active learning (AL) criterion should account for both the estimation uncertainty and the spreadout degree of the selected observation locations. To achieve this goal, we propose a sequential design criterion by

incorporating both the dropout uncertainty measure and the geodesic distance-based max-min design as:

$$\mathbf{s}_{n+1} = \operatorname{argmax}_{\mathbf{s}} \left\{ \frac{\hat{\sigma}(\mathbf{s})}{\max_{\mathbf{s}} \hat{\sigma}(\mathbf{s})} + \lambda \cdot \frac{\min_{i \in \{1, \dots, n\}} d_g(\mathbf{s} - \mathbf{s}_i)}{\max_{\mathbf{s}} \min_{i \in \{1, \dots, n\}} d_g(\mathbf{s} - \mathbf{s}_i)} \right\} \quad (3.17)$$

where $\min_{i \in \{1, \dots, n\}} d_g(\mathbf{s} - \mathbf{s}_i)$ represents the minimum geodesic distance between the unmeasured location $\mathbf{s}$ and the measured ones $\mathbf{s}_i$'s, $i \in \{1, \dots, n\}$. The upper bound of the minimum geodesic distance is denoted as $\max_{\mathbf{s}} \min_{i \in \{1, \dots, n\}} d_g(\mathbf{s} - \mathbf{s}_i)$, which is used to normalize the distance criterion. On the other hand, $\max_{\mathbf{s}} \hat{\sigma}(\mathbf{s})$ stands for the maximum prediction uncertainty over all the unmeasured spatial locations on the heart geometry to normalize the uncertainty criterion. Parameter $\lambda > 0$ is added to balance between the effects of uncertainty estimation and space-filling design. The AL criterion in Eq. (3.17) is expected to collect the most informative training points to not only identify the high uncertainty region but also allow the P-DL to approximate the cardiac electrodynamics globally well.

During the active learning procedure, a small number of spatial points are initially randomly chosen from the heart surface to collect the intracardiac sensor observations. Those initial training points will then be utilized to train the proposed P-DL model. Given the trained P-DL, the AL criterion is further evaluated to select the next h (≥ 1) query locations. The electric potential $\mathbf{u}(\mathbf{s}, t)$ at the chosen spatial locations will be measured and the corresponding time series data $\mathbf{u}_m(\mathbf{s}, t)$ will be added into the training set to re-train the P-DL. This active learning procedure will be iteratively evaluated until the budget for computation or physical data collection is used up.

3.3 Model Validation

We evaluate the performance of the proposed P-DAL framework in modeling both the healthy and diseased ventricular cardiac electrodynamics in a 3D geometry. The modeling performance of P-DL will be compared with the exponential kernel-based STGP (STGP-E) and Matérn 3/2 kernel-based STGP (STGP-M), and the efficacy of the proposed

active learning criterion will be benchmarked with non-informative random searching and pure space-filling-based active learning (i.e., max-min distance criterion as shown in Eq. (3.16)). The heart geometry (i.e., the ventricles) is formed by 1094 discretized nodes and 2184 mesh elements, which is a refined mesh of the geometry obtained from the 2007 PhysioNet Computing in Cardiology Challenge (Goldberger et al. (2000)). The data of ventricular electrodynamics under the healthy condition is simulated by placing a regular-pacing activation source at the apex of the heart geometry (Trayanova (2011); Clayton and Panfilov (2008)), while the diseased data is simulated by adding an additional activation source on the right ventricle which emits electrical impulse more frequently to imitate the cardiac electrical activities under fibrillation (Haissaguerre et al. (1998); Ten Tusscher et al. (2009)).

Note that measurement noise is inevitable in clinical practice. Thus, in this paper, we add different noise levels into the simulation data (denoted as $\mathbf{u}(\mathbf{s}, t)$) to investigate the prediction performance yielded by different models (i.e., STGP-E, STGP-M, P-DL). In other words, the physical measurements are denoted as $\mathbf{u}_m(\mathbf{s}, t) = \mathbf{u}(\mathbf{s}, t) + \epsilon(\mathbf{s}, t)$. We assume that the noise follows a Gaussian distribution as $\epsilon(\mathbf{s}, t) \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The model performance will be evaluated according to the Relative Error (RE), which is defined as

$$RE = \frac{\sqrt{\sum_{s,t} \|\hat{\mathbf{u}}(\mathbf{s}, t) - \mathbf{u}(\mathbf{s}, t)\|^2}}{\sqrt{\sum_{s,t} \|\mathbf{u}(\mathbf{s}, t)\|^2}} \quad (3.18)$$

where $\mathbf{u}(\mathbf{s}, t)$ and $\hat{\mathbf{u}}(\mathbf{s}, t)$ denotes the reference and predicted intracardiac signals, respectively.

3.3.1 *Healthy heart system*

In the present investigation, a feedforward fully-connected DNN is constructed by incorporating the physics law in the P-DL approach to model the cardiac electrodynamics under the healthy condition. The DNN consists of five hidden layers with ten neurons in each layer. The regularization parameter w in Eq. (3.8) is selected as 0.95. (Note that selection of neural network structure and w is described in Appendices B.1.1). Furthermore, we randomly select $N_{ph} = 50,000$ collocation points from cardiac spatiotemporal domain to

encode physics law into the DNN. It is also assumed that 100 time series of cardiac electric potentials are randomly collected, which serve as the measured data, i.e., $\mathbf{u}_m(\mathbf{s}, t)$, to train the predictive models.

Fig. D.13 demonstrates the comparison of RE generated from the three methods, i.e., STGP-E, STGP-M, and P-DL, at different noise levels of $\sigma_\epsilon = 0.01, 0.05$, and 0.1 . As the noise level increases from $\sigma_\epsilon = 0.01$ to $\sigma_\epsilon = 0.1$, RE monotonically rises for all the approaches. Specifically, RE increases from 0.1618 to 0.2007 for STGP-E, from 0.1568 to 0.2117 for STGP-M, from 0.0811 to 0.0849 for P-DL model. It is notable that the proposed P-DL method yields the lowest RE (achieving at least 48.3% reduction in RE) compared to STGP-E and STGP-M regardless of the noise level. Moreover, it is worth mentioning that the RE 's generated from STGP-E and STGP-M increase drastically when bigger noise is applied to the sensor measurements, with a 24.4% and 35.01% growth as the noise level increases from $\sigma_\epsilon = 0.01$ to $\sigma_\epsilon = 0.1$, respectively. In comparison, the proposed P-DL method displays only a 4.68% increase in RE as the noise level reaches $\sigma_\epsilon = 0.1$, demonstrating extraordinary anti-noise performance. Hence, the P-DL model not only achieves superior prediction accuracy, but also exhibits a robust model attribute when dealing with a high level of noise in the physical observations.

Fig. D.14(a) depicts the true electrical potential mapping on the healthy heart surface. Since the electric potential dynamically evolves overtime, Fig. D.14 captures the spatial distribution of the cardiac electrodynamics at one specific time instance: $t = 30$. Note that here the cardiac electrical potential u is normalized and both u and t are unitless. Fig. D.14(b-d) delineates the estimated mapping of cardiac electrodynamics, generated from the three methods with different noise levels, i.e., $\sigma_\epsilon = 0.01, 0.05$, and 0.1 , respectively. The predicted mappings produced by the proposed P-DL approach best preserve the color scale and patterns of the reference mapping as shown in Fig. D.14(a), while the predictions from STGP demonstrate pronounced deviations under all noise levels. The extraordinary predictive performance of the P-DL model owes to the successful integration of prior physics-based knowledge and data constraints collected from intracardiac sensors, presenting a robust prediction of the spatiotemporal cardiac electrodynamics.

We further conduct experiments to evaluate the efficacy of the proposed AL criterion. In the dropout DNN, a certain percentage from the neurons in the hidden layers is randomly dropped out with a dropout rate of $p = 0.025$ to evaluate the prediction uncertainty. With the well-trained dropout DNN, we run the forward propagation $M = 500$ times to create stochastic predictions and further acquire the empirical standard deviation $\hat{\sigma}(\mathbf{s})$ of the prediction at each unmeasured location on the heart geometry. The uncertainty measure $\hat{\sigma}(\mathbf{s})$ will then be combined with the max-min design to evaluate the AL criterion (i.e., Eq. (3.17)) and output the next query points for physical observation. Here, we select $\lambda = 2$ in the AL criterion. (Please find the procedure to select λ in Appendices B.1.2). In order to balance between the algorithm efficiency and the effectiveness of the active learning, after the network training with existing measurements is complete, we will sequentially evaluate the AL criterion 3 times, which means that 3 spatial locations will be selected for the next sensor exploration. It is worth mentioning that once one specific location is chosen, it will be excluded from the set of unmeasured spatial locations and automatically become one of the existing locations for the subsequent evaluation of the AL criterion. Initially, 20 spatial locations are randomly sampled from the heart geometry to train the initial version of the dropout P-DL. The active learning procedure will be evaluated for 10 rounds with 3 spatial locations actively selected for each round. In the end, time-series data from 50 spatial locations are collected for the final P-DL training.

Fig. D.15(a) compares the P-DL model performance trained by data collected from our proposed active learning procedure (P-DAL) with the pure space-filling selection strategy (P-DSL) and the non-informative random selection strategy (P-DRL) with a noise level of $\sigma_\epsilon = 0.01$. The *RE* estimated from our P-DAL approach demonstrates the fastest-decreasing trend in comparison with the *RE* from P-DSL and P-DRL. Specifically, the *RE* generated from the three methods start to grow apart from each other after 3 subsequential rounds of new measurements and the deviation becomes more obvious. When the number of the spatial locations reaches 50, the *RE* is reduced to 0.0876 by P-DAL, achieving a reduction in *RE* of 16.3% and 28.0% compared with P-DSL (0.1047) and P-DRL (0.1217), respectively. In other words, the P-DAL approach can reliably model the cardiac electrodynamics in the

entire geometry (with 1094 discretized spatial locations) at an estimation error of 0.0876 given the noisy sensor observations from only 50 spatial locations.

Fig. D.15(b) further shows the evolvement of electric potential over time for one specific location estimated by the P-DL based on training data from 50 spatial locations selected by P-DAL, P-DSL, and P-DRL, which are benchmarked by the true dynamic evolvement (i.e., the blue curve). All methods yield a fairly good prediction compared with the ground-truth physical observation. However, as shown by the zoom-in figure, the evolvement curve generated by P-DAL is almost overlapped with the physical observation whereas the curves estimated by P-DSL and P-DRL show a relatively more significant discrepancy. As such, the proposed P-DAL method is more effective to select influential sensor measurements to improve the performance in modeling spatiotemporal cardiac electrodynamics.

3.3.2 *Diseased heart system*

In the present investigation, we engage a network structure with five hidden layers each of which contains twenty neurons for P-DL modeling. The regularization parameter w in Eq. (3.8) is selected as 0.13. (Please find more details about the hyper-parameter selection in Appendices B.1.1). Additionally, we randomly select $N_{ph} = 200,000$ collocation points from the cardiac spatiotemporal domain to encode the physics laws into the DNN. 100 time series of the cardiac electric potentials are further randomly selected as the measured data to train the three predictive models (i.e., STGP-E, STGP-M, and P-DL). Fig. D.16 compares the RE yielded by the three methods at different noise levels of $\sigma_\epsilon = 0.01, 0.05$, and 0.1 in the diseased heart system. The estimated RE monotonically increases when the noise level increases from $\sigma_\epsilon = 0.01$ to $\sigma_\epsilon = 0.1$ for all the three methods, i.e., from 0.2694 to 0.3021 for STGP-E, from 0.2600 to 0.3152 for STGP-M, and from 0.1873 to 0.2154 for P-DL. It is worth noting that the proposed P-DL method yields the lowest RE (achieving at least 28.0% reduction in RE) compared to both STGP-E and STGP-M regardless of the noise level for the diseased heart system. As such, similar to the healthy heart system, the P-DL method demonstrates superior performance in spatiotemporal modeling of the diseased cardiac electrodynamics.

Fig. D.17 presents the spatial distribution of the diseased cardiac electrodynamics at the time point of $t = 33$. Fig. D.17(a) delineates the real spatial mapping. Fig. D.17(b-d) illustrates the estimated mappings from the three methods, given the measured data from 100 spatial locations with different noise levels, i.e., $\sigma_\epsilon = 0.01$, 0.05 , and 0.1 , respectively. The predicted spatial distributions generated from the P-DL model demonstrate the smallest discrepancy from the reference potential mapping (i.e., Fig. D.17(a)) compared with GP-based methods under each noise level. Even with a high noise level of $\sigma_\epsilon = 0.1$, the P-DL model still generates the smallest *RE* of 0.2154 for the diseased heart system, which is remarkably lower than the *RE*'s (i.e., 0.3021 and 0.3152) estimated by GP-based methods.

We further evaluate the proposed active learning procedure in fetching the most informative spatial locations for the sensor measurements in order to enhance the predictive accuracy. The value of λ for the diseased system is selected as 0.5. (Please find more details of the selection of λ in Appendices B.1.2.) Similar to the healthy heart system, we initially select 20 spatial locations and assume the maximum allowed number of locations for sensor exploration is 50. As such, the active learning procedure will be evaluated with 10 rounds and 3 spatial locations will be selected according to the AL criterion for each round. The same dropout rate of 0.025 is used in the dropout P-DL and the well-trained network will be run for $M = 500$ times to get stochastic predictions for uncertainty quantification.

Fig. D.18(a) compares the prediction performance of the P-DL model with training data collected from active learning (i.e., P-DAL), pure space-filling selection (i.e., P-DSL), and random selection (i.e., P-DRL) with a noise level of $\sigma_\epsilon = 0.01$. The P-DAL model exhibits the lowest *RE* in general compared with *RE* yielded by the P-DSL and the non-informative P-DRL. The deviation in the *RE* from the three methods starts to manifest after 4 rounds of new training data collection. When the total number of the spatial locations reaches 50, the *RE* decreases to 0.2518 for the P-DAL in comparison with the *RE* of 0.2832 (11.1% reduction) by P-DSL and 0.3194 (21.2% reduction) by P-DRL.

Fig. D.18(b) shows the estimated dynamic evolvement of cardiac electric potentials over time for one specific location using time series data from 50 spatial locations selected from P-DAL, P-DSL, and P-DRL. All methods generate relatively good predictions for the first heart cycle. Starting from the second cycle, the evolvement curves show non-negligible disparity

from the physical observation due to the sparsity of the training data (i.e., the intracardiac signals collected from only 50 spatial locations) and the complex dynamic behavior in the electrical activities of a diseased heart. Nonetheless, the evolvement curve estimated by the P-DAL is closer to the truth whereas the P-DSL and P-DRL-based curves exhibit a more significant deviation. This demonstrates that the proposed active learning strategy is able to effectively reduce the predicted *RE* by the P-DL trained with the cardiac signals collected from limited spatial locations, leading to a better prediction of the diseased cardiac electrodynamics.

3.3.3 Training time

It takes about 12 minutes to train the P-DL for the health heart system with training data from 50 spatial locations for constructing data-driven loss and 50,000 collocation points for calculating physics-based loss. For the diseased heart system, with 50 spatial locations and 200,000 collocation points, the training time is about 1 hour and 10 minutes. The computational time is summarized in Table E.5. The algorithm is processed on the “Pete” supercomputer of Oklahoma State University, which has 10 GPU nodes, each containing dual NVIDIA Quadro RTX6000 GPUS for a total of 20 GPUs.

3.4 Discussion

3.4.1 The main differences in methodological contributions between the proposed P-DAL framework and the P-DL developed in Chapter 2

(1) **Different research problems:** The framework developed in (Xie and Yao (2022b)) focused on the inverse ECG problem. Specifically, Xie and Yao (2022b) proposed a P-DL framework for robust inverse ECG modeling to estimate the heart surface potential mapping from the body surface sensor measurements, which mainly focused on addressing the ill-posed issue of the inverse ECG problem. The proposed P-DAL framework here focuses

on estimating heart electrical signals from the intra-cardiac signals directly instead of from the body-surface signals. The intra-cardiac signals are with low resolution in the spatial domain due to the invasive nature of the data collection procedure, and we mainly focus on addressing the sparseness issue of the data.

(2) **Data-driven loss in P-DL:** The data-driven loss in (Xie and Yao (2022b)) is defined as the transformation loss from the body-surface measurements to the heart-surface signals, i.e., $\|\mathbf{y}(\mathbf{s}, t) - \mathbf{R}_{BH}\mathbf{u}(\mathbf{s}, t)\|$, where $\mathbf{y}(\mathbf{s}, t)$ denotes the body-surface data, $\mathbf{u}(\mathbf{s}, t)$ denotes the unknown heart-surface electrical potential, and $\mathbf{R}_{BH}$ is the transfer matrix between $\mathbf{y}(\mathbf{s}, t)$ and $\mathbf{u}(\mathbf{s}, t)$. On the other hand, the data-driven loss in the P-DAL framework is derived from the difference between the estimated heart-surface signals and the true intra-cardiac measurements, i.e., $\|\hat{\mathbf{u}}_m(\mathbf{s}, t) - \mathbf{u}_m(\mathbf{s}, t)\|$.

(3) **Active learning:** In the P-DAL framework, we propose an active learning strategy for data collection to further cope with the sparseness issue of intracardiac signals, which is another new contribution compared with the P-DL framework in (Xie and Yao (2022b)). Specifically, the active learning criterion consists of two key components:

- *Uncertainty Quantification:* We investigate the uncertainty quantification in the predictive modeling through the neural network dropout technique, which will quantify the prediction uncertainty and guide the active searching for the next sensor location.
- *Geodesic Distance-based Space-filling Design:* In order to guarantee the prediction method to estimate the heart signals globally well, we further combine the uncertainty quantification with space-filling design to actively search for the subsequent measurement locations in the proposed P-DAL framework. Note that the space-filling design is based on geodesic distance instead of simple Euclidean distance to preserve the geometric information of the heart.

3.4.2 Advantages of the P-DAL approach in modeling the spatiotemporal cardiac systems

The advantage of the P-DAL approach in modeling cardiac electrodynamics based on sparse sensor measurements can be summarized as follows:

(1) **The capability of combining both sensor measurements and physics prior knowledge into DNN:** Traditional FEM-based methods can be engaged to numerically solve cardiac electrophysiological models to explicate the dynamic distribution of heart surface potential (Heidenreich et al. (2010), Sundnes et al. (2006), McNerney and Terzopoulos (1995)). However, FEM-based simulation only provides the theoretical estimation of cardiac electrodynamics without considering clinical sensor observations, which makes it ill-suited for real-world application. GP methods can be utilized to interpolate cardiac electrodynamics based on clinical measurements (Hu et al. (2020), Dhamala et al. (2018)). However, the GP-based cardiac modeling demands sufficient training data to generate a trustworthy result. Moreover, GP modeling involves the computation of inversion and determination of a large kernel matrix in detailed cardiac modeling, which is deemed to be computationally expensive. In this study, we engage spatiotemporal GP and exploit the Kronecker product property to deal with the computational issue. However, the predictive results are still non-ideal for precision cardiology because the GP-based method only explores the data correlation while neglecting the underlying physics rules. Compared with traditional FEM and GP modeling, the proposed P-DAL model is capable of easily integrating both sensor observations and physics principles. The clinical sensor measurements can be easily incorporated into deep learning as the data-driven loss. Moreover, the physics-based cardiac model (i.e., the AP model) is encoded into the DNN as a spatiotemporal physical regularization represented by the physics-based loss. As such, the P-DAL method can produce a better prediction of cardiac electrodynamics compared with the spatiotemporal GP, as shown in Figs. D.14-D.15 and D.17-D.18 for both heart systems.

(2) **The capability of integrating uncertainty quantification with space-filling design for active learning:** Due to the clinical difficulty and invasive nature of cardiac catheterization, only a small number of sensors can be placed on the heart surface for data collection. In order to take full advantage of the limited sensor measurements and boost the predictive strength of the P-DL model, an effective active learning strategy is needed to judiciously select the most informative spatial locations on the heart surface with a limited number of sensor placements. The proposed P-DAL method engages dual selective criteria which fuse the uncertainty quantification provided by the network dropout technique and

geodesic distance-based space-filling design to enhance the effectiveness of the spatial location selection. Numerical results show that *RE* generated from the proposed P-DAL is about 16.3% and 28.0% smaller than from the P-DSL and P-DRL methods in the healthy heart system, respectively. Similarly, for the diseased heart condition, *RE* produced by the P-DAL scheme is 11.1% and 21.2% lower compared with the P-DSL and P-DRL methods, respectively. Please also refer to Fig. [D.15](#) and [D.18](#) for more detailed results.

3.5 Conclusions

Heart disease has been consistently ranked among the leading causes of death in the US. To better facilitate disease diagnosis and treatment decision-making, it is of great importance to perceive high-resolution spatiotemporal evolvement of the cardiac electrical signals. However, the current medical procedure (e.g., catheterization) can only distribute limited sensors onto the heart surface, generating sparse sensor observations, which significantly challenges traditional data-driven methods in reconstructing the cardiac electrodynamics. This paper presents a physics-constrained deep active learning (P-DAL) strategy to model the spatiotemporal cardiac electrodynamics from sparse intracardiac sensor measurements. This approach encodes the underlying physics that governs the propagation of cardiac electrical waves into deep learning for robust predictive modeling of the electrical behavior of the heart. In addition, we develop a novel active learning strategy that combines both the prediction uncertainty and the geodesic distance-based space-filling scheme to seek the informative spatial locations on the heart surface that are capable of potentially providing a more comprehensive picture about the heart electrical activity.

We evaluate the proposed framework in both healthy and diseased heart systems. Experimental results show that the P-DL method achieves up to 48.3% and 28.0% reduction in the estimated Relative Error (RE) compared with that from the traditional STGP models in the healthy and diseased heart systems, respectively. We also demonstrate the efficacy of the proposed active learning procedure by comparing it with traditional learning strategies. Specifically, the RE generated from the proposed P-DAL achieves 16.3% and 28.0% (11.1% and 21.2%) reduction compared with REs generated from P-DSL and P-DRL in the healthy

(diseased) heart system, respectively. The proposed P-DAL framework not only has a great potential to realize precision cardiology for optimal heart disease diagnosis and treatment plan, but also is able to be applied in other spatiotemporal dynamic systems such as space-time progression in tumor growth, thermodynamics in advanced manufacturing, and spatiotemporal evolution of water content in agriculture.

One limitation of this work is that we have only evaluated our method with synthetic data. Additional challenges might occur when implementing our method to real clinical data. One possible challenge is the selection of the physics-based prior model. Although the Aliev-Panfilov model has been widely recognized to be effective to describe the propagation principle of cardiac electrical waves, the model discrepancy may still exist compared with reality. However, we want to first highlight that the proposed framework can be easily adapted to different cardiac models including the FitzHugh-Nagumo model (Rocsoreanu et al. (2012)) and the Courtemanche-Ramirez-Nattel model (Courtemanche et al. (1998)). Additionally, we want to emphasize that our framework doesn't require the physics-based model to be perfectly accurate. The physics-based model only serves as the regularization for the network training. We expect the model discrepancy in the physics-based principle will be mitigated by the real-world sensing data. In other words, we expect the physics-based loss and data-driven loss will interact with each other, making the proposed P-DL and P-DAL models robust against both model discrepancy and measurement noise. As a future step, we will test the proposed methodology with real patient data before fully establishing the utility of the proposed framework in clinical applications.

Chapter 4

Hierarchical Active Learning for Defect Localization in 3D Systems

4.1 Introduction

Tremendous advancements in modern sensing techniques have enormously benefited the data-driven modeling of complex systems in various fields, such as mechanical engineering (Rao et al., 2021a,b), biomedical engineering (Xie and Yao, 2022b; Yao, 2021; Dhamala et al., 2017; Wang et al., 2010), and manufacturing (Li et al., 2020; Kang et al., 2021). Multitudinous industrial applications have evolved from the data-rich environment, among which, anomaly or defect detection features as one of the most important issues to ensure a system to proceed without a glitch. For example, reliable defect detection is of key importance for quality assurance in advanced manufacturing and there have been increasing interests in data-driven defect identification in the 3D parts (Thompson et al., 2015; Yao et al., 2018). Additionally, a variety of data-driven disease detection tools have been proposed to advance smart health (Van Wyk et al., 2019; Wang and Yao, 2021; Chen et al., 2022). However, the above mentioned data-driven methods focus on interpreting the system conditions binarily, i.e., answering the Yes (abnormal) or No (normal) questions, which are limited in the ability to provide detailed characterizations of the defect states.

Rapid developments in advanced imaging such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) substantially facilitate defect localization by granting us to non-intrusively access the inner structures of complex systems, providing direct information of defect states (Lian et al., 2018; Wang et al., 2021b; Lahoti et al., 2021). Extensive research has been done to promote automated identification of the lesion region for various diseases. For example, Yan et al. (2018) proposed a weakly supervised deep learning method to classify thoracic disease and localize the suspicious lesion based on chest X-rays. Roy et al. (2020) analyzed lung ultrasound images for the diagnosis of COVID-19 pathologies and proposed to localize pathological artifacts through spatial transformer networks. Additionally, computational modeling also flourishes in defect localization in infrastructure systems and manufacturing. For example, Yin et al. (2020) engaged a CNN-based object detector to identify the sewer pipe damages and their locations using massive amounts of video footage from an underground drainage system. Lin et al. (2019) build a

CNN to localize defect regions of light-emitting diode chips without using region-level human annotations.

The aforementioned approaches for defect localization are provided with direct visual information such as images or videos. However, there exist other types of defects such as flaws in cantilever beam (Johannesson et al., 2005) or early-stage diseased tissue in the human heart (Giffard-Roisin et al., 2016), which are not visually detectable and cannot be directly captured by existing imaging techniques. One typical example is the detection and localization of myocardial infarction (MI). MI is a deadly cardiac disease caused by necrosis of the myocardial tissue, which is usually diagnosed by a multi-lead electrocardiogram (ECG) (Wang et al., 2023). Different computational models have been developed for MI localizations using ECG data, such as deep learning (Tripathy et al., 2019), multi-class support vector machine (SVM) (Sharma et al., 2015), and k-nearest neighbor (KNN) classifier (Arif et al., 2012). However, most existing methods reduce the model complexity by segmenting the heart geometry into several predefined partitioned regions, so that the disease localization can be converted into a classification problem. This reformation considerably affects the prediction accuracy of the disease area because the real-world defect region generally appears in an irregular shape and may cross multiple pre-defined segments, which can greatly hinder the precision treatment planning.

Even though early-stage defects generally cannot be directly captured by existing imaging techniques, such changes in the physical property of complex systems can be reflected in certain observable and quantifiable features (e.g., ECG signals for MI detection) (Dhamala et al., 2017; Zhao et al., 2022), which can be integrated with the underlying system physics to develop personalized simulation modeling for defect localization. In fact, simulation-based model calibration serves as a vital component of the inverse problem to develop personalized computational models and output reliable defect locations. For example, Bousselham et al. (2018) developed a genetic algorithm to estimate the brain tumor location by minimizing the discrepancy between measured signals on the brain surface and those obtained by numerical simulation. Olson and Throne (2010) developed an eigenvector expansion technique to identify the breast tumor location using the tissue stiffness distribution in a finite element model of the breast.

Simulation-based defect localization involves modeling the heterogeneous physical systems and calibrating spatial-varying model parameters. Estimating spatial-varying parameters is a challenging problem in terms of computational burden posed by the large vector of unknowns in the parametric space and a prodigious number of calls to the simulation models (Koutsourelakis, 2009). In existing literature, different hierarchical Bayesian-based optimization techniques have been proposed to solve the computational issue (Koutsourelakis, 2009; Latz et al., 2018; Wan and Zabaras, 2011). Specifically, a hierarchical multiscale Bayesian model in the manner of non-parametric kernel regression is first developed to characterize the prior information of the parametric space, and a sequential or adaptive Monte Carlo technique is then involved to sample from a sequence of probability distributions to estimate the posterior. However, those methods are designed for 2D systems with regular geometric shapes, which are difficult to be implemented in 3D complexly structured systems.

In most real-world practice, defect localization calls upon personalized modeling of 3D complex systems (Chinchapatnam et al., 2008; Agathos et al., 2018). For example, Dhamala et al. (2017) developed a multi-scale optimization framework to adaptively estimate the excitability parameter of myocardium tissue on a 3D heart geometry from low to high spatial resolution for MI localization. However, this method requires iteratively solving a Bayesian optimization problem to identify a parameter vector with the dimensionality associated with the selected spatial resolutions. This will be considerably computational intensive with a very fine spatial resolution for accurate defect localization. Little has been done to develop an innovative computational approach that can handle the complex geometry of a 3D system and further effectively estimate the spatial-varying system parameter for reliable defect localization.

This paper presents a Hierarchical Gaussian Process-based Active Learning (HGPAL) framework for reliable defect localization in 3D systems. We first develop a GP regression model on the 2D embedding of a 3D geometry to capture the geometric information. Second, we propose a Hierarchical GP-based active learning approach composed of master GP and sub-GP modeling to account for the system heterogeneity and further estimate the spatial-varying model parameter for defect localization. We evaluate the performance of the

proposed HGPAL framework in a 3D body-heart system to identify and localize ischemia or infarct in the heart. Experimental results demonstrate the effectiveness of our HGPAL framework in 3D defect localization.

The remainder of the paper is organized as follows: Section 4.2 introduces related works on GP modeling and Bayesian optimization. Section 4.3 presents the problem formulation. Section 4.4 presents the details of the proposed HGPAL algorithm. Section 4.5 demonstrates the numerical implementation of our HGPAL approach to identify the defect location in a 3D body-heart system. Section 4.6 summarizes the present investigation.

4.2 Related Works

The Gaussian process (GP), a powerful and flexible probabilistic framework, is widely used for predictive modeling and uncertainty quantification. However, large-scale GP training requires calculating the inverse and determinant of a high-dimensional kernel matrix at each training iteration, incurring a significant burden on the computational effort. Different advanced techniques have been developed to cope with such computational complexity. For example, Wang et al. (2019c) relieved the computational constraint utilizing multi-GPU parallelization and linear conjugate gradients, which enabled large-scale GP training with million training points. Chen et al. (2020a) further provided a convergence guarantee for mini-batch stochastic gradient descent for large-scale GP training. While these methods indeed solve the scalability issue of GPs, it is important to note that they are designed for situations where a large number of observations are readily available. However, training data collection for simulation calibration involves repeatedly evaluating the simulation model, making it computationally infeasible to obtain numerous observations for expensive simulation modeling that involves a complex 3D geometry.

In addition, to improve the computational efficiency in high-dimensional predictive modeling, several hierarchical GP methods have been developed in the literature by exploiting the hierarchical structure of the problem. For example, Park and Choi (2010) proposed a Hierarchical GP Regression (HGPR) method to model complex data with an inherent hierarchical structure. They developed a two-layer GP model, where the lower-level

GPs capture local variations in the data, while the higher-level GP captures global trends. The training data needs to be partitioned into multiple subsets: each subset is modeled by an individual lower-level GP; prototypes of subsets are involved for coarse modeling by the high-level GP. Shi et al. (2005) proposed a hierarchical GP mixture regression model to simulate the standing-up maneuvers by assuming that different batches of the observations (i.e., standing-up behaviors) can be modeled by a mixture of GPs with different kernel parameter sets to address the heterogeneity among different batches. Lawrence and Moore (2007) developed a hierarchical GP latent variable model to study human motion data. They first projected the dynamic motion data into a low-dimensional manifold, and then built GPs on the manifold space and the temporal domain in a hierarchical way. Datta et al. (2016) developed a hierarchical nearest-neighbor GP model for handling large geostatistical datasets. However, most existing hierarchical GPs focus on exploiting the hierarchical/clustered structure in the input data and further realizing accurate prediction at an arbitrary input. They also generally assume the training data is readily available and cannot be directly used for optimizing expensive black-box functions.

In recent decades, Bayesian optimization has emerged as a powerful technique for the efficient optimization of expensive black-box functions. By using probabilistic surrogate models such as GPs, Bayesian optimization efficiently navigates the search space to identify the optimal solution while minimizing the number of calls to computationally demanding simulation models (Srinivas et al., 2009; Brochu et al., 2010). Bayesian optimization has been applied to solve a variety of inverse problems (Huang et al., 2021; Deng et al., 2020; Winter et al., 2023). Researchers have also extended the standard Bayesian optimization framework to accommodate multi-task problems, incorporating various surrogate structures to better represent the relationships among different tasks, datasets, and functions (Swersky et al., 2013; Bonilla et al., 2007; Astudillo and Frazier, 2019; Shen et al., 2022).

Additionally, projection-based Bayesian optimization methods have recently been developed to cope with the computational complexity when optimizing high-dimensional black-box functions. For instance, Wang et al. (2013) proposed a Bayesian Optimization method based on Random Embeddings by projecting the high-dimensional search space onto a low-dimensional subspace. Ulmasov et al. (2016) proposed a new dimension scheduling algorithm

(DSA) to reduce the computational burden in Bayesian optimization. The DSA was achieved by randomly selecting the subsets of the input parameters and assigning a distinct model for each subset for computation-efficient modeling. [Moriconi et al. \(2020\)](#) developed a scalable quantile GP, which exploited the effective lower dimensionality with axis-aligned projections and solved the optimization problem based on a partition of the input space. However, most existing Bayesian optimization methods are less concerned with simulation calibration problems involving spatial-varying parameters and complex geometries, and thus are not directly applicable for defect localization in 3D systems.

4.3 Problem Formulation

4.3.1 Problem Setting

Simulation-based defect localization in a 3D system calls upon reliably estimating the spatial-varying model parameter such that the output of the 3D simulation closely matches the physical observations. We denote the model parameter as $\boldsymbol{\theta}(\mathbf{s})$, where $\mathbf{s}$ is the spatial location on the 3D geometry, and further denote the spatiotemporal simulation output as $\hat{\mathbf{y}}(\mathbf{s}, t; \boldsymbol{\theta}(\mathbf{s})) = \mathcal{F}(\boldsymbol{\theta}(\mathbf{s}))$, where $\mathcal{F}$ denotes the 3D simulation model. The inverse system for defect localization can be formulated as an optimization problem:

$$\hat{\boldsymbol{\theta}}(\mathbf{s}) = \arg \min_{\boldsymbol{\theta}(\mathbf{s})} J(\boldsymbol{\theta}(\mathbf{s})), \quad (4.1)$$

where the objective function J is defined as the mean squared error (MSE) between the measured ground truth $\mathbf{y}(\mathbf{s}, t)$ and the estimated observation $\hat{\mathbf{y}}(\mathbf{s}, t; \boldsymbol{\theta}(\mathbf{s}))$:

$$J(\boldsymbol{\theta}(\mathbf{s})) = \frac{1}{N_{st}} \sum_{\mathbf{s}} \sum_t \|\mathbf{y}(\mathbf{s}, t) - \hat{\mathbf{y}}(\mathbf{s}, t; \boldsymbol{\theta}(\mathbf{s}))\|^2 \quad (4.2)$$

where N_{st} is the total number of the spatiotemporal observation points. The estimated parameter $\hat{\boldsymbol{\theta}}(\mathbf{s})$ will then be evaluated to identify heterogeneous patterns of the 3D system for defect localization.

4.3.2 Gaussian Process Meta-modeling for Model Calibration

Gaussian Process regression (GPR) is a non-parametric Bayesian approach to learn input-output mappings from a small training dataset. GPR has wide applications in simulation model calibration, where the input of GPR is the parameter setting, i.e., $\boldsymbol{\theta}$, and the output is the model discrepancy between the physical observation and simulation output given parameter setting $\boldsymbol{\theta}$, i.e., $J(\boldsymbol{\theta})$. The key idea behind GPR is to introduce a latent variable $\hat{J}(\boldsymbol{\theta})$ to describe $J(\boldsymbol{\theta})$ for any arbitrary $\boldsymbol{\theta}$. The prior distribution of $\hat{J}(\boldsymbol{\theta})$ is assumed to follow a GP, i.e., $\hat{J}(\cdot) \sim \mathcal{GP}(m(\cdot), \mathcal{K}(\cdot, \cdot))$, where $m(\cdot)$ and $\mathcal{K}(\cdot, \cdot)$ denote the mean and covariance functions. Specifically, we assume:

$$J(\boldsymbol{\theta}) = \hat{J}(\boldsymbol{\theta}) + \delta, \quad \delta \sim \mathcal{N}(0, \sigma_\delta^2) \quad (4.3)$$

where δ denotes the Gaussian noise introduced for numerical stability. Without loss of generality, the mean function is generally set as $m(\boldsymbol{\theta}) = 0$. The behavior of the kernel $\mathcal{K}$ is governed by kernel hyperparameters, which can be determined by maximum likelihood estimation (Brochu et al., 2010).

Given the collected observation pairs $\mathcal{D} = \{\boldsymbol{\theta}_i, J(\boldsymbol{\theta}_i)\}_{i=1}^n = \{\mathbf{\Theta}_{1:n}, \mathbf{J}_{1:n}\}$, where $\mathbf{\Theta}_{1:n} = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n]^T$ and $\mathbf{J}_{1:n} = [J(\boldsymbol{\theta}_1), \dots, J(\boldsymbol{\theta}_n)]^T$, GPR will provide the predictive distribution of $J(\boldsymbol{\theta}_*)$ at an unevaluated instance $\boldsymbol{\theta}_*$ as

$$P(J(\boldsymbol{\theta}_*) | \mathcal{D}) = \mathcal{N}\left(\mu_n(\boldsymbol{\theta}_* | \mathcal{D}), \sigma_n^2(\boldsymbol{\theta}_* | \mathcal{D})\right) \quad (4.4)$$

The posterior mean $\mu_n(\cdot)$ and covariance $\sigma_n(\cdot)$ are evaluated based on the Sherman-Morrison-Woodbury formula (Williams and Rasmussen, 2006) as:

$$\begin{aligned} \mu_n(\boldsymbol{\theta}_* | \mathcal{D}) &= \mathcal{K}(\boldsymbol{\theta}_*, \mathbf{\Theta}_{1:n}) \left(\mathcal{K}(\mathbf{\Theta}_{1:n}, \mathbf{\Theta}_{1:n}) + \sigma_\delta^2 \mathcal{I} \right)^{-1} \mathbf{J}_{1:n} \\ \sigma_n^2(\boldsymbol{\theta}_* | \mathcal{D}) &= \mathcal{K}(\boldsymbol{\theta}_*, \boldsymbol{\theta}_*) - \mathcal{K}(\boldsymbol{\theta}_*, \mathbf{\Theta}_{1:n}) \left(\mathcal{K}(\mathbf{\Theta}_{1:n}, \mathbf{\Theta}_{1:n}) + \sigma_\delta^2 \mathcal{I} \right)^{-1} \mathcal{K}(\mathbf{\Theta}_{1:n}, \boldsymbol{\theta}_*) \end{aligned} \quad (4.5)$$

where $\mathcal{I}$ denotes an $n \times n$ identity matrix. Note that $\mathcal{K}(\mathbf{\Theta}_{1:n}, \boldsymbol{\theta}_*) = \mathcal{K}(\boldsymbol{\theta}_*, \mathbf{\Theta}_{1:n})^T$ is a cross-variance vector that defines the relationship between the observations and the prediction.

$\mathcal{K}(\Theta_{1:n}, \Theta_{1:n})$ is an $n \times n$ matrix capturing the correlation between all the observation pairs. $\mathcal{K}(\theta_*, \theta_*)$ denotes the covariance between the to-be-predicted instance. Eqs. (4.4)-(4.5) provide the predictive distribution of the response variable at any arbitrary inputs to assess the quality of the GPR modeling and further provide guidance to solve the optimization problem in Eq. (4.1) and estimate the model parameter.

4.3.3 Gaussian Process Optimization

The key idea underlying GP optimization is to leverage the GPR as a surrogate to model the input-output mapping and sequentially query the input space for the optimal point of $J(\theta)$. GP optimization is especially useful in situations where collecting experimental observations is expensive, e.g., calibrating large-scale simulation model. The sequential query criterion should balance the trade-off between the exploitation and exploration, i.e., to exploit the solution space of the GP where the predictive mean $\mu_n(\cdot)$ is low and meanwhile explore the parameter space with a high uncertainty $\sigma_n(\cdot)$.

To control the balance between local optimization and global search, GP upper-confidence-bound (GP-UCB) (Srinivas et al., 2009) has been widely used as an acquisition function to actively search for the next query point θ_{n+1} in computer experiment design for model calibration:

$$\theta_{n+1} = \arg \max_{\theta} \{-\mu_n(\theta) + \beta^{1/2} \sigma_n(\theta)\} \quad (4.6)$$

where β is a positive constant to balance between the global search and local optimization. Once the knowledge of the new input-output pair $(\theta_{n+1}, J(\theta_{n+1}))$ becomes available, then, $\Theta_{1:n+1} = [\Theta_{1:n}; \theta_{n+1}]$, $J_{1:n+1} = [J_{1:n}; J(\theta_{n+1})]$, and the GP surrogate model will be updated accordingly:

$$\begin{bmatrix} J_{1:n+1} \\ J(\theta_*) \end{bmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} \mathcal{K}(\Theta_{1:n+1}, \Theta_{1:n+1}) + \sigma_{\delta_\theta}^2 \mathcal{I} & \mathcal{K}(\Theta_{1:n+1}, \theta_*) \\ \mathcal{K}(\theta_*, \Theta_{1:n+1}) & \mathcal{K}(\theta_*, \theta_*) \end{bmatrix} \right) \quad (4.7)$$

This active learning procedure will be repeated until the convergence of θ , where function J is supposed to have the minimum value such that the minimal discrepancy between the simulation output and the physical observation is achieved.

Despite the wide applications of GP optimization in computer model calibration, it can be computationally inefficient to estimate spatial-varying model parameters (i.e., $\theta = \theta(\mathbf{s})$) for defect localization. For example, if the 3D geometry is discretized with 1000 spatial nodes, the corresponding dimensionality of $\theta(\mathbf{s})$ is at least 1000, estimating which requires solving an optimization problem with 1000 unknown variables, posing a significant computational challenge for GP optimization. Koutsourelakis (Koutsourelakis, 2009) proposed to model the spatial-varying parameter using smooth and stationary kernel regression to reduce the dimensionality. However, the model parameter from heterogeneous systems with defects violates the smooth and stationary kernel assumption. Heterogeneity remains one of the most common problems in computational modeling across industrial, engineering, and biological systems, which poses significant challenges in simulation-based abnormality diagnosis. Lee et al. (Lee et al., 2021) proposed a partitioned active learning method for model calibration of heterogeneous systems. This method assumes the heterogeneity partition is a known *priori*, which is not applicable to modeling general heterogeneous systems with defects whose location and extent are unknown. Thus, 3D defect localization calls upon new methods to effectively account for the heterogeneous pattern for model calibration.

4.4 Research Methodology

This section presents the proposed framework of Hierarchical Gaussian Process Active Learning (HGPAL) for defect identification in complexly structured 3D systems, as shown in Fig. D.20. We first develop a GP regression model on the 2D embedding of a 3D surface to capture the geometric information. Second, we propose a Hierarchical GP-based active learning approach to account for the heterogeneity pattern and further estimate the spatial-varying model parameter for defect localization.

4.4.1 Gaussian Process Regression on the 2D Embedding of a 3D Surface

Traditional GP modeling generally assumes the input domain lies in a Euclidean space and computes the kernel correlation based on Euclidean distance. However, simulation modeling of a real system usually involves a complexly-structured 3D surface, which is intrinsically a 2D manifold. Euclidean-based kernels capture the correlation between different locations based on the shortest distance in the Euclidean space without considering the complex geometry or topology of the system manifold. Although, it is natural to think of replacing the Euclidean distance with geodesic distance into the kernel function to model the spatial correlation, directly applying the substitution is not feasible because the geodesic distance metric will lead to a non-positive definite covariance matrix in GP modeling (Kanagawa et al., 2018).

Here, we propose to construct the GP regression model on the 2D embedding of a 3D surface to capture the geometric information. We denote the original input domain (i.e., the geometry of the 3D system) for defect pattern modeling as $\mathcal{X}^{3d} \subset \mathbb{R}^3$. We first map $\mathcal{X}^{3d}$ into a 2D embedded space $\mathcal{X}^{2d}$ through a transformation $\mathcal{M} : \mathcal{X}^{3d} \rightarrow \mathcal{X}^{2d}$, and further build the GP model on $\mathcal{X}^{2d}$. As such, the overall regression procedure for defect pattern modeling consists of two sub-functions: $J(\cdot) = \mathcal{GP}_{\mathcal{M}}(\cdot) \circ \mathcal{M}(\cdot)$, where $\mathcal{GP}_{\mathcal{M}}(\cdot)$ denotes the GP modeling on the 2D embedded space $\mathcal{X}^{2d}$, with a kernel function $\mathcal{K}_{\mathcal{M}}$ defined as

$$\mathcal{K}_{\mathcal{M}}(\mathbf{x}_i, \mathbf{x}_j) = \mathcal{K}(\mathcal{M}(\mathbf{x}_i), \mathcal{M}(\mathbf{x}_j)) \quad (4.8)$$

The predictive distribution of an arbitrary input in $\mathcal{X}^{3d}$ can then be derived by first transforming it into the 2D embedded space and then applying the GPR modeling as given by Eq. (4.4).

In order to preserve the geometric information of the 3D surface, the 2D embedding transformation $\mathcal{M}$ is based on the isometric mapping (Isomap) (Balasubramanian and Schwartz, 2002; Xie and Yao, 2022a). The Isomap scheme seeks a lower-dimensional embedding while maintaining the geodesic distance between all spatial nodes. Specifically, the Isomap embedding includes the following two steps:

(1) **Geodesic distance calculation:** We first construct a k -nearest neighbor graph of the 3D system by connecting each locations $\mathbf{s}_i$ to its k nearest neighbors $\mathbf{s}_{j,i \leftrightarrow j}$, where $i \leftrightarrow j$ indicates node j is one of the k nearest neighbors of node i . Second, the distance between neighboring nodes i and j is initialized as: $d_G(i, j) = \|\mathbf{s}_i - \mathbf{s}_j\|$, where $\|\cdot\|$ denotes the Euclidean norm; if nodes $\mathbf{s}_i$ and $\mathbf{s}_j$ are disconnected, $d_G(i, j) = \infty$. Third, Dijkstra's algorithm (Dijkstra et al., 1959) is engaged to calculate the shortest path between any pair of nodes, which is equivalent to extracting the geodesic information from the system geometry. Specifically, for $k = 1, \dots, N_G$ (where N_G denotes the number of spatial nodes in the 3D system), each $d_G(i, j)$ is updated by $\min_k \{d_G(i, j), d_G(i, k) + d_G(k, j)\}$, which defines the geodesic distance between $\mathbf{s}_i$ and $\mathbf{s}_j$.

(2) **2D embedding construction:** In order to guarantee that the 2D embedding space preserves the 3D geometric property, the objective function for searching the 2D counterpart of the 3D surface is formulated as:

$$\min_{\hat{\mathbf{S}}=\{\hat{\mathbf{s}}_i\}_{i=1}^{N_G}} \sum_{i < j \in \{1, \dots, N_G\}} \left(\|\hat{\mathbf{s}}_i - \hat{\mathbf{s}}_j\| - d_G(i, j) \right)^2 \quad (4.9)$$

where $\hat{\mathbf{s}}_i$ and $\hat{\mathbf{s}}_j$ stand for the locations on the 2D isometric graph. This optimization problem can be solved by introducing the Gram matrix G : $G = -\frac{1}{2}PD_G^2P$, where the matrix P is defined as $P = \{\delta_{ij} - \frac{1}{N_G}\}_{i,j=1}^{N_G}$, δ_{ij} is the Kronecker delta. Each element in the squared matrix D_G^2 is $d_G^2(i, j)$, $\forall i, j \in \{1, \dots, N_G\}$. Then, the optimization problem in Eq. (4.9) can be transformed as $\hat{\mathbf{S}}\hat{\mathbf{S}}^T \approx G$. The gram matrix G can be further decomposed as $G = V\Lambda V^T = V\sqrt{\Lambda}\sqrt{\Lambda}V^T$, where V is a matrix of eigenvectors and Λ is a diagonal matrix of the eigenvalues: $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{N_G})$. As such, the solution $\hat{\mathbf{S}}$ to the objective function in Eq. (4.9) can be obtained based on the decomposition of G : $\hat{\mathbf{S}} = V\sqrt{\Lambda}$.

We can further model the heterogeneous pattern of the 3D system on the 2D embedded manifold for defect localization through GP optimization, which will be detailed in the next subsection.

4.4.2 Hierarchical GP-based Active Learning

To enable the active learning procedure, we first project the 3D complex geometry into a 2D embedding, as discussed in Section 4.4.1, to model the 3D heterogeneity pattern. Specifically, we model the 3D defect pattern as the corresponding 2D area which is defined by the center point at $[x_1, x_2]$ with radius r on the 2D embedded space, whose value will be estimated by GP optimization. We denote $\boldsymbol{\gamma}^{3D}$ as the 3D defect pattern and $\boldsymbol{\gamma} = [x_1, x_2, r]^T$ denotes the corresponding pattern on the 2D manifold, i.e., $\mathcal{M}(\boldsymbol{\gamma}^{3D}) = \boldsymbol{\gamma}$ or $\boldsymbol{\gamma}^{3D} = \mathcal{M}^{-1}(\boldsymbol{\gamma})$, where $\mathcal{M}^{-1}$ denotes the inverse transformation of $\mathcal{M}$. Additionally, it is straightforward to assume the system property is homogeneous within the defect or normal region, and heterogeneity only exists between the two regions. Namely, the system property consists of two distinct phases. Thus, the spatial-varying parameter can be modeled as $\boldsymbol{\theta}(\mathbf{s}) \equiv [\boldsymbol{\theta}_{\mathcal{M}^{-1}(\boldsymbol{\gamma})}; \mathcal{M}^{-1}(\boldsymbol{\gamma})] \equiv [\boldsymbol{\theta}_{\boldsymbol{\gamma}}; \boldsymbol{\gamma}]$. In other words, $\boldsymbol{\theta}(\mathbf{s})$ is a function of the defect pattern $\boldsymbol{\gamma}$, and $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$ is the model parameter that minimizes the discrepancy J between the simulation output and true observation given the defect pattern $\boldsymbol{\gamma}$. Note that $\boldsymbol{\theta}_{\boldsymbol{\gamma}} \equiv \boldsymbol{\theta}_{\mathcal{M}^{-1}(\boldsymbol{\gamma})} = [(\boldsymbol{\theta}^A, \boldsymbol{\theta}^N) | \mathcal{M}^{-1}(\boldsymbol{\gamma})]^T$ denotes the parameter setting given the defect pattern $\boldsymbol{\gamma}$, where $\boldsymbol{\theta}^A \in \mathbb{R}^{\vartheta}$ and $\boldsymbol{\theta}^N \in \mathbb{R}^{\vartheta}$ denote the parameter values for the defect and normal regions respectively, and ϑ represents the total number of the queried attributes of the system property. As such, the optimization problem in Eq. (4.1) can be reformulated as:

$$\min_{\boldsymbol{\gamma}, \boldsymbol{\theta}} \quad J(\mathbf{y}, \hat{\mathbf{y}}; \boldsymbol{\theta}, \boldsymbol{\gamma}) \quad (4.10)$$

$$s.t. \quad \boldsymbol{\theta} = \boldsymbol{\theta}_{\boldsymbol{\gamma}} = \arg \min_{\boldsymbol{\theta}} J(\mathbf{y}, \hat{\mathbf{y}}; \boldsymbol{\theta} | \boldsymbol{\gamma}) \quad (4.11)$$

The new optimization problem in Eq. 4.10 aims to minimize the objective function J with respect to both the location $\boldsymbol{\gamma}$ and the model parameter $\boldsymbol{\theta}$, and Eq. 4.11 imposes a constraint by expressing the dependency of $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$ on $\boldsymbol{\gamma}$. Given the hierarchical formulation in Eq. (4.10)-(4.11), we develop an HGPAL procedure to solve the optimization problem which includes a master GP active learning and a sub-GP active learning.

(1) **Master GP Active Learning:** Note that our major goal is to find the defect location $\boldsymbol{\gamma}$ of a 3D system that is not directly observable and meanwhile, determine the

system properties captured by the model parameter set $\boldsymbol{\theta}$. We first consider the estimation of $\boldsymbol{\gamma}$ given the corresponding optimal parameter $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$ obtained from the sub-GP active learning (detailed in the next sub-section):

$$\hat{\boldsymbol{\gamma}} = \arg \min_{\boldsymbol{\gamma}} J(\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{\gamma}}), \quad (4.12)$$

We use a GP to model the functional relationship between the input of defect pattern $\boldsymbol{\gamma}$'s and the output J 's. Specifically, we assume that

$$J(\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{\gamma}}) = \hat{J}(\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{\gamma}}) + \delta_{\boldsymbol{\gamma}}, \quad \delta_{\boldsymbol{\gamma}} \sim \mathcal{N}(0, \sigma_{\delta_{\boldsymbol{\gamma}}}^2)$$

where $\delta_{\boldsymbol{\gamma}}$ denote a Gaussian white noise, and we further model $\hat{J}(\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{\gamma}})$ as a GP: $\hat{J}(\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{\gamma}}) \sim \mathcal{GP}(0, \mathcal{K}_{\boldsymbol{\gamma}})$. Note that we omit the subscript $\mathcal{M}$ in $\mathcal{GP}_{\mathcal{M}}$, i.e., GP modeling on the 2D embedding, to simplify the notations. We select the exponentiated quadratic kernel to model the correlations between the 2D defect patterns $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_j$:

$$\mathcal{K}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j) = \sigma_{\boldsymbol{\gamma}}^2 \exp\left(-\frac{\|\boldsymbol{\gamma}_i^{(1:2)} - \boldsymbol{\gamma}_j^{(1:2)}\|^2}{2l_c^2}\right) \cdot \exp\left(-\frac{\|\boldsymbol{\gamma}_i^{(3)} - \boldsymbol{\gamma}_j^{(3)}\|^2}{2l_r^2}\right) \quad (4.13)$$

where $\boldsymbol{\gamma}^{(1:2)}$ denotes the first two elements in $\boldsymbol{\gamma}$, i.e., the coordinate of the center point $[x_1, x_2]$ of the 2D embeded defect pattern; $\boldsymbol{\gamma}^{(3)}$ is the third element, i.e., the radius r of the defect area; $\sigma_{\boldsymbol{\gamma}}$, l_c , and l_r represent the kernel hyperparameters. Given the prior observation $\{\boldsymbol{\Gamma}_{1:m}|\boldsymbol{\Theta}_{\boldsymbol{\gamma}_{1:m}}, \boldsymbol{J}_{1:m}\}$, where $\boldsymbol{\Gamma}_{1:m}|\boldsymbol{\Theta}_{\boldsymbol{\gamma}_{1:m}} = [\boldsymbol{\gamma}_1|\boldsymbol{\theta}_{\boldsymbol{\gamma}_1}, \dots, \boldsymbol{\gamma}_m|\boldsymbol{\theta}_{\boldsymbol{\gamma}_m}]^T$, $\boldsymbol{J}_{1:m} = [J_1, \dots, J_m]^T$, the posterior esimation of an unevaluated $\boldsymbol{\gamma}_*$ follows a normal distribution:

$$P(J(\boldsymbol{\gamma}_*)|\{\boldsymbol{\Gamma}_{1:m}|\boldsymbol{\Theta}_{\boldsymbol{\gamma}_{1:m}}, \boldsymbol{J}_{1:m}\}) = \mathcal{N}(\mu_m(\boldsymbol{\gamma}_*|\boldsymbol{\theta}_{\boldsymbol{\gamma}_*}), \sigma_m^2(\boldsymbol{\gamma}_*|\boldsymbol{\theta}_{\boldsymbol{\gamma}_*}))$$

where the predictive mean $\mu_m(\boldsymbol{\gamma}_*|\boldsymbol{\theta}_{\boldsymbol{\gamma}_*})$ and variance $\sigma_m^2(\boldsymbol{\gamma}_*|\boldsymbol{\theta}_{\boldsymbol{\gamma}_*})$ can be derived according to the GP property in Eq. (4.5).

Given the trained GP model, the next query location $\boldsymbol{\gamma}_{m+1}$ will be generated according to the GP-UCB as given by Eq.(4.6) as

$$\boldsymbol{\gamma}_{m+1} = \arg \max_{\boldsymbol{\gamma}} \{-\mu_m(\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{\gamma}}) + \beta^{1/2}\sigma_m(\boldsymbol{\gamma}|\boldsymbol{\theta}_{\boldsymbol{\gamma}})\} \quad (4.14)$$

Once the new location $\boldsymbol{\gamma}_{m+1}$ is acquired, the sub-GP active learning will be explored to search for the corresponding optimal parameter set $\boldsymbol{\theta}_{\boldsymbol{\gamma}_{m+1}}$. Given the new input $\boldsymbol{\gamma}_{m+1}$ and $\boldsymbol{\theta}_{\boldsymbol{\gamma}_{m+1}}$, the 3D simulator will be evaluated to generate the new response of the objective function $J_{m+1} = J(\mathbf{y}, \hat{\mathbf{y}}; \boldsymbol{\theta}_{\boldsymbol{\gamma}_{m+1}}, \boldsymbol{\gamma}_{m+1})$, and the master GP will be further updated according to the GP update rule in Eq. (4.7).

(2) **Sub-GP Active Learning:** Given the new query points for the defect pattern $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}_{m+1}$, we will conduct the sub-GP active learning to search for a $\hat{\boldsymbol{\theta}}_{\tilde{\boldsymbol{\gamma}}}$ that minimizes the model discrepancy:

$$\hat{\boldsymbol{\theta}}_{\tilde{\boldsymbol{\gamma}}} = \arg \min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}|\tilde{\boldsymbol{\gamma}}) \quad (4.15)$$

In the sub-GP modeling, we assume

$$J(\boldsymbol{\theta}|\tilde{\boldsymbol{\gamma}}) = \hat{J}(\boldsymbol{\theta}|\tilde{\boldsymbol{\gamma}}) + \delta_{\boldsymbol{\theta}}, \quad \delta_{\boldsymbol{\theta}} \sim \mathcal{N}(0, \sigma_{\delta_{\boldsymbol{\theta}}}^2) \quad (4.16)$$

$$\hat{J}(\boldsymbol{\theta}|\tilde{\boldsymbol{\gamma}}) \sim \mathcal{GP}(0, \mathcal{K}_{\boldsymbol{\theta}}) \quad (4.17)$$

where $\delta_{\boldsymbol{\theta}}$ is a Gaussian white noise. We also engage the exponentiated quadratic kernel to characterize the correlation of priors on the functions sampled from the sub-GP, which is formulated as:

$$\mathcal{K}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \sigma_{\boldsymbol{\theta}}^2 \exp\left(-\frac{\|\boldsymbol{\theta}_{1_i} - \boldsymbol{\theta}_{1_j}\|^2}{2l_{\boldsymbol{\theta}_1}^2}\right) \cdots \exp\left(-\frac{\|\boldsymbol{\theta}_{\vartheta_i} - \boldsymbol{\theta}_{\vartheta_j}\|^2}{2l_{\boldsymbol{\theta}_{\vartheta}}^2}\right) \quad (4.18)$$

where $\sigma_{\boldsymbol{\theta}}, l_{\boldsymbol{\theta}_1}, \dots, l_{\boldsymbol{\theta}_{\vartheta}}$ are the kernel hyperparameters. Given the observation set $\{\boldsymbol{\Theta}_{1:n}|\tilde{\boldsymbol{\gamma}}, \mathbf{J}_{1:n}^{\text{sub}}\}$, where $\boldsymbol{\Theta}_{1:n}|\tilde{\boldsymbol{\gamma}} = [\boldsymbol{\theta}_1|\tilde{\boldsymbol{\gamma}}, \dots, \boldsymbol{\theta}_n|\tilde{\boldsymbol{\gamma}}]$ and $\mathbf{J}_{1:n}^{\text{sub}} = [J_1^{\text{sub}}, \dots, J_n^{\text{sub}}]$, the posterior distribution of

$J^{\text{sub}}(\theta_*|\tilde{\gamma})$ given an unevaluated parameter set $\theta_*|\tilde{\gamma}$ follows a Gaussian distribution:

$$P(J^{\text{sub}}(\theta_*|\tilde{\gamma})|\{\Theta_{1:n}|\tilde{\gamma}, J_{1:n}^{\text{sub}}\}) = \mathcal{N}(\mu_n(\theta_*|\tilde{\gamma}), \sigma_n^2(\theta_*|\tilde{\gamma})) \quad (4.19)$$

where, the predictive mean $\mu_n(\theta_*|\tilde{\gamma})$ and variance $\sigma_n^2(\theta_*|\tilde{\gamma})$ can be derived according to Eq. (4.5).

Given the trained sub-GP model, the GP-UCB in Eq. (4.6) will be applied to search for the next query point $\theta_{n+1}|\tilde{\gamma}$:

$$\theta_{n+1}|\tilde{\gamma} = \arg \max_{\theta} \{-\mu_n(\theta|\tilde{\gamma}) + \beta^{1/2} \sigma_n(\theta|\tilde{\gamma})\} \quad (4.20)$$

Upon the selection of a new set of model parameters $\theta_{n+1}|\tilde{\gamma}$, the new objective function J_{n+1}^{sub} will be reevaluated, and the sub-GP will be updated once the new input-output pair becomes available according to the GP update rule in Eq. (4.7). The procedure for actively learning of $\theta_{\tilde{\gamma}}$ is run iteratively until convergence to obtain the optimal model parameter set at the given $\tilde{\gamma}$. As such, the sub-GP active learning procedure enables us to find a parameter set $\hat{\theta}_{\tilde{\gamma}}$ of the system that can potentially minimize the objective function J given an arbitrary guess of the defect whereabouts $\tilde{\gamma}$.

We employ the L-BFGS-B algorithm to solve the maximum likelihood estimates for the kernel hyperparameters in Eqs. (4.13, 4.18) and solve the optimization problems in Eqs. (4.14, 4.20) (Byrd et al., 1995). The L-BFGS-B algorithm is a quasi-Newton method that approximates the inverse Hessian matrix of the objective function using limited memory. It combines the benefits of both the BFGS method and the limited-memory aspect to enable efficient memory usage for solving high-dimensional optimization problems.

The Master and sub-GP active learning procedures will be evaluated alternately and iteratively until γ does not change significantly over iterations. The converged defect pattern in the 2D embedding will then be transformed back to the 3D through $\mathcal{M}^{-1}$ for localizing the 3D defect pattern. The overall procedure of HGPAL is summarized in Algorithm 1. As such, we build the HGPAL framework that can help to approximate the defect whereabouts and simultaneously infer the system parameters that cannot be observed directly.

Algorithm 1: Hierarchical Gaussian Process Active Learning (HGPAL)

Input: $\zeta, \xi, M, N; m = 1, \tau = \tau_0 > \zeta, n = 1, \rho = \rho_0 > \xi;$

$\Theta_0 = \{\theta_0\} \leftarrow$ random initialize, $\Gamma_0 = \{\gamma_0\} \leftarrow$ random initialization

while $\tau \geq \zeta$ and $m \leq M$ **do**

begin Master GP construction for the objective function $J(\gamma)$

 GP predictive modeling given observed data $\{\Gamma_{1:m-1} | \Theta_{\gamma_{1:m-1}}, J_{1:m-1}\}$ for $J(\gamma)$:

$$P\left(J(\gamma) \mid \{\Gamma_{1:m-1} | \Theta_{\gamma_{1:m-1}}, J_{1:m-1}\}\right) = \mathcal{N}\left(\mu_{m-1}(\gamma \mid \Theta_{\gamma_{m-1}}), \sigma_{m-1}^2(\gamma \mid \Theta_{\gamma_{m-1}})\right)$$

end

 Select the next defect location according to the GP-UCB criterion:

$$\gamma_m = \arg \max_{\gamma} \{-\mu_{m-1}(\gamma \mid \Theta_{\gamma_{1:m-1}}) + \beta^{1/2} \cdot \sigma_{m-1}(\gamma \mid \Theta_{\gamma_{1:m-1}})\}$$

while $\rho \geq \xi$ and $n \leq N$ **do**

begin Sub-GP construction for the objective function $J^{\text{sub}}(\theta | \gamma_m)$

 GP predictive modeling given observed data $\{\Theta_{1:n-1} | \gamma_m, J_{1:n-1}^{\text{sub}}\}$ for $J^{\text{sub}}(\theta | \gamma_m)$:

$$P\left(J^{\text{sub}}(\theta | \gamma_m) \mid \{\Theta_{1:n-1} | \gamma_m, J_{1:n-1}^{\text{sub}}\}\right) = \mathcal{N}\left(\mu_{n-1}(\theta \mid \gamma_m), \sigma_{n-1}^2(\theta \mid \gamma_m)\right)$$

end

 Select the next training point according to the GP-UCB criterion:

$$\theta_n \mid \gamma_m = \arg \max_{\theta} \{-\mu_{n-1}(\theta \mid \gamma_m) + \beta^{1/2} \cdot \sigma_{n-1}(\theta \mid \gamma_m)\}$$

 Run the simulation model with parameter setting $\theta_n \mid \gamma_m$ to calculate the new objective function J_n^{sub} using Eq. (4.2)

$$\rho = \|\theta_n - \theta_{n-1}\|; \{\Theta_{1:n} \mid \gamma_m, J_{1:n}^{\text{sub}}\} = \{\Theta_{1:n-1} \mid \gamma_m, J_{1:n-1}^{\text{sub}}\} \cup \{\theta_n \mid \gamma_m, J_n^{\text{sub}}\}; n \leftarrow n + 1;$$

end

$$\theta_{\gamma_m} \leftarrow \arg \min_{n_m \in \{1, 2, \dots, n\}} \hat{J}(\theta_{n_m} | \gamma_m)$$

 Run the simulator with θ_{γ_m} to calculate the new objective function J_m according to Eq. (4.2)

$$\tau = \|\gamma_m - \gamma_{m-1}\|; \{\Gamma_{1:m} \mid \Theta_{\gamma_{1:m}}, J_m\} = \{\Gamma_{1:m-1} \mid \Theta_{\gamma_{1:m-1}}, J_{m-1}\} \cup \{\gamma_m \mid \theta_{\gamma_m}, J_m\}; m \leftarrow m + 1;$$

end

Output: $\{\hat{\gamma}, \hat{\theta}\} = \arg \min_{\{\gamma, \theta\} \in [\{\gamma_1, \theta_{\gamma_1}\}, \{\gamma_m, \theta_{\gamma_m}\}]} J(\mathbf{y}, \hat{\mathbf{y}}; \gamma, \theta).$

It is worth noting that the HGPAL framework assumes the defect region consists of a set of spatial points within a round shape characterized by γ in the 2D embedding. However, in a real-world case, the defect area comes with an irregular shape, whose 2D projection, most likely, will not be a perfectly round shape. The inaccurate assumption of defect region can apparently deteriorate the estimation accuracy for both the defect whereabouts γ and the system parameter set θ . To overcome this issue, we develop a refinement strategy, i.e., HGPAL-RFF, that is able to selectively adjust the estimated defect pattern and adaptively update the model parameters for more accurate estimation (See more details in the Appendix).

4.5 Experimental Design and Results

In this section, we implement our HGPAL-REF framework in a 3D cardiac simulation to identify the myocardial infarction (MI) in the heart based on ECG observations on the body surface. In this study, we choose the Aliev-Panfilov (AP) model (Aliev and Panfilov, 1996b) to simulate the evolvement of cardiac electrodynamics:

$$\frac{\partial \mathbf{u}}{\partial t} = \nabla \cdot (d \nabla \mathbf{u}) + k \mathbf{u}(\mathbf{u} - a)(1 - \mathbf{u}) - \mathbf{u} \mathbf{v} \quad (4.21)$$

$$\frac{\partial \mathbf{v}}{\partial t} = \xi(\mathbf{u}, \mathbf{v})(-\mathbf{v} - k \mathbf{u}(\mathbf{u} - a - 1)) \quad (4.22)$$

where $\mathbf{u}$ stands for the transmembrane action potential, and $\mathbf{v}$ is the latent recovery current to model the local depolarization behavior. Parameter $\xi(\mathbf{u}, \mathbf{v}) = e_0 + \mu_1 \mathbf{v} / (\mathbf{u} + \mu_2)$ characterizes the coupling effect between $\mathbf{u}$ and $\mathbf{v}$. k controls the repolarization. Parameter d denotes the diffusion conductivity, where a smaller value of d leads to lower myocardial conductivity, i.e., higher severity of MI. Parameter a controls the excitability of heart cells, i.e., the ability to be excited by an electrical pulse. Note that an increased value of a results in lower excitability, i.e., higher severity of MI.

It is recognized that the health status of the cardiac muscular tissue can be largely affected by the parameters a and d , whose values are closely affiliated with the severity level of the cardiac infarct (Aliev and Panfilov, 1996a; Moreau-Villéger et al., 2006; Dhamala et al., 2017). This heterogeneous property on the myocardium can further reflect on the ECG observation on the torso surface. Thus, the model parameters of interest for this study lie in a and d that characterize the excitative and conductive properties of heart tissue. It is natural to assume the myocardium tissue property is homogeneous within the MI or healthy region, and heterogeneity only exists between the two regions. As such, we assign a_h , d_h and a_d , d_d to denote excitability and conductivity on the healthy and diseased regions, respectively. We assume the reminder parameter set $[k, e_0, \mu_1, \mu_2]$ is constant, whose value is adopted from existing study and set as $[8, 0.002, 0.3, 0.3]$ on the entire domain (Aliev and Panfilov, 1996b).

Given the vector $\boldsymbol{\theta} = [a_h, a_d, d_h, d_d]^T$ that defines the healthy and abnormal tissue properties, the spatiotemporal evolvement of the heart surface potential $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{s}, t)$ can be numerically computed by solving the AP model (Eq. (4.21-4.22)) with finite element method (FEM), which can be further related to the body-surface electrical signal $\mathbf{y}_{\boldsymbol{\theta}}$ through forward ECG modeling:

$$\mathbf{y}_{\boldsymbol{\theta}}(\mathbf{s}, t) = \mathbf{R}_{BH}\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{s}, t) + \epsilon \quad (4.23)$$

where $\mathbf{s}$ and t denote the spatial and temporal coordinates, respectively, $\mathbf{R}_{BH}$ is the transfer matrix derived from the divergence theorem and Green’s Second Identify (Yao and Yang, 2020; Barr et al., 1977; Yao and Yang, 2016), and ϵ is the noise term. We will leverage the proposed HGPAL-REF algorithm to minimize the discrepancy between the physical observations $\mathbf{y}(\mathbf{s}, t)$ and the simulated ECG $\mathbf{y}_{\boldsymbol{\theta}}(\mathbf{s}, t)$ to estimate $\boldsymbol{\theta}$ for disease localization.

We evaluate the performance of the HGPAL-REF algorithm with a 3D body-heart system. The heart geometry is discretized into 1094 nodes and 2184 mesh elements, and the body surface is formed by 352 nodes with 677 mesh elements, which are obtained from the 2007 PhysioNet Computing in Cardiology Challenge (Dawoud et al., 2008; Goldberger et al., 2000).

4.5.1 Experimental Design

As illustrated in Fig. D.21, we will investigate three levels of MI severity characterized by 3 sets of the model parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_3$, the true values of which are listed in Table E.7. Additionally, three diseased locations, i.e., Loc_1, Loc_2, and Loc_3, with two different sizes will be investigated. In the experiment, the true diseased regions $\boldsymbol{\gamma}^{3D*}$ ’s and model parameter sets $\boldsymbol{\theta}^*$ ’s are first assigned to the simulation model to generate the ground-truth body-surface potential (BSP) $\mathbf{y}^*(\mathbf{s}, t)$. A Gaussian noise $\delta \sim (0, 0.05)$ is added to the true BSP, i.e., $\mathbf{y}(\mathbf{s}, t) = \mathbf{y}^*(\mathbf{s}, t) + \delta$, which serves as the physical observation.

The proposed HGPAL-REF algorithm is implemented to identify the MI region as well as the normal and diseased heart tissue properties. We select the balance parameter as $\beta = 9$ in both the master- and sub-GP active learning. The performance of HGPAL-REF

is benchmarked with the multi-resolution and multi-task GP upper confidence bound (MT-GPUCB) approaches. The detailed introduction of the two benchmark methods are provided in the Appendix.

The model performance will be evaluated according to three metrics: *MSE*, size discrepancy percentage (*SDP*), and overlap percentage (*OVP*). The *MSE* shares the same definition of the objective function J (i.e., Eq. (4.2)) to evaluate the discrepancy between the estimated and true BSPs. In order to measure the MI size, once the MI region γ in the 2D space is determined, we project it back to the 3D space to identify the 3D MI region by invoking the inverse Isomap transformation: $\gamma^{3D} = \mathcal{M}^{-1}(\gamma)$. Then, the MI size is calculated as the ratio between the number of discretized nodes within γ^{3D} and the total number of discretized nodes on the heart geometry. Based on this, the *SDP* and *OVP* are defined as follows:

$$\begin{aligned} SDP &= \frac{|\text{Estimated MI size} - \text{True MI size}|}{\text{True MI size}} \times 100\% \\ OVP &= \frac{\text{Estimated MI region} \cap \text{True MI region}}{\text{True MI region}} \times 100\% \end{aligned} \quad (4.24)$$

which evaluate the discrepancies in the size and location of the estimated and true MI regions.

4.5.2 Experimental Results

Table E.6 compares the results of MI identification by our HGPAL-REF scheme, the multi-resolution, and MT-GPUCB methods for both small and large MI sizes at three locations. Note that the *MSE* generated by HGPAL-REF exhibits a lower value compared to the *MSE*'s from the multi-resolution search and MT-GPUCB methods for all the MI regions with different parameter sets. Specifically, the *MSE* from the HGPAL-REF algorithm has a mean value of 1.1659×10^{-4} and a standard deviation (std) of 0.1799×10^{-4} , whereas the multi-resolution method produces an *MSE* with the mean of 13.7837×10^{-4} and std of 20.3264×10^{-4} , and MT-GPUCB yields a mean of 9.4471×10^{-4} and std of 8.6720×10^{-4} . This indicates that our HGPAL-REF approach provides more accurate and stable estimated BSPs compared with the multi-resolution and MT-GPUCB methods. Additionally, the proposed HGPAL-REF algorithm yields smaller *SDP*'s between 0 to 40% with the mean of 17.96%, which increases to 34.95% by the multi-resolution method and 59.57% by the MT-GPUCB.

It should be noted that *SDP* only reflects one aspect of the prediction in terms of the size discrepancy between the estimated and true MI's. For example, both HGPAL-REF and multi-resolution methods accomplish 0 *SDP* for small MI at Loc_1 with parameter set θ_1 . However, the multi-resolution and MT-GPUCB methods yield the average *OVP*'s of 32.36% and 52.63%, respectively, which is significantly smaller compared to 79.73% provided by our HGPAL-REF method. Note that the *OVP* generated by HGPAL-REF ranges from 63.16% to 97.3%, which is significantly higher than that of the multi-resolution method, i.e., 5.26% to 55.31%, and MT-GPUCB, i.e., 8.1% to 88.88%.

Fig. D.22 illustrates the comparison of estimated diseased regions on the heart surface with parameter set θ_2 . Specifically, Fig. D.22 (a-b), (c-d), and (e-f) show small and large diseased regions at three different heart locations, respectively. The diseased regions estimated by the multi-resolution method and MT-GPUCB present significant spatial discrepancies compared with the true regions. On the other hand, it can be visually perceived that the overlap between the true area and estimated MI region by our HGPAL-REF approach is consistently larger for both small and larger sizes at different locations.

Apart from the reliable estimations for the diseased area, our HGPAL-REF algorithm is also able to accurately estimate the key parameters of the system that cannot be directly measured, as shown in Table E.7. Three sets of cardiac electrophysiological parameters, i.e., θ_1 , θ_2 , and θ_3 are employed to reflect different severity levels of MI. Note that the true excitability and conductivity of the healthy tissue (a_h and d_h) keep the same for θ_1 , θ_2 , and θ_3 . According to Table 2, the variations of the estimated a_h and d_h never exceed 0.9% and 1.4% of the truth (i.e., $a_h = 0.1$ and $d_h = 10$), respectively.

It is worth mentioning that the estimations of a_d and d_d by HGPAL-REF in θ_1 demonstrate non-negligible deviations from the truth (i.e., $a_d = 0.25$ and $d_d = 0$), as shown in Table E.7. This is because the estimated a_d is already greater than the excitability threshold of a_d for severe MI, i.e., $a_d \geq 0.20$, and the resulted diseased tissue will be inexcitable and will not conduct electrical signals even with a non-zero conductivity d_d . In other words, given a fixed diseased region, the diseased parameter combination of $(a_d, d_d) \in \{(a_d, d_d) \mid a_d > 0.2, d_d \geq 0\}$ will generate the same BSP simulation results. For less severe MI conditions (i.e., θ_2 and θ_3), where $a_d < 0.2$ and $d_d > 0$, HGPAL-REF is

capable of providing close approximations to the true excitability a_d and conductivity d_d . For example, with the parameter set of θ_2 , the estimated a_d and d_d vary between 0.1389 to 0.1575 and 4.5819 to 5.0, respectively, given the true model parameters $a_d = 0.15$ and $d_d = 5$. Such experimental results demonstrate the great quantification ability of our HGPAL-REF method for identifying electrophysiological parameters in the diseased and healthy heart regions. Additionally, the estimations of parameter set yielded from our HGPAL-REF show a significant accuracy enhancement compared to the estimations generated from MT-GPUCB. This is because the hierarchical active learning structure can effectively capture the complex relationships between the defect location and the system parameters. On the other hand, the separable kernel function for defect pattern γ and system parameter θ in MT-GPUCB (see the Appendix) may not be able to effectively model the complex interactions between them, leading to inferior performance in defect localization in 3D systems.

Fig. D.23 shows the comparison of the estimated BSPs, $\hat{y}$, evolving over time at one specific spatial location on the body surface produced by HGPAL-REF, multi-resolution search, and MT-GPUCB with different MI regions on the heart. Note that θ_2 is chosen as the true electrophysiological parameter setting to generate the ground-truth physical observations. The evolvement curves generated by the multi-resolution method and MT-GPUCB display significant pattern deviations compared with the physical observations. In contrast, our HGPAL-REF algorithm produces reliable BSP estimations that almost overlap with the true signals.

4.6 Conclusions

This paper proposes a novel Hierarchical Gaussian Process Active Learning framework for reliable defect localization in 3D systems. We first develop a GP regression model on the 2D embedding of a 3D geometry to capture the geometric information. Second, we propose a Hierarchical GP-based active learning approach including master GP and sub-GP active learning to account for the system heterogeneity and further estimate the spatial-varying model parameter for defect localization. We evaluate the performance of the proposed HGPAL framework in a 3D body-heart system to identify and localize infarct in the heart.

Numerical experiments demonstrate the effectiveness of our HGPAL framework for 3D defect localization. The proposed HGPAL framework has a strong potential to realize optimal defect detection for precision engineering system design and healthcare treatment plan.

Chapter 5

Conclusions

With the rapid progression in sensor technology, we find ourselves in a data-abundant environment that significantly enhances predictive modeling and decision-making in intricate systems. However, to fully tap into the potential of this complex sensor data, we need innovative and trustworthy analytical models and tools for system informatics. The improvements in sensor technology pave the way for unrivaled opportunities to scrutinize system dynamics and enhance decision-making, particularly in the realm of smart health. However, sensor data is often characterized by high dimensionality and complex structures, thus necessitating innovative analytical methods and tools for effective information processing.

The purpose of this dissertation is to establish cutting-edge methodologies that are augmented with physical principles, aiming to model, supervise, and optimize complex systems. The spotlight of the research is on the creation of machine learning models that are informed by physics, aiding in optimal decision-making within complex healthcare systems. The execution of this research will make possible of 1) reliable predictive modeling of spatiotemporal systems; 2) extraction of vital feature information about system dynamics; 3) enhancement of decision-making in the face of uncertainty and sparse sensor observations. The key achievements of my research include:

- In Chapter 2, a new physics-constrained deep learning (P-DL) strategy is developed to predict the evolvement of the heart-surface electrical signals from body-surface sensor measurements. The P-DL model innovatively encodes the well-developed physics model that describes the spatiotemporal dynamics of cardiac signals into the deep learning infrastructure to improve the robustness of inverse prediction. This strategy enables the parallel computing to increase computational efficiency in addressing the inverse ECG problem. The introduced P-DL framework holds broad applicability for exploring other high-dimensional spatiotemporal systems, including those related to thermodynamics and tumor growth.
- In Chapter 3, a novel physics-constrained deep active learning (P-DAL) framework is developed to model the spatiotemporal cardiac electrodynamics from the limited catheterization sensor observations. We first engage the P-DL framework presented in Chapter 2 to model the cardiac electrodynamics. Moreover, we propose a unique active learning approach that integrates both prediction uncertainty and a scheme based on geodesic distance for space-filling. This strategy is aimed at identifying informative spatial locations on the heart surface to enhance the predictive power of P-DL. The introduced active learning scheme has great potential to be implemented in other spatiotemporal systems for sequential optimization, such as soil moisture systems and thermal systems.
- In Chapter 4, a novel hierarchical Gaussian process (GP)-based active learning (HGPAL) framework is proposed for reliable defect localization in complex 3D systems. This framework not only can identify the defect region but also provide the estimation of spatial-varying parameters in a heterogeneous inverse physical system. The proposed HGPAL framework holds significant promise for achieving optimal defect detection, which can be crucial for the design of precision engineering systems and the formulation of healthcare treatment plans.

Building on my past research endeavors outlined above, my upcoming research intentions encompass:

- *Robust modeling of cardiac electrodynamics based on local spatiotemporal measurements:* The local spatiotemporal sensor measurements occur when sensor exploration is only applicable in a localized spatial domain of the complex systems, or the sensor signals are incomplete in the time domain. The local spatiotemporal measurements can lead to inaccurate prediction of spatiotemporal dynamics and patterns even when the model is constrained by physics law. The research objective is to develop a robust deep learning model that integrates the physical laws of the cardiac electric wave propagation, and further address the inaccurate prediction issue originating from the local sensor measurements by prompting effective physics enforcement.
- *Patient-specific detection and characterization of cardiac diseases for clinical decision support:* The research objective is to build an innovative physics-augmented algorithm that can identify and characterize cardiac diseases by taking the advantage of both sensing data and the physics-based electrophysiology model. Specifically, this research aims to extend my previous research in [2] to incorporate mechanistic modeling and patient-specific sensing data into machine learning for personalized disease detection and characterization.
- *Physics-augmented modeling in other 3D complex systems:* The research objective is to leverage the framework of physics-constrained deep learning and broaden its applications for robust modeling of various spatiotemporal 3D systems.
 - **Tumor growth systems:** The kinetics of tumor growth are closely related to current tumor density, cell carrying capacity, as well as the diffusion and proliferation rate of tumor cells. These properties are governed by a well-known mechanistic law called the Fisher-Kolmogorov model. However, the deterministic physics models with a small set of model parameters are insufficient to capture the patient-specific properties of the tumor growth dynamics. The objective of this research is to develop a novel physics-augmented deep learning model that is able to incorporate patient-specific tumor data into the mechanistic modeling for the precision prediction of tumor growth trajectory. The physics-augmented prediction of tumor growth trajectory can immensely benefit the

accurate evaluation of tumor aggressiveness, patient survival rate, and optimal therapeutical planning.

- **Soil moisture systems:** Water retention and hydraulic conductivity are important characteristics in soil moisture dynamics, which can be described by Richard’s equation. Involving such physics-based knowledge in deep learning can enforce a more accurate prediction of soil moisture, which can significantly benefit agricultural water management and the prediction of natural disasters. However, the limited sensor availability and distribution uncertainty can lead to non-ideal neural network prediction. The research goal is to propose a physics-augmented deep learning algorithm to reconstruct the water content distribution and develop a novel sequential design scheme that can help to make the wise selection of the hydraulic sensor location to promote the robust prediction.
- **3D-printing systems:** Fused filament fabrication (FFF) is one of the most widely used extrusion-based technology of additive manufacturing processes. The temperature and viscosity distribution of the extruding polymer is closely related to the coalescence and interdiffusion between deposited layers and beads, which has a critical impact on the resulting microstructure and mechanical resistance. Modeling polymer extrusion processes involves a multi-physics process which can be described by Navier-Stokes equations and advection-diffusion equation. However, traditional approaches for extrusion process modeling aim to numerically solve partial differential equations without accounting for experimental observations. My research objective is to develop physics-augmented machine learning models that account for both physical laws and sensor measurement to investigate the thermal effect and viscosity variation in a 3D FFF for effective process control.

Bibliography

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pages 265–283. [23](#)
- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*. [46](#), [47](#)
- Agathos, K., Chatzi, E., and Bordas, S. (2018). Multiple crack detection in 3d using a stable xfem and global optimization. *Computational mechanics*, 62(4):835–852. [63](#)
- Ahmad, F. B. and Anderson, R. N. (2021). The leading causes of death in the us for 2020. *JAMA*, 325(18):1829–1830. [2](#), [36](#)
- Aliev, R. R. and Panfilov, A. V. (1996a). Modeling of heart excitation patterns caused by a local inhomogeneity. *Journal of theoretical biology*, 181(1):33–40. [77](#)
- Aliev, R. R. and Panfilov, A. V. (1996b). A simple two-variable model of cardiac excitation. *Chaos, Solitons & Fractals*, 7(3):293–301. [14](#), [15](#), [77](#)

- Althoff, D., Rodrigues, L. N., and Bazame, H. C. (2021). Uncertainty quantification for hydrological models based on neural networks: the dropout ensemble. *Stochastic Environmental Research and Risk Assessment*, 35(5):1051–1067. [46](#)
- Arif, M., Malagore, I. A., and Afsar, F. A. (2012). Detection and localization of myocardial infarction using k-nearest neighbor classifier. *Journal of medical systems*, 36(1):279–289. [62](#)
- Astudillo, R. and Frazier, P. (2019). Bayesian optimization of composite functions. In *International Conference on Machine Learning*, pages 354–363. PMLR. [65](#)
- Ba, S. and Joseph, V. R. (2012). Composite gaussian process models for emulating expensive functions. *The Annals of Applied Statistics*, pages 1838–1860. [37](#)
- Balasubramanian, M. and Schwartz, E. L. (2002). The isomap algorithm and topological stability. *Science*, 295(5552):7–7. [70](#)
- Barr, R. C., Ramsey, M., and Spach, M. S. (1977). Relating epicardial to body surface potential distributions by means of transfer coefficients based on geometry measurements. *IEEE Transactions on biomedical engineering*, (1):1–11. [11](#), [16](#), [78](#)
- Baydin, A. G., Pearlmutter, B. A., Radul, A. A., and Siskind, J. M. (2018). Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18. [18](#), [109](#)
- Berg, J. and Nyström, K. (2019). Data-driven discovery of pdes in complex datasets. *Journal of Computational Physics*, 384:239–252. [13](#)
- Bond, R., Finlay, D., Nugent, C., and Moore, G. (2009). An xml format for storing body surface potential map recordings. In *2009 36th Annual Computers in Cardiology Conference (CinC)*, pages 233–236. IEEE. [9](#)
- Bonilla, E. V., Chai, K., and Williams, C. (2007). Multi-task gaussian process prediction. *Advances in neural information processing systems*, 20. [65](#)

- Both, G.-J., Choudhury, S., Sens, P., and Kusters, R. (2020). Deepmod: Deep learning for model discovery in noisy data. *Journal of Computational Physics*, page 109985. [13](#)
- Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. [21](#)
- Bourassa, M. G. (2005). The history of cardiac catheterization. *The Canadian journal of cardiology*, 21(12):1011–1014. [36](#)
- Bousselham, A., Bouattane, O., Youssfi, M., and Raihani, A. (2018). 3d brain tumor localization and parameter estimation using thermographic approach on gpu. *Journal of Thermal Biology*, 71:52–61. [62](#)
- Brochu, E., Cora, V. M., and De Freitas, N. (2010). A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*. [65](#), [67](#), [121](#)
- Brooks, D. H., Ahmad, G. F., MacLeod, R. S., and Maratos, G. M. (1999). Inverse electrocardiography by simultaneous imposition of multiple constraints. *IEEE Transactions on Biomedical Engineering*, 46(1):3–18. [10](#), [12](#)
- Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5):1190–1208. [75](#)
- Camara, O., Sermesant, M., Lamata, P., Wang, L., Pop, M., Relan, J., De Craene, M., Delingette, H., Liu, H., Niederer, S., et al. (2011). Inter-model consistency and complementarity: Learning from ex-vivo imaging and electrophysiological data towards an integrated understanding of cardiac physiology. *Progress in Biophysics and Molecular Biology*, 107(1):122–133. [14](#)
- Chan, T. M. (1996). Optimal output-sensitive convex hull algorithms in two and three dimensions. *Discrete & Computational Geometry*, 16(4):361–368. [118](#)
- Chen, H., Zheng, L., Al Kontar, R., and Raskutti, G. (2020a). Stochastic gradient descent in correlated settings: A study on gaussian processes. *Advances in neural information processing systems*, 33:2722–2733. [64](#)

- Chen, S., Wang, Z., Yao, B., and Liu, T. (2022). Prediction of diabetic retinopathy using longitudinal electronic health records. In *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, pages 949–954. IEEE. [61](#)
- Chen, X., Chen, X., Zhou, W., Zhang, J., and Yao, W. (2020b). The heat source layout optimization using deep learning surrogate modeling. *Structural and Multidisciplinary Optimization*, 62(6):3127–3148. [13](#)
- Chen, Y. and Yang, H. (2015). Sparse modeling and recursive prediction of space–time dynamics in stochastic sensor networks. *IEEE Transactions on Automation Science and Engineering*, 13(1):215–226. [9](#)
- Chen, Y. and Yang, H. (2016). Numerical simulation and pattern characterization of nonlinear spatiotemporal dynamics on fractal surfaces for the whole-heart modeling applications. *The European Physical Journal B*, 89(8):1–16. [48](#)
- Chen, Z., Liu, Y., and Sun, H. (2020c). Deep learning of physical laws from scarce data. *arXiv preprint arXiv:2005.03448*. [38](#)
- Chinchapatnam, P., Rhode, K. S., Ginks, M., Rinaldi, C. A., Lambiase, P., Razavi, R., Arridge, S., and Sermesant, M. (2008). Model-based imaging of cardiac apparent conductivity and local conduction velocity for diagnosis and planning of therapy. *IEEE Transactions on Medical Imaging*, 27(11):1631–1642. [63](#)
- Clayton, R. and Panfilov, A. (2008). A guide to modelling cardiac electrical activity in anatomically detailed ventricles. *Progress in biophysics and molecular biology*, 96(1-3):19–43. [50](#)
- Corral-Acero, J., Margara, F., Marciniak, M., Rodero, C., Loncaric, F., Feng, Y., Gilbert, A., Fernandes, J. F., Bukhari, H. A., Wajdan, A., et al. (2020). The “digital twin” to enable the vision of precision cardiology. *European heart journal*, 41(48):4556–4564. [10](#)
- Courtemanche, M., Ramirez, R. J., and Nattel, S. (1998). Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model. *American Journal of Physiology-Heart and Circulatory Physiology*, 275(1):H301–H321. [59](#)

- Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812. [65](#)
- Daw, A., Thomas, R. Q., Carey, C. C., Read, J. S., Appling, A. P., and Karpatne, A. (2020). Physics-guided architecture (pga) of neural networks for quantifying uncertainty in lake temperature modeling. In *Proceedings of the 2020 siam international conference on data mining*, pages 532–540. SIAM. [32](#), [40](#)
- Dawoud, F., Wagner, G. S., Moody, G., and Horáček, B. M. (2008). Using inverse electrocardiography to image myocardial infarction?reflecting on the 2007 physionet/computers in cardiology challenge. *Journal of electrocardiology*, 41(6):630–635. [11](#), [22](#), [78](#)
- Deng, Z., Tutunnikov, I., Averbukh, I. S., Thachuk, M., and Krems, R. (2020). Bayesian optimization for inverse problems in time-dependent quantum dynamics. *The Journal of Chemical Physics*, 153(16):164111. [65](#)
- Dhamala, J., Arevalo, H. J., Sapp, J., Horáček, B. M., Wu, K. C., Trayanova, N. A., and Wang, L. (2018). Quantifying the uncertainty in model parameters using gaussian process-based markov chain monte carlo in cardiac electrophysiology. *Medical image analysis*, 48:43–57. [38](#), [57](#)
- Dhamala, J., Arevalo, H. J., Sapp, J., Horacek, M., Wu, K. C., Trayanova, N. A., and Wang, L. (2017). Spatially adaptive multi-scale optimization for local parameter estimation in cardiac electrophysiology. *IEEE transactions on medical imaging*, 36(9):1966–1978. [61](#), [62](#), [63](#), [77](#)
- Dijkstra, E. W. et al. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271. [48](#), [71](#)
- Dower, G. E., Yakush, A., Nazzal, S. B., Jutzy, R. V., and Ruiz, C. E. (1988). Deriving the 12-lead electrocardiogram from four (easi) electrodes. *Journal of electrocardiology*, 21:S182–S187. [3](#)

- Ezzat, A. A., Jun, M., and Ding, Y. (2019). Spatio-temporal short-term wind forecast: A calibrated regime-switching method. *The annals of applied statistics*, 13(3):1484. [37](#)
- Feng, Y., Guo, Z., Dong, Z., Zhou, X.-Y., Kwok, K.-W., Ernst, S., and Lee, S.-L. (2017). An efficient cardiac mapping strategy for radiofrequency catheter ablation with active learning. *International journal of computer assisted radiology and surgery*, 12(7):1199–1207. [38](#)
- Gal, Y. and Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR. [40](#), [46](#)
- Ge, R., Huang, F., Jin, C., and Yuan, Y. (2015). Escaping from saddle points?online stochastic gradient for tensor decomposition. In *Conference on Learning Theory*, pages 797–842. [21](#)
- Ghosh, S. and Rudy, Y. (2009). Application of l1-norm regularization to epicardial potential solution of the inverse electrocardiography problem. *Annals of biomedical engineering*, 37(5):902–912. [11](#), [31](#)
- Giffard-Roisin, S., Jackson, T., Fovargue, L., Lee, J., Delingette, H., Razavi, R., Ayache, N., and Sermesant, M. (2016). Noninvasive personalization of a cardiac electrophysiology model from body surface potential mapping. *IEEE Transactions on Biomedical Engineering*, 64(9):2206–2218. [62](#)
- Goldberg, Y. (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57:345–420. [38](#)
- Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220. [22](#), [50](#), [78](#)
- Graham, R. L. (1972). An efficient algorithm for determining the convex hull of a finite planar set. *Info. Pro. Lett.*, 1:132–133. [118](#)

- Greensite, F. (2003). The temporal prior in bioelectromagnetic source imaging problems. *IEEE transactions on biomedical engineering*, 50(10):1152–1159. [10](#), [12](#)
- Gulrajani, R. M. (1998). The forward and inverse problems of electrocardiography. *IEEE Engineering in Medicine and Biology Magazine*, 17(5):84–101. [10](#), [36](#)
- Haissaguerre, M., Jaïs, P., Shah, D. C., Takahashi, A., Hocini, M., Quiniou, G., Garrigue, S., Le Mouroux, A., Le Métayer, P., and Clémenty, J. (1998). Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. *New England Journal of Medicine*, 339(10):659–666. [50](#)
- Han, Z., De, S., et al. (2019). A deep learning-based hybrid approach for the solution of multiphysics problems in electrosurgery. *Computer Methods in Applied Mechanics and Engineering*, 357:112603. [38](#)
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778. [38](#)
- Heidenreich, E. A., Ferrero, J. M., Doblaré, M., and Rodríguez, J. F. (2010). Adaptive macro finite elements for the numerical solution of monodomain equations in cardiac electrophysiology. *Annals of biomedical engineering*, 38(7):2331–2345. [57](#)
- Herring, N. and Paterson, D. (2006). Ecg diagnosis of acute ischaemia and infarction: past, present and future. *Journal of the Association of Physicians*, 99(4):219–230. [3](#), [9](#)
- Hu, Z., Du, D., and Du, Y. (2020). Gaussian process-based spatiotemporal modeling of electrical wave propagation in human atrium. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 2602–2605. IEEE. [38](#), [57](#)
- Huang, C., Ren, Y., McGuinness, E. K., Losego, M. D., Lively, R. P., and Joseph, V. R. (2021). Bayesian optimization of functional output in inverse problems. *Optimization and Engineering*, 22:2553–2574. [65](#)

- Iquebal, A. S. and Bukkapatnam, S. T. (2020). Consistent estimation of the max-flow problem: Towards unsupervised image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. [9](#)
- Jenkins, K. J., Walsh, E. P., Colan, S. D., Bergau, D. M., Saul, J. P., and Lock, J. E. (1993). Multipolar endocardial mapping of the right atrium during cardiac catheterization: description of a new technique. *Journal of the American College of Cardiology*, 22(4):1105–1110. [4](#), [36](#)
- Jin, X., Cai, S., Li, H., and Karniadakis, G. E. (2021). Nsfnets (navier-stokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations. *Journal of Computational Physics*, 426:109951. [13](#)
- Johannesson, G., Glaser, R., Lee, C., Nitao, J., and Hanley, W. (2005). Multi-resolution markov-chain-monte-carlo approach for system identification with an application to finite-element models. Technical report, Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States). [62](#)
- Kanagawa, M., Hennig, P., Sejdinovic, D., and Sriperumbudur, B. K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*. [70](#)
- Kang, S., Jin, R., Deng, X., and Kenett, R. S. (2021). Challenges of modeling and analysis in cybermanufacturing: a review from a machine learning and computation perspective. *Journal of Intelligent Manufacturing*, pages 1–14. [61](#)
- Karoui, A., Bear, L., Migerditichan, P., and Zemzemi, N. (2018). Evaluation of fifteen algorithms for the resolution of the electrocardiography imaging inverse problem using ex-vivo and in-silico data. *Frontiers in physiology*, 9:1708. [10](#)
- Keskar, N. S. and Socher, R. (2017). Improving generalization performance by switching from adam to sgd. *arXiv preprint arXiv:1712.07628*. [21](#)

- Kim, M. and Liu, K. (2020). A bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics. *IISE Transactions*, 53(3):326–340. [46](#)
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. [21](#)
- Kissas, G., Yang, Y., Hwuang, E., Witschey, W. R., Detre, J. A., and Perdikaris, P. (2020). Machine learning in cardiovascular flows modeling: Predicting arterial blood pressure from non-invasive 4d flow mri data using physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 358:112623. [13](#), [39](#)
- Koutsourelakis, P.-S. (2009). A multi-resolution, non-parametric, bayesian framework for identification of spatially-varying model parameters. *Journal of computational physics*, 228(17):6184–6211. [63](#), [69](#)
- Lacroix, D., Dubuc, M., Kus, T., Savard, P., Shenasa, M., and Nadeau, R. (1991). Evaluation of arrhythmic causes of syncope: correlation between holter monitoring, electrophysiologic testing, and body surface potential mapping. *American heart journal*, 122(5):1346–1354. [9](#)
- Lahoti, G., Chen, J., Yue, X., Yan, H., Ranjan, C., Qian, Z., Zhang, C., and Wang, B. (2021). Image decomposition-based sparse extreme pixel-level feature detection model with application to medical images. *IISE Transactions on Healthcare Systems Engineering*, 11(4):338–354. [61](#)
- Latz, J., Papaioannou, I., and Ullmann, E. (2018). Multilevel sequential monte carlo for bayesian inverse problems. *Journal of Computational Physics*, 368:154–178. [63](#)
- Lawrence, N. D. and Moore, A. J. (2007). Hierarchical gaussian process latent variable models. In *Proceedings of the 24th international conference on Machine learning*, pages 481–488. [65](#)
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444. [13](#)

- Lee, C., Wang, K., Wu, J., Cai, W., and Yue, X. (2021). Partitioned active learning for heterogeneous systems. *arXiv preprint arXiv:2105.08547*. 69
- Li, X., Jia, X., Yang, Q., and Lee, J. (2020). Quality analysis in metal additive manufacturing with deep learning. *Journal of Intelligent Manufacturing*, 31(8):2003–2017. 61
- Lian, C., Liu, M., Zhang, J., and Shen, D. (2018). Hierarchical fully convolutional network for joint atrophy localization and alzheimer’s disease diagnosis using structural mri. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):880–893. 61
- Lin, H., Li, B., Wang, X., Shu, Y., and Niu, S. (2019). Automated defect inspection of led chip using deep convolutional neural network. *Journal of Intelligent Manufacturing*, 30(6):2525–2534. 61
- Lin, Y., Liu, S., and Huang, S. (2018). Selective sensing of a heterogeneous population of units with dynamic health conditions. *IIEE Transactions*, 50(12):1076–1088. 9
- Liu, D. and Wang, Y. (2019). Multi-fidelity physics-constrained neural network and its application in materials modeling. *Journal of Mechanical Design*, 141(12). 13
- MacKay, D. J. (1992). Bayesian interpolation. *Neural computation*, 4(3):415–447. 39
- Maddox, W. J., Izmailov, P., Garipov, T., Vetrov, D. P., and Wilson, A. G. (2019). A simple baseline for bayesian uncertainty in deep learning. *Advances in Neural Information Processing Systems*, 32:13153–13164. 39
- Mao, H. (2015). Integrated cardiac electromechanics: Modeling and personalization. 37
- Markowitz, S. M. and Lerman, B. B. (2006). How to interpret electroanatomic maps. *Heart Rhythm*, 3(2):240–246. 44
- McInerney, T. and Terzopoulos, D. (1995). A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4d image analysis. *Computerized medical imaging and graphics*, 19(1):69–83. 37, 57

- Messnarz, B., Tilg, B., Modre, R., Fischer, G., and Hanser, F. (2004). A new spatiotemporal regularization approach for reconstruction of cardiac transmembrane potential patterns. *IEEE transactions on Biomedical Engineering*, 51(2):273–281. [12](#), [31](#)
- Miyato, T., Maeda, S.-i., Koyama, M., and Ishii, S. (2018). Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993. [46](#)
- Moreau-Villéger, V., Delingette, H., Sermesant, M., Ashikaga, H., McVeigh, E., and Ayache, N. (2006). Building maps of local apparent conductivity of the epicardium with a 2-d electrophysiological model of the heart. *IEEE Transactions on Biomedical Engineering*, 53(8):1457–1466. [77](#)
- Moriconi, R., Kumar, K. S., and Deisenroth, M. P. (2020). High-dimensional bayesian optimization with projections using quantile gaussian processes. *Optimization Letters*, 14:51–64. [66](#)
- Neal, R. M. (2012). *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media. [39](#)
- Niederer, S. A., Lumens, J., and Trayanova, N. A. (2019). Computational models in cardiology. *Nature Reviews Cardiology*, 16(2):100–111. [37](#)
- Olson, L. G. and Throne, R. D. (2010). Numerical simulation of an inverse method for tumour size and location estimation. *Inverse Problems in Science and Engineering*, 18(6):813–834. [62](#)
- Oster, H. S., Taccardi, B., Lux, R. L., Ershler, P. R., and Rudy, Y. (1997). Noninvasive electrocardiographic imaging: reconstruction of epicardial potentials, electrograms, and isochrones and localization of single and multiple electrocardiac events. *Circulation*, 96(3):1012–1024. [10](#)
- Oster, H. S., Taccardi, B., Lux, R. L., Ershler, P. R., and Rudy, Y. (1998). Electrocardiographic imaging: noninvasive characterization of intramural myocardial

- activation from inverse-reconstructed epicardial potentials and electrograms. *Circulation*, 97(15):1496–1507. [3](#), [9](#)
- Park, J., Lee, B., Kang, S., Kim, P. Y., and Kim, H. J. (2016). Online learning control of hydraulic excavators based on echo-state networks. *IEEE Transactions on Automation Science and Engineering*, 14(1):249–259. [13](#)
- Park, S. and Choi, S. (2010). Hierarchical gaussian process regression. In *Proceedings of 2nd Asian conference on machine learning*, pages 95–110. JMLR Workshop and Conference Proceedings. [64](#)
- Pastore, C. A., Tobias, N., Samesima, N., Martinelli Filho, M., Pedrosa, A., Nishioka, S., Douglas, R. A., Moreira, L. F. P., and Ramires, J. F. (2007). Ventricular electrical activation in cardiac resynchronization as characterized by body surface potential mapping. *Arquivos Brasileiros de Cardiologia*, 88:251–257. [3](#)
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic differentiation in pytorch. [46](#)
- Piri, S., Delen, D., Liu, T., and Zolbanin, H. M. (2017). A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble. *Decision Support Systems*, 101:12–27. [9](#)
- Plumlee, M., Joseph, V. R., and Yang, H. (2016). Calibrating functional parameters in the ion channel models of cardiac cells. *Journal of the American Statistical Association*, 111(514):500–509. [37](#)
- Raissi, M. (2018a). Deep hidden physics models: Deep learning of nonlinear partial differential equations. *The Journal of Machine Learning Research*, 19(1):932–955. [13](#)
- Raissi, M. (2018b). Deep hidden physics models: Deep learning of nonlinear partial differential equations. *The Journal of Machine Learning Research*, 19(1):932–955. [38](#)
- Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707. [38](#), [39](#), [109](#)

- Raissi, M., Yazdani, A., and Karniadakis, G. E. (2020). Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030. [39](#)
- Rao, C. and Liu, Y. (2020). Three-dimensional convolutional neural network (3d-cnn) for heterogeneous material homogenization. *Computational Materials Science*, 184:109850. [38](#)
- Rao, C., Sun, H., and Liu, Y. (2021a). Physics-informed deep learning for computational elastodynamics without labeled data. *Journal of Engineering Mechanics*, 147(8):04021043. [13](#), [61](#)
- Rao, C., Sun, H., and Liu, Y. (2021b). Physics-informed deep learning for computational elastodynamics without labeled data. *Journal of Engineering Mechanics*, 147(8):04021043. [39](#), [61](#)
- Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer School on Machine Learning*, pages 63–71. Springer. [19](#), [42](#)
- Rocsoreanu, C., Georgescu, A., and Giurgiteanu, N. (2012). *The FitzHugh-Nagumo model: bifurcation and dynamics*, volume 10. Springer Science & Business Media. [59](#)
- Roy, S., Menapace, W., Oei, S., Luijten, B., Fini, E., Saltori, C., Huijben, I., Chennakeshava, N., Mento, F., Sentelli, A., et al. (2020). Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound. *IEEE transactions on medical imaging*, 39(8):2676–2687. [61](#)
- Rudy, Y. and Burnes, J. E. (1999). Noninvasive electrocardiographic imaging. *Annals of Noninvasive electrocardiology*, 4(3):340–359. [9](#), [10](#), [11](#)
- Sadoughi, M. and Hu, C. (2019). Physics-based convolutional neural network for fault diagnosis of rolling element bearings. *IEEE Sensors Journal*, 19(11):4181–4192. [32](#)
- Sahli Costabal, F., Yang, Y., Perdikaris, P., Hurtado, D. E., and Kuhl, E. (2020). Physics-informed neural networks for cardiac activation mapping. *Frontiers in Physics*, 8:42. [13](#), [39](#), [40](#)

- Schulman, K. A., Berlin, J. A., Harless, W., Kerner, J. F., Sistrunk, S., Gersh, B. J., Dube, R., Taleghani, C. K., Burke, J. E., Williams, S., et al. (1999). The effect of race and sex on physicians’ recommendations for cardiac catheterization. *New England Journal of Medicine*, 340(8):618–626. [13](#)
- Seeger, M. (2004). Gaussian processes for machine learning. *International journal of neural systems*, 14(02):69–106. [18](#)
- Sermesant, M., Delingette, H., and Ayache, N. (2006). An electromechanical model of the heart for image analysis and simulation. *IEEE transactions on medical imaging*, 25(5):612–625. [14](#)
- Sharma, L., Tripathy, R., and Dandapat, S. (2015). Multiscale energy and eigenspace approach to detection and localization of myocardial infarction. *IEEE transactions on biomedical engineering*, 62(7):1827–1837. [62](#)
- Shen, B., Gnanasambandam, R., Wang, R., and Kong, Z. J. (2022). Multi-task gaussian process upper confidence bound for hyperparameter tuning and its application for simulation studies of additive manufacturing. *IISE Transactions*, pages 1–13. [65](#)
- Shi, J. Q., Murray-Smith, R., and Titterington, D. M. (2005). Hierarchical gaussian process mixtures for regression. *Statistics and computing*, 15:31–41. [65](#)
- Shi, S., Wang, Q., Xu, P., and Chu, X. (2016). Benchmarking state-of-the-art deep learning software tools. In *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, pages 99–104. IEEE. [23](#)
- Shou, G., Xia, L., Liu, F., Jiang, M., and Crozier, S. (2010). On epicardial potential reconstruction using regularization schemes with the l1-norm data term. *Physics in Medicine & Biology*, 56(1):57. [11](#), [31](#)
- Si, B., Yakushev, I., and Li, J. (2017). A sequential tree-based classifier for personalized biomarker testing of alzheimer’s disease risk. *IISE Transactions on Healthcare Systems Engineering*, 7(4):248–260. [9](#)

- Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. (2009). Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995*. [20](#), [65](#), [68](#), [114](#)
- Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. W. (2012). Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265. [20](#)
- Sundnes, J., Nielsen, B. F., Mardal, K. A., Cai, X., Lines, G. T., and Tveito, A. (2006). On the computational complexity of the bidomain and the monodomain models of electrophysiology. *Annals of biomedical engineering*, 34(7):1088–1097. [37](#), [57](#)
- Swersky, K., Snoek, J., and Adams, R. P. (2013). Multi-task bayesian optimization. *Advances in neural information processing systems*, 26. [65](#), [120](#)
- Ten Tusscher, K. H., Mourad, A., Nash, M., Clayton, R. H., Bradley, C. P., Paterson, D. J., Hren, R., Hayward, M., Panfilov, A. V., and Taggart, P. (2009). Organization of ventricular fibrillation in the human heart: experiments and models. *Experimental physiology*, 94(5):553–562. [50](#)
- Thompson, S. M., Bian, L., Shamsaei, N., and Yadollahi, A. (2015). An overview of direct laser deposition for additive manufacturing; part i: Transport phenomena, modeling and diagnostics. *Additive Manufacturing*, 8:36–62. [61](#)
- Throne, R. D. and Olson, L. G. (2000). Fusion of body surface potential and body surface laplacian signals for electrocardiographic imaging. *IEEE Transactions on biomedical engineering*, 47(4):452–462. [11](#)
- Trayanova, N. (2019). From genetics to smart watches: developments in precision cardiology. *Nature Reviews Cardiology*, 16(2):72–73. [10](#)
- Trayanova, N. A. (2011). Whole-heart modeling: applications to cardiac electrophysiology and electromechanics. *Circulation research*, 108(1):113–128. [50](#)

- Tripathy, R. K., Bhattacharyya, A., and Pachori, R. B. (2019). Localization of myocardial infarction from multi-lead ecg signals using multiscale analysis and convolutional neural network. *IEEE Sensors Journal*, 19(23):11437–11448. [62](#)
- Ulmasov, D., Baroukh, C., Chachuat, B., Deisenroth, M. P., and Misener, R. (2016). Bayesian optimization with dimension scheduling: Application to biological systems. In *Computer Aided Chemical Engineering*, volume 38, pages 1051–1056. Elsevier. [65](#)
- Urbich, M., Globe, G., Pantiri, K., Heisen, M., Bennison, C., Wirtz, H. S., and Di Tanna, G. L. (2020). A systematic review of medical costs associated with heart failure in the usa (2014–2020). *Pharmacoeconomics*, 38:1219–1236. [2](#)
- Van Wyk, F., Khojandi, A., and Kamaleswaran, R. (2019). Improving prediction performance using hierarchical analysis of real-time data: a sepsis case study. *IEEE journal of biomedical and health informatics*, 23(3):978–986. [61](#)
- Wan, J. and Zabaras, N. (2011). A bayesian approach to multiscale inverse problems using the sequential monte carlo method. *Inverse Problems*, 27(10):105004. [63](#)
- Wang, D., Kirby, R. M., and Johnson, C. R. (2011). Finite-element-based discretization and regularization strategies for 3-d inverse electrocardiography. *IEEE Transactions on Biomedical Engineering*, 58(6):1827–1838. [11](#)
- Wang, D., Liu, K., and Zhang, X. (2019a). Modeling of a three-dimensional dynamic thermal field under grid-based sensor networks in grain storage. *IIEE Transactions*, 51(5):531–546. [37](#)
- Wang, D., Liu, K., and Zhang, X. (2021a). A generic indirect deep learning approach for multisensor degradation modeling. *IEEE Transactions on Automation Science and Engineering*. [13](#)
- Wang, F., Zhuang, Y., Gu, H., and Hu, H. (2019b). Octreenet: a novel sparse 3-d convolutional neural network for real-time 3-d outdoor scene analysis. *IEEE Transactions on Automation Science and Engineering*, 17(2):735–747. [13](#)

- Wang, J., Ma, Y., Zhang, L., Gao, R. X., and Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, 48:144–156. [15](#)
- Wang, K., Pleiss, G., Gardner, J., Tyree, S., Weinberger, K. Q., and Wilson, A. G. (2019c). Exact gaussian processes on a million data points. *Advances in neural information processing systems*, 32. [64](#)
- Wang, L., Hawkins-Daarud, A., Swanson, K. R., Hu, L. S., and Li, J. (2021b). Knowledge-infused global-local data fusion for spatial predictive modeling in precision medicine. *IEEE Transactions on Automation Science and Engineering*. [13](#), [61](#)
- Wang, L., Wong, K. C., Zhang, H., Liu, H., and Shi, P. (2010). Noninvasive computational imaging of cardiac electrophysiology for 3-d infarct. *IEEE Transactions on biomedical engineering*, 58(4):1033–1043. [61](#)
- Wang, L., Zhang, H., Wong, K. C., Liu, H., and Shi, P. (2009). Physiological-model-constrained noninvasive reconstruction of volumetric myocardial transmembrane potentials. *IEEE Transactions on Biomedical Engineering*, 57(2):296–315. [10](#), [12](#), [22](#), [27](#), [28](#), [31](#), [110](#), [111](#)
- Wang, R., Walters, R., and Yu, R. (2020). Incorporating symmetry into deep dynamics models for improved generalization. *arXiv preprint arXiv:2002.03061*. [32](#)
- Wang, Z., Stavrakis, S., and Yao, B. (2023). Hierarchical deep learning with generative adversarial network for automatic cardiac diagnosis from ecg signals. *Computers in Biology and Medicine*, 155:106641. [62](#)
- Wang, Z. and Yao, B. (2021). Multi-branching temporal convolutional network for sepsis prediction. *IEEE Journal of Biomedical and Health Informatics*. [15](#), [38](#), [61](#)
- Wang, Z., Zoghi, M., Hutter, F., Matheson, D., De Freitas, N., et al. (2013). Bayesian optimization in high dimensions via random embeddings. In *IJCAI*, volume 13, pages 1778–1784. [65](#)

- Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V. (2020). Integrating physics-based modeling with machine learning: A survey. *arXiv preprint arXiv:2003.04919*. [32](#)
- Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA. [67](#)
- Winter, J., Abaidi, R., Kaiser, J., Adami, S., and Adams, N. (2023). Multi-fidelity bayesian optimization to solve the inverse stefan problem. *Computer Methods in Applied Mechanics and Engineering*, 410:115946. [65](#)
- Xia, H., Ding, Y., and Wang, J. (2008). Gaussian process method for form error assessment using coordinate measurements. *Iie Transactions*, 40(10):931–946. [37](#)
- Xiao, B., Xu, W., Guo, J., Lam, H.-K., Jia, G., Hong, W., and Ren, H. (2020). Depth estimation of hard inclusions in soft tissue by autonomous robotic palpation using deep recurrent neural network. *IEEE Transactions on Automation Science and Engineering*, 17(4):1791–1799. [15](#)
- Xie, J. and Yao, B. (2022a). Physics-constrained deep active learning for spatiotemporal modeling of cardiac electrodynamics. *Computers in Biology and Medicine*, 146:105586. [35](#), [37](#), [44](#), [70](#)
- Xie, J. and Yao, B. (2022b). Physics-constrained deep learning for robust inverse ecg modeling. *IEEE Transactions on Automation Science and Engineering*. [8](#), [14](#), [39](#), [40](#), [43](#), [44](#), [55](#), [56](#), [61](#), [114](#)
- Yan, C., Yao, J., Li, R., Xu, Z., and Huang, J. (2018). Weakly supervised deep learning for thoracic disease classification and localization on chest x-rays. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, pages 103–110. [61](#)
- Yang, H., Kan, C., Liu, G., and Chen, Y. (2013). Spatiotemporal differentiation of myocardial infarctions. *IEEE Transactions on Automation Science and Engineering*, 10(4):938–947. [3](#), [9](#)

- Yang, L., Meng, X., and Karniadakis, G. E. (2021). B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913. [40](#)
- Yao, B. (2021). Spatiotemporal modeling and optimization for personalized cardiac simulation. *IISE Transactions on Healthcare Systems Engineering*, 11(2):145–160. [10](#), [37](#), [61](#), [119](#)
- Yao, B., Chen, Y., and Yang, H. (2021). Constrained markov decision process modeling for optimal sensing of cardiac events in mobile health. *IEEE Transactions on Automation Science and Engineering*, pages 1–13. [9](#)
- Yao, B., Chen, Y., and Yang, H. (2021). Constrained markov decision process modeling for optimal sensing of cardiac events in mobile health. *IEEE Transactions on Automation Science and Engineering*. [36](#)
- Yao, B., Imani, F., Sakpal, A. S., Reutzel, E., and Yang, H. (2018). Multifractal analysis of image profiles for the characterization and detection of defects in additive manufacturing. *Journal of Manufacturing Science and Engineering*, 140(3). [61](#)
- Yao, B., Pei, S., and Yang, H. (2016). Mesh resolution impacts the accuracy of inverse and forward ecg problems. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4047–4050. IEEE. [10](#)
- Yao, B. and Yang, H. (2016). Physics-driven spatiotemporal regularization for high-dimensional predictive modeling: A novel approach to solve the inverse ecg problem. *Scientific reports*, 6(1):1–13. [12](#), [78](#)
- Yao, B. and Yang, H. (2018). Constrained markov decision process modeling for sequential optimization of additive manufacturing build quality. *IEEE Access*, 6:54786–54794. [37](#)
- Yao, B. and Yang, H. (2020). Spatiotemporal regularization for inverse ecg modeling. *IISE Transactions on Healthcare Systems Engineering*, pages 1–13. [10](#), [11](#), [12](#), [16](#), [78](#)

- Yao, B., Zhu, R., and Yang, H. (2017). Characterizing the location and extent of myocardial infarctions with inverse ecg modeling and spatiotemporal regularization. *IEEE journal of biomedical and health informatics*, 22(5):1445–1455. [9](#)
- Yin, X., Chen, Y., Bouferguene, A., Zaman, H., Al-Hussein, M., and Kurach, L. (2020). A deep learning-based framework for an automated defect detection system for sewer pipes. *Automation in construction*, 109:102967. [61](#)
- Yu, L., Zhou, Z., and He, B. (2015). Temporal sparse promoting three dimensional imaging of cardiac activation. *IEEE transactions on medical imaging*, 34(11):2309–2319. [10](#), [12](#)
- Yu, Q., Shang, W., Zhao, Z., Cong, S., and Li, Z. (2020). Robotic grasping of unknown objects using novel multilevel convolutional neural networks: From parallel gripper to dexterous hand. *IEEE Transactions on Automation Science and Engineering*, pages 1–12. [13](#)
- Zhang, D., Lu, L., Guo, L., and Karniadakis, G. E. (2019a). Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. *Journal of Computational Physics*, 397:108850. [40](#)
- Zhang, R., Chen, Z., Chen, S., Zheng, J., Büyüköztürk, O., and Sun, H. (2019b). Deep long short-term memory networks for nonlinear structural seismic response prediction. *Computers & Structures*, 220:55–68. [38](#)
- Zhang, R., Liu, Y., and Sun, H. (2020). Physics-informed multi-lstm networks for metamodeling of nonlinear structures. *Computer Methods in Applied Mechanics and Engineering*, 369:113226. [39](#)
- Zhao, S. and Huang, B. (2020). Trial-and-error or avoiding a guess? initialization of the kalman filter. *Automatica*, 121:109184. [12](#), [27](#)
- Zhao, X., Yan, H., Hu, Z., and Du, D. (2022). Deep spatio-temporal sparse decomposition for trend prediction and anomaly detection in cardiac electrical conduction. *IJSE Transactions on Healthcare Systems Engineering*, 12(2):150–164. [62](#)

- Zheng, W., Liu, H., Wang, B., and Sun, F. (2019). Cross-modal material perception for novel objects: a deep adversarial learning method. *IEEE Transactions on Automation Science and Engineering*, 17(2):697–707. [13](#)
- Zheng, Z., Yan, H., Setzer, F. C., Shi, K. J., Mupparapu, M., and Li, J. (2020). Anatomically constrained deep learning for automating dental cbct segmentation and lesion detection. *IEEE Transactions on Automation Science and Engineering*, 18(2):603–614. [13](#)
- Zhu, R., Yao, B., Yang, H., and Leonelli, F. (2018). Optimal sensor placement for space–time potential mapping and data fusion. *IEEE sensors letters*, 3(1):1–4. [9](#), [36](#)

Appendices

Appendix A

Supplement Materials of Chapter 2

A.1 Automatic Differentiation

The partial differential terms in the AP model can be easily calculated by automatic differentiation with no approximation error (Baydin et al., 2018; Raissi et al., 2019), which is widely used in DL to compute the derivatives by applying the chain rule repeatedly. Automatic differentiation consists of two major steps: first compute the values of all variables through a forward pass and second calculate the derivatives through a backward pass. We use a simple network example in Fig. D.24 to demonstrate this technique. The network consists of one hidden layer with inputs x and t , and one output u . Table E.8 shows the forward and backward pass of automatic differentiation to compute the partial derivatives of $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial t}$ at the point $(x, t) = (2, 3)$.

A.2 The optimal regularization parameter w_{opt}

Table E.9 lists the optimal parameter value w_{opt} and the iteration number to converge for each NN structure. The GP-UCB algorithm takes 20 iterations to converge for the structure NN1, NN2, and NN3, yielding the optimal parameter value of 0.23, 0.29, and 0.44,

respectively. The GP-UCB for the P-DL with the structure of NN4 and NN5 takes 17 and 19 iterations to converge, with the value of 0.56 and 0.33 for w_{opt} , respectively. Fig. D.25 shows an example of the convergence of GP-UCB. The regularization parameter w converges after 20 iterations for the structure of NN2.

A.3 Physiological-Model-constrained Kalman filter (P-KF)

Wang et al. (2009) proposed a P-KF statistical framework to estimate the cardiac electrical potential from BSPM. The state space representation of the inverse ECG model in the P-KF framework is given as:

$$\begin{aligned} (\mathbf{u}_t, \mathbf{v}_t) &= \mathcal{F}(\mathbf{u}_{t-1}, \mathbf{v}_{t-1}) + \boldsymbol{\phi}_t \\ \mathbf{y}_t &= \mathbf{R}\mathbf{u}_t + \boldsymbol{\varphi}_t \end{aligned} \tag{A.1}$$

where $\mathcal{F}$ denotes the AP model, and $\boldsymbol{\phi}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_\phi)$ and $\boldsymbol{\varphi}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{M}_\varphi)$ stand for the model uncertainty and data error at time t , respectively. In order to solve the inverse ECG problem, this method sequentially estimated the electric potential of the heart (i.e., $\mathbf{u}_t$) by iteratively assimilating the body-surface data (i.e., $\mathbf{y}_t$) into the solution procedure.

Specifically, at each iteration, the P-KF first tackles the nonlinearity of the physics-based cardiac model using the unscented transformation (UT) that is adapted from the unscented Kalman filter (UKF). A set of sigma points $\{\mathbf{u}_{t-1,i}\}_0^{2N}$ (where N is the dimension of the state vector $\mathbf{u}_{t-1}$, i.e., the total number of the discretized spatial nodes on the heart geometry) is generated according to the estimates of HSP $\hat{\mathbf{u}}_{t-1}$, and covariance matrix $\hat{\mathbf{P}}_{\mathbf{u}_{t-1}}$ at the previous time instance $t-1$ in a deterministic manner defined by UT (see step 2 in Algorithm 1). The sigma points at time $t-1$ are then dynamically propagated to generate the *priori* prediction of the cardiac electrodynamics $(\mathbf{u}_{t|t-1,i}, \mathbf{v}_{t|t-1,i})$ at time t by numerically solving the AP model using finite element method. The statistical prediction $\bar{\mathbf{u}}_t^-$ is then formed by taking the weighted sample mean of the new ensemble set $\{\mathbf{u}_{t|t-1,i}\}_0^{2N}$ and the statistical covariance $\mathbf{P}_{\mathbf{u}_t}^-$ is computed as the weighted sample covariance added by a noise matrix $\mathbf{Q}_\phi$ which is adopted from Wang et al's work (see step 3 in Algorithm 1). The *posteriori* correction (i.e.,

the data assimilation) given the BSPM $\mathbf{y}_t$ at time t and the noise matrix $\mathbf{M}_\varphi$ (adopted from Wang et al's work) is realized by implementing the KF updates (see step 4 in Algorithm 1). The above process is iterated T times, where T is the total number of the discretized time instances.

Algorithm 2: P-KF algorithm (adapted from Wang et al. (2009))

- 1: Initialization: $\hat{\mathbf{u}}_0(\mathbf{s})$
- 2: Generate sigma points and the corresponding weights
 - Sigma points $\mathbf{u}_{t-1,i} (i = 0, 2, \dots, 2N)$

$$\{\mathbf{u}_{t-1,i}\}_0^{2N} = \left(\hat{\mathbf{u}}_{t-1} \quad \hat{\mathbf{u}}_{t-1} \pm \sqrt{(N + \gamma) \hat{\mathbf{P}}_{\mathbf{u}_{t-1}}} \right)$$

- Weights W_i

$$\begin{aligned} W_0^m &= \lambda / (N + \gamma) \\ W_0^c &= \lambda / (N + \gamma) + (1 - p^2 + q) \\ W_i^m &= W_i^c = 1 / (2(N + \gamma)) \quad i = 1 \dots 2N \end{aligned}$$

- 3: AP model-based statistical *priori* prediction

- AP model applicaton

$$\left(\mathbf{u}_{t|t-1,i}, \mathbf{v}_{t|t-1,i} \right) = \mathcal{F} \left(\mathbf{u}_{t-1,i}, \bar{\mathbf{v}}_{t-1} \right)$$

- Statistical Prediction

$$\begin{aligned} \bar{\mathbf{u}}_t^- &= \sum_{i=0}^{2N} W_i^m \mathbf{u}_{t|t-1,i}, \quad \bar{\mathbf{v}}_t = \sum_{i=0}^{2N} W_i^m \mathbf{v}_{t|t-1,i} \\ \mathbf{P}_{\mathbf{u}_t}^- &= \sum_{i=0}^{2N} W_i^c \left(\mathbf{u}_{t|t-1,i} - \bar{\mathbf{u}}_t^- \right) \left(\mathbf{u}_{t|t-1,i} - \bar{\mathbf{u}}_t^- \right)^T + \mathbf{Q}_\phi \end{aligned}$$

- 4: Kalman filter-based *posteriori* correction

- Kalman gain: $\mathbf{K}_t = \mathbf{P}_{\mathbf{u}_t}^- \mathbf{R}^T \left(\mathbf{R} \mathbf{P}_{\mathbf{u}_t}^- \mathbf{R}^T + \mathbf{M}_\varphi \right)^{-1}$
 - Mean estimation: $\hat{\mathbf{u}}_t = \bar{\mathbf{u}}_t^- + \mathbf{K}_t (\mathbf{y}_t - \mathbf{R} \bar{\mathbf{u}}_t^-)$
 - Covariance update: $\hat{\mathbf{P}}_{\mathbf{u}_t} = (\mathbf{I} - \mathbf{K}_t \mathbf{R}) \mathbf{P}_{\mathbf{u}_t}^-$
-

A.4 Statistical test for the comparison of RE generated from P-DL and P-KF

We formulate the following hypothesis testing to evaluate if the difference in predictive performance between our P-DL method and the P-KF algorithm with the perfect initial condition (i.e., $\hat{\mathbf{u}}_0(\mathbf{s}) = \mathbf{u}_0(\mathbf{s})$) is statistically significant:

$$H_0 : \mu_A = \mu_B \quad (\text{A.2})$$

$$H_1 : \mu_A \neq \mu_B \quad (\text{A.3})$$

where μ_A and μ_B denote the population means of RE yielded from P-DL and P-KF, respectively. H_0 assumes the RE from both methods does not have a significant difference, while H_1 believes otherwise. We engage a two-sample T -test to calculate the test statistics by assuming that the calculated RE follows a Gaussian distribution and has an unknown population variance. As such, the test statistics is given as

$$T_0^* = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{S_A^2/n_A + S_B^2/n_B}} \quad (\text{A.4})$$

where $\bar{X}_A = 0.1490$ and $\bar{X}_B = 0.1426$ denote the sample mean of RE yielded from P-DL and P-KF respectively, $S_A = 0.0123$ and $S_B = 1e - 5$ denote the sample standard deviations, and $n_A = n_B = 10$ is the repetition number for the experiment. If H_0 is true, T_0^* will follow a t -distribution with the degrees of freedom given by

$$\nu = \frac{(S_A^2/n_A + S_B^2/n_B)^2}{(S_A^2/n_A)^2/(n_A - 1) + (S_B^2/n_B)^2/(n_B - 1)} \quad (\text{A.5})$$

The resulted p-value is 0.1343, which is greater than 0.01. This implies the null hypothesis cannot be rejected with a significance level of 0.01. As such, it is legitimate to conclude that there is no significant difference between the performance of our P-DL approach and the P-KF method with the perfect initial condition when solving the inverse ECG problem.

Appendix B

Supplement Materials of Chapter 3

B.1 Hyperparameter Section

B.1.1 Hyperparameter selection for neural network structures and regularization parameter w

It should be noted that the heart geometry is formed by 1094 discretized nodes, which means there are 1094 spatial locations for possible sensor measurements. In order to characterize the sparse nature of sensor observations, data from 75 randomly selected spatial locations out of the 1094 locations (i.e., 75 time series) are utilized for hyper-parameter selection. Specifically, 80% of this dataset is used to train our P-DL model, and the remaining 20% serves as the validation set, generating a validation relative error (i.e., RE_v) to evaluate the predictive performance for hyper-parameter optimization.

- *The selection of neural network structures:* We have investigated the influence of the neural network structure on the predictive performance in both the healthy and diseased heart systems. Specifically, five different feed-forward neural network structures are evaluated. The structural detail is shown in Table [E.10](#).

The resulted RE_v for the healthy and diseased heart systems is shown in Fig. D.26 (a) and (b), respectively. Note that each experiment is repeated 6 times, and the resulted RE_v is shown with a corresponding error bar (i.e., $\pm$ the standard deviation). As the network structure becomes more complex, RE_v drops for both systems. Specifically, when the network complexity increases from NN1 to NN5, RE_v decreases from 0.1471 ± 0.0150 and 0.4517 ± 0.0641 to 0.1254 ± 0.0139 and 0.3083 ± 0.0332 for the healthy and diseased conditions, respectively. No salient reduction of RE_v is observed when further increasing network complexity from NN3 to NN5 for both systems. Thus, we engage NN3 as the neural network structure of both systems for later experiments.

- *Regularization parameter w* : Hyperparameter w in Eq. (3.8) controls the regularization effect of the physics law, which is actively chosen by GP-based Bayesian optimization. Specifically, the model input is the regularization parameter, i.e., w 's, and the output is the corresponding RE_v 's of the validation set estimated by P-DL. Here, RE_v is considered as an indicator to evaluate the performance of the P-DL, which can be viewed as a black-box function of w , i.e., $RE_v(w)$'s. It is assumed $RE_v(w) = \widehat{RE}_v(w) + \delta_{re}$, where δ_{re} is a noise term added for numerical stability. We further model $\widehat{RE}_v(w)$ with a GP, i.e., $\widehat{RE}_v(w) \sim \mathcal{GP}(0, \mathcal{K})$, where $\mathcal{K}$ denotes a commonly used square-exponential kernel function to capture the correlation between any two w 's.

Given a set of n observation pairs $\{\mathbf{w}_{1:n}, \mathbf{RE}_{v1:n}\}$, where $\mathbf{w}_{1:n} = [w_1 \dots w_n]^T$, $\mathbf{RE}_{v1:n} = [RE_{v1} \dots RE_{vn}]^T$, the prediction of $RE_v(w_*)$ given a new w_* follows a Gaussian distribution: $RE_v(w_*) | \mathbf{w}_{1:n}, \mathbf{RE}_{v1:n}, w_* \sim \mathcal{N}(\mu_n(w_*), \sigma_n^2(w_*))$, where I denotes the $n \times n$ identity matrix. To find the next query point of w_{n+1} that can potentially further decrease RE_v , we engage GP-upper-confidence-bound (GP-UCB) as the acquisition function to balance between the exploitation of the known information and exploration of the uncertainty. The GP-UCB selection rule can be mathematically expressed as $w_{n+1} = \operatorname{argmax}_w \{-\mu_n(w) + \beta^{1/2} \sigma_n(w)\}$, where β is set as 9 in this study, μ_n and σ_n stand for the posterior mean and variation at the n th GP evaluation. The detailed GP-UCB algorithm is demonstrated in Xie and Yao (2022b) and Srinivas et al. (2009).

The GP-UCB evaluation is repeated until the regularization parameter w converges. As shown in Fig. D.27 (a) and (b), it takes about 27 iterations for the algorithm to converge in both healthy and diseased heart systems, where the regularization parameter w converges at 0.95 and 0.13, respectively.

B.1.2 The influence of parameter λ on the active learning performance

The active learning criterion of the proposed P-DAL combines both uncertainty estimation and the dispersion level of the observed spatial locations on the heart surface, as formulated in Eq. (3.17), whose effects are balanced by the hyperparameter λ . In the current investigation, we empirically search for the proper λ value. Specifically, five candidates, i.e., $\lambda \in \{0.1, 0.5, 1, 2, 3\}$, are chosen to test their effectiveness on the identification of informative spatial locations from the heart surface. Fig. D.28 depicts the estimated RE 's of the P-DAL model with different λ 's in the healthy (Fig. D.28(a)) and diseased (Fig. D.28(b)) heart systems as the number of observed spatial locations increases. Table E.11 further listed the final $RE \pm 1$ standard deviation based on observations from 50 spatial locations selected by P-DAL with different λ values for both systems. For the healthy condition, the P-DAL model demonstrates the lowest RE (i.e., 0.0919) when $\lambda = 2$ after the data collection completes. As λ decreases from 2 to 0.1 or increases from 2 to 3, the predictive performance becomes worse, due to the over-consideration (when λ is too small) or under-consideration (when λ is too big) of predictive uncertainty. In the diseased heart system, the P-DAL model generates a significantly lower RE (i.e., 0.2588) when $\lambda = 0.5$ compared with the other candidates. It is worth noting that the optimal value for λ is smaller in the diseased system (i.e., $\lambda = 0.5$) compared with the healthy system ($\lambda = 2$). In other words, the predictive uncertainty is weighed more than the space-filling design in the diseased system. This is due to the fact that the electrical dynamic process is more chaotic in a diseased heart, generating a bigger predictive uncertainty, which leads to a greater impact of uncertainty on the location selection.

B.2 Statistical test to compare the results from P-DAL with P-DSL and P-DRL

The experiments in the active learning process have been repeated 6 times to test the statistical significance. Specifically, in each experiment, twenty initial spatial locations are randomly selected from the heart geometry and the AL criterion is engaged to make a judicious sequential selection. Fifty time series of the HSP data, corresponding to the electrical signals from 50 locations on the heart surface, are eventually collected in each experiment. We compare the *RE* generated from P-DAL, P-DSL, and P-DRL with a noise level of $\sigma_\epsilon = 0.01$ added to the observations. The resulted *RE* ± 1 standard deviation for the health and diseased heart systems are summarized in Table E.12 .

We further implement the two-sample T-test to identify if significant differences exist between *RE* generated from P-DAL with P-DSL and P-DRL for both healthy and diseased heart systems. In order to statistically verify that the *RE* produced by P-DAL is smaller than *RE* from P-DSL and P-DRL. We formulate the hypothesis testing as:

$$H_0 : \mu_A \geq \mu_B \tag{B.1}$$

$$H_1 : \mu_A < \mu_B \tag{B.2}$$

where μ_A and μ_B denote the population means of *RE* generated from P-DAL and P-DSL (or P-DRL), respectively. The null hypothesis (H_0) assumes *RE* generated from the P-DAL is greater or equal than that from the P-DSL and P-DRL. The alternative hypothesis H_1 indicates otherwise. It is presumed that the estimated *RE* generated from each method follows Gaussian distribution and has unknown population variance.

The resulted *p*-values for each comparison are displayed in Table E.13. All the *p*-values are lower than 0.01, indicating that the null hypothesis H_0 should be rejected at a confidence level larger than 99%. Therefore, we can statistically conclude that *RE* generated from the P-DAL approach is smaller than the one estimated by the P-DSL or P-DRL method for both healthy and diseased heart systems. The statistical test strongly proves that the proposed

P-DAL method indeed provides an effective instruction for searching the most informative spatial locations on the heart surface to model spatiotemporal cardiac electrodynamics.

Appendix C

Supplement Materials of Chapter 4

C.1 Refinement Procedure in HGPAL

We first denote the estimated defect location, the corresponding system parameter, and the value of the objective function from the HGPAL framework as $\boldsymbol{\gamma}_0^*$, $\boldsymbol{\theta}_0^*$, and J_0^* . We further denote the iterable defect area of refinement as R_{re} , which contains a set of defect spatial coordinates projected from the 3D defect nodes on the 2D embedded space. R_{re} is initialized by $R_{re,0} = R_0$ at iteration $q = 0$, where $R_0 = \{\hat{\mathbf{s}}_i \mid \|\hat{\mathbf{s}}_i - \boldsymbol{\gamma}_0^{*,1:2}\| \leq \boldsymbol{\gamma}_0^{*,3}, \forall i \in \{1, \dots, N_G\}\}$. The iterable objective function value for refinement is denoted as J_{re} , which is initiated as $J_{re,0} = J_0^*$. Similarly, the iterable model parameter $\boldsymbol{\theta}_{re}$ is initialized as $\boldsymbol{\theta}_{re,0} = \boldsymbol{\theta}_0^*$.

During the refinement procedure, at iteration q , we first identify the circumference of the defect area $R_{re,q}$ and its corresponding model parameter set $\boldsymbol{\theta}_{re,q}$. The circumference consists of a set of spatial points on the defect border, denoted as $\mathbf{E} = \{e_1, \dots, e_p\}$, where p is the total number of the edge points and $\mathbf{E}$ can be determined by Graham's scan algorithm (Graham, 1972; Chan, 1996). Next, for each node e_i ($i = \{1, \dots, p\}$) in $\mathbf{E}$, we locate its w nearest neighbors: $\eta_{e_i,\kappa}$'s, where $\kappa = \{1, \dots, w\}$. Second, we adjust the defect area $R_{re,q}$ by updating the defect status of each neighbor $\eta_{e_i,\kappa}$. For example, if $\eta_{e_i,\kappa} \in R_{re,q}$, we remove it from the defect area, i.e., defining the transient defect area as $R'_{q,e_i,\kappa} = \{\hat{\mathbf{s}} : \hat{\mathbf{s}} \in R_{re,q}, \hat{\mathbf{s}} \neq$

$\eta_{e_i, \kappa}$ }; if $\eta_{e_i, \kappa} \notin R_{re, q}$, we add it into the defect area, i.e., $R'_{q, e_i, \kappa} = R_{re, q} \cup \eta_{e_i, \kappa}$. Given $R'_{q, e_i, \kappa}$ and $\theta_{re, q}$, we run the simulation model to calculate the value of the transient objective function $J'_{q, e_i, \kappa}$ based on Eq. (4.2). If this adjustment of the defect area decreases the objective function value, i.e., $J'_{q, e_i, \kappa} < J_{re, q}$, $R_{re, q}$ and $J_{re, q}$ will be substituted with $R'_{q, e_i, \kappa}$ and $J'_{q, e_i, \kappa}$, respectively; otherwise, $R_{re, q}$ and $J_{re, q}$ stay same. This procedure will be repeated for each border point at iteration q . After $R_{re, q}$ and $J_{re, q}$ finish their updates at iteration q , $\theta_{re, q}$ given $R_{re, q}$ will be reevaluated through the sub-GP active learning to acquire the updated model parameter for the $(q + 1)$ iteration, i.e., $\theta_{re, q+1}$. This process will be iterated for Q times to eventually obtain a refined defect area R_{re} and its corresponding parameter set θ_{re} . The refinement (REF) procedure is summarized in Algorithm 3.

C.2 Benchmark methods

C.2.1 Multi-resolution method

In the multi-resolution method, the heart geometry is iteratively partitioned into multiple segments to search for the diseased region. Note that in the multi-resolution approach, we assume the true parameters for diseased and healthy regions are provided. In other words, the parameter estimation in the multi-resolution approach is exempted to reduce the computational burden. For example, assume the target region at i th iteration in the multi-resolution approach is denoted as seg_i , which is further divided into two sub-segments, seg_{i1} and seg_{i2} . The true diseased model parameters a_{d0} and d_{d0} are alternately endowed to these two segments, while the rest of the heart geometry is deemed as healthy with true normal parameters a_{h0} and d_{h0} . The cardiac simulation model given the two different parameter settings will be executed to evaluate the corresponding values of the objective function $J(seg_{i1})$ and $J(seg_{i2})$. If $J(seg_{i1}) < J(seg_{i2})$, the diseased region is assumed to lie in seg_{i1} and vice versa. The resulted diseased segment will be further divided into two sub-segments for a more accurate estimation of the diseased region. This process will be repeated until the value of the objective function converges. The algorithm of the multi-resolution approach is detailed in (Yao, 2021).

Algorithm 3: Refinement (REF)

Initialization: $\theta_{re,0} = \theta_0^*$, $J_{re,0} = J_0^*$, $R_{re,0} = R_0$, $q = 0$, Q, ξ, N

while $q \leq Q$ **do**

Find the circumference of $R_{re,q}$: $E_q = \{e_1, \dots, e_p\}$

Find w nearest neighbors for each instance in E_q : $\eta_{e_i,\kappa}, \iota = \{1, \dots, p\}, \kappa = \{1, \dots, w\}$

$\kappa = 1; \iota = 1$

while $\iota \leq p$ **do**

while $\kappa \leq w$ **do**

if $\eta_{e_i,\kappa} \in R_{re,q}$ **then**

$R'_{q,e_i,\kappa} = \{\hat{s} : \hat{s} \in R_{re,q}, \hat{s} \neq \eta_{e_i,\kappa}\}$

else

$R'_{q,e_i,\kappa} = R_{re,q} \cup \eta_{e_i,\kappa}$

end

Run the numerical simulation with $\theta_{re,q}$ and $R'_{q,e_i,\kappa}$ to calculate the objective function $J'_{q,e_i,\kappa}$ according to Eq. (4.2).

if $J'_{q,e_i,\kappa} < J_{re,q}$ **then**

$R_{re,q} = R'_{q,e_i,\kappa}, J_{re,q} = J'_{q,e_i,\kappa}$

end

$\kappa = \kappa + 1$

end

$\iota = \iota + 1$

end

while $\rho \geq \xi$ and $n \leq N$ **do**

begin GP construction for the objective function $J(\theta|R_{re,q})$

 GP predictive modeling given observed data $\{\Theta_{1:n-1}|R_{re,q}, J_{1:n-1}\}$ for $J(\theta|R_{re,q})$:

$$P\left(J(\theta|R_{re,q}) \mid \{\Theta_{1:n-1}|R_{re,q}, J_{1:n-1}\}\right) = \mathcal{N}\left(\mu_{n-1}(\theta \mid R_{re,q}), \sigma_{n-1}^2(\theta \mid R_{re,q})\right)$$

end

Select the next training point according to the GP-UCB criterion:

$$\theta_n = \arg \max_{\theta} \{-\mu_{n-1}(\theta \mid R_{re,q}) + \beta^{1/2} \cdot \sigma_{n-1}(\theta \mid R_{re,q})\}$$

Run the simulator with θ_n and $R_{re,q}$ to calculate the new objective function J_n according to Eq. (4.2)

$$\rho = \|\theta_n - \theta_{n-1}\|; \{\Theta_{1:n} \mid R_{re,q}, J_{1:n}\} = \{\Theta_{1:n-1} \mid R_{re,q}, J_{1:n-1}\} \cup \{\theta_n \mid R_{re,q}, J_n\}; n \leftarrow n + 1;$$

end

$q \leftarrow q + 1, \theta_{re,q} \leftarrow \theta_n, J_{re,q} \leftarrow J_n$

end

Output: R_{re} and θ_{re}

C.2.2 Multi-task Gaussian Process Upper Confidence Bound (MT-GPUCB)

The MT-GPUCB method is adapted from the multi-task Bayesian optimization approach (Swersky et al., 2013). Specifically, the inverse system for defect localization can be formulated as an optimization problem to search for $\mathbf{z} = [\boldsymbol{\gamma}, \boldsymbol{\theta}]$, where $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ denote

defect location and system parameter set, respectively:

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} J(\mathbf{z}) \quad (\text{C.1})$$

To address the computational issue induced by the large parameter space, we propose to utilize a separable kernel function to model the correlations among $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, and further formulate the overall kernel function as the Kronecker product of two sub-covariance matrices, $\mathcal{K}_{\boldsymbol{\gamma}}$ and $\mathcal{K}_{\boldsymbol{\theta}}$, that characterize $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ separately:

$$K[(\boldsymbol{\gamma}, \boldsymbol{\theta}), (\boldsymbol{\gamma}', \boldsymbol{\theta}')] = K_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}, \boldsymbol{\gamma}') K_{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{\theta}') \rightarrow K = K_{\boldsymbol{\gamma}} \otimes K_{\boldsymbol{\theta}} \quad (\text{C.2})$$

We engage Matern 3/2 kernel to accommodate varying degrees of function smoothness, i.e.,

$$\begin{aligned} \mathcal{K}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j) &= \sigma_{m\boldsymbol{\gamma}} \cdot \left(\frac{\sqrt{3}|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j|}{l_{m\boldsymbol{\gamma}_1}} \right) \exp \left(-\frac{\sqrt{3}|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_j|}{l_{m\boldsymbol{\gamma}_2}} \right) \\ \mathcal{K}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) &= \sigma_{m\boldsymbol{\theta}} \cdot \left(\frac{\sqrt{3}|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j|}{l_{m\boldsymbol{\theta}_1}} \right) \exp \left(-\frac{\sqrt{3}|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j|}{l_{m\boldsymbol{\theta}_2}} \right) \end{aligned} \quad (\text{C.3})$$

where $\{\sigma_{m\boldsymbol{\gamma}}, l_{m\boldsymbol{\gamma}_1}, l_{m\boldsymbol{\gamma}_2}\}$ and $\{\sigma_{m\boldsymbol{\theta}}, l_{m\boldsymbol{\theta}_1}, l_{m\boldsymbol{\theta}_2}\}$ denote the kernel hyperparameters for $\mathcal{K}_{\boldsymbol{\gamma}}$ and $\mathcal{K}_{\boldsymbol{\theta}}$, respectively, which will be optimized via the maximum likelihood method (Brochu et al., 2010).

We define the location-based and parameter-based observations as $\boldsymbol{\Gamma}_m = [\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_m]$, $\boldsymbol{\Theta}_m = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_m]$ respectively, and the corresponding values of the objective function as $\mathbf{J}_{1:m \times m} = [J(\boldsymbol{\gamma}_i, \boldsymbol{\theta}_j)] \Big|_{\boldsymbol{\gamma}_i \in \boldsymbol{\Gamma}_m, \boldsymbol{\theta}_j \in \boldsymbol{\Theta}_m}$. Given the observations, the posterior mean of $J(\boldsymbol{\gamma}_*, \boldsymbol{\theta}_*)$ at arbitrary parameter set $(\boldsymbol{\gamma}_*, \boldsymbol{\theta}_*)$ is derived as $\mu_k(\boldsymbol{\gamma}_*, \boldsymbol{\theta}_*) = \mathbf{K}_k^* \boldsymbol{\Sigma}^{-1} \mathbf{J}_{1:m \times m}$, where $\boldsymbol{\Sigma} = \mathcal{K}[(\boldsymbol{\Gamma}_m, \boldsymbol{\Theta}_m), (\boldsymbol{\Gamma}_m, \boldsymbol{\Theta}_m)] + \sigma_{\delta}^2 \mathbf{I}$, $\mathcal{K}_k^* = \mathcal{K}[(\boldsymbol{\gamma}_*, \boldsymbol{\theta}_*), (\boldsymbol{\Gamma}_m, \boldsymbol{\Theta}_m)] = \mathcal{K}_{\boldsymbol{\gamma}}^* \otimes \mathcal{K}_{\boldsymbol{\theta}}^*$, $\mathbf{I}$ is the identity matrix with the dimension of $(m * m) \times (m * m)$, and σ_{δ}^2 denotes the variance of Gaussian noises.

We further exploit the Kronecker property to enhance computational efficiency when calculating the precision matrix $\boldsymbol{\Sigma}^{-1}$. Specifically, we first perform the singular value decomposition (SVD) on $\mathcal{K}_{\boldsymbol{\gamma}}$ and $\mathcal{K}_{\boldsymbol{\theta}}$ as $\mathcal{K}_{\boldsymbol{\gamma}} = U_{\boldsymbol{\gamma}} \Lambda_{\boldsymbol{\gamma}} U_{\boldsymbol{\gamma}}^T$ and $\mathcal{K}_{\boldsymbol{\theta}} = U_{\boldsymbol{\theta}} \Lambda_{\boldsymbol{\theta}} U_{\boldsymbol{\theta}}^T$, respectively.

Following this, the SVD of Σ can be expressed as:

$$\Sigma = (U_\gamma \otimes U_\theta) \left(\Lambda_\gamma \otimes \Lambda_\theta + \sigma_\delta^2 I \right) (U_\gamma \otimes U_\theta)^\top = (U_\gamma \otimes U_\theta) \Lambda (U_\gamma \otimes U_\theta)^\top$$

where $\Lambda = \Lambda_\gamma \otimes \Lambda_\theta + \sigma_\delta^2 I$ is a diagonal matrix. The predicted posterior mean μ_k can then be written in the form of Kronecker product:

$$\begin{aligned} \mu_k(\gamma_*, \theta_*) &= \left(\mathcal{K}_\gamma^* \otimes \mathcal{K}_\theta^* \right) \Sigma^{-1} \mathbf{J}_{1:m \times m} = \left(\mathcal{K}_\gamma^* \otimes \mathcal{K}_\theta^* \right) (U_\gamma \otimes U_\theta) \Lambda^{-1} (U_\gamma \otimes U_\theta)^\top \mathbf{J}_{1:m \times m} \\ &= \left(\mathcal{K}_\gamma^* U_\gamma \otimes \mathcal{K}_\theta^* U_\theta \right) \Lambda^{-1} \text{vec} \left(U_\theta^\top \mathbf{J}_{m \times m} U_\gamma \right) = \left(\mathcal{K}_\gamma^* U_\gamma \otimes \mathcal{K}_\theta^* U_\theta \right) \mathbf{J}'_{1:m \times m} \end{aligned}$$

where $\mathbf{J}_{1:m \times m} = \text{vec}(\mathbf{J}_{m \times m})$, and $\mathbf{J}'_{1:m \times m} = \Lambda^{-1} \text{vec} \left(U_\theta^\top \mathbf{J}_{m \times m} U_\gamma \right)$, which is a vector with the same length as $\mathbf{J}_{1:m \times m}$. $\text{vec}(A)$ stands for the vectorization of the matrix A by stacking the columns of A into a single-column vector. Then, the posterior mean at γ_*, θ_* is given as

$$\mu_k(\gamma_*, \theta_* \mid \mathbf{\Gamma}_m, \mathbf{\Theta}_m) = \text{vec} \left(\mathcal{K}_\theta^* U_\theta \mathbf{J}'_{m \times m} U_\gamma \mathcal{K}_\gamma^* \right)$$

where $\mathbf{J}'_{m \times m} = \text{vec}^{-1} \left(\mathbf{J}'_{1:m \times m} \right)$ and $\text{vec}^{-1}(\cdot)$ is the inverse operation of $\text{vec}(\cdot)$. Similarly, the posterior variance can be derived using Kronecker product as:

$$\begin{aligned} \sigma_k^2(\gamma_*, \theta_* \mid \mathbf{\Gamma}_m, \mathbf{\Theta}_m) &= \mathcal{K}[(\gamma_*, \theta_*), (\gamma_*, \theta_*)] - \mathcal{K}[(\gamma_*, \theta_*), (\mathbf{\Gamma}_m, \mathbf{\Theta}_m)] \Sigma^{-1} \mathcal{K}[(\mathbf{\Gamma}_m, \mathbf{\Theta}_m), (\gamma_*, \theta_*)] \\ &= \mathcal{K}_\gamma^{**} \otimes \mathcal{K}_\theta^{**} - \left(\mathcal{K}_\gamma^* U_\gamma \otimes \mathcal{K}_\theta^* U_\theta \right) \Lambda^{-1} \left(\mathcal{K}_\gamma^* U_\gamma \otimes \mathcal{K}_\theta^* U_\theta \right)^\top \end{aligned}$$

As such, we are able to model and predict the defect location and system parameters simultaneously in a computation-efficient way. With the GP model trained by the observations $\mathbf{\Gamma}_m$ and $\mathbf{\Theta}_m$, the subsequent query location $\mathbf{z}_{m+1}$ is determined according to the following GPUCB:

$$\mathbf{z}_{m+1} = \text{argmax}_{\mathbf{z}} \left\{ -\mu_k(\mathbf{z} \mid \mathbf{\Gamma}_m, \mathbf{\Theta}_m) + \beta_m^{\frac{1}{2}} \sigma_k(\mathbf{z} \mid \mathbf{\Gamma}_m, \mathbf{\Theta}_m) \right\}$$

After acquiring the new parameter pair $\mathbf{z}_{m+1}$, the 3D simulator will be run and generate the corresponding response of the objective function value, i.e., $J(\mathbf{z}_{m+1})$. The multitask GP will be updated when the new input-output pair, i.e., $(\mathbf{z}_{m+1}, J(\mathbf{z}_{m+1}))$, becomes available.

The MT-GPUCB procedure will be iterated until the objective function value J converges, generating estimations of the optimal defect location and system parameter value, $\hat{\mathbf{z}} = [\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\theta}}]$.

Appendix D

Figures

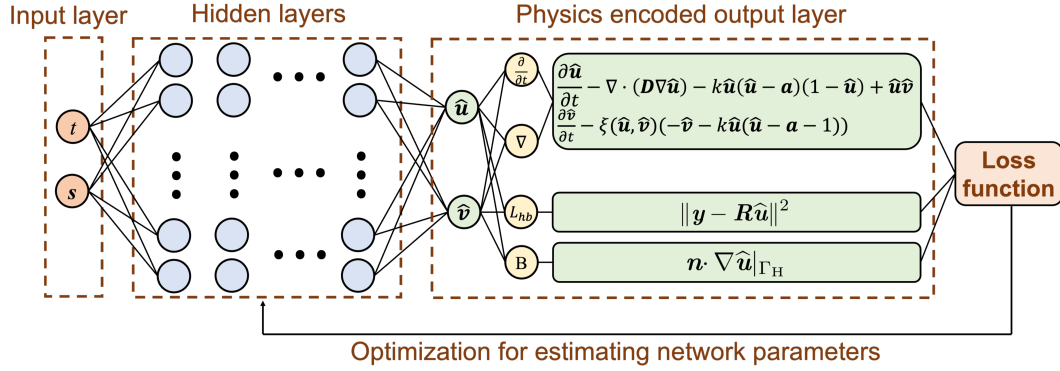


Figure D.1: Illustration of the proposed P-DL framework for HSP prediction. Note that $\hat{u}$ and $\hat{v}$ are the predicted HSP and recovery current, respectively. The physical constraint of AP model and boundary condition are converted into residuals as physical loss $\mathcal{L}_{ph}$ and added to the total loss function of the DNN for respecting the underlying physics principle.

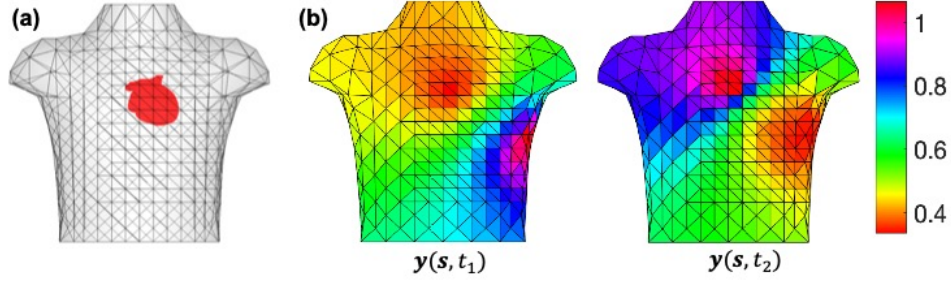


Figure D.2: (a) Illustration of the torso-heart geometry; (b) BSPM at two different time points t_1 and t_2 .

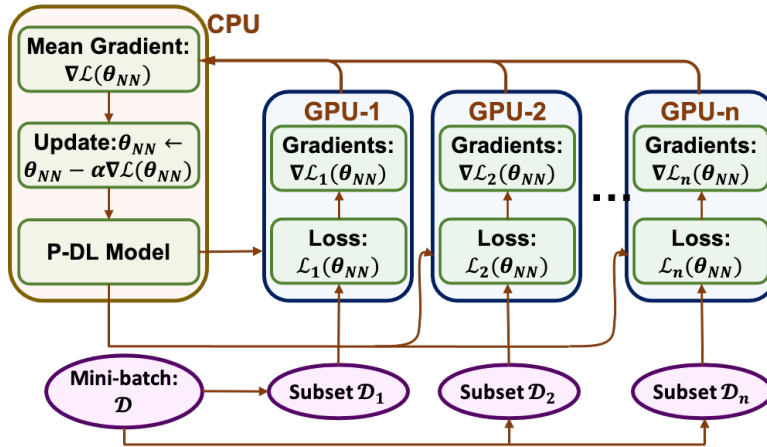


Figure D.3: Scheme for parallel computing.

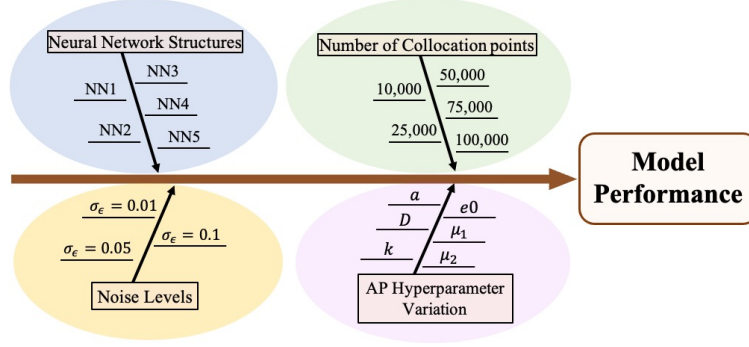


Figure D.4: Experimental design for evaluating the performance of the proposed P-DL approach.

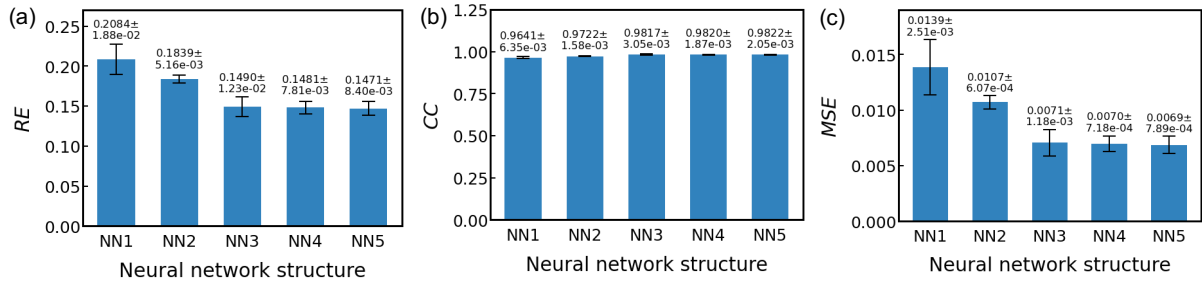


Figure D.5: The variation of (a) RE , (b) CC , and (c) MSE with respect to neural network structure.

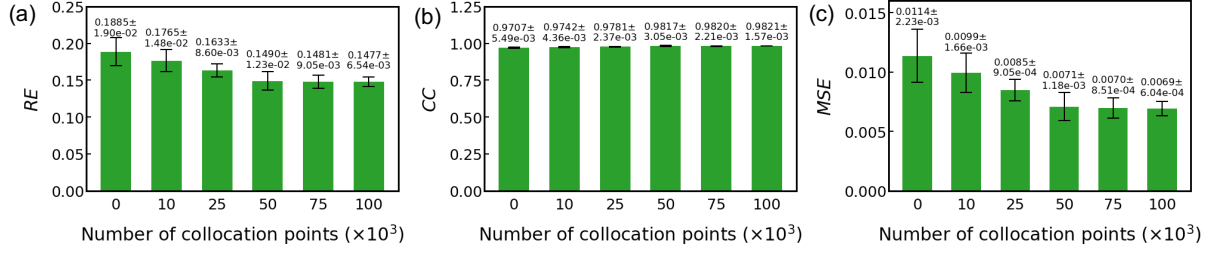


Figure D.6: The variation of (a) RE , (b) CC , and (c) MSE with respect to the number of collocation points.

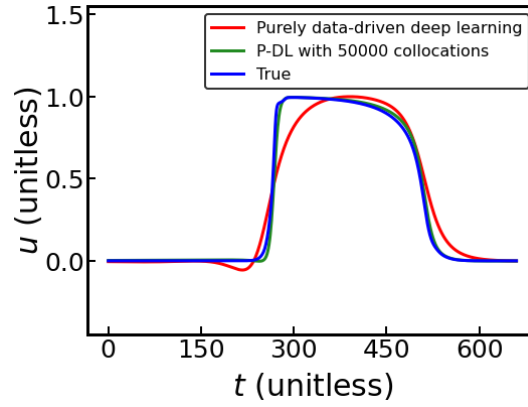


Figure D.7: The estimated HSP evolution $\hat{u}(s, t)$ from an arbitrary spatial location in the heart generated from pure data-driven DL and P-DL with 50000 collocation points compared with the true evolution.

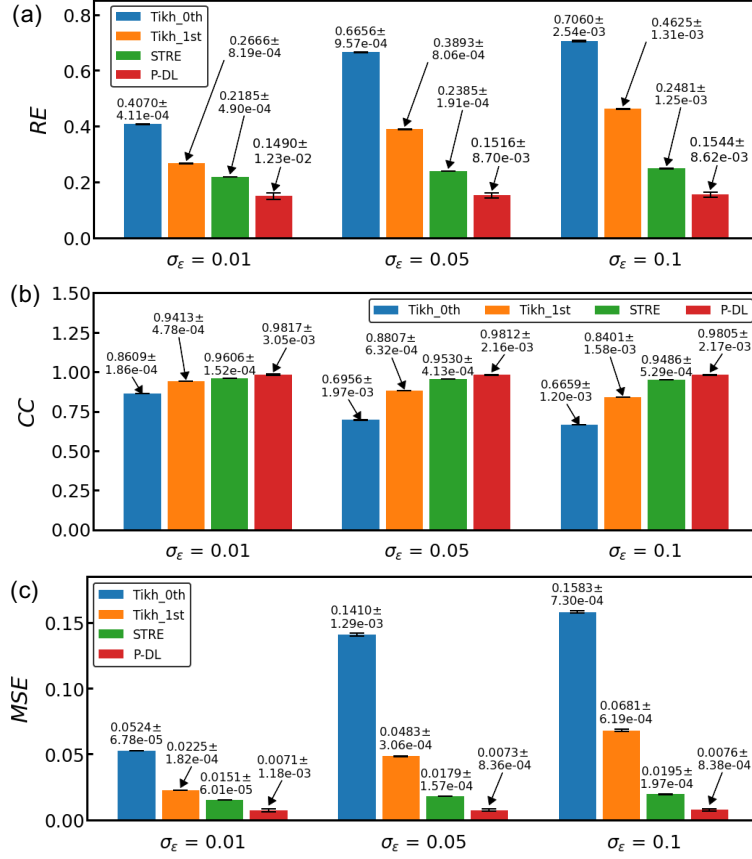


Figure D.8: The comparison of (a) RE , (b) CC , and (c) MSE between the proposed P-DL method and other regularization models (i.e., Tikh_0th, Tikh_1st, and STRE) when different noise levels $\sigma_\epsilon = 0.01, 0.05, 0.1$ are added to the BSPM $\mathbf{y}(\mathbf{s}, t)$.

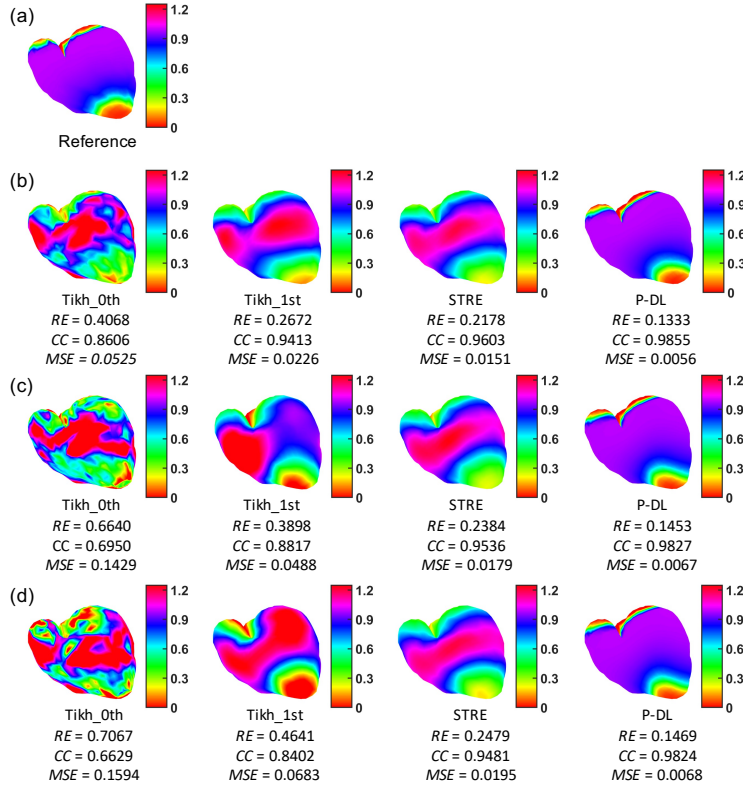


Figure D.9: (a) Reference HSP mapping $\mathbf{u}(\mathbf{s}, t)$ at $t = 300$; Estimated mappings $\mathbf{u}(\mathbf{s}, t)$ at $t = 300$ by P-DL and traditional methods (i.e., Tikh_0th, Tikh_1st, and STRE) with noise level of (b) $\sigma_\epsilon = 0.01$ (c) $\sigma_\epsilon = 0.05$ and (d) $\sigma_\epsilon = 0.1$ in BSPM $\mathbf{y}(\mathbf{s}, t)$.

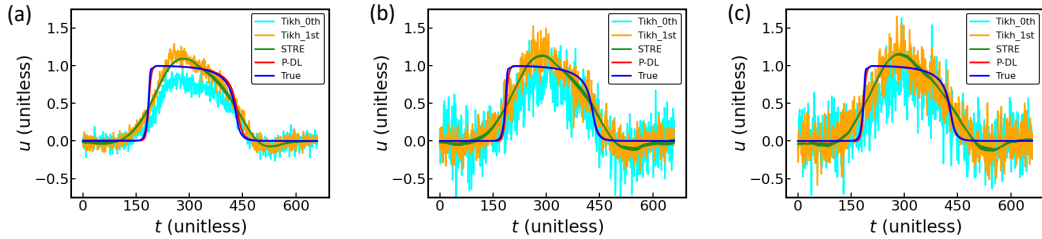


Figure D.10: The comparison of HSP evolution $\mathbf{u}(\mathbf{s}, t)$ over time for one specific spatial location, between the true HSP data and predictions by Tikh_0th, Tikh_1st, STRE, and P-DL under the noise level of (a) $\sigma_\epsilon = 0.01$, (b) $\sigma_\epsilon = 0.05$, (c) $\sigma_\epsilon = 0.1$.

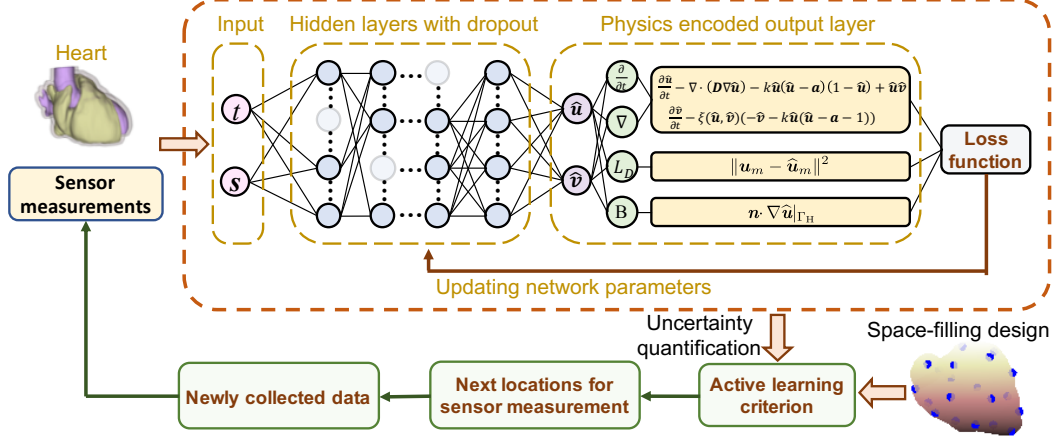


Figure D.11: The flowchart of the proposed P-DAL framework. Note that $\hat{u}$ and $\hat{v}$ are the predicted electrical potential and recovery current on the heart geometry, respectively. u_m and $\hat{u}_m$ denotes the measured data at specific spatial locations and the corresponding predictions, respectively. The physical constraint of AP model and boundary condition are transformed into the form of residuals as the physical loss $\mathcal{L}_{ph}$ and incorporated into the total loss function $\mathcal{L}$ of the DNN to respect the underlying physics-based principle.

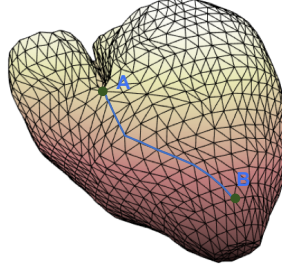


Figure D.12: The illustration of the geodesic distance between points A and B, which is the length of the shortest path on the 3D geometry between the two points.

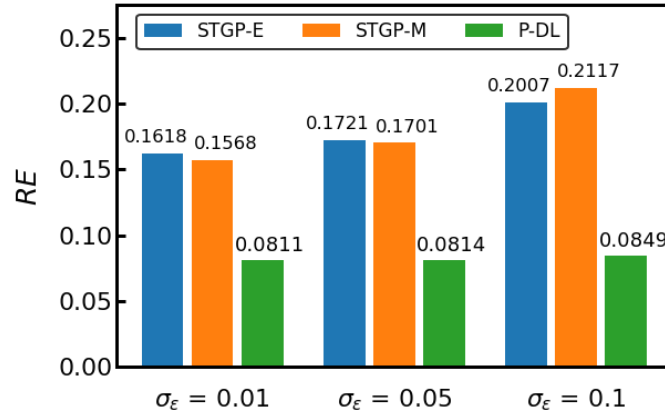


Figure D.13: The comparison of RE generated from STGP-E, STGP-M, and P-DL models in the healthy heart system with different noise levels of $\sigma_\epsilon = 0.01$, 0.05 , and 0.1 .

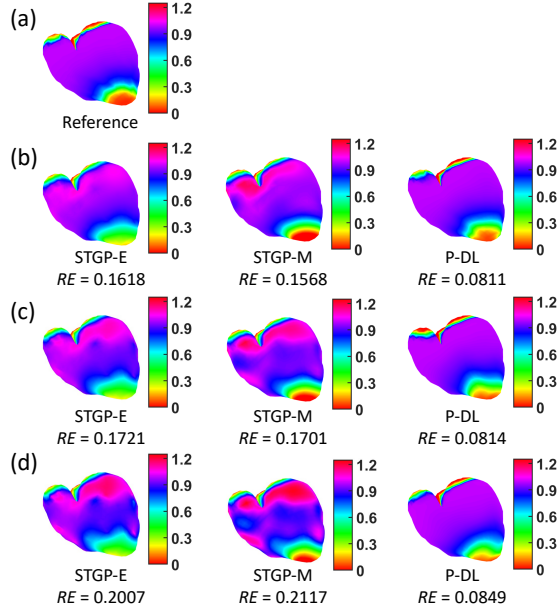


Figure D.14: (a) Reference potential mapping $u(s, t)$ on the healthy heart surface at $t = 30$; Predictive potential mappings $\hat{u}(s, t)$ at $t = 30$ by STGP-E, STGP-M, and P-DL approaches with noise level of (b) $\sigma_\epsilon = 0.01$ (c) $\sigma_\epsilon = 0.05$ and (d) $\sigma_\epsilon = 0.1$.

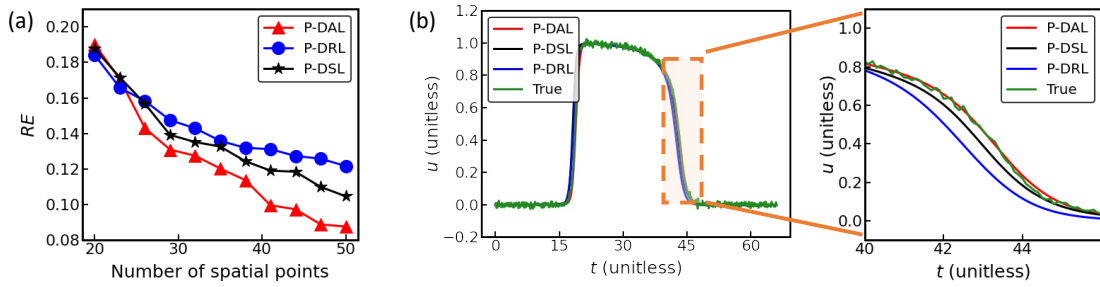


Figure D.15: (a) The RE estimated by P-DL model based on training data collected from random-sampling (P-DRL), space-filling (P-DSL), and active learning (P-DAL); (b) The estimated evolvement of cardiac electrodynamics over time for the healthy heart system.

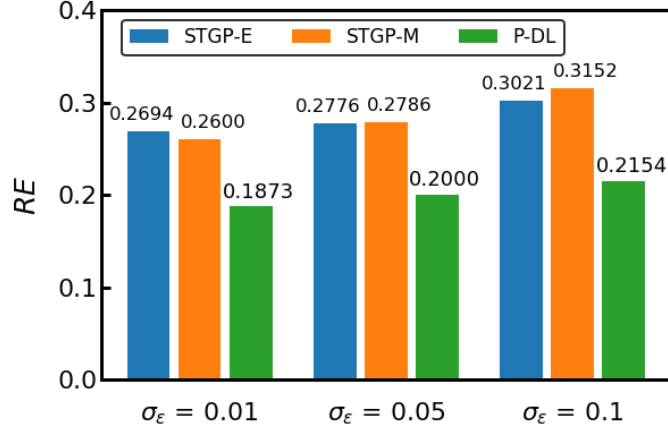


Figure D.16: The comparison of RE generated from STGP-E, STGP-M, and P-DL models in diseased heart system with different noise levels of $\sigma_\epsilon = 0.01$, 0.05 , and 0.1 .

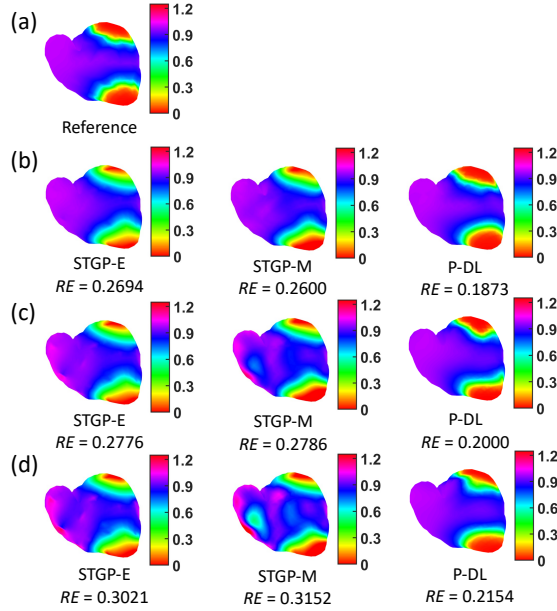


Figure D.17: (a) Reference potential mapping $\mathbf{u}(\mathbf{s}, t)$ on the diseased heart surface at $t = 33$; Predictive potential mappings $\hat{\mathbf{u}}(\mathbf{s}, t)$ at $t = 33$ by STGP-E, STGP-M, and P-DL with noise level of (b) $\sigma_\epsilon = 0.01$ (c) $\sigma_\epsilon = 0.05$ and (d) $\sigma_\epsilon = 0.1$ for the diseased heart system.

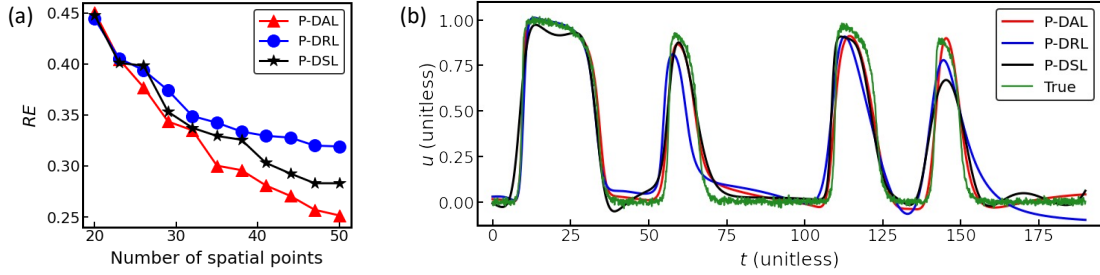


Figure D.18: (a) The RE estimated by P-DL based on training data collected from random-sampling (i.e., P-DRL), space-filling (i.e., P-DSL), and active learning (i.e., P-DAL); (b) The estimated evolvement of cardiac electrodynamics over time in the diseased heart system.

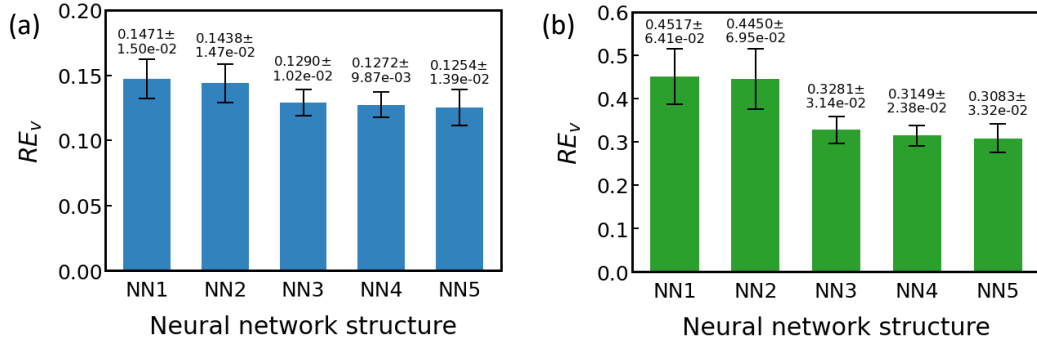


Figure D.19: The RE estimated by P-DL model with different neural network structures for (a) healthy and (b) diseased heart systems. The experiments are repeated 6 times to generate the standard deviation of RE_v (i.e., the error bar).

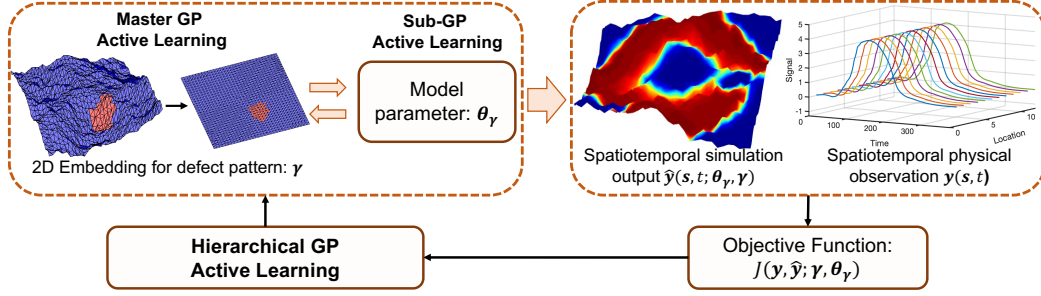


Figure D.20: The flowchart of the proposed research methodology.

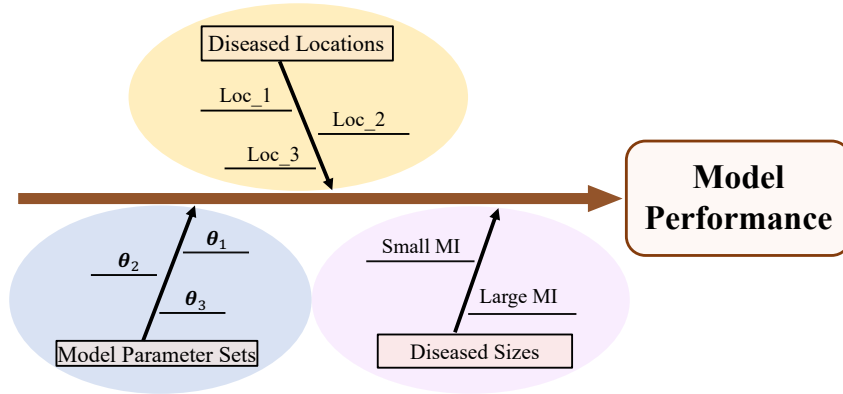


Figure D.21: Illustration of the experimental design for evaluating the performance of our HGPAL-REF method.

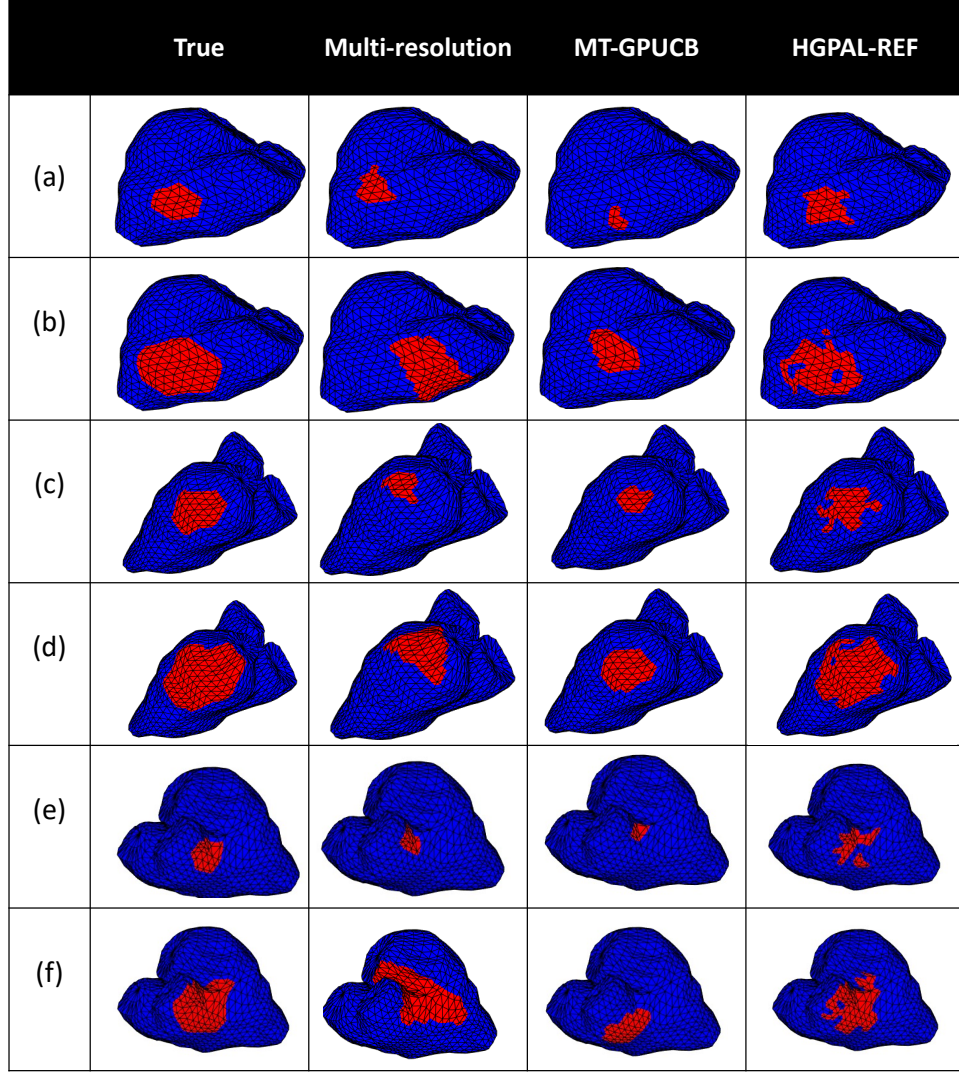


Figure D.22: The different diseased regions (marked as red): (a) small MI at Loc_1, (b) large MI at Loc_1, (c) small MI at Loc_2, (d) large MI at Loc_2, (e) small MI at Loc_3, (f) large MI at Loc_3 on the heart surfaces estimated by the multi-resolution approach, MT-GPUCB, and the proposed HGPAL-REF method. Note that parameter set θ_2 is used to simulate the true diseased locations.

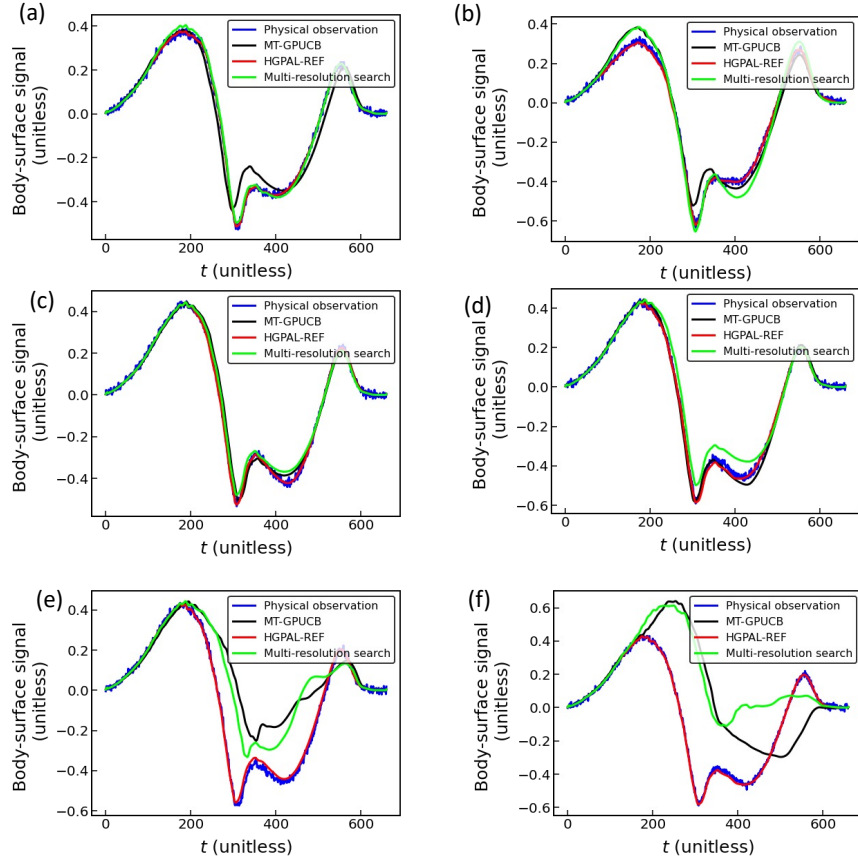


Figure D.23: Evolvement curves of estimated BSPs produced by different methods with (a) small MI at Loc_1, (b) large MI at Loc_1, (c) small MI at Loc_2, (d) large MI at Loc_2, (e) small MI at Loc_3, (f) large MI at Loc_3, at one specific spatial point on the body surface.

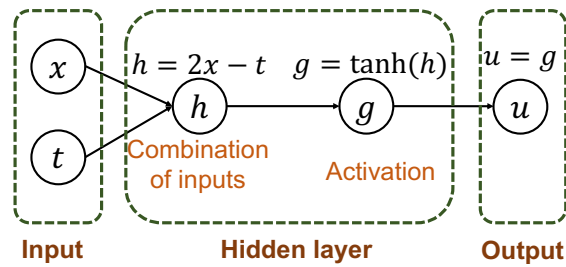


Figure D.24: A simple network example to illustrate automatic differentiation.

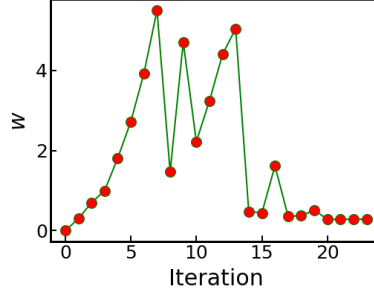


Figure D.25: The illustration of the convergence of GP-UCB.

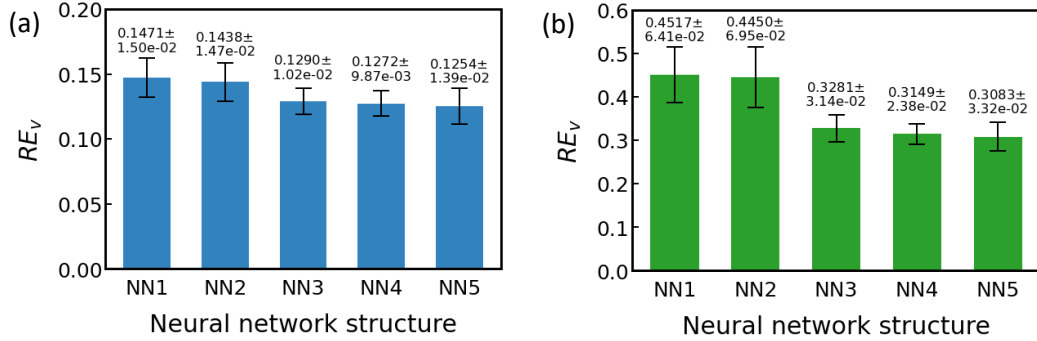


Figure D.26: The RE estimated by P-DL model with different neural network structures for (a) healthy and (b) diseased heart systems. The experiments are repeated 6 times to generate the standard deviation of RE_v (i.e., the error bar).

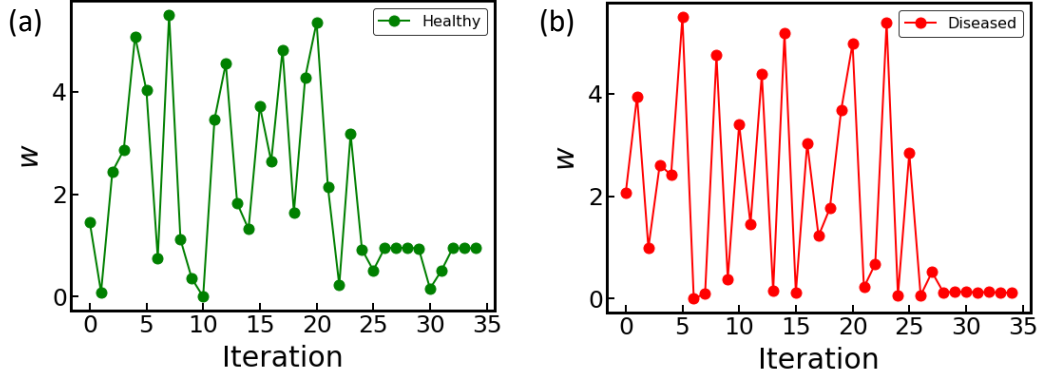


Figure D.27: The illustration of GP-UCB convergence for (a) healthy and (b) diseased heart systems.

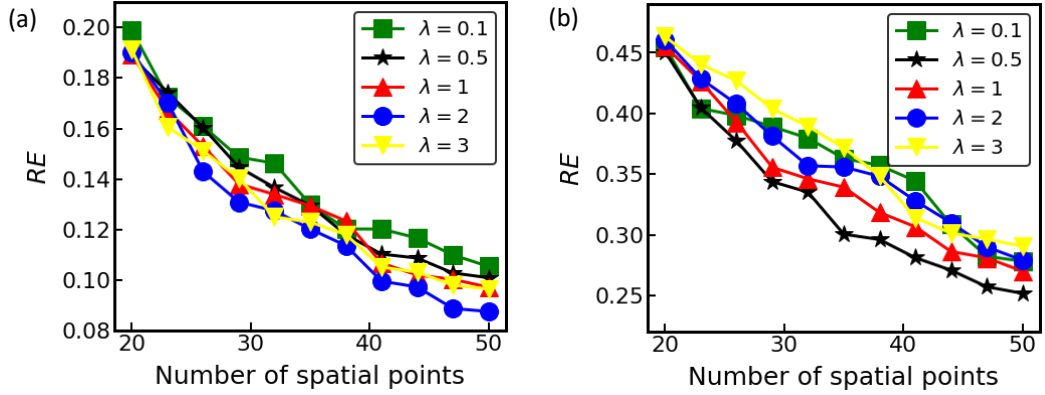


Figure D.28: The RE 's estimated by P-DAL model with different λ 's for (a) healthy and (b) diseased heart systems as the number of training points increases.

Appendix E

Tables

Table E.1: Structure details of different neural networks

	NN1	NN2	NN3	NN4	NN5
Layers	5	3	5	8	5
Neurons	5	10	10	10	20

Table E.2: The RE , CC , and MSE from the P-KF algorithm with different initial conditions. Here, $\hat{\mathbf{u}}_0(\mathbf{s})$ and $\mathbf{u}_0(\mathbf{s})$ denote the specified and true initial condition, respectively; the initialization noise follows a normal distribution, i.e., $\boldsymbol{\epsilon}_{\mathbf{u}_0}(\mathbf{s}) \sim N(0, 0.0025)$.

Initial condition for HSP (i.e., $\hat{\mathbf{u}}_0(\mathbf{s})$)	$\mathbf{u}_0(\mathbf{s})$	$\mathbf{u}_0(\mathbf{s}) + \boldsymbol{\epsilon}_{\mathbf{u}_0}(\mathbf{s})$	Random location initialization	Zero initialization
RE	$0.1426 \pm 1e-5$	0.1626 ± 0.0027	0.8940 ± 0.0458	$0.9960 \pm 7e-6$
CC	$0.9848 \pm 3e-6$	0.9803 ± 0.0006	0.3247 ± 0.0712	0.1563 ± 0.0007
MSE	$0.0064 \pm 1e-6$	0.0084 ± 0.0003	0.2539 ± 0.0260	0.3143 ± 0.0028

Table E.3: The computational time for different methods.

Method	Tikh_0th	Tikh_1st	STRE	P-KF	P-DL
Running Time	0.31 s	0.46 s	2 h 26 min	3 h 10 min	19 min

Table E.4: The influence of variations in AP model parameters ($\pm 10\%$) on the *RE*, *CC*, and *MSE* provided by the P-DL.

		a	k_r	D	$e0$	μ_1	μ_2
<i>RE</i>	-10%	0.1681	0.1525	0.1665	0.1620	0.1625	0.1777
	+10%	0.1550	0.1797	0.1647	0.1615	0.1949	0.1701
<i>CC</i>	-10%	0.9768	0.9809	0.9773	0.9785	0.9783	0.9742
	+10%	0.9804	0.9738	0.9778	0.9786	0.9690	0.9763
<i>MSE</i>	-10%	0.0090	0.0074	0.0088	0.0083	0.0084	0.0100
	+10%	0.0076	0.0102	0.0086	0.0082	0.0120	0.0092

Table E.5: Training time for P-DL modeling

	Number of spatial locations	Number of collocation points	Training time
Healthy system	50	50,000	12 min
Diseased system	50	200,000	1 h and 10 min

Table E.6: The comparison of MSE , SDP , and OVP generated by the HGPAL-REF approach, MT-GPUCB, and the multi-resolution method.

Size	Location	Parameter set	HGPAL-REF			Multi-resolution search			MT-GPUCB		
			$MSE(e^{-4})$	SDP (%)	OVP (%)	$MSE(e^{-4})$	SDP (%)	OVP (%)	$MSE(e^{-4})$	SDP (%)	OVP (%)
Small MI	Loc_1	θ_1	1.1073	0	78.94	4.3283	52.63	5.26	3.2161	36.84	47.36
		θ_2	1.0341	0	84.21	1.3294	26.32	26.32	4.6521	73.86	10.52
		θ_3	1.0766	26.31	63.16	1.8805	0	52.63	2.1208	31.57	57.89
	Loc_2	θ_1	1.0995	32.43	97.29	6.2413	59.46	18.92	3.4888	10.81	86.48
		θ_2	1.0136	13.51	75.68	2.2587	62.16	13.51	2.0091	89.18	8.10
		θ_3	1.0497	18.92	94.59	1.6690	27.03	40.54	3.3943	84.48	13.51
	Loc_3	θ_1	1.3457	40	95	7.0958	35	50	24.6346	70	25
		θ_2	1.1049	20	75	2.8825	35	50	4.0657	65	25
		θ_3	1.0653	30	70	13.6903	30	20	11.7177	75	10
Large MI	Loc_1	θ_1	1.2971	13.33	78.33	10.5307	33	35	2.9483	5	73.33
		θ_2	1.0943	21.67	73.33	5.7438	8.33	21.66	4.5889	63.33	33.33
		θ_3	1.2605	26.67	70	9.9443	5	23.33	6.0448	40	51.67
	Loc_2	θ_1	1.0174	7.45	86.17	23.5734	62.67	23.40	15.3521	64.89	35.11
		θ_2	1.0521	14.89	82.98	5.1661	53.19	34.04	3.3931	64.89	35.11
		θ_3	1.1076	21.28	70.21	5.7982	39.36	55.31	7.7292	62.76	37.23
	Loc_3	θ_1	1.7608	8.33	83.33	64.7031	13.89	36.11	23.8423	83.33	88.98
		θ_2	1.1299	12.50	77.78	7.2863	52.78	47.22	27.1124	55.56	27.78
		θ_3	1.3696	1.39	79.16	73.7159	33.33	29.16	19.6486	95.83	88.88

Table E.7: The estimated parameter sets for both healthy and diseased domains on the heart surface by the proposed HGPAL-REF approach and the MT-GPUCB method.

Method			HGPAL-REF				MT-GPUCB			
Parameters			a_h	a_d	d_h	d_d	a_h	a_d	d_h	d_d
Real			0.1	0.25	10	0	0.1	0.25	10	0
θ_1	Small MI	Loc_1	0.1001	0.2917	10.0139	3.4457	0.0965	0.2241	10.2863	4.7095
		Loc_2	0.0992	0.2255	10.0536	4.7889	0.1056	0.3954	10.1300	2.6725
		Loc_3	0.0999	0.2077	10.1462	4.1271	0.0954	0.3361	10.6323	1.1299
	Large MI	Loc_1	0.1000	0.2866	10.0184	2.6876	0.1149	0.3974	9.0399	4.7666
		Loc_2	0.1002	0.2883	9.9936	0.8631	0.0994	0.2490	9.8548	1.6411
		Loc_3	0.1001	0.2424	10.0729	5.0	0.1000	0.2417	11.0136	4.4565
Real			0.1	0.15	10	5	0.1	0.15	10	5
θ_2	Small MI	Loc_1	0.0999	0.1471	9.9909	4.8481	0.0937	0.3999	10.2142	1.1531
		Loc_2	0.0999	0.1525	10.0163	4.6147	0.1010	0.4368	9.7971	1.3169
		Loc_3	0.0997	0.1389	10.0938	5.0	0.1052	0.4093	9.7194	3.6160
	Large MI	Loc_1	0.1001	0.1575	10.0103	4.5819	0.0975	0.2165	10.6269	4.5777
		Loc_2	0.1003	0.1566	10.0229	4.7041	0.0982	0.1500	10.1149	1.6156
		Loc_3	0.0999	0.1549	10.0114	4.7424	0.1009	0.1837	10.4727	2.8527
Real			0.1	0.15	10	2	0.1	0.15	10	2
θ_3	Small MI	Loc_1	0.0999	0.1734	10.0081	2.1053	0.0987	0.3701	10.2456	3.2105
		Loc_2	0.0998	0.1446	10.0289	2.3500	0.1024	0.4313	9.9377	4.7545
		Loc_3	0.1001	0.1622	9.9829	2.1586	0.1053	0.3889	9.8626	1.4038
	Large MI	Loc_1	0.0999	0.1715	10.0250	2.0271	0.1027	0.3221	10.1345	3.7948
		Loc_2	0.1000	0.1582	10.0178	1.8711	0.1009	0.3969	9.8645	2.9685
		Loc_3	0.0991	0.1447	10.1268	1.9311	0.0938	0.1311	11.0324	4.3987

Table E.8: The computation of the forward and backward pass of the automatic differentiation example as shown in Fig. D.24.

Forward pass	Backward pass
$x = 2, t = 3$	$\frac{\partial u}{\partial u} = 1$
$h = 2x - t = 1$ $g = \tanh(h) = 0.7616$	$\frac{\partial u}{\partial g} = 1$ $\frac{\partial u}{\partial h} = \frac{\partial u}{\partial g} \frac{\partial g}{\partial h} = \frac{\partial u}{\partial g} \text{sech}^2(h) = 0.4184$
$u = g = 0.7616$	$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial h} \frac{\partial h}{\partial x} = 0.4184 \times 2 = 0.8368$ $\frac{\partial u}{\partial t} = \frac{\partial u}{\partial h} \frac{\partial h}{\partial t} = 0.4184 \times (-1) = -0.4184$

Table E.9: The optimal regularization parameter w_{opt} and the iteration number in GP-UCB for each NN structure.

NN structure	NN1	NN2	NN3	NN4	NN5
w_{opt}	0.23	0.29	0.44	0.56	0.33
Iterations	20	20	20	17	19

Table E.10: The structure details of different neural network.

Neural network structure		NN1	NN2	NN3	NN4	NN5
Healthy system	Layers	5	3	5	8	5
	Neurons	5	10	10	10	20
Diseased system	Layers	5	3	5	8	5
	Neurons	10	20	20	20	30

Table E.11: The variation of RE by P-DAL that is trained based on sensor observations from 50 spatial locations with different λ 's in the healthy and diseased heart systems. The experiments are repeated 6 times to generate the standard deviation of RE .

λ	0.1	0.5	1	2	3
Healthy system	0.1122 ± 0.0053	0.1019 ± 0.0093	0.0978 ± 0.0059	0.0919 ± 0.0029	0.0959 ± 0.0045
Diseased system	0.2800 ± 0.0113	0.2588 ± 0.0183	0.2720 ± 0.0257	0.2809 ± 0.0237	0.2837 ± 0.0310

Table E.12: The RE 's generated by P-DAL, P-DSL, and P-DRL. The experiments are repeated 6 times to generate the standard deviation of RE .

Method	P-DAL	P-DSL	P-DRL
Healthy system	0.0919 $\pm$ 0.0026	0.0983 $\pm$ 0.0034	0.1377 $\pm$ 0.0151
Diseased system	0.2588 $\pm$ 0.0183	0.2950 $\pm$ 0.0160	0.3481 $\pm$ 0.0246

Table E.13: p -value for the two-sample statistical test

Method for μ_A and μ_B	P-DALand P-DSL	P-DAL and P-DRL
p-value of healthy system	2.28e-3	2.90e-4
p-value of diseased system	1.66e-3	2.39e-5

Vita

Jianxin Xie received her B.S. in Physics from Southeast University, China, and her M.S. from Florida State University, USA. She is currently pursuing a PhD with the Department of Industrial Systems Engineering, University of Tennessee Knoxville. Her current research interests lie in advanced data analytics, data mining, and physical-statistical modeling with applications in various complex systems. She is the recipient of several awards such as IISE Gilbreth Memorial Fellowship, QCRE Best Poster Award at 2023 IISE Annual Conference, QCRE and DAIS Track Best Poster Award and Runner-up Best Paper Award in Operation Research Track at the 2021 IISE Annual Conference. Jianxin is a student member of IISE, INFORMS, and IISE. After graduation, she hopes to dedicate her career to research and innovation in the healthcare field, where she can contribute to advancements in smart diagnosis and treatment planning. Additionally, she aspires to engage in education, fostering the next generation of scholars and professionals, and inspiring them with the transformative power of healthcare research.