Title
Computational Methods for the Imputation and Prediction of Digital Health Data

Author
Hill, Brian Lawrence

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Methods for the Imputation and Prediction of Digital Health Data

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Computer Science

by

Brian Lawrence Hill

2021

ABSTRACT OF THE DISSERTATION

Computational Methods for
the Imputation and Prediction
of Digital Health Data

by

Brian Lawrence Hill
Doctor of Philosophy in Computer Science
University of California, Los Angeles, 2021
Professor Eran Halperin, Chair

Advances in both technology and medicine have enabled monumental progress toward the realization of precision medicine. In particular, machine learning algorithms – powered by electronic health records, genomic information, wearable sensors, and medical images – are positioned to become an integral part of the clinical workflow. While a tremendous amount of biomedical data is being generated and collected on a daily basis, plenty of data are still not routinely captured due to invasiveness, inconvenience, or cost. In this dissertation, we first describe the development and validation of a machine learning model that uses pre-operative data readily available in the electronic health record to predict post-operative in-hospital mortality. We then present multiple novel computational methods for accurately imputing unobserved health data using several different types of observed data, including physiological waveforms, genomics, and videos.

The dissertation of Brian Lawrence Hill is approved.

Sriram Sankararaman

Cho-Jui Hsieh

Eleazar Eskin

Eran Halperin, Committee Chair

University of California, Los Angeles

2021

To my parents, Dave and Linda,
for their unconditional love and support,
and to my wife, Kayla,
for always being up for an adventure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I want to acknowledge and thank my advisor, Eran Halperin. Choosing the right advisor can have a tremendous impact on your PhD experience, and I am very lucky Eran decided to join UCLA when he did. It's been an incredible privilege to learn from Eran during the past four years, and I consider him both a mentor and close friend. His relentless pursuit of scientific excellence, coupled with his mission to make a positive impact in healthcare, has been an inspiration to me and is something I strive to replicate. I deeply appreciate the combination of guidance and freedom that Eran has given me during my PhD. He allowed me to learn in my own way, and he was there when I needed honest and pragmatic feedback and help. As a leader, he clears any and all roadblocks and is willing to provide his lab with the help and tools needed to succeed. I'll be forever grateful for all of his help, time, knowledge, and support.

As my initial advisors at UCLA, Eleazar Eskin and Jason Cong both deserve a special thanks. They gave me the opportunity to learn from and work with them at UCLA, and when I wanted to change directions, they both were extremely supportive. Eleazar has been an impressive role model for creating a highly collaborative cross-disciplinary research environment and community, and I'm excited to see the Computational Medicine department grow under his leadership.

I want to especially thank my dissertation committee members Eleazar Eskin, Sriram Sankararaman, Cho-Jui Hsieh, and Eran Halperin for their guidance, patience, and consistent support throughout my PhD. Maxime Cannesson has also been instrumental as both a collaborator and a mentor. He has always been willing to share his deep clinical expertise and his honest sage advice, and for that I'm extremely grateful.

I've been extremely lucky to have been surrounded by a set of incredible friends and collaborators throughout my time at UCLA. In particular, my friends and lab mates Elior Rahmani, Nadav Rakocz, Liat Shenhav, Jeff Chiang, Akos Rudas, Mike Thompson, Zeyuan

(Johnson) Chen, Brandon Jew, Leah Briscoe, Ulzee An, Berkin Durmus, and Igor Mandric made the lab environment such a fun and exciting place to work and learn. It was also a pleasure to be located so close to so many talented students in Zar's lab (Nathan LaPierre, Rob Brown, Serghei Mangul, Kodi Collins, Lisa Gai, Jennifer Zou, Harry Yang, Dat Duong) and Sriram's lab (Alec Chiu, Chris Robles, Ruthie Johnson, Arun Durvasula, Yue (Ariel) Wu, Ali Pazokitoroudi, Boyang Fu).

I enjoyed working with so many wonderful collaborators at UCLA including Carlos Cinelli, Noah Zaitlen, Ira Hofer, Bogdan Pasaniuc, Jae-Hoon Sul, Chris Denny, Jonathan Flint, Chad Hazlett, David (Ami) Wulf, and many others. Daniel McDuff and Xin Liu were fantastic collaborators at MSR.

I'm very appreciative of Sim-Lin Lau, Stephania Kay, and Joseph Brown for their help and patience with administrative tasks over the years, often going out of their way to help.

The earliest supporter in my PhD journey was Ganapati (Gans) Srinivasa. As a young intern at Intel, Gans took me under his wing and introduced me to the world of academic research, and in particular, genomics and machine learning. He selflessly pushed me to pursue my PhD, and I can't articulate how much his friendship and mentorship has meant to me.

Without the unwavering support of my family, I'm not sure I would have been able to survive the PhD program. My parents, Linda and Dave, have supported my educational journey from the beginning, always going above and beyond to ensure that I had everything I needed to succeed. Both have been constant sources of unconditional love, encouragement, and patience. They are my role models for kindness, hard work, perseverance, and doing your best. The brothers, Jesse and Scott, have had my back every step of the way, happy to help out in any way possible.

Last but certainly not least, I want to thank my wife, Kayla, for being my foundation that holds me together. The PhD program can be a roller coaster at times, but she has held

on through the ups and the downs. She has helped and supported me in every way possible with incredible patience, enthusiasm, and love. I can't put into words how much it's meant to have her by my side during this journey through life.

Chapter 2 is a version of Brian L. Hill, Robert Brown, Eilon Gabel, Nadav Rakocz, Christine Lee, Maxime Cannesson, Pierre Baldi, Loes Olde Loohuis, Ruth Johnson, Brandon Jew, Uri Maoz, Aman Mahajan, Sriram Sankararaman, Ira Hofer, Eran Halperin; "An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data." British Journal of Anaesthesia (2019). Chapter 3 is a version of a work that was submitted for publication by Mike Thompson, Brian L. Hill, Nadav Rakocz, Jeffrey N. Chiang, Sriram Sankararaman, Ira Hofer, Maxime Cannesson, Noah Zaitlen, and Eran Halperin, tentatively titled "The methylation risk score is an informative biomarker within electronic health record systems." Chapter 4 is a version of Brian L. Hill, Nadav Rakocz, Akos Rudas, Jeffrey N. Chiang, Sidong Wang, Ira Hofer, Maxime Cannesson, Eran Halperin; "Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning." Scientific Reports (2021). Chapter 5 is a version of a work that was submitted for publication by Brian L. Hill, Xin Liu, and Daniel McDuff, tentatively titled "Learning Higher-Order Dynamics in Video-Based Cardiac Measurement."

VITA

2016–2021	Ph.D. Candidate Computer Science University of California, Los Angeles, CA
2021	Research Intern Microsoft Research, Remote
2016–2019	M.S. Computer Science University of California, Los Angeles, CA
2016–2019	Principal Software Engineer Omics Data Automation, Beaverton, OR
2012–2016	B.S. Electrical and Computer Engineering Oregon State University, Corvallis, OR
2012–2016	Intern Intel, Hillsboro, OR

SELECTED PUBLICATIONS

* denotes equal contribution

Beat-to-beat cardiac pulse rate measurement from video Brian L. Hill, Xin Liu, Daniel McDuff; Proceedings of the IEEE/CVF International Conference on Computer Vision (2021).

Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning Brian L. Hill, Nadav Rakocz, Akos Rudas, Jeffrey N. Chiang, Sidong Wang, Ira Hofer, Maxime Cannesson*, Eran Halperin*; Scientific Reports (2021).

Automated identification of clinical features from sparsely annotated 3-dimensional medical imaging Nadav Rakocz*, Jeffrey N. Chiang*, Muneeswar G. Nittala, Giulia Corradetti, Liran Tiosano, Swetha Velaga, Michael Thompson, Brian L. Hill, Sriram Sankararaman, Jonathan L. Haines, Margaret A. Pericak-Vance, Dwight Stambolian, Srinivas R. Sadda*, Eran Halperin*; npj Digital Medicine (2021).

Bladder Cancer Immunotherapy by BCG Is Associated with a Significantly Reduced Risk of Alzheimers Disease and Parkinsons Disease Danielle Klinger, Brian L. Hill, Noam Barda, Eran Halperin, Ofer N. Gofrit, Charles L. Greenblatt, Nadav Rappoport, Michal Linial, Herv Bercovier; Vaccines (2021).

A machine learning algorithm to increase COVID-19 inpatient diagnostic capacity David Goodman-Meza, Akos Rudas, Jeffrey N. Chiang, Paul C. Adamson, Joseph Ebinger, Nancy Sun, Patrick Botting, Jennifer A. Fulcher, Faysal G. Saab, Rachel Brook, Eleazar Eskin, Ulzee An, Misagh Kordi, Brandon Jew, Brunilda Balliu, Zeyuan Chen, Brian L. Hill, Elior Rahmani, Eran Halperin, Vladimir Manuel; PLOS ONE (2020).

An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data Brian L. Hill*, Robert Brown*, Eilon Gabel, Nadav Rakocz, Christine Lee, Maxime Cannesson, Pierre Baldi, Loes Olde Loohuis, Ruth Johnson, Brandon Jew, Uri Maoz, Aman Mahajan, Sriram Sankararaman, Ira Hofer, Eran Halperin; British Journal of Anaesthesia (2019).

CHAPTER 1

Introduction

1.1 Scope of Research

Improvements in both compute and data storage capabilities have ushered in the information age, creating rich "big data" resources. These data sources can be used to drive large-scale data analysis, where machine learning methods can be used to detect complex patterns and accurately predict events and trends. In particular, software and hardware advancements in general purpose graphics processing units (GPGPUs) combined with deep learning algorithms have enabled state-of-the-art performance in multiple domains including computer vision, natural language processing, and even genomics and drug discovery [LBH15, KSH12, PCA18]. Given enough data, these deep neural networks are able to learn feature representations from underlying structure that accurately map a set of input values to a set of outputs.

The twenty-first century has seen major innovations at the intersection of technology and medicine, leading to the generation of invaluable digital health data. The healthcare industry has undergone a huge transition from paper-based medical records to electronic medical records, resulting in a treasure trove of digitized health information such as laboratory results, vital sign measurements, medications that were ordered, and disease diagnoses. Medical imaging scans such as magnetic resonance imaging (MRI), computed tomography (CT), or optical coherence tomography (OCT) also contribute high-fidelity data for understanding a patient's health state. Completion of the initial sequencing of the human genome in

2001 [LLB01] has accelerated incredible research in genomics over the past two decades, and there is ongoing work to translate findings into clinical practice. Finally, wearables and other continuous health sensors are capable of generating high-frequency measurements of various vital signs, which allows a patient to be monitored around the clock by a computer in the hospital or at home. Each of these different data types provide a piece of the puzzle for understanding a patient's wellbeing.

These sources of health data can be used for two related but different tasks: prediction and imputation. We define prediction as the estimation of a future value based on data currently available at a particular time. For example, in the pre-operative period, we can use a machine learning model to predict the probability of a particular patient dying in the hospital after surgery. In the prediction task, measuring the value being estimated is theoretically infeasible at the current point in time because, by definition, the value is a future value. However, we can use retrospective data to build machine learning models for predictive tasks, and hope that they generalize to future data points.

Imputation is concerned with estimating an unobserved (missing) value from the past or current time using observed (available) data. As an example, a clinician may have ordered a complete blood count lab panel for a patient, but not a basic metabolic lab panel. Therefore we observe blood cell counts for that patient at the time the lab panel was taken, but we do not observe sodium or potassium levels. Measuring the unobserved or missing values is theoretically feasible, but for various reasons (safety, privacy, cost, etc.) the data is not available for analysis. Both prediction and imputation can only leverage data currently available, but where they differ is the point in time where the estimated value occurs.

1.2 Contributions and Overview

In this dissertation, we propose methods for solving several prediction- and imputation-related problems, leveraging various data sources such as electronic health records, genomics,

physiological waveforms, and videos.

Chapter 2 describes a machine learning model that was developed to predict the risk of postoperative in-hospital mortality using electronic health record data readily available and automatically extracted prior to surgery. We compared the performance of our model to existing clinical risk scores, and integrated our model with a previously published postoperative mortality risk score [LHG18] to quantify the change in risk during the perioperative period.

In Chapter 3 we consider the epigenetic analog of the polygenic risk score, the methylation risk score (MRS), and show that because methylation state is affected by both genetic and environmental factors it can be a useful biomarker for various phenotypes such as diagnoses, clinical lab tests, and medication prescriptions. Incorporating MRS into an imputation procedure, along with a set of baseline clinical features extracted from the electronic health record, improves the imputation performance over the baseline features alone, demonstrating the utility of collecting methylation data as versatile biomarker.

In Chapter 4, we present a novel method for imputing the continuous arterial blood pressure waveform from non-invasive signals. Using 150,000 minutes of waveforms collected from ICU patients in two separate hospital systems, we show that a deep learning model can be trained to impute the blood pressure waveform for all hospital patients using data currently collected as part of the standard-of-care.

Prior work has demonstrated that it is possible to accurately obtain an estimate of the photoplethysmogram (PPG) waveform by detecting small color changes using a camera capturing video [WBS00, PMP10, VSN08, TO07]. These computer vision methods typically optimize for first-order dynamics (such as optical flow), however, in many cases the properties of interest are subtle variations in higher-order changes, such as acceleration. In Chapter 5 we investigate several ways of incorporating higher-order dynamics into neural models for a more accurate estimation of PPG waveform morphology.

CHAPTER 2

An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data

2.1 Introduction

A small number of high-risk patients comprise the majority of patients with surgical complications [PHJ06]. Many studies have demonstrated that early interventions can help reduce or even prevent perioperative complications [LTP17, KKK12]. In the current value-based care environment, it is critical to have methods to rapidly identify patients who are at the highest risk for perioperative complications and most likely to benefit from labour or cost-intensive interventions. Unfortunately, many current methods of risk stratification either lack patient-level precision or require a trained clinician to review each patients medical record and assess a score.

Existing preoperative patient risk scores generally fall into one of two groups. The first leverage International Statistical Classification of Diseases and Related Health Problems (ICD) codes in order to create models of risk [LCR16, SSM10, CSP94]. Unfortunately, ICD codes are not available until after patient discharge. While these scores tend to perform well at the population level, they rely on data not available prior to surgery, and have been repeatedly shown to lack precision at the patient level [SLM16]. The second group of models

relies on subjective clinician judgment, as seen with the American Society of Anaesthesiologists Physical Status Score (ASA score) alone or when incorporated into another model (such as the NSQIP risk calculator) [BLP13]. While these scores tend to have increased precision compared to ICD codes, they cannot be fully automated due to the need for a highly trained clinician to manually review the patients chart prior to calculation.

Recently, attempts have been made to leverage machine learning techniques using healthcare data in order to improve the predictive ability of various models [ROC18, FK18]. These methods have shown progress in leveraging increasingly complex data while still allowing for the full automation of the scoring system.

In this work, we hypothesized that machine learning methods can be used to predict in-hospital post-surgical mortality using only features from the electronic medical record (EMR) readily available and automatically extracted before surgery. We compare the performance of our model to existing clinical risk scores (ASA score, POSPOM score [LCR16], and Charlson Comorbidity Score [CSP94]). Lastly, we aim to integrate our model with a previously published model11 that estimates in-hospital mortality at the end of surgery to quantify the change in risk during the perioperative period.

2.2 Methods

2.2.1 Data source and extraction

All data for this study were extracted from the Perioperative Data Warehouse (PDW), a custom built, robust data warehouse containing all patients who have undergone surgery at UCLA Health since the implementation of UCLAs EMR (EPIC Systems, Madison, WI) in March 2013. We have previously described the creation of the PDW, which has a two-stage design [HGP16]. Briefly, in the first stage, data are extracted from EPICs Clarity database into 29 tables organized around three distinct concepts: patients, surgical procedures and

health system encounters. These data are then used to populate a series of 4000 distinct measures and metrics such as procedure duration, readmissions, admission ICD codes, and others. All data used for this study were obtained from this data warehouse and IRB approval (IRB 16-001768) was obtained with exemption status for this retrospective review.

2.2.2 Model Endpoint Definition

We trained classification models to predict in-hospital mortality as a binary outcome. This classification was extracted from the PDW and was set to true if a death date was noted during the hospitalization, or the final disposition was set to expired and there were no future admissions for the patient and a clinician death note existed. Due to the concern about the need to eliminate false positive results, the resulting labels using this definition were validated by trained clinicians in a subset of patients.

2.2.3 Inclusion and Exclusion Criteria

Cases were included in the study if they underwent a surgical procedure with general anaesthesia between April 1, 2013 and December 10, 2018. The type of anaesthesia was extracted from the post-anaesthesia hand-off note documented by the anaesthesia provider at the end of the case. Cases were excluded if they had an American Society of Anaesthesiologists (ASA) Physical Status score of 6 (indicating organ donors), were not discharged at the time of data analysis or were aged less than 18, and patients older than 89 had their age redacted (due to institutional restrictions on data security). A CONSORT [MSA01] diagram is shown in in Supplementary Figure 1 in [HBG18].

Some patients, particularly those of highest risk, underwent more than one surgery during the course of their hospital admission. In these cases all surgeries that met the above criteria were included. We performed a subsequent analysis to ensure that their inclusion would not unduly affect the results of the entire population. This analysis is shown and described in

Supplementary Appendix 1 in [HBG18].

2.2.4 Model Input Features

The model was created using a set of features including basic patient information such as age, sex, BMI, blood pressure, and pulse rate; lab tests frequently obtained prior to surgery such as sodium, potassium, creatinine, and blood cell counts; and surgery specific information such as the surgical procedure codes. In total, 58 preoperative features (including ASA status) were selected by clinicians consensus (I.H., E.G.) as potentially useful for predicting the outcome, and a full list is available in Supplementary Table 1 in [HBG18]. For all variables only the most recent value prior to surgery was included.

In order to help elucidate the relative predictive value of different types of features, five models were created. The first model (1) included all the input features including the ASA physical status score. The second model (2) included all input features except the ASA physical status score as this score would not be able to be fully automated prior to review by a trained anaesthesia provider. In order to overcome this limitation of automation, the third model (3) included all of the input features with an automated surrogate for the ASA score. The details of the generation of this surrogate score can be found below. The last two models were variations of models 1 and 3, however they excluded the timestamps of the preoperative lab results (relative to the admission start time), though they included the actual results themselves. Since the time between a lab result and a surgery is not a marker of the patient illness, we excluded this information so that the model would not incorrectly weight the significance of this feature.

2.2.5 Comparison of Model Performance

In order to assess the performance of our models against currently used risk stratification systems we also tested the performance of three baseline models: a model containing only

the ASA physical status score, a model containing only the POSPOM score [LCR16], and a model containing only the Charlson comorbidity score [CSP94]. Using a model with a single feature such as these has the effect of producing the same result format as our more complex models and allows a direct comparison.

2.2.6 Data Pre-processing

Data points greater than 4 standard deviations from the mean were removed as they were assumed to be erroneous outliers. Categorical features were converted into indicator variables, and the first variable was dropped. Thus, if a categorical variable takes on k values, only k-1 values are converted into indicator variables because the kth variable becomes the reference value. The cohort was divided into a training dataset and a testing dataset by selecting all surgeries that occurred between April 1, 2013 and February 28, 2018 for training and surgeries from March 1, 2018 and December 10, 2018 as the test set. Any patients that appeared in the test set were removed from the training set to prevent information leakage. Temporally splitting the cohort allows us to estimate model performance on future surgical cases. The training data features were rescaled to have a mean of 0 and standard deviation of 1, and the test data were rescaled using the training data means and standard deviations. Missing data were imputed in the training and testing sets separately using the SoftImpute algorithm [MHT10], which leverages the similarity of groups of patients to estimate missing values. The SoftImpute algorithm was implemented by the fancyimpute Python package [RFO17], with a maximum of 200 iterations.

The number of inpatient mortalities was much smaller than the number of survivors, resulting in extreme class imbalance (2.01% mortality rate). To overcome this issue, the training set was oversampled using the SMOTE algorithm [CBH02], implemented in the imblearn Python package [LNA17], using 3 nearest neighbours and the baseline1 method to create a balanced class distribution. The testing set was not oversampled and, therefore, maintained the natural outcome frequency.

2.2.7 Generating a surrogate for ASA scores

While the ASA status is a strong predictor of patient health status [WWS96, Daa11, VVH70, HDJ15], this classification requires a clinician to look through the patients chart and subjectively determine the score, consuming valuable time and requiring clinical expertise. In order to balance the value of this score with the desire for automation, we sought to generate a similar metric using readily available data from the EMR - a surrogate ASA score. Recent works have similarly attempted to develop machine learning approaches to predict ASA scores [EES13, ZFW16]. However, these methods have difficulty differentiating ASA scores of 4 and 5 due to the low frequency of occurrence of 5 scores, and resort to either grouping classes together or ignoring patients with an ASA status of 5. The goal in our work is not to predict the ASA score but to estimate a measure of general patient health for use as a feature in our model to predict in-hospital mortality, without needing the time-consuming clinician chart review.

Using the existing ASA physical status classification extracted from the EMR data, we trained a gradient boosted tree regression model to predict the ASA status of new patients using preoperative features unrelated to the surgery. The model was implemented using the XGBoost package [CG16] with 2000 trees and a max tree depth of 7. We used 5-fold cross-validation to generate predictions. This surrogate-ASA value is a continuous number, unlike the actual ASA status which is limited to integers. We call it the ASA surrogate score to distinguish it from an actual ASA score. This score is a continuous score of patient risk that uses the ASA score to supervise parameter learning in the model.

2.2.8 Model Creation, Training, and Testing

We evaluated four different classification models: logistic regression, Elastic Net [ZH05] logistic regression, random forests, and gradient boosted trees. Logistic regression is a statistical model that assumes a binary outcome can be predicted as a weighted combination of inde-

pendent variables. The Elastic Net [ZH05] logistic regression adds additional constraints to a linear prediction model by forcing the weights to be both small and sparse. A random forest classifier uses an ensemble of independently-trained decision trees, which classify data based on a series of binary questions about the values of particular features, to determine the most likely outcome based on a majority vote. Like random forests, gradient boosted tree classifiers predict using an ensemble of decision trees, but instead of building each decision tree independently, the trees are created sequentially such that each new tree is fit to the residual error remaining after the previous step.

Model hyperparameters were chosen using 5-fold cross-validation on the training dataset, where surgeries from the same patient were grouped together such that they appeared only in a training fold or a testing fold, not both. In 5-fold cross-validation, the dataset is divided into 5 partitions, where 4/5 of the data is used to train the models and the remaining 1/5 is used as the testing set. This process is repeated such that each partition is used as a testing set only once, and a training set 4 times. Cross-validation provides a better assessment of model performance by averaging metrics over multiple trials. Logistic regression classifiers were trained with both an L2 penalty and an ElasticNet [ZH05] penalty, where alpha (regularization constant) and the L1/L2 mixing parameter were set using 5-fold cross-validation. The random forest classifiers were trained with 2000 estimators, Gini impurity as the splitting criterion, and no maximum tree depth was specified. The gradient boosted tree classifiers were trained using 2000 estimators and a max tree depth of 5. The logistic regression and random forest classifiers were implemented using Scikit-learn [PVG11], and the gradient boosted tree classifiers were implemented using the XGBoost package [CG16]. All performance metrics were calculated on the held-out test set using methods implemented by Scikit-learn [PVG11].

We generate confidence intervals for the test set performance metrics using block bootstrapping of the predictions. Since patients in the test set can undergo multiple surgeries, their risk predictions for each surgery are correlated. However, the general bootstrapping

procedure typically samples cases randomly, and assumes each case is independent, but under this assumption the correlation structure would be lost. Therefore, instead of randomly sampling cases, we randomly sample patients, and include all predictions in the bootstrap sample. This block bootstrap procedure is repeated 1000 times. For each bootstrap sample we calculate performance metrics, these metrics are then sorted, and we select the 25th and 975th values of the sorted list of metrics to determine the 95% confidence interval.

As described above, we compared our method with the Charlson Comorbidity Index scores [CSP94], a well-known and proven existing method for prediction of risk of postoperative mortality, for each patient in the cohort. We used the updated weights as described by Quan and colleagues [QLC11]. Scores were calculated using the R package icd [Was18] on all ICD10 codes associated with each surgery admission.

Another respected preoperative risk score is the POSPOM score [LCR16]. While the POSPOM risk score was shown to have excellent discriminative ability, the features used in the model present an issue when trying to implement such a model in a medical center that does not use the French classification for medical procedures (Classification Commune des Actes Mdicaux or CCAM). The POSPOM score groups CCAM surgery codes to 25 categories, where each category has an associated risk score as determined by their model. In order to replicate their model on our dataset, HCUP (or CPT) surgery codes must be mapped to CCAM codes. However, a mapping between HCUP (or CPT) codes to CCAM codes currently does not exist. Therefore, we created an approximate mapping between the case service group and the POSPOM surgical category. For each patient, we generated the ICD-based POSPOM score and the approximate surgical POSPOM score to compare the predictive capability of the POSPOM risk score to our method.

To determine which features were most important to the classification models, we examined the model weights for linear models, the feature (Gini) importance for the random forest models, and the feature weight (number of times a feature appears in a tree) for the gradient boosted tree models.

2.2.9 Model Calibration

A well-calibrated binary classification model outputs probabilities that are close to the true label (in our case, either a 1 for patients who die in the hospital, or 0 for survivors). Model calibration is often measured using the Brier score, which is the average squared distance between the predicted probability of the outcome and the true label; thus, a lower Brier score usually indicates a better performing model. We used this metric to assess the calibration of our models.

2.2.10 Precision and Recall Calculations

ROC curves are very informative of binary classification prediction performance in general, as they illustrate how the performance changes as the discriminative threshold varies. However, precision-recall (PR) curves can be more informative when the classes are highly imbalanced [SR15]. ROC curves show the true positive rate (recall, sensitivity) as a function of the false positive rate (1 - specificity), but for imbalanced datasets, the false positive rate can be misleading. The false positive rate is inversely related to the total number of negative samples and, therefore, a model that predicts a large number of false positives (relative to the number of true positives) may still achieve a small false positive rate. Therefore, precision (or positive predictive value) which penalizes a model for a large number of false positives relative to the number of true positives, is a useful metric. PR curves show the precision of a classifier as a function of recall. An optimal model would reach the point in the upper-right corner of the PR plot (i.e. perfect recall and perfect precision).

2.2.11 Integration of Preoperative Risk with Postoperative Risk

Previous work [LHG18] has shown that integrating a measure of preoperative risk, such as the ASA score, into a postoperative mortality risk prediction model increases the model performance. We aimed to conduct a similar approach, but instead of using the ASA status

as a measure of preoperative risk, we replaced it with the preoperative predictions from our model. First, we used the deep neural network architecture and features described by Lee and colleagues [LHG18]. However, we replaced the ASA status feature with the preoperative risk scores which were generated using the random forest model, which was trained using the preoperative features and surrogate ASA scores as described in the previous section. Next, we trained the postoperative model using the training cohort used for preoperative risk prediction using 5-fold cross-validation, where the intraoperative data was pre-processed in the same manner as described by Lee and colleagues [LHG18]. We then compared the area under the ROC of the postoperative model trained using the ASA status and intraoperative features to the model that was trained using our preoperative risk score and intraoperative features. Lastly, in order to attempt to assess the degree to which risk changes during the intraoperative period, we compared on a per-patient basis the risk scores generated by our preoperative model with those generated by the incorporation of our results with the model described by Lee and colleagues.

2.3 Results

2.3.1 Patient Demographics

The patient dataset contained 66,294 surgical records encompassing 52,894 patients. Patients were between the ages of 18 and 89, with a mean age of 56.0, and were classified as either inpatients, same-day admits, emergencies, or overnight recoveries. The frequency of mortality in the dataset was approximately 2.01%. An ASA status of 3 was the most common, comprising 47% of the dataset. Detailed information on patient demographics can be found in Table 2.1.

Table 2.1: Patient Demographics - Patient demographics for the cohort used for training and testing models. Number of patients and percent of the cohort are shown. The selected surgical services represent the top 4 most frequent surgical services.

Property	Training Data	Testing Data
Number of patients	46400	6494
Number of admissions	54813	6853
Number of surgeries	58916	7378
Average number of surgeries per patient	1.27	1.14
Average number of admissions per patient	1.18	1.06
Average number of surgeries per admission	1.07	1.08
Number of patients with more than 1 admission	6400 (13.79%)	328 (5.05%)
Number of admissions with more than 1 surgery	2817 (5.14%)	351 (5.12%)
Number of mortalities	1243 (2.11%)	124 (1.68%)
Mean age	55.99 (17.01 std. dev.)	56.07 (16.93 std. dev.)
Female patients	29770 (50.53%)	3680 (49.88%)
ASA status 1	3592 (6.10%)	383 (5.19%)
ASA status 2	21093 (35.80%)	2412 (32.69%)
ASA status 3	27395 (46.50%)	3751 (50.84%)
ASA status 4	6432 (10.92%)	779 (10.56%)
ASA status 5	404 (0.69%)	53 (0.72%)
Ronald Reagan Operating Room	39599 (67.21%)	4935 (66.89%)
Santa Monica Operating Room	19317 (32.79%)	2443 (33.11%)
Types of Surgery		
- Orthopaedics	9113 (15.47%)	1083 (14.68%)
- General Surgery	7456 (12.66%)	958 (12.98%)
- Urology	7255 (12.31%)	931 (12.62%)
- Neurological Surgery	6404 (10.87%)	843 (11.43%)
- Other	28688 (48.69%)	3563 (48.29%)

Table 2.2: Area under the ROC (AUROC) curve values for each model and each of the eight input feature sets on the held-out test set. Models with the highest AUROC are shown in bold. The mean value of the AUROC is shown, along with the 95% confidence interval from bootstrapping the test predictions 1000 times shown in parenthesis. When using the ASA status or the Charlson comorbidity score as the only input feature, the linear models (Logistic regression, ElasticNet) outperform the non-linear models (Random Forest, XGBoost). However, for the other feature sets, the non-linear models outperform the linear models. In particular, the Random Forest has the highest AUROC compared to the other models.

Model / AUC (95% CI)	Logistic Regression	ElasticNet Classifier	Random Forest	XGBoost Classifier
POSPOM	0.653 (0.602-0.705)	0.653 (0.602-0.705)	0.660 (0.598-0.722)	0.660 (0.598-0.722)
Charlson Comorbidity	0.742 (0.658-0.812)	0.742 (0.658-0.812)	0.740 (0.658-0.811)	0.740 (0.658-0.811)
ASA Status	0.866 (0.829-0.897)	0.866 (0.829-0.897)	0.855 (0.819-0.888)	0.855 (0.819-0.888)
Preop Features	0.900 (0.863-0.931)	0.919 (0.891-0.942)	0.925 (0.900-0.947)	0.920 (0.894-0.944)
Preop + ASA Status	0.913 (0.880-0.940)	0.924 (0.895-0.947)	0.936 (0.915-0.956)	0.922 (0.894-0.948)
Preop + surrogate-ASA	0.908 (0.872-0.937)	0.923 (0.895-0.946)	0.931 (0.909-0.952)	0.929 (0.907-0.948)
Preop (No Time) + ASA Status	0.919 (0.887-0.944)	0.932 (0.908-0.951)	0.936 (0.917-0.955)	0.923 (0.895-0.950)
Preop (No Time) + surrogate-ASA	0.911 (0.877-0.941)	0.924 (0.898-0.948)	0.932 (0.910-0.951)	0.915 (0.887-0.940)

Table 2.3: Random Forest model performance metrics for predicting in-hospital mortality using different sets of features. Confidence intervals derived by bootstrapping the predictions using 1000 samples shown in parenthesis. True positives: TP, False positives: FP, True negatives: TN, False negatives: FN. Accuracy = (TP+TN)/(TP+TN+FP+FN). Precision = TP/(TP+FP). Recall = TP/(TP+FN). Specificity = TN/(TN+FP). F1 Score = 2/((1/Recall) + (1/Precision)).

Model	Accuracy	F1 Score	Precision	Recall	Specificity
POSPOM	0.861 (0.851-0.869)	0.047 (0.021-0.078)	0.026 (0.012-0.045)	0.201 (0.097-0.318)	0.872 (0.864-0.881)
Charlson Comorbidity	0.895 (0.885-0.904)	0.112 (0.064-0.165)	0.065 (0.037-0.098)	0.390 (0.240-0.538)	0.904 (0.895-0.913)
ASA Status	0.897 (0.889-0.906)	0.160 (0.110-0.222)	0.093 (0.061-0.133)	0.587 (0.472-0.709)	0.903 (0.895-0.911)
Preoperative Features	0.985 (0.981-0.988)	0.275 (0.115-0.446)	0.610 (0.333-0.814)	0.179 (0.069-0.315)	0.998 (0.997-0.999)
Preop Features + ASA Status	0.984 (0.980-0.988)	0.284 (0.119-0.464)	0.590 (0.333-0.810)	0.189 (0.074-0.329)	0.998 (0.997-0.999)
Preop + surrogate-ASA	0.984 (0.980-0.988)	0.280 (0.125-0.452)	0.541 (0.294-0.750)	0.191 (0.078-0.331)	0.997 (0.996-0.998)
Preop + ASA status, w/o Lab Times	0.982 (0.977-0.986)	0.302 (0.172-0.449)	0.420 (0.245-0.615)	0.239 (0.127-0.379)	0.994 (0.992-0.997)
Preop + surrogate-ASA status, w/o Lab Times	0.980 (0.976-0.985)	0.258 (0.127-0.412)	0.358 (0.180-0.551)	0.204 (0.094-0.342)	0.994 (0.992-0.996)

2.3.2 Model Performance

Area under the receiver operating characteristic (ROC) curve The area under the ROC curve values for each model are shown in Table 2.2 and ROC curves are shown for the random forest model in Figure 1a and for all models in Supplementary Figure 2 in [HBG18]. For all models except the ASA status alone, the random forest model produced the best results, although these differences often did not reach statistical significance. Models using the preoperative features have higher area under the ROC values (0.925, 95%CI 0.900-0.947) than the models that use the Charlson comorbidity score (0.742, 95%CI 0.658-0.812), the POSPOM score (0.660, 95% CI 0.598-0.722), or the ASA status (0.866, 95%CI 0.829-0.897) alone. Adding the surrogate ASA status values to the preoperative features did not improve the area under the ROC (0.931, 95%CI 0.909-0.952) as compared to the preoperative features alone (0.925, 95%CI 0.900-0.947). While adding the true ASA value assigned by anaesthesiologists to the preoperative features (0.936, 95%CI 0.915-0.956, Wilcoxon signed-rank test, p¿0.05 as compared to preop features with surrogate ASA score) and reducing the preoperative feature set by removing variables indicating when the lab tests resulted [(0.932, 95%CI 0.910-0.951) and the preoperative features and true ASA status (0.936, 95%CI 0.917-0.955)] increased the AUC, these increases were not statistically significant (Wilcoxon signed-rank test, p-value¿0.05). Table 2.3 contains the accuracy, F1 score, precision, recall and specificity for all five random forest models.

2.3.3 Calibration

The non-linear models (Random Forest, XGBoost) had much lower (better) Brier scores compared to the linear models (Logistic Regression, ElasticNet). When using either the POSPOM score, the Charlson comorbidity score, or the ASA status as the only feature, the Random Forest and XGBoost classifiers had the lowest Brier score (0.098, 0.091, and 0.086 respectively). For the other five feature sets the XGBoost models obtained the lowest

Brier scores (0.015, 0.015, 0.016, 0.016, and 0.017 respectively). These data are shown in Supplementary Table 2 in [HBG18].

2.3.4 Precision-Recall

Using the random forest model, precision and recall (PR) curves for each of the sets of features are shown in Figures 1b and, for all models, in Supplementary Figure 3 in [HBG18]. Overall the various sets of preoperative features had better performance than the ASA score, the Charlson co-morbidity score, and the POSPOM score.

Hospitals have limited resources and must decide how to allocate those resources. One option is to allocate prioritized care to individuals who are at the highest risk of adverse outcomes, particularly mortality. A hospital could choose to use the ASA score, the Charlson comorbidity score, the POSPOM score, or the score generated by the random forest model as an estimate of the risk. Our score is continuous and therefore has a definitive ordering of patients, while the ASA score, the Charlson comorbidity score, and the POSPOM score, being discrete, have random intra-score level ordering. To assess the effectiveness of the ordering based on the proposed score compared to the ASA score, the Charlson comorbidity score, and the POSPOM score, in Figure 2.2 we order the individuals by their risk of mortality and calculate the number of mortalities in our set of high-risk patients as we vary the size of the set. In other words, if we have a fixed set of resources such that we can allocate additional care to n patients, we would like to know how many of the n patients are true positives. While receiving prioritized care does not imply that a specific individual will not die, we argue that a population should have improved outcomes as care levels are better matched to patients.

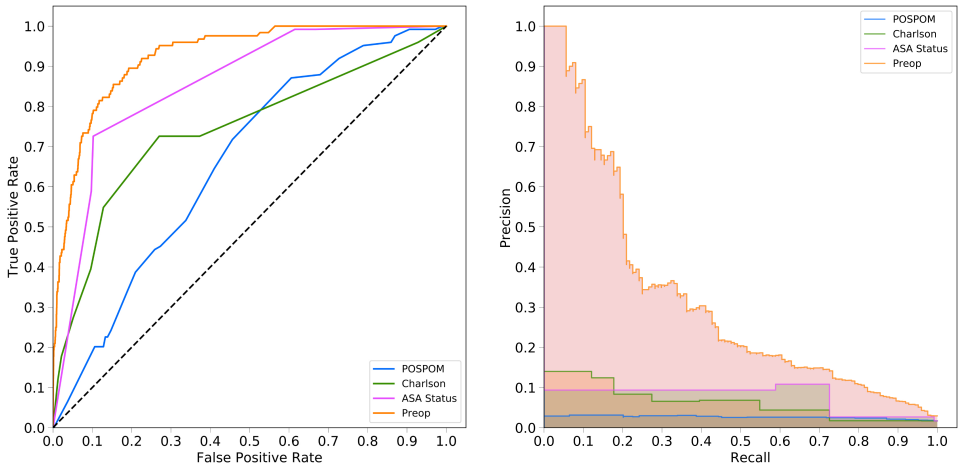

Figure 2.1: Receiver operating characteristic (ROC) and precision recall curves for the random forest model. Plots were generated using the predictions from the held-out test dataset. ROC curves (a, left) show the false positive rate on the x-axis and the true positive rate on the y-axis. The optimal point is the upper-left corner. Precision-recall curves (b, right) show the recall on the x-axis and precision on the y-axis. The optimal point is in the upper-right corner.

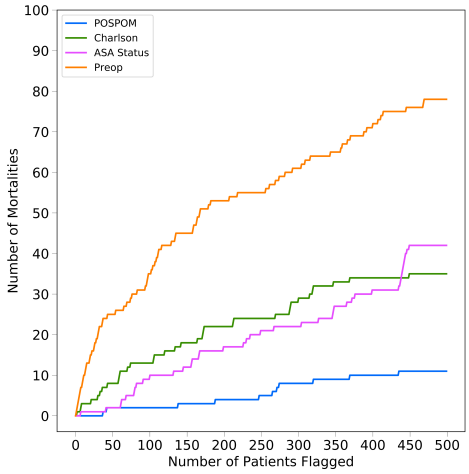

Figure 2.2: Number of in-hospital mortalities captured as a function of the number of patients flagged as high-risk. Using the random forest predicted probabilities for each set of features, surgeries were ranked from highest to lowest risk. For each feature set, we count the number of mortalities captured as we vary the number of high-risk patients flagged for additional resources.

2.3.5 Feature Importance

To determine the most important features for each of the models, we examined the feature weights of the linear models and feature importance of the non-linear models. In Supplementary Table 3 in [HBG18], the feature importance for the random forest model is shown using four different sets of input features. For the feature sets that include lab result timestamps, many of the most important features are the lab result time-stamp features for labs such as BNP and bicarbonate. However, when these features are removed, the feature importance shifts to the lab results themselves, for example, albumin, INR, prothrombin time, haemoglobin, and total bilirubin. Surgery-specific features such as the patient class (inpatient, same-day admission) and the location of the patient in the hospital before surgery are also highly informative. Additionally, the ASA score is the most important feature in every model where it is contained.

2.3.6 Integrating Preoperative Risk with Postoperative Risk

Replacing the ASA status with the preoperative risk predictions in the postoperative risk prediction model generated similar results to what were previously published by Lee and colleagues. The postoperative risk model that was trained using the preoperative risk scores had an area under the ROC of 0.943 (95%CI 0.934-0.953), whereas the postoperative model trained using the ASA status had an area under the ROC of 0.935 (95%CI 0.926-0.947) (Wilcoxon signed-rank test, p-value¿0.05). This is in line with the previously published results of this model [LHG18].

In order to examine how mortality risk changes from immediately before surgery to after surgery, the preoperative and postoperative risk scores for all patients were grouped by percentiles and the counts of each grouping are displayed in Figure 2.3a. For the majority of patients, we see a slight increase or decrease in their postoperative risk compared to the initial preoperative risk, as demonstrated by the colouring just above/below the diagonal line

in Figure 2.3a. Figure 2.3b demonstrates the same plot but contains only those patients who eventually died during that admission. Most of these patients fall above the line indicating that their risk increased during the intraoperative period, and all patients in this cohort who had a preoperative risk below the 50th percentile had a postoperative risk that was substantially increased. Supplementary Table 4a and Supplementary Table 4b in [HBG18] quantify this change in risk for the entire cohort and the in-hospital mortalities, respectively.

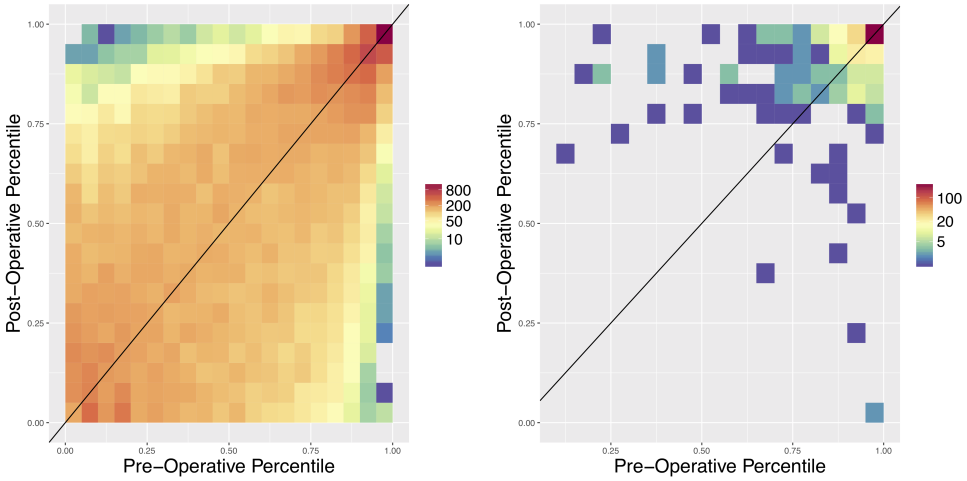

Figure 2.3: Heatmap of Preoperative Risk vs. Postoperative Risk. Preoperative (x-axis) and postoperative (y-axis) risk scores were binned by percentile, and the counts per bin visualized as a heatmap in log scale. Preoperative risk predictions were generated using the random forest model trained on the preoperative features, including lab times, and the surrogate-ASA status. In (a, left) all patients are displayed, and in (b, right) only the in-hospital mortalities are shown. 78% of patients who die and have a preoperative risk percentile below 95% have an increased postoperative risk percentile. This is substantially greater than the percent of matched patients from a null distribution who have an increased percentile.

2.4 Discussion

In this manuscript we were able to successfully create a fully-automated preoperative risk prediction score that can better predict in-hospital mortality than the ASA score, the POSPOM score, and the Charlson comorbidity score. In contrast to the ASA score, the POSPOM score,

or the Charlson comorbidity scores, this score was built using purely objective clinical information that was readily available from the EMR prior to surgery and does not require a clinicians assistance for score calculation. Unlike previous models [LHG18], the results indicate that inclusion of the ASA score in the model did not improve the predictive ability. We were additionally able to integrate the results of our model into a previously developed postoperative risk prediction model and achieve a performance that was comparable to the use of the ASA physical status score in that model. Lastly, when using the preoperative and postoperative scores together we were able to demonstrate that, on a patient level, risk does change during the perioperative period - indicating that choices made in the operating room may have profound implications for our patients.

The challenge of perioperative risk stratification is certainly not new. In fact, the presence of so many varied risk scores (ASA Score, Charlson Comorbidity Index [CSP94], POSPOM [LCR16], RQI [DKT11], NSQIP Risk Calculator [BLP13]) speaks to the importance with which clinicians view this problem. A major limitation of many of these models has been that they either rely on data not available at the time of surgery (i.e. ICD codes), or they require an anaesthesiologist to review the chart (those that contain the ASA score). Thus, the creation of a model that can be fully automated and perform better than these models implies that it may have broad applicability. Of note, in this study, the non-linear machine learning models outperformed logistic regression both regarding AUC and calibration (Brier score). This is different than what has been shown in other work [LHG18] where the logistic regression performed similarly to non-linear machine learning approaches.

As demonstrated in many previous studies [WWS96, Daa11, VVH70, HDJ15], the ASA score itself remains a good predictor of postoperative outcomes. This is likely because the ASA score is essentially a predictor generated by the most advanced neural network known the human brain. However, the ASA score alone did not perform as well as our baseline model. The discrepancy may have several possible explanations. One possibility is that the introduction of the EMR has led to an explosion of information, making it challenging for a

clinician to consume everything. It would be essentially impossible for an anaesthesiologist to review every note, lab result and pathology report prior to surgery. A second possible explanation is that the ASA score is not a predictor of mortality per se, but rather a marker of overall patient complexity. Thus, a score that is designed to predict a specific complication such as mortality will perform better than a measure of overall patient complexity. Regardless of the exact reason, we believe this highlights the advantage of an automated scoring system such as this - not as a replacement for physicians - but as a tool to help them better focus their efforts on those patients most likely to benefit.

Another advantage of an automated model such as this is that it allows for the continuous recalculation of risk longitudinally over time. As shown in Figure 2.3, most patients have either a minor increase or decrease in risk in the time from before to after surgery and, unsurprisingly, in patients who eventually die, that risk tends to increase. Given the challenges of continually monitoring the risk of all patients in the hospital, advanced analytical models such as the one demonstrated in this manuscript have great potential to act as early warning systems alerting clinicians to sudden changes in risk profiles and facilitating the use of rapid response teams.

More importantly, the frequency with which risk changed substantially during the operative period highlights the effect to which intraoperative interventions may have implications far beyond the operating room. Multiple specific interventions, including the avoidance of intraoperative hypotension and hypothermia, have been shown to have effects on longer-term outcomes, and currently enhanced recovery after surgery (ERAS) pathways have promoted the standardization of intraoperative interventions. We believe that our findings should add to the evidence that a well-prescribed anaesthetic plan may be of significant long-term benefit to patient outcomes.

One potential promise of the use of machine learning in medicine is the ability to leverage these models in order to better understand what features are truly driving outcomes. In an effort to better understand this, we extracted the weights of the features in both the linear

and non-linear models. Removing some features, specifically the relative time of lab tests, actually improved the results of our model (though not to a level of statistical significance). This could potentially be caused by multiple correlated features tagging an underlying cause, and the correlation introduces noise in the model as the importance is distributed among multiple features rather than focused on a single feature. In theory, a machine learning model should be able to remove these features by setting a coefficient to zero. However, in practice, this may not always be the case - as illustrated here, where we force this behaviour by manually removing the features from the model. We believe that this finding highlights the importance of having collaborative relationships between experts in machine learning and clinicians who are able to help guide which features to include in a model. Simply entering large amounts of data from an EMR, without proper clinical context, is unlikely to create the most effective or efficient models.

There are several key limitations of this study. The most significant is the low frequency of the outcome in question - in-hospital mortality. The incidence of mortality in the testing set was less than 2% - implying that a model that blindly reports survives every time will have an accuracy greater than 98%. Predicting such a rare outcome makes it highly challenging to produce results with very high precision. Nonetheless, the models presented in this paper do outperform other models currently in use, as measured by area under the ROC curve, and had precision-recall curves that were superior to the ASA score, POSPOM score, or Charlson comorbidity scores alone. Secondly, the large amount of missing data in the EMR makes imputation a complex task, in particular because the data are not necessarily missing at random. Many of the missing values are due to systematic reasons, such as forgoing a set of lab tests because the clinician believes the patients lab values are relatively normal. In fact, creating optimal imputation algorithms is whole field of work on its own and suboptimal imputation algorithms will reduce the prediction performance. However, given the sparsity of the data, some form of imputation is necessary and our choice of imputation algorithm, while not optimal, is better than a trivial method such as mean imputation (see Supplementary

Figures 4, and 5a-c in [HBG18]). In fact, our algorithm performed better than the same algorithm using mean imputation (see Supplementary Figure 6a and b in [HBG18]). We believe that the overall strong performance of our models despite these limitations indicates the value of machine learning in predicting postoperative outcomes. Thirdly, the data used here are from a single large academic medical centre. Thus, it is possible, though unlikely, that this model will not perform similarly at another institution. More likely is that the model would require recalibration in order to be transferred from one institution to another. However, with such a recalibration the exact weights of the various features might change. One last limitation lies not necessarily with the study itself but with the overall landscape of EMR data. While the promises of fully-automated risk scores are great, the reality remains that most institutions still have trouble accessing the data stored in the EMRs. Thus, in order to truly automate processes such as these, robust data interoperability standards (such as Fast Healthcare Interoperability Resources (FHIR)) will be needed in order to allow access to data.

The promise of using machine learning techniques in healthcare is great. In this work we have presented a novel set of easily-accessible (via EMR data) preoperative features that are combined in a machine learning model for predicting in-hospital post-surgical mortality, which outperforms current clinical risk scores, however, a model that incorporates both physician judgement (via the ASA score) with machine learning produces the best results. We have also shown that the risk of in-hospital mortality changes over time. It is our hope and expectation that the next few years will produce a plethora of research leveraging data obtained during routine patient care to improve care delivery models and outcomes for all of our patients.

CHAPTER 3

The methylation risk score is an informative biomarker within electronic health record systems

3.1 Introduction

Widespread adoption of electronic health record systems coupled with an increasing interest in hospital biobanking systems has spurred research efforts spanning machine-learning and genomics communities[SGA15a, MWG05, RPB08, BHH18, HJM19, LTN19, CHB19]. These efforts have produced increasingly accurate imputation (current state) and prediction (future state) of patient phenotypes from medical records [CKL18, HBG19] and polygenic risk scores [SGA15a, MWG05, RPB08, KCA18, MMD19, LH19, LV20, KMH21], and are already being investigated in translational contexts [HJB18, WGH20, GPC16, BBH20]. For example, recent work has shown that machine learning can leverage high-dimensional data to aid in the prediction of a multitude of clinical phenotypes including cardiac function and arrhythmia [GOA20, RRP20, HRH19], post-operative complications [CKL18, HBG19], sepsis [KCB18], breast cancer [MBJ16, MMD19], and prostate cancer [SAB18]. Nonetheless, a genetics-based predictor such as the polygenic risk score may be limited in predictive utility as it does not account for changes in disease risk throughout one's lifespan [LV20].

In this work we examine the potential for epigenetic information to improve phenotype inference in combined biobank-EHR systems. As methylation is affected by both genetics and environment—such as lifestyle choices, diet, exercise, and smoking status—it captures multi-factorial information about predispositions to clinical conditions [LP13,

GGO17, HHA19, WSB13, ZMC11, DNT14]. Moreover, methylation is readily available for use in existing biobanks that collect DNA samples, and recent advancements in methylation profiling technologies have enabled an abundance of large-scale studies of methylation and its role as a biomarker for a variety of phenotypes and health-related outcomes [Lev10, KNK19, CTS17, LAP13, RBD11, HGT14, DNT14]. It is therefore a natural candidate for an extension of PRS, and we hypothesized that methylation can be used to complement genetics as a clinical prediction tool. To that end, we have generated and evaluated methylation risk scores (MRS), which are linear combinations of CpG methylation states.

To comprehensively investigate the utility of MRS and characterize its properties, we conducted a study of 651 EHR-derived phenotypes spanning medications (e.g. vasopressers, glucocorticosteroids, fluoroquinolones), labs (e.g. creatinine, glucose, prothrombin time), and diagnoses (e.g. T2D, bacterial pneumonia, anemia) that were available for a sufficient number of patients in the cohort. The cohort contained 826 patients—to the best of our knowledge, the largest epigenetic biobank dataset to date (including genetics, methylation, and EHR)—from the UCLA Health ATLAS cohort across a wide range of ages (18-90), racial and ethnic groups, and overall health (including patients ascertained on kidney and heart disease, with matched controls), with corresponding genetic and EHR data. This provides the opportunity to study the potential contribution of methylation to larger biobanks and in multiple clinical contexts. We find that the MRS-based imputations were more informative compared to PRS in 78 (87%) medications, 34 (91%) labs, and 128 (83%) diagnoses, more than doubling the imputation accuracy in over half of the outcomes considered. We also show that the MRS improves the imputation accuracy over PRS for cases in which the PRS is trained on very large external biobanks (roughly 3 orders of magnitude larger), as opposed to 826 samples that are available in this study. We observe that MRS improves over PRS learned from large biobanks in 50% of the tested phenotypes. Further, as our cohort was ethnically diverse, we performed replicability analyses within each racial and ethnic subset of our data. We broadly showed the replicability of the five best-predicted medications,

labs, and diagnoses —53% and 100% of which replicated in (n=123) non-white Hispanic-Latino- and (n=561) white non-Hispanic-Latino-identifying individuals respectively. Finally, we demonstrate the ability of MRS to transfer between methylation arrays and cohorts by conducting an association study of kidney-related MRS in an external diabetic nephropathy EWAS [BTR10], where the minimum replication p-value was 6.36×10^{-7}.

These results provide evidence for the utility of methylation in phenotype imputation in general, and in biobank settings in particular. However, the promise of clinical translation of genomic risk scores, including PRS or MRS, is highly dependant on the clinical context of the patient. There is a large body of work investigating phenotype imputation and prediction in clinical settings using EHR data alone, typically with machine learning techniques, without any genomic data. To the best of our knowledge, the question of whether genomic data can be used to complement such algorithms has not been studied. Since the application of MRS or PRS to clinical data without taking into account the EHR data provides a limited clinical utility, this is a natural question.

Here, we demonstrate that MRS can be used in conjunction with EHR data to improve the imputation of clinical data of patients. Critically, most machine learning approaches rely on imputation because of the inability of such algorithms to process missing data, making accurate imputation a crucial step. We found that the combination of MRS with a gold standard imputation approach—SoftImpute [MHT10]—for clinical data imputation, provides improved accuracy in 42.4% of the examined phenotypes with a median increase of 71.5%. This result provides the potential to improve machine learning algorithms that use the EHR data, by complementing the data with methylation information for the patients.

In summary, our results quantify the contribution of methylation information in clinical settings, both in isolation and in conjunction with the EHR data, and they demonstrate the potential utility of epigenetic biobanks in clinical settings.

3.2 Results

Risk model description Analogous to the PRS, we defined the MRS by a linear combination of m CpG site beta values c and weights w:

$$\text{MRS} = \sum_{i=1}^{m} w_i c_i \tag{3.1}$$

To ensure the methylation risk score added predictive value over commonly captured features (e.g. age and sex), we created a baseline predictive model that included patients' age, sex, reference-based methylation cell-type composition estimates [HAK12], self-reported race-ethnicity, and the first ten genetic principal components [GGO17] (see Supplementary Table A.1 for cohort demographic data). We fit the baseline model using a linear or logistic regression model depending on whether the outcome was continuous or binary. We compared the baseline model to models that included the baseline features as well as either methylation or genotype data. For both the MRS and PRS, we used regression with LASSO, elastic net, and ridge regularization over the genomic features while treating the baseline features as fixed covariates. We fit all models using 10-fold double cross-validation, wherein each training set an additional cross-validation was performed for hyperparameter selection, then this training-set cross-validated model was used to predict the held-out test set. We tested for significance using an association test (via linear regression) between the cross-validated predicted outcome (i.e. the concatenated predictor across all folds) and the true outcome. For full details see Methods.

Methylation risk scores significantly outperform the baseline and PRS models From our EHR database, we extracted diagnosis codes, medication orders, and the most recent lab results, all of which occurred before the methylation samples were collected. We aggregated the ICD codes into higher-level phenotypes according to the phenotype code (Phecode) mapping proposed by Denny et al. [DRB10, DBR13] and grouped individual

medications by pharmaceutical subclass to increase generalizability and power.

We trained penalized linear models to predict clinical phenotypes for which there was a sufficient number of patient data available, which included 172 medication subclasses, 69 lab values, and 369 Phecodes. Using a Bonferroni-adjusted association test, the baseline and MRS models significantly predicted the usage of 72 and 84 medications, 19 and 35 labs, and 117 and 142 Phecodes respectively (Figure A.1). We compared the performance of the MRS to a model that used both the PRS and baseline features on the same set of individuals, which significantly predicted the usage of 48 medications, 20 lab results, and 89 Phecodes. Notably, the baseline model predicted a greater number of medications, labs, and Phecodes than models that leveraged a PRS, which suggests that including genomic features may either add noise or our sample size may not have been sufficient to discover their effects for certain outcomes.

Next, we investigated outcomes for which genomic features add predictive power to the baseline features and, in such cases, the extent to which their inclusion improves predictive accuracy. We conducted a likelihood ratio test comparing an association test of the true outcome using the cross-validated baseline predictor alone, to a model that included the cross-validated baseline predictor as well as the cross-validated predictor that included both baseline and genomic features. The methylation significantly improved the baseline predictor for 60 medications, 30 labs, and 64 Phecodes, and led to a median increase of 10.42%, 459.25%, and 15.35% over the baseline predictor's accuracy (AUC, R^2) in each outcome, respectively (Figure 3.1). The genotypes significantly improved the baseline predictor for 13 medications, 11 labs, and 14 Phecodes, and led to a median increase of 5.10%, 33.79%, and 1.24% increase over the baseline in the AUC and R^2 of each outcome respectively (Figure 3.1).

The medications that improved the greatest using methylation corresponded to drugs often prescribed to individuals with neutropenia (hematopoeitic growth factors, AUC baseline .688 95% CI [.644,.729] to AUC methylation .845 [.811,.877]) or chronic kidney disease

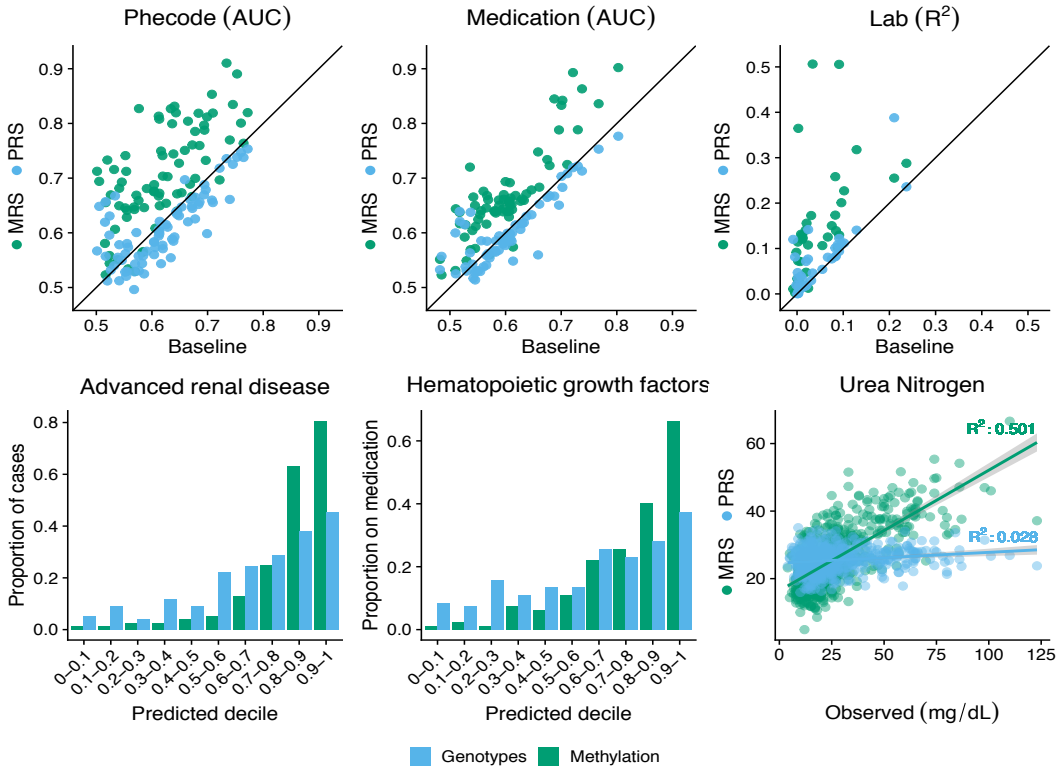

Figure 3.1: MRS increases prediction accuracy on a variety of outcomes. (Top) The performance of the PRS (blue) and MRS (green) predictions on the y-axis with the baseline model predictions on the x-axis. The performance of binary phenotypes (Phecodes, medications) is measured using area under the ROC curve (AUC) and the performance of continuous phenotypes (lab results) is measured using proportion of variance explained (R^2). (Bottom) The disease incidence as a function of the PRS (blue) and MRS (green) binned by deciles (left, middle); and the observed Urea Nitrogen lab result value plotted against its predicted value (right).

(phosphate binder agents AUC from .721 [.672, .767] to .893 [.858, .921]). The lab panels best improved with the addition of the methylation-based predictor included those related to kidney function as well as cell counts (Urea nitrogen baseline adjusted R^2 .034 [.020,.047] compared to .506 [.475,.538] with methylation, hemoglobin .129 [.105,.153] to .318 [.289,.347]). The addition of the genotype-based predictor improved greatly the prediction QRS duration (adjusted R^2 of .020 [.005,.034] to .072 [.042,.101]) and forced expiratory volume (.210 [.134,.284] to .388 [.301,.475]), both of which are influenced by ancestry [RBS17, HMW04]. In the context of Phecodes, methylation greatly increased the prediction of advanced renal

disease over the baseline and genotype models (for example, AUC baseline .734 [.688,.776] to 0.910 [.880,.936] with methylation), and the genotype model increased the prediction of heart-related conditions such as AV block (AUC from .524 [.476,.571] to .596 [.553,.643]).

Overall, when looking at the intersection of medications significantly predicted by either the methylation and genotypes or methylation and baseline, 88% were better predicted by methylation sites than genotypes (median 9.45% increase) and 80% were better predicted by methylation compared to the baseline (median 6.90% increase). In the context of significantly predicted lab values, methylation explained more variability than the baseline (median 185% increase) and genotype (median 195% increase) predictors in 97% and 91% of the respective union of significantly predicted labs. Methylation was more accurate than the baseline (median 3.80% increase) or genotypes (median 7.51%) for 65% and 83% of each respective union of Phecodes. For the prediction performance on the full list of phenotypes, see Supplemental Tables A.2, A.3, and A.4.

Importantly, cell-type composition, age, sex, and ancestry provide sufficient power for the prediction of many EHR outcomes. In our analyses, we directly compared the power gained by methylation over this set of baseline features. However, we note that obtaining these baseline features may be unnecessary as the methylation alone may capture their signal [Hor13a, Lev10, GGO17, ZMC11, SSW15]. Further, previous reports have suggested that approaches that fit all methylation probes simultaneously with regularization may perform better when excluding latent confounders, such as cell type composition [TMP20]. We therefore suggest that using the methylation alone is sufficient to replicate a substantial proportion of the associations generated from the baseline features alone.

Using methylation risk scores improves imputation approaches Due to significant heterogeneity in patient populations, the diagnosis and treatment process can vary widely between patients, causing many variables to be left unobserved. This sparse structure in the data must be reconciled before performing many downstream analyses, and the imputation

accuracy of these unobserved variables is therefore crucial to subsequent steps. A commonly-used approach for imputation is matrix completion, for example, SoftImpute [MHT10], where the data matrix is reconstructed from a low-rank representation. Often, one would jointly use demographic information, diagnosis codes, lab results, and medications to generate an estimate of the unobserved EHR values using an imputation method such as SoftImpute, and therefore we used this as our baseline imputation estimate [BLS18].

To investigate whether methylation can add additional useful information to the imputation, we included the MRS values as part of imputation procedure and compared the performance to the estimates that do not take methylation data into account (see Methods). Specifically, we included cross-validated MRS values for diagnosis codes, lab results, medications, and demographics that were significantly imputed as 261 additional features (i.e. columns of the input matrix) in the imputation procedure. We randomly removed a subset of the observed lab results, including other labs that are ordered as part of the same lab panel(s), and imputed the masked values using the remaining observed values. The imputed values were then compared to the held-out, masked values to assess the quality of the imputation. In Figure 3.2, we show the imputation accuracy (R^2 between the masked true and imputed values) for labs where the addition of cross-validated MRS to the baseline SoftImpute procedure explained significantly more variability. Of the 66 lab results considered, 28 (42.4%) were significantly better imputed by including the MRS values. Including the MRS values led to a median increase of 71.5% (95% CI 28.8%-99.6%) in the imputation R^2 values.

Methylation risk scores will improve with larger sample sizes In this study, our analyses of prediction accuracy were performed on 826 individuals' methylation and genetic features. For many phenotypes, the genetic effects are relatively small and require large sample sizes to identify associations between genomic features and the outcome of interest. Consequently, in many biobanks the number of individuals with measured genomic features is several orders of magnitude larger than our sample size [SGA15a, MWG05, RPB08]. While

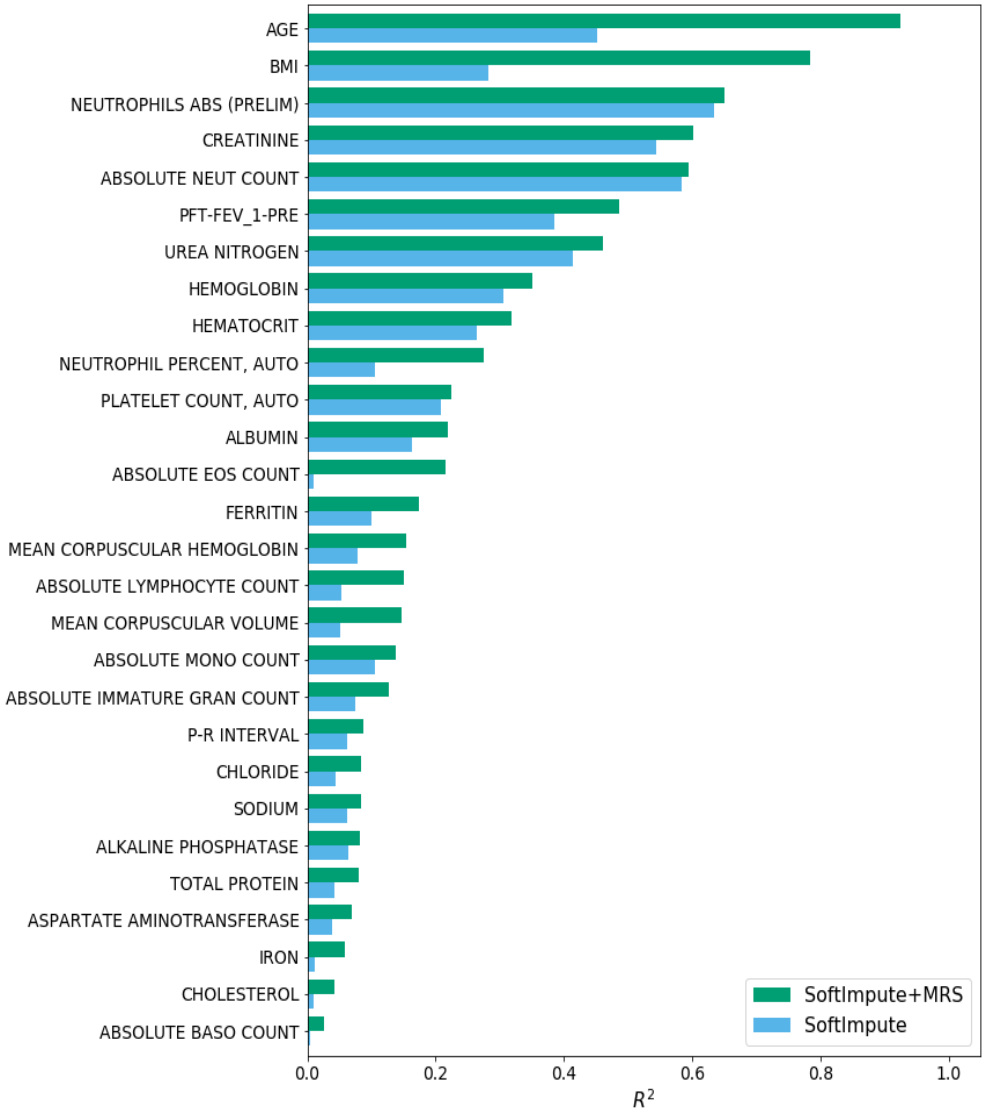

Figure 3.2: Improvement in lab result imputation performance by including MRS. For lab results that were significantly better imputed using a matrix completion imputation procedure that included the MRS values, we compare the quality of the imputed values (R^2) using only the EHR data (SoftImpute) to the values generated when including the MRS values in addition to the EHR data (SoftImpute+MRS).

the methylation data provided sufficient power to significantly predict numerous outcomes, there may remain much power to be gained by increasing the number of methylation samples to numbers approaching biobank-scale.

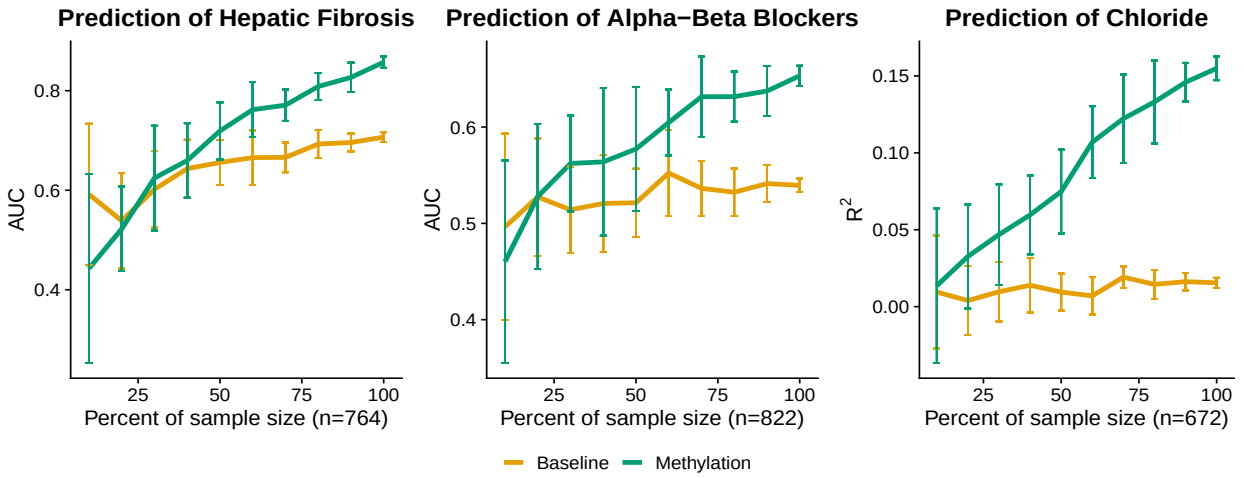

Figure 3.3: Prediction accuracy may improve with additional samples. We downsampled the number of individuals to evaluate the prediction performance as a function of sample size using a well-predicted medication and lab value. The performance is significantly affected by the number of individuals, suggesting that there is additional power to be gained with the addition of more methylation samples.

To determine the role of sample size in our prediction accuracy, we performed an experiment in which we downsampled the number of individuals in our data and trained models on the subsampled data. From the best methylation-predicted outcomes, we chose one medication, lab, and Phecode on which to perform 10-fold cross-validation. For each sample size, we repeated the procedure 20 times to attempt to mitigate variance due to ascertainment effect. Though we selected features that had high accuracy using the full set of data, our results suggest that our models may become more accurate as the sample size increases (Figure 3.3). We further posit that there may be additional outcomes that will be significantly predicted as the number of methylation samples increases.

Comparing MRS to UKBiobank PRS As expected, due to a small sample size and the likely small effects of SNPs on phenotypes, the PRS developed using the UCLA cohort did not add substantial predictive power over the baseline features. Studies leveraging biobanks with sample sizes several magnitudes larger than the cohort at UCLA however, have shown non-zero heritability for a variety of phenotypes [SGA15a]. Therefore, we sought to compare

the MRS and PRS generated with the UCLA data to a polygenic risk score created using the UKBiobank data [SGA15b]. To do so, we obtained the genotype weights corresponding to 10 polygenic risk scores trained on the UKBiobank [SGA15b, STA21, VBA20] data and imputed the external risk scores into our health record system using PLINK [PNT07]. We included in the comparison labs that were significantly predicted by the baseline model and excluded labs that corresponded to cell counts or labs for which the internal PRS outperformed the external PRS (indicating a mismatch in the phenotypes or cryptic population structure that was unaccounted for by principal components). While the external polygenic risk score improved substantially the imputation performance relative to the internal polygenic risk score, it did not significantly outperform the methylation for any of the tested phenotypes (Figure 3.4). The methylation remained the best predictor in general—even when trained on fewer than 1000 samples—significantly outperforming the other models in the prediction of urea nitrogen, creatinine, hematocrit, albumin, and mean corpuscular volume. The externally-derived polygenic risk score greatly outperformed both the internally-derived PRS and the MRS when predicting glycated hemoglobin and HDL levels, however the improvement was not significant.

Evaluation of methylation risk scores across ancestral populations Previous reports have suggested that a significant confounder to the application and versatility of polygenic risk scores is population structure, where a population-specific bias is induced that affects generalizability of PRS to different ancestries [DSG19, KMK19, MKK19]. The collection of samples analyzed throughout this study is ethnically heterogeneous—individuals were self-identified as non-Hispanic/Latino European, Hispanic/Latino, Black, or Asian. Methylation data is also influenced by differences in population [RSS17], and in particular the first several methylation principal components sufficiently capture population structure in European and African groups [BAK14, MZM13]). Consequently, we examined the performance of the methylation risk scores within and across ancestral populations.

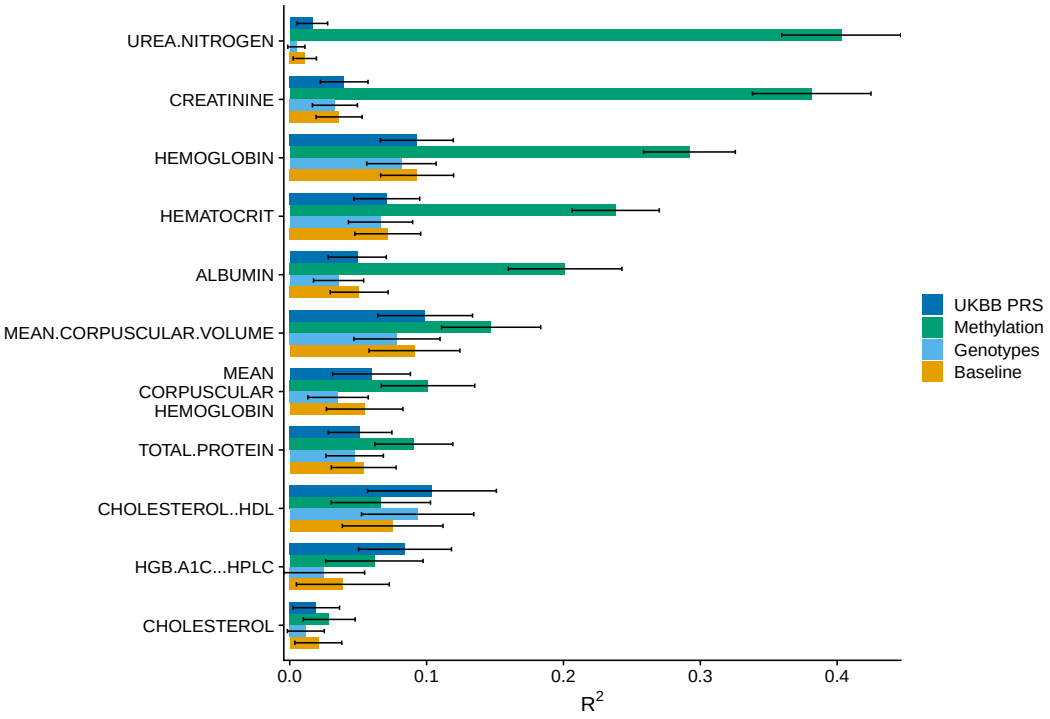

Figure 3.4: Labs as predicted by methylation, genotypes, and an externally-trained polygenic risk score. The cross-validated R^2 between the true and imputed lab value on 541 unrelated patients of non-Hispanic-Latino white-identifying individuals using a baseline predictor as well as a baseline predictor with methylation, genotypes, and a PRS externally-trained from UKBiobank summary statistics.

Primarily, after training the models on the entire heterogeneous set of samples, we examined the predictive performance within each ancestral population. When we examined the top 10 best-predicted (across the entire set of individuals) lab panels, medications, and Phecodes, only 7 of the entire 180 possible comparisons ($\binom{4}{2}$ comparisons across 30 outcomes) displayed significant differences between the predictive performance within each population separately (Figures 3.5, A.3, A.4).

In a second replication analysis we trained predictive models within ancestral groupings separately. As the individuals self-identified as either Black or Asian comprised less than 100 individuals in both groupings, we focused our analyses on Hispanic/Latino- and white-non-Hispanic/Latino-identifying individuals. We retrained models for the top 5 best-

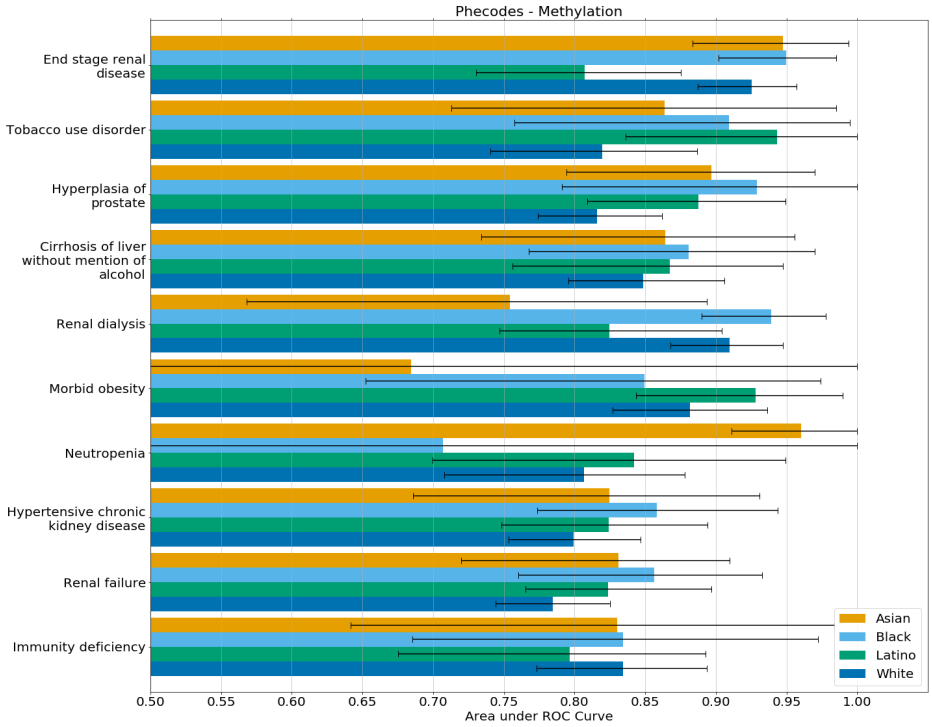

Figure 3.5: Best methylation-predicted Phecodes within ancestral populations. After training a model on the entire heterogeneous population of individuals, we evaluated the predictive performance within each population separately. We observed only 4 (of 60) significant differences between self-reported ancestral groupings.

predicted medications, lab panels, and unique Phecodes on the Hispanic/Latino individuals and white non-Hispanic/Latino individuals alone and treated a prediction as significant if its association p-value was lower than .01. Creatinine, hematocrit, mean corpuscular hermmoglobin, and urea nitrogen replicated across both groupings, however, hemoglobin did not replicate in the Latino/Hispanic grouping (Table 3.1). In the context of medications, CMV agents, osmotic diuretics, phosphate binder agents, hematopoietic growth factors, and immunosuppressive agents replicated within the white non-Hispanic/Latino population but only CMV and phosphate binder agents replicated within the Hispanic/Latino population (Table 3.1). Finally, Phecodes corresponding to immunity deficiency, and end-stage renal failure replicated within both groupings, however, obesity, anemia, and hypertensive renal disease replicated only within the white non-Hispanic/Latino set of individuals (Table 3.1).

Table 3.1: Replication statistics within ethnic groupings. Predictive accuracy (R^2 and AUC) for MRS trained within only Latino/Hispanic- or white-non-Latino/Hispanic-identifying individuals

Outcome	**Metric**	**Accuracy, pvalue Hispanic/Latino**	**Accuracy, pvalue white, non-Hispanic/Latino**
Creatinine	R^2	.138, 7.23e-05	.366, 6.97e-48
Hematocrit	R^2	.070, 5.49e-03	.255, 2.63e-29
Hemoglobin	R^2	.043, 3.10e-02	.267, 7.42e-31
Mean corpuscular hemoglobin	R^2	.096, 1.14e-03	.132, 1.04e-14
Urea nitrogen	R^2	.223, 2.64e-07	.444, 3.06e-59
CMV Agents	AUC	.872, 3.35e-05	.904, 2.93e-19
Osmotic Diuretics	AUC	.715, .0387	.841, 1.51e-16
Phosphate binder agents	AUC	.644, 7.14e-03	.841, 3.25e-22
Hematopoietic growth factors	AUC	.645, .0902	.795, 6.09e-19
Immunosuppressive agents	AUC	.601, .0863	.840, 7.84e-24
Obesity	AUC	.626, 5.03e-02	.813, 1.01e-21
Immunity deficiency	AUC	.779, 1.20e-05	.814, 2.97e-18
Anemia	AUC	.552, .295	.747, 1.64e-12
Hypertensive renal disease	AUC	.585, .378	.767, 6.27e-18
End-stage renal failure	AUC	.700, 2.30e-04	.867, 6.45e-31

Replication of methylation risk scores across external datasets To evaluate the transferability of the MRS to a different population, we performed an experiment in which we imputed the MRS into an external dataset. The dataset used the HumanMethylation27k array to measure the methylation of 194 individuals who had Type 1 Diabetes, 49.7% of whom had nephropathy (cases). We re-trained the models for Phecodes corresponding to end-stage renal disease and chronic renal disease as well as labs corresponding to creatinine and urea nitrogen on our data, limiting our analysis to the 27,000 sites that belonged to the external dataset. The imputed chronic renal disease was significantly associated with nephropathy in the external dataset (p=7.25e-05, AUC=.676 [.602,.748] as was end-stage renal disease (p=.0262, AUC=.629 [.554,.705]). Further, both of the imputed values for

creatinine and urea nitrogen were significantly associated with nephropathy (p=6.36e-07, AUC=.731 [.661,.802] and p=1.31e-06 AUC=.723 [.652,.794], respectively). Importantly, when limiting our internal analysis to sites only on the 27k array, the association signal decreased (for end-stage renal disease, from p=1.02e-74 to p=2.78e-53, chronic renal disease from p=4.52e-53 to p=4.88e-35, creatinine p=8.22e-103 to p=6.28e-65, and urea nitrogen p=9.06e-104 to p=4.94e-49). However, likely due to correlation between CpGs, the association tests for outcomes trained on the smaller set of sites were still significant.

3.3 Methods

Electronic Health Record Data De-identified electronic health record data for this study was extracted from the perioperative data warehouse (PDW), a custom-built, robust data warehouse containing all patients who have undergone surgery at UCLA Health since the implementation of UCLA's EMR (EPIC Systems, Madison, WI, USA) in March 2013. The PDW, which has been described previously [HGP16], has a two-stage design. First, data are extracted from EPIC's Clarity database into 29 tables organised around three distinct concepts: patients, surgical procedures, and health system encounters. Then, these data are used to populate a series of 4000 distinct measures and metrics such as procedure duration, admission ICD codes, lab results, and medication orders.

Patient Ascertainment Methylation and genotype samples were collected using blood from 826 patients as part of the UCLA ATLAS precision health initiative between October 26, 2016 and December 10, 2018. The samples were collected from patients before undergoing surgery with general anesthesia at UCLA Health, and the patients had not undergone surgery in the 30 days prior to blood sample collection. Of these patients, 302 were selected for inclusion based on the presence of acute kidney injury (AKI), defined as an Acute Kidney Injury Network (AKIN) classification of one or greater, after undergoing surgery. The

remaining 558 patients were risk-matched controls, with either glomerular filtration rate (GFR) less than or equal to 38 (210 patients), or GFR greater than 38 and an AKI risk score that matched the AKI cases (348 patients). Demographics of the patient population are further described in Table A.1.

Medication Usage For each medication, a patient was labeled as using a medication if the electronic health record contained a medication order that occurred before the methylation sample collection date. Medications were grouped by pharmaceutical subclass using the Generic Product Identifier (GPI) hierarchical classification system codes. Any medications that were ordered in fewer than 5% of the patients were excluded from the analysis. In total, 171 pharmaceutical subclasses were considered in our analysis. The number of patients using medications from each subclass is shown in Supplemental Table A.5. In Supplemental Table A.6, we show for each pharmaceutical subclass the specific medication that patients in our cohort received.

Lab Results The most recent lab result prior to the methylation sample collection was extracted from the PDW for each patient. Any labs with a result date that occurred more than 365 days before the methylation sample collection date were excluded from the analyses. Additionally, labs for which there were less than 50 patients with valid results were excluded.

Diagnosis Codes International Classification of Diseases, Ninth Revision (ICD-9) and International Classification of Diseases, Tenth Revision (ICD-10) codes are a standard set of diagnosis codes, primarily used for billing purposes. While these codes provide a standardized methodology for describing a diagnosis, they are very specific. To map these specific diagnosis codes into meaningful, distinct diseases/traits, Denny et al. aggregated the ICD codes into phenotype codes (Phecodes) [DRB10, DBR13]. Specifically, for each patient, we queried all diagnoses prior to the methylation sample collection date, and used the Phecode (version 1.2) mapping to aggregate ICD-9 and ICD-10 codes to unique, meaningful phenotypes. If

a patient's diagnosis record had both ICD-9 and ICD-10 labels, the ICD-10 to Phecode mapping was used instead of the ICD-9 to Phecode mapping. Each Phecode was treated as a binary variable, indicating the presence or absence of a relevant diagnosis code at any point in time before sample collection. We excluded rare Phecodes (occurrence less than 5% of the patients) and, in total, our cohort contained 207 unique Phecode phenotypes.

Preprocessing of genotype data We measured the genotypes for 826 individuals based on their DNA sampled from whole blood using the ATLAS genotype array. We preprocessed the genotype data using Beagle (d20) [BB07], PLINK (1.07) [PNT07], and GCTA (1.93.2) [YLG11]. We restricted the genotypes to autosomal variants and removed rare variants (MAF $< .05$). Using Beagle, we imputed any missing values, but did not impute to an external dataset. In total we were left with 301,790 SNPs.

Preprocessing of methylation array data We measured methylation data for 826 individuals based on their DNA sampled from whole blood using the EPIC Illumina array. To generate beta-normalized methylation levels at each CpG, we ran the default pipeline of ENmix (1.22.0) [XNL16] on the the raw probe data (IDAT files), which performs background correction, RELIC dye bias correction, and RCP probe-type bias adjustment. We removed from our analysis CpGs that coincided with SNP loci as well as CpGs on the sex chromosome. We also filtered out outlier samples, defined as having a PC score more than 4 standard deviations away from the average PC score in the first two principal components. In the prediction tasks, we removed sites with low variability (standard deviation < 0.02) leading to a total of 554,761 sites.

Prediction using baseline medical features To establish a baseline level of prediction performance, we constructed a set of features derived from basic patient information. We trained a simple linear (or logistic) model with 10-fold cross validation using patients' age, sex, BMI, methylation-based cell-type proportions (from the reference-based method of

Houseman et al. [HAK12]), self-reported ancestry, and first ten genetic principal components. Importantly, we wished to establish how well an outcome (medication, Phecode, or lab value) could be captured by using information that would be available to the clinician.

Prediction using a single penalized linear model After establishing a baseline level of prediction performance, we performed penalized logistic and linear regression using either individuals' methylation, genotypes, or both. More concretely, we fit 10-fold cross-validation using LASSO, elastic net and ridge regularization under the following two models:

$$y = G\beta_G + C\beta_C + \epsilon \tag{3.2}$$

$$y = M\beta_M + C\beta_C + \epsilon \tag{3.3}$$

where y corresponds to the outcome, G the $n \times s$ genotypes, M the $n \times c$ methylation data, β the vector of length-s or -c effect sizes for the given explanatory variable, C and β_C the covariates from the baseline model and their corresponding effect sizes, and ϵ the length n noise vector. After fitting all three penalized linear models for a given datatype and outcome, we selected a final model as determined by the model with the highest cross-validated metric (AUC or R^2 if the outcome was binary or continuous, respectively). We share MRS weights for outcomes that were significantly predicted at `https://github.com/cozygene/EHR_MRS_UCLA`.

Imputing lab results using EHR data and MRS values Imputing a partially-observed matrix of values is often formulated as a matrix-completion problem. In a matrix completion problem, the observed values of the matrix are used to estimate the values of the unobserved values by assuming that there is some underlying structure that is responsible for generating the data. For example, in the popular SoftImpute method [MHT10], the data is assumed to be well-approximated by a low-rank representation, and the error between the observed values and the reconstructed values is minimized through a convex optimization procedure. However, since the unobserved values are, by definition, not observed, and therefore cannot

be used to assess the imputation performance, the primary method for measuring the performance involves masking (removing) observed values and comparing the imputed values to these held-out, true values.

The EHR data used in the imputation procedure included demographic information, diagnosis codes, medication usage, and lab results, which were extracted from the EHR database using the previously described criteria. In addition to the EHR data, we also ran the imputation procedure while including relevant MRS values. Specifically, we included the MRS values for demographics, diagnosis codes, medication usage, and lab results that were predicted at a statistically significant level. These MRS values were added as additional observed features to the EHR matrix.

To estimate the imputation performance, we randomly masked 10% of the observed lab result values, and performed the imputation procedure (SoftImpute matrix completion) to generate estimates of the missing values. However, since labs are most often ordered in panels, for example a metabolic panel, if a lab is missing then typically other labs that are part of the same panel are also missing. We simulated a more realistic missingness scenario by, instead of masking out values only from a specific lab l, masking out all labs that are ordered as a panel that include lab l. This masking procedure was done per lab, using 10-fold cross-validation, such that 10% of the non-missing values of a particular lab result (and its associated lab panels) were masked (removed), and the remaining 90% of the observed values were used to complete the matrix. Matrix completion was performed using the SoftImpute algorithm, as implemented in the *fancyimpute* [RFO17] python package (version 0.5.5). The proportion of variance explained (R^2) of the true lab values by the imputed lab values was used to measure the imputation performance. Confidence intervals were derived using bootstrapping.

Ethical approval and patient consent Retrospective data collection and analysis was approved by the UCLA IRB. All research was conducted in accordance with the tenets set

forth in the Declaration of Helsinki.

Data availability The MRS weights for outcomes that were significantly predicted are located at: `https://github.com/cozygene/EHR_MRS_UCLA`. The raw UCLA datasets generated during and/or analyzed during the current study are not publicly available due to institutional restrictions on data sharing and privacy concerns.

3.4 Discussion

In this study, we provide a comprehensive investigation of the utility of methylation risk scores in a clinical setting. We used (to our knowledge) the largest methylation biobank cohort produced to date, which includes methylation, genotype, and comprehensive EHR data for all patients. We find that the MRS improved prediction performance over a baseline model by 10.88%, 189.41% and 11.75% when predicting medication usage, lab panel values, and diagnosis codes respectively. These contributions are significantly more substantial than those obtained by PRS.

The vision of genomic biobanks is that the genomic data will be translated into improved clinical diagnosis and treatment decisions [LV20, Bel17, LH19]. In practice, clinical decisions are not expected to be based solely on genomic information, but rather on the combination of the genomic, medical, and demographic information of the patient. While previous studies have used a limited number of key features as a baseline for imputation of a phenotype (e.g., age, sex, and major comorbidities) [BLS18, JLY20, MLK16, ZXX17], to the best of our knowledge, these studies did not take into account the entire familial-genetic or environmental history of the patients. Thus, the question of whether genomic data (methylation or genetics) can be used to improve imputation over the EHR data is critical in order to claim clinical relevance. Our results demonstrate that adding MRS to existing EHR-based imputation frameworks improve prediction accuracy by over 29% in a clinical context.

It is well appreciated that PRS are sensitive to the studied population, and it is often the case that a PRS developed for one ethnic group performs poorly on others [MKK19, DSG19]. It is therefore important to evaluate the population effect on MRS performance. For this reason, we measured the transferability of our results across different populations, and we observe that the accuracy of the MRS was robust to population structure. This is likely driven by the diversity of the training cohort used.

While our study was focused on methylation, there are many other possibilities for the introduction of genomic data in clinical settings. First and foremost, genetic data has been heavily studied by others and large biobanks including genetic data of patients already exists. However, other measurements such as RNA, microbiome, metabolomics, or proteomics may also be relevant. Some of these have logistic and cost considerations at scale. One of the advantages of methylation is that DNA biobanks already exist in large numbers, and the cost of measuring methylation is close to that of measuring genetic data. Moreover, different genomic measurements provide different snapshots of the patient's data or health status. For example, while polygenic risk scores provide a lifetime risk for a patient, methylation risk scores may provide the current risk of the patient over the last few months [RZB16, Hor13b, FHH21], and other genomic information may provide risk with the resolution of days or hours (e.g., RNA or certain metabolomics [LGH15, CGE18, CMT16, AS15]). Nonetheless, owing to the dynamic nature of methylation, it is currently unclear what the range or duration of the methylation risk score are. Furthermore, while methylation patterns are associated with outcomes, it is generally unknown if they cause a disease or are a response to a disease [RD12].

To assist the research community in investigating methylation in the context of disease, we provide the MRS predictors for all significantly predicted outcomes at `https://github.com/cozygene/EHR`. The database may be used by researchers and clinicians in different ways. For example, in many epigenome-wide association studies (EWAS), in which associations between specific methylation CpG sites and a phenotype are studied, one may wish to account for patients'

comorbidities and medications, which are often not available to the study. Using the MRS database, the researchers leveraging EWAS will be able to incorporate such covariates into their model.

There are multiple potential next steps for the examination of methylation in clinical contexts. First, in this work we focused our attention on the imputation of the phenotypes, or in other words, the inference as to whether the patient is currently diagnosed with a disease. We hope that our findings will be able to be translated to the inference of future clinical events, i.e., prediction of future deterioration or disease occurrence. Second, although our evaluation is across the largest dataset which includes both EHR, methylation, and genotype data, the sample size of our study is still moderate compared to genetic studies that are performed on biobanks. Indeed, we demonstrate that for some of the phenotypes, an increase in sample size will likely lead to a substantially improved prediction accuracy (Figure 3.3). Moreover, larger sample size data may be able to reveal the quantity or contribution of genetics verses methylation to the MRS prediction accuracy [TMP20]. We therefore recommend that future biobanks consider measuring methylation in addition to the genotypes across a large number of patients.

CHAPTER 4

Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning

4.1 Introduction

Each year in the United States, 5.7 million patients are admitted to an intensive care unit (ICU) and nearly 50 million patients undergo surgery. Hypotension and hypertension in the ICU and perioperative period are associated with adverse patient outcomes including stroke [BPP12], myocardial infarction [WKW16], acute kidney injury [WDG13], and death [LST13]. Recent studies even suggest that only a few minutes of hypotension in the acute care setting increases the incidence of these complications [MNM18]. These observations strongly suggest that continuous blood pressure monitoring is critical in the acute care setting to identify periods of hypertension and/or hypotension as early as possible. Today, the gold standard for blood pressure monitoring is the invasive arterial line, a small catheter inserted into an artery, which enables continuous blood pressure monitoring [LST13]. However, this technique is highly invasive and is associated with significant complications such as bleeding, hematoma, pseudoaneurysm, infection, nerve damage, and distal limb ischemia [BLL09, KLS14], and thus it is only applied to very high-risk patients.

On the other hand, the widely used non-invasive blood pressure monitoring system using cuff-based devices is both inaccurate and intermittent, only allowing for the monitoring

of blood pressure every three or five minutes [BHV03]. More recently, devices allowing for continuous and non-invasive blood pressure monitoring have been introduced. These devices, however, are sensitive to patient movement, they are expensive, and they cause continuous pressure on the finger that can interfere with blood circulation [YVG18]. Additionally, the accuracy of the device can deteriorate in patients with severe vasoconstriction, peripheral vascular disease, or distorted fingers due to arthritis [MS18].

The need for measuring the continuous blood pressure non-invasively suggests that the physiological waveform of the blood pressure should be imputed from other data. However, previous attempts at blood pressure imputation have primarily focused on the imputation of discrete systolic and diastolic blood pressure measurements taken intermittently by cuff [TZ03, XMZ19, XS16, ZRH19] or at the resolution of a heartbeat [SDZ17, KLG13, DYZ17], while high-risk patients need to be monitored using the continuous arterial line blood pressure waveform, and thus current methods are not adequate for usage in critical care settings. Moreover, many of the previously-studied cohorts consisted of healthy patients at rest, and therefore the blood pressure variability is not as dynamic compared to patients in the ICU. This calls into question the utility of these approaches in real life settings. Finally, many of the existing methods developed and tested their models in the same set of patients by training the model on an earlier part of each patient record and testing the model on the remaining data. Additionally, the number of patients used was often only several dozen, all coming from the same health system. This calls into question the generalizability of the models to unseen patients. For example, Sideris et al. impute the arterial blood pressure waveform by training the model on the same patient for which arterial blood pressure waveforms are provided [SKN16]. However, such a scenario is not clinically useful, since applying such an approach will require invasive monitoring of the patients for at least part of the time.

In this paper, we present the development, training, and validation of a novel non-invasive and continuous deep learning method for predicting the arterial blood pressure waveform using the ECG waveform, the pulse oximeter (PPG) waveform, and non-invasive blood pressure

cuff measurements. These measurements are collected as part of the current standard-of-care, and therefore no additional patient monitoring devices are needed. Our method leverages a well-known deep learning model architecture originally designed for image segmentation (V-Net [MNA16]), and we adapted it for 1D physiological waveform signals. A key aspect of our preprocessing pipeline includes a manual labeling of PPG quality in a large subset of the training data to improve the signal-to-noise ratio and remove artifacts. The manual labeling was used to train a deep neural network to predict the signal quality of the waveform, resulting in a high-quality preprocessing pipeline that can be used beyond the scope of this study. We demonstrate that the modified 1D V-Net approach provides a highly accurate prediction of continuous arterial blood pressure waveform (root mean square error 5.823 (95% CI 5.806-5.840) mmHg), as well as the derived systolic (mean difference 2.398 $\pm$ 5.623 mmHg) and diastolic blood pressure (mean difference -2.497 $\pm$ 3.785 mmHg).

As opposed to previous studies, here we show that the modified 1D V-Net approach successfully generalizes to new patients. Particularly, we validate the approach on non-healthy populations from ICUs in two different health systems, and our training and validation cohorts include different sets of patients.

4.2 Methods

This manuscript follows the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View [LPT16]

4.2.1 Study participants and sampling procedures.

The first retrospective, de-identified dataset consisted of 309 randomly sampled ICU patients with ECG (lead II) waveforms, photo-plethysmographic (PPG) waveforms, arterial blood pressure (ABP) waveforms, and at least one non-invasive blood pressure measurement from the MIMIC-III waveform database [JPS16, GAG00] matched subset. The 309 patients were

randomly divided into a training set (206 patients, 66% of the cohort) and a testing set (103 patients, 33% of cohort). Of these patients, 31 patients from the training set and 14 from the testing set were removed because none of the data were found to pass the quality filtering process (see Dataset Creation), leaving 175 patients in the training set and 89 in the testing set. The second retrospective dataset consisted of 150 randomly sampled UCLA ICU patients with recorded ECG, PPG, and ABP waveforms. The UCLA dataset was divided into 40 patients for calibrating the model, and 110 patients for testing the model. Of these patients, 12 were removed from the calibration set and 23 were removed from the testing set due to issues with waveform data quality, for a final training set of 28 patients and testing set of 87 patients.

The ABP waveform prediction model was first trained using patients from the MIMIC training set, and model performance results were computed using patients from the held-out MIMIC testing set. Then, the model parameters were fine-tuned using patients from the UCLA calibration set, and model performance results were computed using patients from the held-out UCLA testing set.

4.2.2 Dataset Creation.

4.2.2.1 Demographic data

Cohort demographic data were extracted for the MIMIC patients using the MIMIC clinical database. The features extracted included patient age, height, weight, BMI and sex. For de-identification, patients older than 90 have their age encoded as 300 in the clinical database. Since we do not know the exact age of these patients, we set their age to 90 years old. The same demographic information for the UCLA patients was retrieved from the UCLA Clinical Data Mart, a data warehouse system that extracts data from UCLA's electronic medical record system (EPIC Systems, Madison, WI, USA). As was done for the MIMIC data, any UCLA patients older than 90 years had their age set to a maximum of 90 years

for de-identification purposes.

4.2.2.2 Normalization of waveforms

In both cohorts, if the signal sampling rate was greater than 100 Hz, each waveform signal was downsampled to 100 Hz. Each signal was low-pass filtered with a cutoff frequency of 16 Hz to remove high-frequency noise. Since the range of the ECG and PPG signals differed for each patient, we scaled each 32-second window by subtracting the running median and dividing by the difference between the upper and lower quartiles.

4.2.2.3 Derived non-invasive blood pressure

Intermittent non-invasive blood pressure (NIBP) measurements were extracted for each patient from the MIMIC-III database. However, since NIBP measurements were only recorded, on average, once per hour, the frequency of non-invasive blood pressure measurement was insufficient. Therefore, we created derived NIBP measurements by sampling the invasive blood pressure waveform (i.e. median systolic, diastolic BP, and mean ABP in a 4-second window) every five minutes to simulate the frequency of NIBP measurement that would be used when deploying the algorithm in practice. Since the derived non-invasive blood pressure was measured every five minutes, but the waveforms were sampled at 100 Hz, we used the most recent derived NIBP measurement to fill in missing NIBP values. As an additional feature, we also included the time (in milliseconds) from the most recent NIBP measurement to each sample in our input window.

4.2.2.4 Correction of signal drift

As mentioned by the authors of the MIMIC dataset [CSV], issues with clock synchronization can cause waveform signals to drift. We corrected for signal drift between the PPG signal and the arterial blood pressure signal by computing the cross-correlation of the PPG signal

Table 4.1: Window filtering statistics for each cohort

	MIMIC	UCLA
Total minutes, No., mins	1,535,413	240,241
Valid minutes, No., mins	115,388	35,601
Total heartbeats, No.	9,791,870	2,935,846
Total windows per patient, median (IQR)	8376. (4509.3-16656.3)	4087.0 (2627.5-5414.5)
Valid windows per patient, median (IQR)	411.5 (90.8-1076.0)	254.0 (67-743)
Total record length per patient, median (IQR), mins	4467.2 (2404.9-8883.3)	2179.7 (1401.3-2887.7)
Valid record length per patient, median (IQR), mins	219.5 (48.4-573.9)	135 (35.7-396.5)
Median (IQR), %	4.8 (1.1-12.9)	8.5 (2.2-24.9)
Mean (SD), %	9.6 (12.4)	15.5 (17.3)
Min/Max %	0.01/63.8	0.02/71.5

under consideration with the arterial blood pressure signal. Once the cross-correlation was computed, the location of the highest cross-correlation was used to correct the PPG signal drift by shifting the signal in time, up to a maximum of 4 seconds in either direction.

4.2.2.5 Creation of valid windows

After completing the above preprocessing steps, we selected valid 32-second windows from the record using a sliding window approach with a 16-second step size. A window size of 32 seconds was chosen since it is long enough to give temporal context for several heartbeats, yet short enough to ensure that there would be a sufficient number of windows that did not contain artifacts. See Table 4.1 for filtering rate across each cohort. For each 32-second window, the following process was used to determine whether the window will be included in the development or validation of the algorithm.

4.2.2.6 Filtering of windows with artifacts

Each waveform (ECG, PPG, ABP) was checked for signal quality and windows with technical artifacts or invalid parameters were removed. Windows were excluded if they met any of the following criteria for the ECG or PPG data: the variance of the signal was less than a small value (1e-4 for ECG and PPG), the number of peaks in a window was greater than a threshold (4 peaks/sec x 32 seconds for both ECG, PPG; 4 peaks/sec results in maximum allowable HR of 240 beats/min), or the number of peaks in a window was less than a threshold (0.5 peak/sec x 32 seconds for ECG, PPG; 0.5 peak/sec results in minimum allowable HR of 30 beats/min).

For the arterial blood pressure waveform, windows were excluded if they met any of the following criteria: the mean signal value was less than 30 mmHg or greater than 200 mmHg, the maximum signal value was greater than 300 mmHg or less than 60 mmHg, the minimum signal value was less than 20 mmHg, the variance of the signal was less than 80, we could not find any systolic or diastolic blood pressure values using the $find\_peaks$ function from the scipy [VGO20] Python package, the difference between two consecutive systolic or diastolic values was greater than 50 mmHg, the waveform signal was flat (i.e. did not change value for 2 or more consecutive samples), a pulse pressure value in the window was greater than 70 mmHg, the difference between the systolic BP and the most recent NIBP was greater than 40 mmHg, or the time delay between a diastolic blood pressure measurement and the subsequent systolic blood pressure measurement was greater than 0.5 seconds.

If a window contained signals that passed all of the above criteria, we performed two additional filtering steps, and excluded windows that failed either of these criteria: the number of PPG peaks was different than the number of arterial blood pressure peaks, or the mean absolute time difference between arterial blood pressure peaks and PPG peaks was greater than 0.15 seconds (after correcting for signal drift). Finally, outlier windows were excluded if the mean ECG, PPG, or ABP signal in the window was greater than the 99.9%

quantile or less than the 0.01% quantile.

Assuming a window under consideration passed the above criteria, it was included in our training or testing dataset. The input features were then scaled to have a mean of zero and standard deviation of one using a running mean and standard deviation.

4.2.2.7 Filtering with PPG Quality Index

Since the quality of the PPG waveform is crucial to the performance of our algorithm, we developed an additional filtering step to remove windows containing artifacts in the PPG signal. A CNN model was trained to classify four-second PPG windows as valid or invalid using the PPG waveform as input. To train the model, 4000 four-second PPG windows from the 206 patients in the MIMIC training set were hand-labeled (by B.L.H.) as valid (2682, 67.1%) if the window was free of artifacts, or invalid (1318, 32.9%) if the window contained artifacts, using visual inspection. From these 4000 windows, we randomly sampled a subset of 100 windows and an expert clinician (M.C.) then labeled the windows to estimate the initial classification quality. In 94% of the sampled windows, the clinicians classification matched the initial labeling (Cohens kappa: 0.857). We then trained a 3 layer CNN model (see Supplemental Note 2 in [HRR21] for additional details) on 70% of the patients to predict the window classification. The models predicted probability of being a valid window was then used as a quality index (QI) to exclude windows highly likely to contain artifacts. The remaining 30% of patients (separate from the training patients) from the MIMIC training set were held out for model validation. The quality index threshold for filtering was chosen as the minimum threshold (value: 0.811) achieving a positive predictive value (precision) of at least 0.95 in the validation set. This method of setting the threshold was chosen to reduce the number of windows containing artifacts (false positives) when training and testing the ABP waveform prediction models, while minimizing the threshold to be as sensitive as possible. Since the ABP waveform prediction model uses thirty-two second windows as input, yet the PPG QI model uses four-second windows, the PPG QI model was applied to the eight non-

overlapping four-second windows contained within a thirty-two second window, generating a total of eight PPG QI values per window. The minimum PPG QI value in a thirty-two second ABP prediction window was used to determine if the signal quality was greater than or less than the PPG QI filtering threshold.

4.2.2.8 Waveform features

To improve the ABP waveform imputation we included two features derived from the non-invasive signals that have been previously shown to be predictive of blood pressure values: pulse arrival time (PAT) and heart rate (HR) [XMZ19, SDZ17, KLG13]. Calculation of the pulse arrival time was achieved by first identifying the ECG R wave using the peak finding algorithm implemented in the scipy [VGO20] Python package, and then identifying the PPG systolic peaks using the peak finding algorithm. The number of seconds between the ECG R wave peak and the subsequent PPG systolic peak were then calculated. PAT values were excluded from a window if the value was deemed to be unreasonably small (¡0.1s) or unreasonably large (¿1.0s). Finally, for each window, the median (log-transformed) and standard deviation of the PAT were used as two additional input feature channels. If PAT values were excluded or unavailable, the missing values were imputed using the median observed value.

Additionally, the HR was calculated for each window using the PPG signal. The scipy peak finding algorithm was similarly used to detect the PPG systolic peaks. Then, the number of these peaks found in a given window was divided by the window length (in seconds) and multiplied by the number of seconds in a minute to give the resulting HR in beats per minute. The HR was also included as an additional input feature channel for each window. Any missing HR values were imputed using the median observed value.

4.2.3 Comparison with other methods

The performance of two other methods were used as a comparison: PPG scaling and the LSTM model of Sideris et al. [SKN16]

4.2.3.1 PPG Scaling

Previous work has shown the utility of the PPG signal for predicting blood pressure measurements [TZ03, XMZ19, XS16, ZRH19, SDZ17, KLG13, DYZ17, SKN16]. These approaches use manual feature extraction [TZ03, XMZ19, ZRH19, DYZ17] or learn features from the data using machine learning methods [XS16, SDZ17, KLG13, SKN16]. Since the PPG waveform is correlated with the ABP waveform, one approach for predicting the ABP waveform uses the PPG waveform as a template shape and scales the magnitude of the PPG signal in a given window to match the most recent systolic and diastolic pseudo-NIBP measurements. Specifically, the PPG waveform is stretched such that the maximum value in the window is equal to the most recent systolic NIBP measurement, and the minimum value in the window is equal to the most recent diastolic NIBP measurement. The PPG window was scaled using the transformation

$$PPG_{scaled} = (PPG - min(PPG))\frac{NIBP_{sys} - NIBP_{dias}}{max(PPG) - min(PPG)} + NIBP_{dias} \tag{4.1}$$

4.2.3.2 LSTM Model

Sideris et al. [SKN16] proposed training a patient-specific LSTM model (i.e. one model per patient) to impute the ABP waveform, using the PPG signal from the same window of time as input. While the results were promising, a critical issue is that the model requires that each patient first receive invasive blood pressure monitoring so a patient-specific model can

be trained. However, this means that only a fraction of the patient population can benefit from such a model (the subpopulation that receives invasive blood pressure monitoring), and only after they have already undergone invasive monitoring. Therefore, the algorithm may not generalize to the majority of the patient population who do not receive invasive monitoring. To fairly compare the LSTM model to our proposed model, we trained the model as described in the paper, using 128 nodes as a default since the number of nodes was not described.

4.2.4 Algorithm development.

We developed a deep learning model that takes as input a window of two signals, ECG and PPG, and several constant values encoded as additional channels by repeating the value for each timestep: the most recent non-invasive systolic, diastolic, and mean blood pressure measurements prior to the window, the time since the most recent NIBP measurement, the median and standard deviation of the pulse arrival time, and heart rate. The model is trained to minimize the residual difference between the PPG scaling method and the true ABP waveform to compensate for the difference in waveform morphology, as well as the change in BP over time. This forces the network to focus on windows where the PPG scaling significantly differs from the ABP waveform, and therefore improve on the PPG scaling method. With enough data and a large enough model, the neural network should be able to similarly learn the scaling method. However, to accelerate the learning process we designed the method to learn the residual error. The model output is a prediction of the residual difference between the continuous ABP waveform and the baseline PPG scaling waveform, and this predicted residual difference is added to the PPG scaling waveform to generate the 1D V-Net waveform prediction. The deep learning model architecture was based on the V-Net CNN architecture, which has been proven to be useful in the field of image segmentation [MNA16] (see Supplemental Note 1 in [HRR21] for description). However, instead of 2D or 3D image segmentation, we leveraged the V-Net architecture for

1D signal-to-signal transformation. The motivation behind the V-Net architecture is that it learns a compressed representation of the input data to identify global features, and then reconstructs the signal from this representation. During the reconstruction process, local features are learned to modify the waveform at a finer scale. Our architecture is the same as described in the V-Net paper [MNA16], except instead of 3D volumes with multiple channels our data is represented as a 1D signal with multiple channels. Otherwise, the architecture (number of layers, convolutions per layer, kernel size, etc.) remained the same. An additional L2 penalty was added to the activation of the final network layer to force the network to prioritize modification of the PPG scaling residual waveform.

To train the network, we used a custom loss function consisting of two parts. The first was the mean squared error between the true ABP waveform and the predicted waveform, which forces the network to learn an accurate prediction of the entire waveform. The second part of the loss was the mean squared error between the true and predicted waveforms at the locations of the systolic and diastolic points, to encourage the network to be particularly accurate at these locations.

The deep-learning model was implemented using Keras [Co15] and was trained for a maximum of 100 epochs using random weight initialization and the Nadam [Doz16] optimizer with default parameters beta1 of 0.9, beta2 of 0.999. The learning rate used was 0.001 with a schedule decay of 0.004, and the mini-batch size was 32. For the last layer of the network, an L2 activation penalty with weight 0.0005 was added. Ten percent of the training data was held out for validation, and these data were used for choosing hyperparameters. If the validation loss did not improve after 8 epochs, the model training process was stopped early.

4.2.5 Algorithm evaluation.

4.2.5.1 Bland-Altman

To evaluate the agreement between the gold standard invasive blood pressure measurements (the arterial catheter) and the DNN predictions, we used the Bland and Altman method [MA86] as this is the standard method for comparing the agreement of two medical devices. Accuracy and precision of the predictions were described as the mean ± standard deviation of the differences between the predicted and true blood pressure values, and the differences are considered acceptable by the Association for the Advancement of Medical Instrumentation (AAMI) criteria if less than 5 ± 8 mmHg. The method was implemented as follows. For each window under consideration, we extracted the systolic and diastolic blood pressure measurements from the ABP waveform using the peak finding algorithm described previously. These measurements were used as the gold standard reference values. We then used the peak finding algorithm on the DNN-generated waveform to determine the systolic and diastolic blood pressure measurements and used these as the comparison values. If the number of systolic or diastolic points identified by the peak finding algorithm differed between the true and predicted waveforms, we performed local alignment by minimizing the sum of the differences between the indices of the true and predicted waveform points. We then took the difference between the reference blood pressure measurement and the predicted blood pressure measurement pairs, and plotted these differences as a function of the average of the reference and predicted value pairs. The 95% confidence intervals (CI) were calculated using bootstrapping.

4.2.5.2 Waveform Metrics

Two metrics were used to quantitatively compare the quality of the predicted waveform and the true ABP waveform: root mean square error (RMSE) and correlation. These values were calculated for all windows per patient. The 95% confidence intervals (CI) were calculated

Table 4.2: Cohort Characteristics of MIMIC and UCLA Data

Characteristic	MIMIC (n = 309)	UCLA (n = 150)
Male, No. (%)	178 (56.9)	80 (53.3)
Age, mean (SD), years	63.4 (16.2)	46.5 (20.1)
BMI, mean (SD), kg/m^2	30.3 (9.3)	24.9 (4.7)
Height, mean (SD), cm	168.7 (10.4)	172.8 (11.1)
Weight, mean (SD), kg	85.0 (25.6)	73.8 (19.0)
Systolic BP, mean (SD), mmHg	106.4 (13.5)	102.6 (11.9)
Diastolic BP, mean (SD), mmHg	57.9 (11.5)	54.1 (9.1)
Mean BP, mean (SD), mmHg	74.8 (12.1)	71.3 (8.9)

using bootstrapping.

4.2.5.3 Error as a function of time since NIBP measurement

RMSE was calculated as a function of time from the most recent NIBP measurement. The time between the window and the most recent NIBP measurements were binned into ten second intervals, and for each bin the mean and SD of the RMSE between the true and predicted waveforms was calculated.

4.3 Results

4.3.1 Description of dataset and features

Two separate cohorts of ICU patients were used in this study to train and validate the method. The first cohort consisted of randomly sampled ICU patients from the Medical Information Mart for Intensive Care version III (MIMIC-III) [JPS16] waveform database who had ECG waveforms, photo-plethysmographic (PPG) waveforms, arterial blood pressure (ABP) waveforms, and at least one non-invasive blood pressure measurement. After exclusion of patients with insufficient data, the remaining 264 patients were then randomly

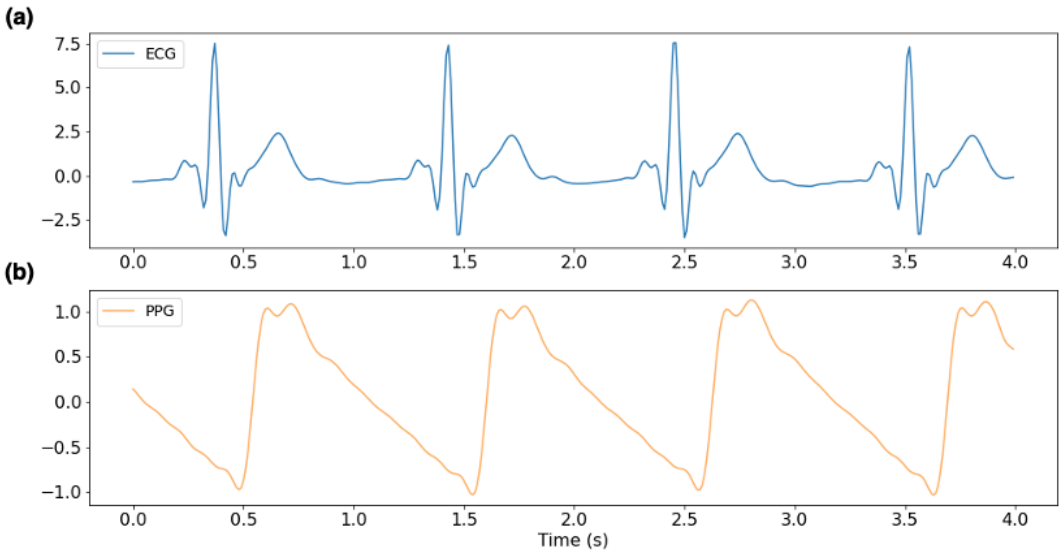

Figure 4.1: Examples of input waveforms for 1D V-Net model (a) 4-second sample of electrocardiogram (ECG) waveform and (b) a 4-second sample photo-plethysmograph (PPG) waveform.

Table 4.3: Root Mean Square Error (mean (95% CI)) for each cohort

Method	MIMIC	UCLA
PPG Scaling	6.895 (6.876-6.914)	9.108 (9.078-9.137)
Sideris et al.	13.940 (13.901-13.978)	13.111 (13.072-13.151)
1D V-Net	5.823 (5.806-5.840)	6.961 (6.937-6.985)

Table 4.4: Correlation (mean (95% CI)) between true and predicted blood pressure for each cohort

Method	MIMIC	UCLA
PPG Scaling	0.938 (0.938-0.938)	0.926 (0.925-0.926)
Sideris et al.	0.939 (0.939-0.939)	0.940 (0.940-0.940)
1D V-Net	0.957 (0.957-0.957)	0.947 (0.947-0.948)

separated into disjoint training and testing sets, with 175 and 89 patients, respectively (see Methods). The MIMIC dataset was used for primary training of the model. The second cohort of patients consisted of 115 ICU patients (after excluding patients with invalid records, see Methods) from the UCLA Health hospital system who had ECG, PPG, and ABP waveform records. These patients were randomly separated into two disjoint groups: 28 patients for secondary fine-tuning (calibration) of the MIMIC-based model, and 87 patients for testing the method. Summary cohort demographic information is shown in Table 4.2.

The ECG and PPG waveforms, the most recent NIBP measurements, the time since the most recent NIBP measurement, the pulse arrival time, and the heart rate were used as input to a deep learning model that was trained to predict the continuous blood pressure waveform that occurred during the timeframe of the input window. An example window of ECG and PPG waveforms used as input to the algorithm is shown in Figure 4.1. The true blood pressure waveform and the predicted waveform that corresponds to the window shown in Figure 4.1 are shown in Figure 4.2a. Figure 4.2b and Figure 4.2c show examples of how the predicted ABP waveform compares to the arterial line for much longer timeframes.

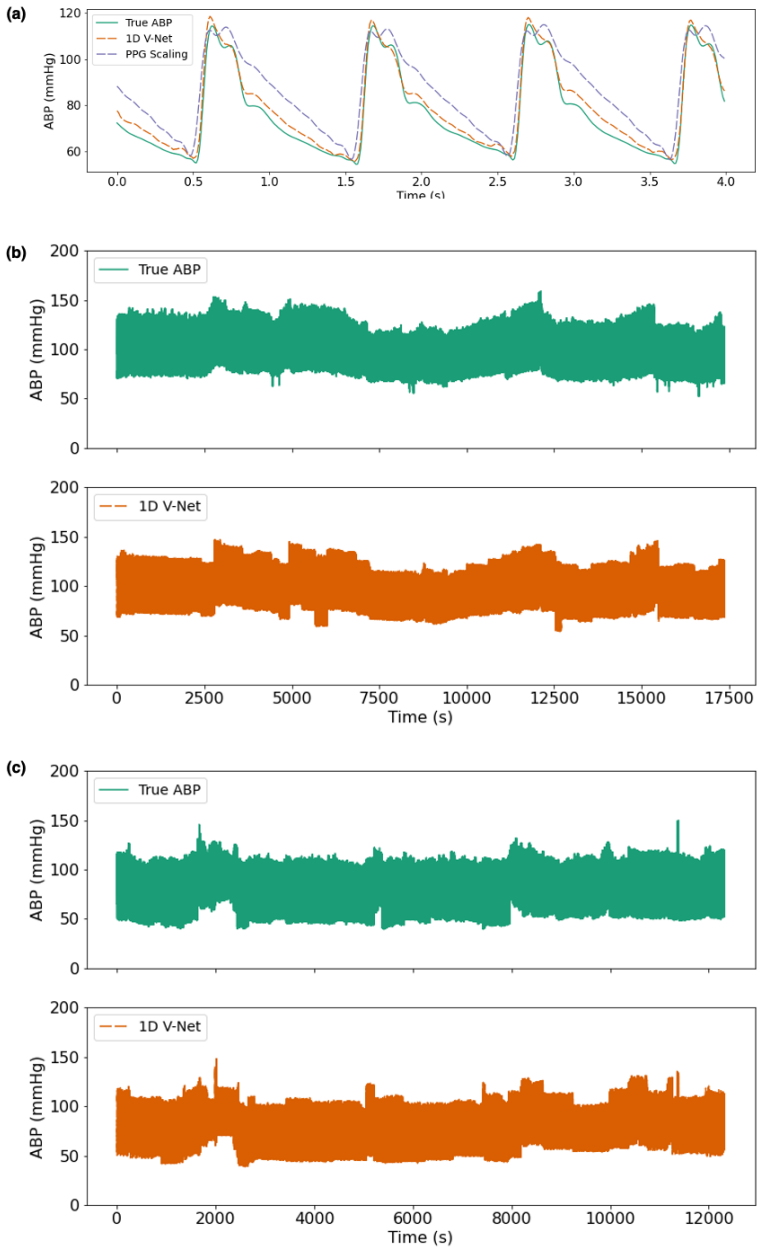

Figure 4.2: Example ground truth & predicted waveforms (a) 4-second window (for the input data shown in Figure 1) and >3 hour records (b) and (c). The true continuous blood pressure waveform is shown above in green, and the predicted blood pressure waveform shown below in red.

4.3.2 Waveform quality evaluation

One of the unique features of our approach is that it provides a prediction of the continuous blood pressure waveform, and not simply summary statistics such as systolic or diastolic blood pressure in a window. To evaluate the quality of the waveform generated by 1D V-Net, we compared the predicted ABP waveform to the ground-truth ABP waveform obtained from the arterial line. Time was split into windows of constant duration (32 seconds). For measuring the method performance, we used the root mean square error (RMSE) and the correlation between the true and the predicted signals within each window. As shown in Table 4.3, for both cohorts, we observed a low RMSE value (MIMIC RMSE 5.823 mmHg, 7.8% of true MAP; UCLA RMSE 6.961 mmHg, 9.8% of true MAP) when comparing the true and predicted waveforms. Additionally, the correlation between the arterial line waveform and the 1D V-Net predicted waveform was high across both sets of patients (MIMIC correlation 0.957, UCLA correlation 0.947) (see Table 4.4).

As an additional waveform quality analysis, we compared the systolic and diastolic values derived from the arterial line and the predicted ABP waveforms. For each heartbeat in a window, we computed the systolic and diastolic blood pressure, and compared those values to the imputed blood pressure waveform generated by our algorithm. In Figure 4.3, Bland-Altman plots were used to show the differences between the predicted and true measurements across a range of blood pressure values (see Methods) for each patient in the MIMIC test set (Figure 4.3a) and UCLA test set (Figure 4.3b). The algorithms predicted waveform accurately tracks the true blood pressure values in both the MIMIC cohort (mean difference systolic BP 4.297 $\pm$ 6.527 mmHg, diastolic BP mean difference -3.114 $\pm$ 4.570 mmHg) and the UCLA cohort (mean difference systolic BP 2.398 $\pm$ 5.623 mmHg, diastolic BP -2.497 $\pm$ 3.785 mmHg). In both cohorts, the 1D V-Net model performance meets the Association for the Advancement of Medical Instrumentation (AAMI) criteria with mean differences less than 5 $\pm$ 8 mmHg.

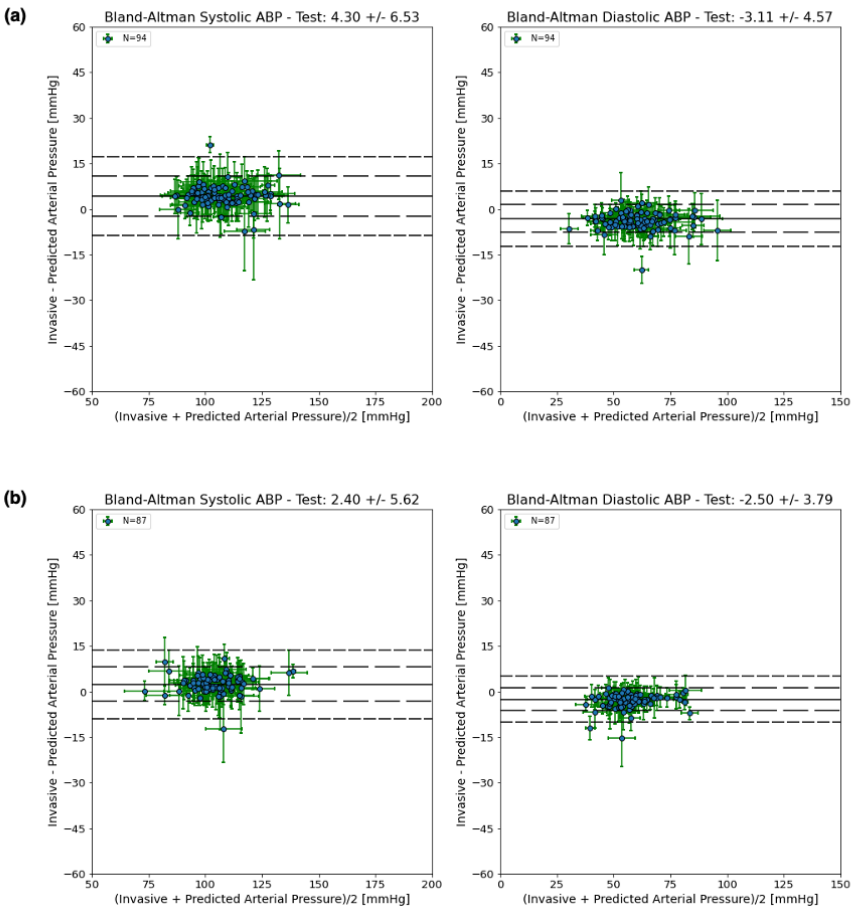

Figure 4.3: Bland-Altman plots for the MIMIC and UCLA ICU test cohorts. Systolic BP measurements per patient (left), and Diastolic BP measurements per patient (right) using a thirty-two second window; horizontal error bars represent the standard deviation of the blood pressure values, vertical error bars represent the standard deviation of the differences; solid lines indicate the mean difference values, dashed lines indicate the mean difference values +/- 1 and 2 times the standard deviation of the differences. Results for MIMIC are shown in (a), and UCLA in (b).

Table 4.5: Bland-Altman accuracy and precision (mean (95% CI) +/- SD (95% CI)) for each cohort

	Method	MIMIC	UCLA
Systolic BP	PPG Scaling	6.133 (6.128-6.139) ± 6.870 (6.864-6.876)	2.668 (2.662-2.674) ± 5.692 (5.684-5.699)
	Sideris et al.	11.474 (11.462-11.486) ± 13.020 (13.011-13.029)	8.899 (8.887-8.912) ± 11.418 (11.409-11.427)
	1D V-Net	4.297 (4.291-4.303) ± 6.527 (6.522-6.533)	2.398 (2.392-2.404) ± 5.623 (5.616-5.629)
Diastolic BP	PPG Scaling	-4.848 (-4.852–4.843) ± 4.975 (4.970-4.981)	-3.595 (-3.600–3.591) ± 3.978 (3.973-3.983)
	Sideris et al.	-12.821 (-12.831–12.811) ± 11.174 (11.166-11.182)	-15.620 (-15.630–15.610) ± 9.154 (9.146-9.162)
	1D V-Net	-3.114 (-3.118–3.110) ± 4.570 (4.565-4.576)	-2.497 (-2.501–2.493) ± 3.785 (3.781-3.789)

4.3.3 Method comparison

Next, we compared the performance of 1D V-Net with the performance of the long short-term memory (LSTM) model described in Sideris et al. and with PPG scaling. PPG scaling uses the PPG waveform as a template shape and scales the magnitude of the PPG signal in a given window to match the most recent systolic and diastolic NIBP measurements. We observe that 1D V-Net achieves the lowest RMSE and the highest correlation across both cohorts (see Table 4.3 and 4.4) and, additionally, the PPG scaling performs better than the LSTM model for both metrics. To demonstrate the improvement in the imputed waveform quality, in Supplemental Figure 4 in [HRR21] we compared the residual error for PPG scaling and 1D V-Net in a 4-second window. The waveform generated by the 1D V-Net model provides not only a more accurate prediction of the systolic and diastolic values, but also the overall waveform shape compared to the PPG scaling.

Table 4.5 contains the distribution of differences between the true and predicted systolic and diastolic BP for each method across both cohorts (see also Supplemental Figures 1, 2,

and 3 in [HRR21] for the Bland-Altman plots that correspond to Table 4.5). In the MIMIC cohort, the PPG scaling method (mean difference systolic BP 6.133 ± 6.870 mmHg, diastolic BP mean difference -4.848 ± 4.975 mmHg) performed better than the LSTM model (mean difference systolic BP 11.474 ± 13.020 mmHg, diastolic BP mean difference -12.821 ± 11.174 mmHg), and 1D V-Net outperformed both of the other methods. Similarly, in the UCLA cohort, the PPG scaling method (mean difference systolic BP 2.668 ± 5.692 mmHg, diastolic BP mean difference -3.595 ± 3.978 mmHg) was more accurate than the LSTM model (mean difference systolic BP 8.899 ± 11.418 mmHg, diastolic BP mean difference -15.620 ± 9.154 mmHg), and 1D V-Net outperformed both methods.

4.3.4 Dependence of model on NIBP measurements

As the model is intermittently calibrated with NIBP measurements, we wanted to measure the error as a function of time since the most recent NIBP value was obtained. Since the LSTM model does not rely on NIBP measurements at all, the RMSE remains relatively constant over time. However, the LSTM model has the highest RMSE of all methods (13.940 (95% CI 13.901-13.978) mmHg for MIMIC, 13.111 (95% CI 13.072-13.151) mmHg for UCLA). Both the PPG Scaling and 1D V-Net do see a minor increase in error as time from the NIBP measurement increases (see Supplemental Figures 5 and 6 in [HRR21] for MIMIC and UCLA cohorts, respectively). However, our algorithm achieves a lower RMSE compared to both the Sideris et al. model (mean difference 8.05 ± 0.45 for MIMIC, 6.13 ± 0.27 for UCLA) and PPG Scaling (mean difference 1.07 ± 0.05 for MIMIC, 2.14 ± 0.02 for UCLA) across all time points.

4.4 Discussion

We have presented a novel method for imputing the arterial blood pressure waveform that is continuous, non-invasive, accurate, and does not require any additional hardware beyond

what is standard monitoring in the acute care setting (ECG, pulse oximeter, non-invasive blood pressure cuff). This predicted waveform would allow clinicians to monitor blood pressure continuously in patients that otherwise receive blood pressure measurement intermittently. With this continuous and non-invasive monitoring, clinicians would be able to rapidly identify changes in patient state and intervene, which is crucial when even a short span of hypotension or hypertension can lead to poor health outcomes. Additionally, we have developed a preprocessing pipeline that leverages machine learning to identify windows containing low-quality signals with high precision and sensitivity. The preprocessing pipeline can be generally useful in applications that utilize physiological waveforms.

Our approach leverages the V-Net deep learning architecture (see Supplementary Note 1 in [HRR21]), which has been previously used successfully for image segmentation. Deep learning techniques have led to the development of many predictive models focusing on diagnostic applications in medicine [GPC16, EKN17, ROC18, HRH19], resulting in promising applications including diagnosis tools for diabetic retinopathy [GPC16], skin cancer [EKN17] and arrhythmia detection using electrocardiograms (ECG) [HRH19]. Notably, the majority of methods using deep learning in medicine have focused on classification problems, such as diagnosis, and not regression. Our approach differs from previous approaches, since it provides regression results as opposed to diagnosis, and to the best of our knowledge, the modified 1D V-Net approach has never been applied to physiological waveforms in this context.

The vast majority of ICU patients would benefit from the proposed system. Since two-thirds of patients in the ICU do not receive continuous blood pressure monitoring, imputing the ABP waveform allows clinicians to better monitor these patients without the need for any additional devices. The proposed system utilizes measurements that are currently part of the standard of care for all ICU patients, and therefore would not disrupt the current clinical workflows. Furthermore, in the remaining one-third of patients that do receive invasive blood pressure monitoring, the imputed waveform can be utilized as a secondary data source in

case of instrument failure or technical artifacts, or might obviate the need for an invasive monitor that may cause complications.

An additional advantage of a continuous blood pressure monitoring system based on machine-learning software is that the model has the potential to be further improved over time with additional training data. Hospitals will continue to collect data on patients that undergo invasive blood pressure monitoring and this data can be used to update the model. This would allow the model to learn any hospital-specific instrumentation or patient population differences. For device-based continuous and non-invasive blood pressure monitoring, the performance is most often fixed at the time of deployment and cannot be updated or improved over time. Therefore, a machine-learning based software approach is at an advantage.

While other computational methods for predicting beat-to-beat blood pressure measurements have been developed, they predict only the systolic/diastolic blood pressure measurements and not the actual continuous waveform. Methods for imputation of systolic and diastolic blood pressure non-invasively utilize handcrafted features including pulse transit time [XMZ19, SDZ17], heart rate [XMZ19, SDZ17, KLG13], perfusion index [XMZ19], stiffness index [XMZ19], reflection index [SDZ17], systolic/diastolic volume [SDZ17], and PPG intensity ratio [DYZ17]. These features are then used as input to machine learning models [TZ03, XMZ19, XS16, ZRH19, SDZ17, KLG13, DYZ17, SKN16]. The arterial blood pressure waveform can be used to estimate important cardiac parameters such as stroke volume (SV), cardiac output (CO), cardiac power output (CPO), vascular resistance, and pulse pressure variation, which can only be calculated using the ABP waveform, not the beat-to-beat measurements. Measurements like CPO are clinically relevant, and have been shown to be predictive of outcomes such as mortality [MCP07, FHL04]. Additionally, the imputed blood pressure waveform can be used as input to predictive algorithms which use the arterial blood pressure waveform as input, such as predicting hypotensive events up to 15 minutes before they occur [HJB18]. Another limitation of previous studies is the limited

number of patients used to train and evaluate models. The small sample sizes, most often only several dozen patients and from a single health system, calls into question the generalizability of the methods. To demonstrate the wide applicability of our proposed method, we used data from over four hundred patients and two separate health systems to show that our model successfully predicts the continuous blood pressure waveform in new patients.

CHAPTER 5

Learning Higher-Order Dynamics in Video-Based Cardiac Measurement

5.1 Introduction

Many of the properties of dynamical systems only become apparent when they move or change as the result of forces applied to them. In most applications we are interested in behavior in terms of positions, velocities, and accelerations, and in some cases the properties of interest may only be observed in subtle variations in the higher-order dynamics (e.g., acceleration). Whether monitoring the flight of a drone to create a control mechanism for stabilization or analyzing the fluid dynamics of the cardiovascular system in the human body, there can be a need to recover these dynamics accurately. However, most video-based systems are trained on lower-order signals, such as position in the case of landmark tracking or velocity/rate-of-change (optical flow) in the case of visual odometry [NNB04]. Thus, they optimize for lower (zeroth or first) order dynamics. Does this harm their ability to estimate higher order changes? We hypothesize that networks trained to predict temporal signals will benefit from combined multi-derivative learning objectives. To test this hypothesis, we explore video-based cardiac measurement as an example application with a complex dynamical system (the cardiovascular system) and introduce simple but effective changes to the inputs and outputs to significantly improve the measurement of clinically relevant parameters.

Photoplethysmography (PPG) is a low-cost and non-invasive method for measuring the

cardiovascular blood volume pulse (BVP). There are many clinical applications for PPG as the signal contains substantial information about health state and risk of cardiovascular diseases [EFL19, RSM08, PTG20]. In the current world, an acutely relevant application of PPG is for pulse oximetry (i.e. measuring pulse rate and blood oxygen saturation) as it can be used to detect low blood oxygen levels associated with the onset of COVID-19 [GKI21]. The COVID-19 pandemic has accelerated the adoption of teleheath systems [APH20] with more and more clinical consultations being conducted virtually. Therefore, techniques for remotely monitoring physiological vital signs are becoming increasingly important [GDM21, RRN21]. As one might expect, with many clinical applications the precision with which the PPG signal can be recovered is of critical importance when it comes to accurate inference of downstream conditions and the confidence of practitioners in the technology.

To date, in video-based PPG measurement the primary focus of analysis and evaluation has been on features extracted from the raw waveform or its first derivative. However, the second derivative of the PPG signal highlights subtle features that can be difficult to discern from those in the lower derivatives. Since the second derivative reflects the acceleration [Tak93] or the rate-of rate-of change of the blood volume, it is more closely related to the change in pressure applied by the heart on blood vessels and its relation to vascular health.

An example of a particular feature accentuated in the second-derivative (i.e. acceleration) PPG is the dicrotic notch (see Fig. 5.1), which occurs when the heart's aortic valve closes due to the pressure gradient between the aorta and the left ventricle. The dicrotic notch may only manifest as an inflection in the raw PPG wave; however, in the second derivative this inflection is a maxima. [IKY17] found that the second derivative of the PPG signal can be used as an indicator of arterial stiffness - which itself is an indicator of cardiac disease. [TTF98] evaluated the second derivative of the PPG waveform and found that its characteristic shape can be used to estimate vascular aging, which was higher in subjects with a history of diabetes mellitus, hypertension, hypercholesterolemia, and ischemic heart

disease compared to age-matched subjects without.

While the second derivative of a signal can be a rich source of information, often the zeroth- or first-order dynamics are given priority. For example, [CM18] observed that training video- or imaging-based PPG (iPPG) models using first-derivative (difference) frames as input with an objective function of minimizing the mean squared error between the prediction and the *first derivative* of the target BVP signal was effective. This approach was used because the authors were designing their system to measure systolic time intervals only, which are most prominent in the lower order signals. However, they did not combine this with higher-order derivatives nor did they do any systematic comparison across derivative objectives.

We argue that a model trained with an explicit second-derivative (acceleration) objective should produce feature representations that better preserve/recover these dynamics than methods that simply derive acceleration from velocity. We observe that providing the model with a second derivative input also helps the network to better predict both the first and second derivative signals.

Finally, as labeled data for training supervised models for predicting dynamical signals is often difficult to come by, we build on promising work in simulation to obtain our training data. Our results show that models trained with synthetic data can learn parameters that successfully generalize to real human subjects. While this is not a central focus of our paper, we believe that it presents a promising proof-of-concept for future work.

To summarize, in this paper, we 1) demonstrate that directly incorporating higher-order dynamics into the loss function improves the quality of the estimated higher-order signals in terms of waveform morphology, 2) show that adding second-derivative inputs additionally improves performance, and 3) we describe a novel deep learning architecture that incorporates the second derivative input frames and target signals and evaluate it against clinical-grade contact sensor measurements.

5.2 Background

Learning Higher-Order Motion from Videos. Despite its significance in many tasks, acceleration is often not explicitly modeled in many computer vision methods. However, there is a small body of literature that has considered how to recover [EJ17] and amplify optical acceleration [ZPV17, TOM18]. Given that acceleration can be equally as important as position and velocity in understanding dynamical systems, we argue that this topic deserves further attention.

A particularly relevant problem to ours is identifying small changes in videos [WRS12, ZPV17, CM20, TOM18], and specifically in acceleration in the presence of relatively large motion. As an example, in the iPPG prediction task the aim is to identify minor changes in skin coloring due to variation in blood flow patterns, while ignoring major pixel changes due to subject or background motion. One method proposed by [ZPV17] for overcoming this signal separation problem is Video Acceleration Magnification, in which large motions are assumed to be linear on the temporal scale of small changes while small changes deviate from this linearity. An extension to this method focused on making it more robust to sudden motions [TOM18]. In both cases, a combination of Eulerian and Lagrangian approaches was used, rather than utilizing a supervised learning paradigm. Of relevance here is also work magnifying subtle physiological changes using neural architectures [CM20], which have been shown to effectively separate signal and noise in both the spatial and temporal domains.

Our work might be most closely related to prior research into feature descriptors for optical acceleration [EJ17]. One example uses histograms of optical acceleration to effectively encode the motion information. However, this work also defined handcrafted features, rather than learning representations from data. Our work is also related conceptually to architectures such as SlowFast [FFM19] in that it utilizes multiple "pathways" to learn different properties of the dynamics within a video. We were inspired by this approach; however, unlike SlowFast, we focus specifically on higher-order pathways rather than slower and faster

frame sequences.

Video-based Cardiac Measurement. Diffuse reflections from the body vary depending on how much light is absorbed in the peripheral layers of the skin and this is influenced by the volume of blood in the capillaries. Digital cameras can capture these very subtle changes in light which can then be used to recover the PPG signal [WBS00, TO07, VSN08, PMP10]. The task then becomes separating pixel changes due to blood flow from those due to body motions, ambient lighting variation, and other environmental factors that we consider noise in this context. While earlier methods leveraged source separation algorithms [WBS16], such as ICA [PMP10] or PCA [LRK11], neural models provide the current state-of-the-art in this domain [CM18, LFP20, LNC21, SCC21, LHZ21]. These architectures support learning spatial attention and source-specific temporal variations and separating these from various sources of noise. Typically, the input to these models are normalized video frames and the output is a 1-D time series prediction of the PPG waveform or the heart rate. A vast majority of work has evaluated these methods based errors in heart rate estimation, which considers the dominant or "systolic" frequency alone. Only a few papers have used more challenging evaluation criteria, such as the estimation of systolic to diastolic peaks [MGP14].

5.3 Optical Basis

We start by providing an optical basis for the measurement of the pulse wave using a camera and specifically the second derivative signal. Starting with Shafer's Dichromatic Reflection Model (DRM)[WBS16, CM18, LFP20], we want to understand how higher order changes in the blood volume pulse impact pixel intensities to motivate the design of our inputs and loss function. Based on the DRM model the RGB values captured by the cameras as given by:

$$\boldsymbol{C}_k(t) = I(t) \cdot (\boldsymbol{v}_s(t) + \boldsymbol{v}_d(t)) + \boldsymbol{v}_n(t) \tag{5.1}$$

where $I(t)$ is the luminance intensity level, modulated by the specular reflection $\boldsymbol{v}_s(t)$ and the diffuse reflection $\boldsymbol{v}_d(t)$. Quantization noise of the camera sensor is captured by $\boldsymbol{v}_n(t)$. $I(t)$ can be decomposed into stationary and time-varying parts $\boldsymbol{v}_s(t)$ and $\boldsymbol{v}_d(t)$ [WBS16]:

$$\boldsymbol{v}_d(t) = \boldsymbol{u}_d \cdot d_0 + \boldsymbol{u}_p \cdot p(t) \tag{5.2}$$

where $\boldsymbol{u}_d$ is the unit color vector of the skin-tissue; d_0 is the stationary reflection strength; $\boldsymbol{u}_p$ is the relative pulsatile strengths caused by hemoglobin and melanin absorption; $p(t)$ represents the physiological changes. Let us assume for simplicity in this case that the luminace, I (i.e., illumination in the video) is constant, not time varying, which is a reasonable assumption for short videos and those in which the subject can control their environment (e.g., indoors). Then differentiating twice with respect to time, t:

$$\frac{\partial^2 \boldsymbol{C}_k(t)}{\partial t^2} = I \cdot (\frac{\partial^2 \boldsymbol{v}_s(t)}{\partial t^2} + \frac{\partial^2 \boldsymbol{u}_d}{\partial t^2} + \frac{\partial^2 \boldsymbol{u}_p(t)}{\partial t^2} + \frac{\partial^2 \boldsymbol{v}_n(t)}{\partial t^2}) \tag{5.3}$$

The non-time varying part $\boldsymbol{u}_d \cdot d_0$ becomes zero. Thus simplifying the equation to:

$$\frac{\partial^2 \boldsymbol{C}_k(t)}{\partial t^2} = I \cdot (\frac{\partial^2 \boldsymbol{v}_s(t)}{\partial t^2} + \frac{\partial^2 \boldsymbol{u}_p(t)}{\partial t^2} + \frac{\partial^2 \boldsymbol{v}_n(t)}{\partial t^2}) \tag{5.4}$$

Furthermore, if specular reflections do not vary over time (e.g., if the camera and subject are stationary), the $\boldsymbol{v}_s(t)$ term will also become zero. This means that the second derivative changes in pixel intensities are a sum of second derivative changes in PPG and camera noise. With current camera technology, and little video compression, image noise is typically much smaller than the PPG signal. Therefore, we would expect the pixel changes to be dominated by second derivative variations in the blood volume pulse:

$$\frac{\partial^2 \boldsymbol{C}_k(t)}{\partial t^2} = I \cdot \frac{\partial^2 \boldsymbol{u}_p(t)}{\partial t^2} \tag{5.5}$$

As such, we can infer that when attempting to estimate the second derivative of the PPG signal from videos without very large motions or illumination changes, second derivative changes in the pixel space would appear helpful and that minimizing the loss between the second derivative prediction and ground truth will be the simplest learning task for the algorithm when the input is second-derivative pixel changes.

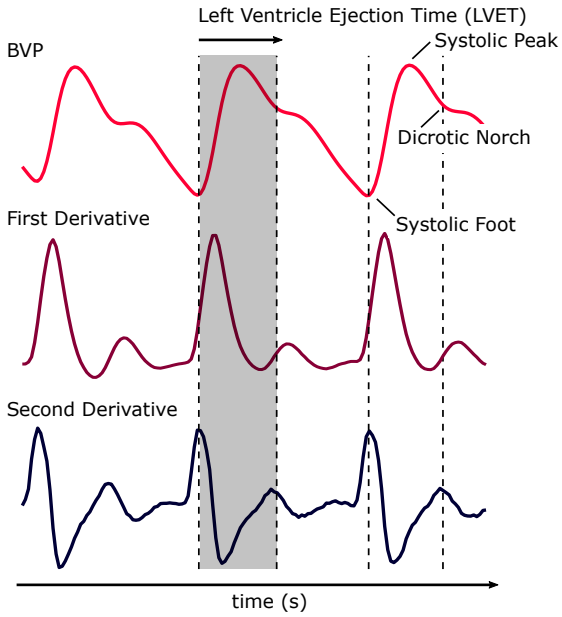

Figure 5.1: The Left Ventricle Ejection Time (LVET) is the duration between the beginning and end of the systolic phase. This interval corresponds to the opening and closing of the heart's aortic valve, during which the left ventricle ejects blood into the system. In the PPG waveform, this interval begins at the diastolic point and ends with the dicrotic notch.

5.4 Our Model

To test our hypothesis we incorporate higher-order dynamics into a model via the loss function and as inputs to see if they will provide a better estimation of the higher-order dynamics of the target signal. We propose a multi-derivative convolutional attention network architecture (see Fig. 5.2) that operates separately on each set of derivative input frames to extract sequences of features, and then maps these sequences to the target derivative signals. The

convolutional attention network (CAN) paradigm was first described by [CM18] to capture the spatial-temporal changes with convolutional operations and uses soft-attention masks to help segment the subject from the background and focus on relevant regions of the body. In our case, we used a CAN architecture with three branches: (1) a first-derivative branch, (2) a second-derivative branch, and (3) an attention branch. The first-derivative branch extracts features from the differences between consecutive video frames, and similarly the second-derivative branch that extracts features from the difference-of-difference frames. The attention branch uses the raw video frames to learn attention masks (one per frame) that encourage the network to prioritize regions of the image that contain useful signal (e.g. participant's skin) and ignore noisy regions (e.g. background). These attention masks are shared between the first-derivative branch and the second-derivative branch as we expect the same spatial regions to contain first and second derivative information. After feature representations are extracted from frames within each derivative-input branch, the features are concatenated together for each time step and the target signals are then generated using recurrent neural network (RNN) layers. A diagram depicting the architecture used for our experimentation is shown in Fig. 5.2.

5.4.1 Predicting multi-derivative target signals

The goal of iPPG is to obtain an estimate of the underlying PPG signal $p(t)$ (as in Eq. 5.2), while only observing video frames $X(t)$ containing a subject's skin (in this case the face). Mathematically, this can be described as learning a function:

$$\hat{p}(t) = f(X(t)) \tag{5.6}$$

Or, because we are interested in changes in blood volume *changes*, estimating the first derivative of the PPG signal:

$$\hat{p}'(t) = f(X(t), X'(t)) \tag{5.7}$$

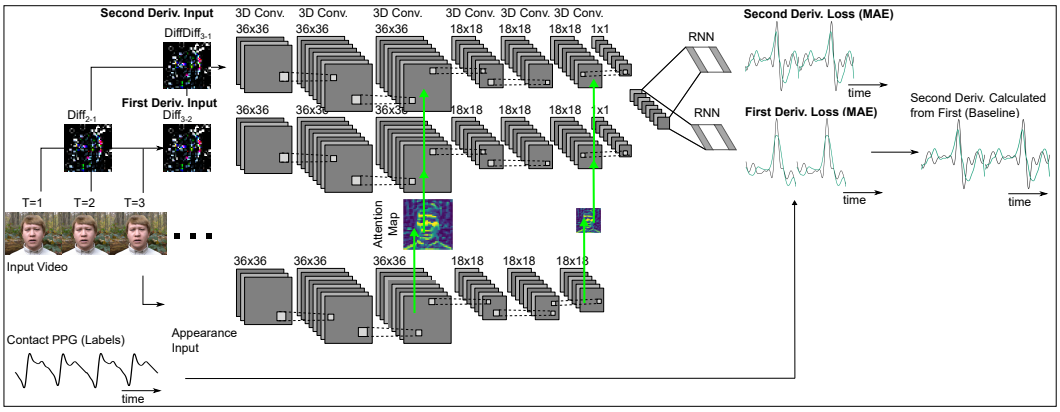

Figure 5.2: Our multi-derivative architecture used for experimentation. Spatial features are extracted separately for each set of derivative frames using 3D convolutional layers and mean pooling layers. Once feature representations are extracted, the temporal features from each branch are concatenated together and recurrent layers are used for modeling the temporal signals. The first- and second-derivative losses are calculated as the mean absolute error between the predictions and the synchronized ground-truth PPG signal.

Where the first derivative PPG signal is defined as:

$$p'(t) = p(t) - p(t-1) \tag{5.8}$$

Using prior methods, to obtain an estimate of the PPG signal's second derivative, one would either differentiate the predicted PPG signal twice, or differentiate the predicted first-derivative PPG once, rather than calculate the acceleration PPG directly. In contrast, we explicitly predict the acceleration PPG waveform as a target signal. We define the second derivative waveform as the difference between consecutive first-derivative time points:

$$p''(t) = p'(t) - p'(t-1) \tag{5.9}$$

Then we train our model to predict the second derivative waveform $\hat{p}''(t) = f(X(t), X'(t))$ given a set of input video frames $X(t)$ and the corresponding normalized difference frames $X'(t)$. To optimize our model parameters we minimize the mean squared difference between

the true and predicted second derivative waveforms:

$$L = \frac{1}{T}\sum_{t=1}^{T}(p''(t) - \hat{p}''(t))^2 \tag{5.10}$$

5.4.2 Leveraging multi-derivative inputs

It has been previously shown that the normalized difference frames are useful for predicting the first derivative PPG waveforms. Therefore, we hypothesized that incorporating the second derivative of the raw video frames $X''(t) = X'(t) - X'(t-1)$ (i.e. the difference-of-difference frames) may also be useful for predicting the PPG signal and its derivatives. Similar to the difference frames, we added a separate convolutional attention branch, where the attention mask is shared between both branches (see Fig. 5.2). Sharing the attention mask is a reasonable assumption as we would expect skin regions to all exhibit the signal and similar dynamics. After the feature maps in each branch are pooled into a single value per feature at each time step, the learned representations are concatenated together. These concatenated features over time are used as input sequences to the recurrent layers that generate the target waveforms.

Given that difference frames $X'(t)$ are useful for predicting the first derivative PPG waveforms, features learned from the difference-of-difference frames $X''(t)$ may be beneficial for predicting the second derivative PPG signal. In theory, if difference-of-difference features are indeed useful for predicting the acceleration PPG, then the CAN network should be able to learn those features from the difference frames due to the 3D convolutional operations. However, manually adding the difference-of-difference frames could help guide the model. To examine the effect of combining higher-order inputs and target signals, we fit a model $\hat{p}''(t) = f(X(t), X'(t), X''(t))$ to predict the second-derivative PPG.

5.5 Experiments

In this section we will describe the data used to train and evaluate our method and perform a systematic ablation study in which we test different combinations of inputs and outputs.

5.5.1 Data

Training To train our models using a large and diverse set of subjects, we leverage recent work that uses highly-parameterized synthetic avatars to generate videos containing simulated subjects with various movements and backgrounds [MHW20]. To drive changes in the synthetic avatars' appearance, the PPG signal is used to manipulate the base skin color and the subsurface radius [MHW20]. The subsurface scattering is spatially weighted using an artist-created subsurface scattering radius texture that captures variations in the thickness of the skin across the face. Using physiological waveforms signals from the MIMIC Physionet [GAG00] database, we randomly sampled windows of PPG waveforms from real patients. The physiological waveform data were sampled to maximize examples from different patients. Using the synthetic avatar pipeline and MIMIC waveforms, we generated 2,800 6-second videos, where half of the videos were generated using hand-crafted facial motion/action signals, and the other half using facial motion/action signals extracted using landmark detection on real videos. Examples of the avatars can be found in Appendix B.1.1.

Testing Given that we are focusing on recovering *very* subtle changes in pixel intensities due to the blood volume pulse, we use a highly controlled and very accurately annotated dataset of real videos for evaluation. The AFRL dataset [EBM14] consists of 300 videos from 25 participants (17 male and 8 female). Each video in the dataset has a resolution of 658x492 pixels sampled at 30 Hz. Ground truth PPG signals were recorded using a contact reflective PPG sensor attached to the subject's index finger. Each participant was instructed to perform three head motion tasks including rotating the head along the horizontal axis, rotating the head along the vertical axis, and rotating the head randomly once every second

Table 5.1: Quantitative performance comparison between different architecture configurations. Values shown are (mean ± standard deviation). Beats-per-minute (BPM); First Derivative (FD); Heart Rate (HR); Mean Absolute Error (MAE); Second Derivative (SD); Left Ventricle Ejection Time (LVET).

	Input Frames		Target Signals		HR MAE (BPM)	LVET MAE (ms)
	FD	SD	FD	SD		
(FD-Optimized)	✓	✗	✓	✗	0.66 ± 2.07	108.26 ± 56.19
	✓	✗	✓	✓	1.63 ± 4.21	64.77 ± 31.04
	✓	✗	✗	✓	1.68 ± 4.85	57.29 ± 22.68
	✓	✓	✓	✗	0.88 ± 2.75	75.41 ± 43.47
	✓	✓	✓	✓	1.16 ± 3.67	65.99 ± 31.88
(SD-Optimized)	✓	✓	✗	✓	3.41 ± 8.77	52.07 ± 19.53

to one of nine predefined locations. Since our goal in this work was to compare methods for estimating subtle waveform dynamics, which can be more difficult to do in the presence of large motion, we focused here on the first two AFRL tasks where participant motion is minimal. Examples of AFRL participants can be found in Appendix B.1.1.

5.5.2 Implementation details

We trained our models using a large dataset of generated synthetic avatars and evaluated model performance on the AFRL dataset, which consists of real human subjects. For each video, we reduced the resolution of the video to 36x36 pixels to reduce noise and computational requirements while maintaining useful spatial signal. The input to the attention branch was T raw video frames. The input to the first-derivative branch was a set of T normalized difference frames, calculated by subtracting consecutive frames and normalizing by the sum. The input to the second-derivative branch was a set of $T-1$ difference-of-difference frames (second derivative frames), calculated by subtracting consecutive normalized difference frames (i.e. the T frames used as input to the motion branch). In our experiments, we used a window size of $T = 30$ video frames to predict the target signals for the corre-

sponding 30 time points. During training, a sliding window of 15 frames (i.e. 50% overlap between consecutive windows) was used to increase the total number of training examples. The model was implemented using Tensorflow [AAB16] and trained for eight epochs using the Adam [KB17] optimizer with a learning rate of 0.001, and a batch size of 16.

5.5.3 Systematic Evaluation

To measure the effect of using multi-derivative inputs and outputs, we systematically removed the second-derivative parts of the model and used quantitative and qualitative methods to examine the change in model performance. To quantitatively measure the quality of the predicted signal, we calculated two clinically important parameters - heart rate (HR) and the left ventricular ejection time (LVET) interval (see Appendix B.1.3 for details). Video-based HR prediction has been a major focus of iPPG applications, with many methods showing highly-accurate results. HR can be determined through peak detection or by determining the dominant frequency in the signal (e.g. using fast Fourier transform). Since current iPPG methods are able to achieve sufficiently-low error rates on the HR estimation task, we believe that metrics that capture the quality of waveform morphology should also be considered.

The LVET interval is defined as the time between the opening and closing of the heart's aortic valve, i.e. the systolic phase when the heart is contracting (see Fig. 5.1). In the PPG waveform, this interval begins at the diastolic point (i.e. the global minimum pressure within a heartbeat cycle) and ends with the dicrotic notch (i.e. local minimum occurring after systolic peak, marking the end of the systolic phase and the beginning of the diastolic phase). LVET typically is correlated with cardiac output (stroke volume $\times$ heart rate)[HHK90], and has been shown to be an indicator of future heart failure as the time interval decreases with left-ventricle dysfunction [BQH18].

Calculating LVET requires identification of the diastolic point and the dicrotic notch. The diastolic point is a (global) minimum point within a heart beat, meaning it corresponds to a positive peak in the second derivative signal according to the second-derivative test.

Similarly, the dicrotic notch is a (local) minimum in the PPG signal, and appears as a positive peak in the second derivative following the diastolic peak in time. Because the dicrotic notch can often be a subtle feature, it is much easier to identify in the PPG's second derivative compared to the raw signal. Therefore, it is a good example of clinically-important waveform morphology that is best captured by higher-order dynamics.

Removing the second-derivative frames In Table 5.1, quantitative evaluation metrics (HR and LVET) are shown for all experiments in our ablation study, using tasks 1 and 2 from the AFRL dataset. Removing the second-derivative (SD) frames results in the model configurations in the top three rows of Table 5.1. When SD frames are removed, the result is a general decrease in the HR error. However, there is also a general increase in LVET interval prediction error, which suggests that including the SD frames leads to improved estimation of waveform morphology.

Removing the first-derivative target signal Intuitively, models that are optimized using a loss function specifically focusing on a single objective will perform better in terms of that objective compared to models trained with loss functions containing multiple objectives. By removing the first-derivative target signal from the training objective, the model is focused to exclusively focus on the second-derivative (SD) objective. Empirically, this leads the SD-Optimized model to have the lowest LVET MAE of any model configuration (last row of Table 5.1). While the SD-Optimized model achieves the lowest LVET error, the HR error is the highest of any configuration. These results suggest that there are performance trade-offs to consider when designing a system for particular downstream tasks.

Removing the second-derivative target signal When the second-derivative target signal is removed from the model, the optimization procedure is purely focused on improving the prediction of the first derivative. The FD-Optimized model (first row of Table 5.1) serves as a form of baseline, since previous works have focused on using first-derivative (FD) frames to predict the first-derivative PPG signal. Fig. 5.4 shows a Bland-Altman plot [MA86] comparing the FD-Optimized and SD-Optimized error distributions as a function of the

ground-truth values both HR and LVET intervals.

Perhaps unsurprisingly, our results show the FD-Optimized model achieves the lowest HR MAE (0.66 $\pm$ 2.07 BPM) of any model configuration examined and, in particular, improves HR estimation compared to models without the first derivative target signal. However, the FD-Optimized model also has the worst performance in terms of the LVET MAE (108.26 $\pm$ 56.19 ms) of any model configuration. This suggests that while the configuration provides an accurate assessment of the heartbeat frequency, the quality of predicted waveform morphology can be improved by incorporating second-derivative information.

Qualitative comparisons For a qualitative comparison, in Fig. 5.3 we plot the ground-truth, FD-Optimized, and SD-Optimized PPG, first derivative, and second derivative. Additionally, in the bottom panel of Fig. 5.3 we overlay the true and predicted LVET intervals for each signal to demonstrate model performance. For additional qualitative comparisons, see Appendix B.2.

5.6 Conclusions

Using the task of video-based cardiac measurement we have shown that when learning representations for dynamical systems that appropriately designing inputs, and optimizing for derivatives of interest can make a significant difference in model performance. Specifically, there is a trade-off between optimizing for lower-order and higher-order dynamics. Given the importance of second-derivatives (i.e., acceleration) in this, and many other video understanding tasks, we believe it is important to understand the trade-off between optimizing for targets that capture different dynamic properties. In cardiac measurement in particular, the LVET is one of the more important clinical parameters and can be better estimated using higher-order information. While we have investigated the importance of higher-order dynamics in the context of video-based cardiac measurement, this paradigm is generally applicable. We believe future work will continue to showcase the importance of explicitly

incorporating higher-order dynamics.

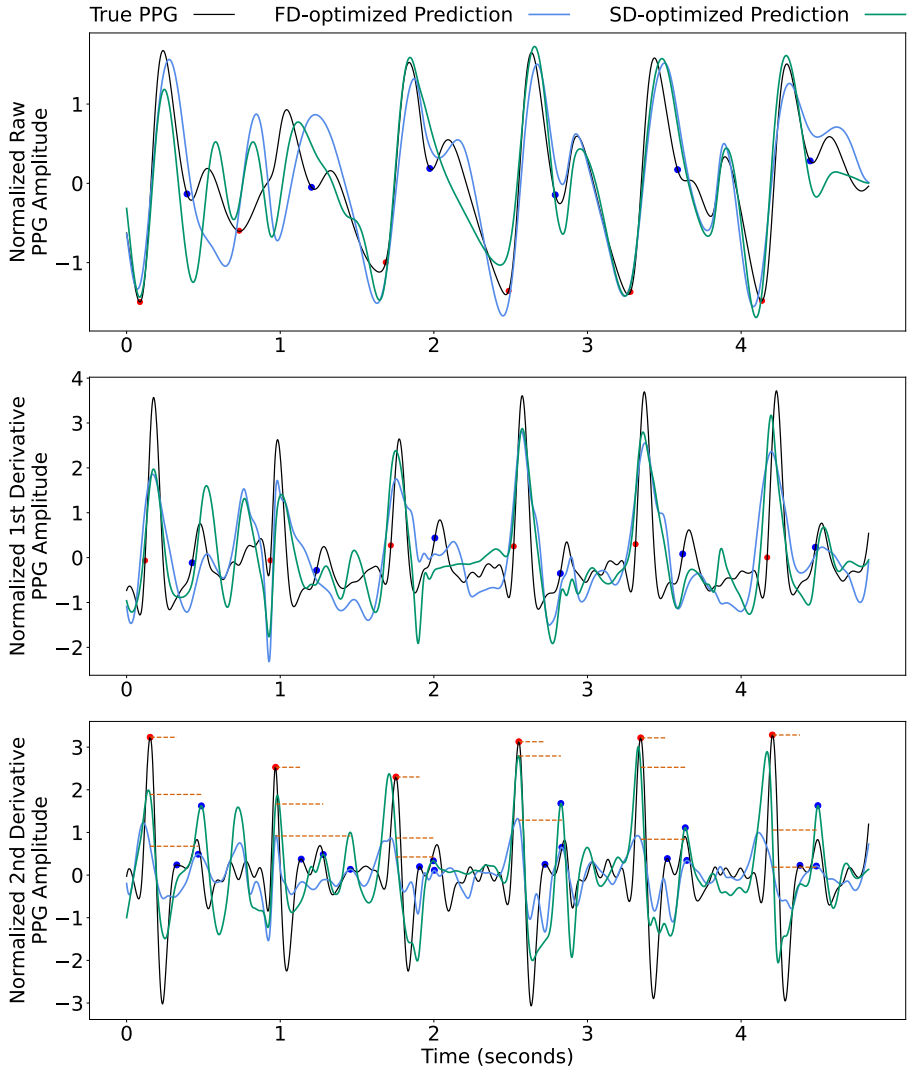

Figure 5.3: Comparison of true (black) and predicted (blue and green) raw or zeroth order (top), first order derivative (middle), and second order derivative (bottom) waveforms. The blue and green lines reflect two models: predicting the first derivative (blue) and predicting the second derivative (green). Diastolic points are labeled with red dots, and dicrotic notches are labeled with blue dots. LVET intervals are labeled by dotted red lines. Notice how the points of interest are generally more obvious in the second derivative waveforms, as they are maxima rather than inflections. Also note that the LVET time intervals for the second derivative model are generally more similar to those from the contact (true) PPG.

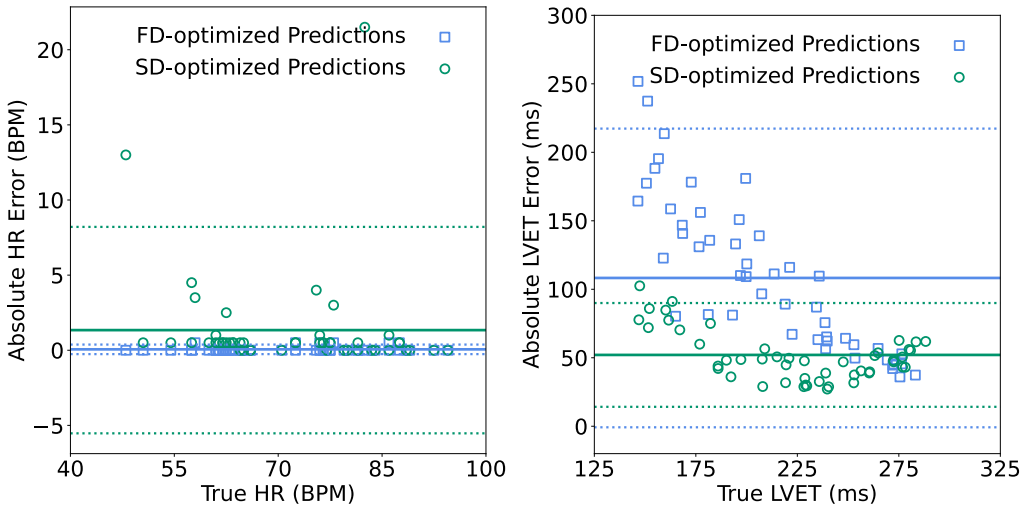

Figure 5.4: Bland-Altman plots comparing error distributions for average task heart rate (left) and LVET intervals (right) for the model optimized for first-derivative prediction (blue) and the model optimized for second-derivative prediction (green). (left) The absolute difference between the true and predicted heart rate for each subject/task. (right) The absolute difference between the predicted and true values (y-axis, in milliseconds) is plotted as a function of the true LVET (x-axis, in milliseconds). The solid line represents the mean error, and the dotted lines represent the 95% confidence intervals ($\pm$ 1.96 $\times$ standard deviation).

APPENDIX A

Supplementary Material - The methylation risk score is an informative biomarker within electronic health record systems

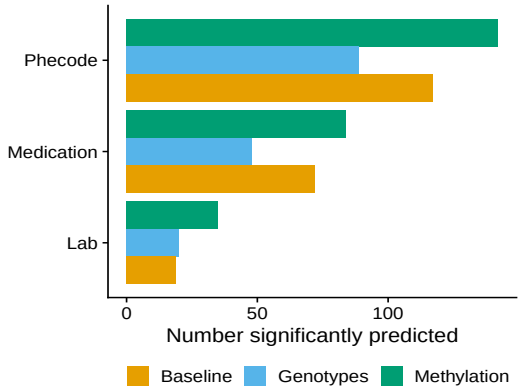

Figure A.1: Significantly predicted outcomes per data type. Total number of significantly predicted outcomes when using the baseline alone, as well as including either set of genomic features in addition to the baseline. We used an association test of the cross-validated predictors and the true outcome and adjusted for multiple testing using Bonferroni correction at a nominal threshold of 0.05.

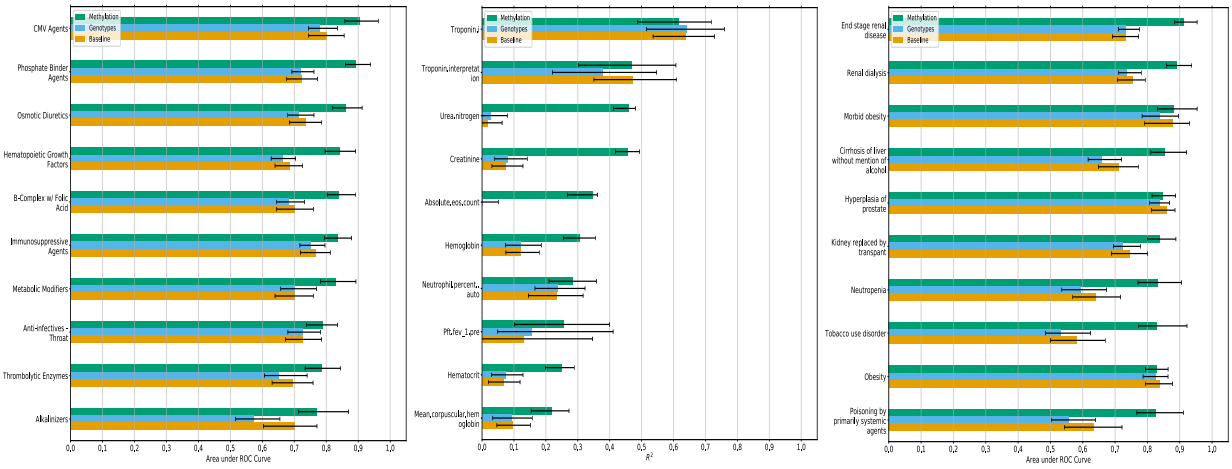

Figure A.2: Significantly-predicted outcomes per data type. The top 10 methylation-predicted (Left) medications, (Middle) labs, and (Right) Phecodes, with Baseline and Genotype prediction performance results for comparison.

		Missing	Overall
n			860
Age, mean (SD)		0	61.0 (15.9)
Sex, n (%)	F	0	367 (42.7)
	M		493 (57.3)
BMI, mean (SD)		1	27.2 (6.5)
AKIN Classification, n (%)	0.0	0	558 (64.9)
	1.0		190 (22.1)
	2.0		28 (3.3)
	3.0		84 (9.8)
GFR $>$ 38, n (%)	False	0	385 (44.8)
	True		475 (55.2)
Self-Reported Ethnicity, n (%)	Cuban	0	1 (0.1)
	Hispanic or Latino		130 (15.1)
	Hispanic/Spanish origin Other		12 (1.4)
	Mexican, Mexican American, Chicano/a		37 (4.3)
	Not Hispanic or Latino		671 (78.0)
	Patient Refused		5 (0.6)
	Puerto Rican		3 (0.3)
	Unknown		1 (0.1)
Self-Reported Race, n (%)	American Indian	0	1 (0.1)
	Asian		77 (9.0)
	Black		74 (8.6)
	Declined to Specify		7 (0.8)
	Other Race		141 (16.4)
	Pacific Islander		3 (0.3)
	White or Caucasian		557 (64.8)

Table A.1: Cohort patient demographics. AKIN is the Acute Kidney Injury Network Classification, BMI is Body Mass Index, GFR is glomerular filtration rate.

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
CMV Agents	0.80 (0.75-0.86)	**0.90 (0.86-0.94)**	0.78 (0.72-0.84)
Phosphate Binder Agents	0.72 (0.68-0.77)	**0.89 (0.86-0.92)**	0.72 (0.67-0.77)
Osmotic Diuretics	0.74 (0.69-0.79)	**0.86 (0.82-0.90)**	0.71 (0.66-0.77)
Hematopoietic Growth Factors	0.69 (0.65-0.73)	**0.84 (0.80-0.88)**	0.66 (0.62-0.71)
B-Complex w/ Folic Acid	0.70 (0.65-0.76)	**0.84 (0.80-0.88)**	0.68 (0.63-0.73)
Immunosuppressive Agents	0.77 (0.72-0.81)	**0.83 (0.79-0.87)**	0.75 (0.71-0.80)
Metabolic Modifiers	0.70 (0.63-0.76)	**0.83 (0.78-0.88)**	0.70 (0.64-0.76)
Anti-infectives - Throat	0.73 (0.67-0.79)	**0.79 (0.74-0.84)**	0.73 (0.67-0.77)
Thrombolytic Enzymes	0.70 (0.61-0.76)	**0.79 (0.73-0.83)**	0.65 (0.59-0.71)
Alkalinizers	0.70 (0.62-0.77)	0.77 (0.71-0.83)	0.57 (0.47-0.67)
Prostatic Hypertrophy Agents	0.75 (0.70-0.79)	0.75 (0.71-0.80)	0.74 (0.70-0.78)
Potassium Removing Agents	0.66 (0.60-0.70)	**0.75 (0.70-0.79)**	**0.56 (0.50-0.63)**
Antacids - Bicarbonate	0.77 (0.71-0.83)	0.74 (0.66-0.81)	0.63 (0.56-0.71)
Plasma Proteins	0.68 (0.63-0.72)	**0.74 (0.69-0.78)**	0.66 (0.62-0.70)
Cephalosporins - 4th Generation	0.66 (0.57-0.74)	0.73 (0.64-0.82)	0.65 (0.58-0.73)
HMG CoA Reductase Inhibitors	0.71 (0.68-0.75)	**0.72 (0.68-0.76)**	0.71 (0.67-0.74)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Imidazole-Related Antifungals	0.68 (0.62-0.73)	**0.72 (0.66-0.77)**	0.67 (0.61-0.72)
Vasodilators	0.53 (0.49-0.58)	**0.72 (0.68-0.76)**	**0.46 (0.41-0.51)**
Salicylates	0.70 (0.67-0.74)	0.71 (0.68-0.74)	0.70 (0.66-0.73)
Bone Density Regulators	0.70 (0.63-0.76)	0.71 (0.66-0.77)	0.66 (0.59-0.73)
Cycloplegic Mydriatics	0.79 (0.72-0.84)	0.71 (0.64-0.77)	0.68 (0.62-0.75)
Specialty Vitamins Products	0.61 (0.52-0.68)	0.71 (0.64-0.79)	0.60 (0.53-0.67)
Gallstone Solubilizing Agents	0.62 (0.54-0.72)	0.71 (0.63-0.79)	0.60 (0.51-0.71)
Antiseptics - Mouth/Throat	0.60 (0.53-0.66)	**0.69 (0.63-0.76)**	0.58 (0.51-0.65)
Impotence Agents	0.70 (0.65-0.74)	0.69 (0.63-0.75)	0.64 (0.56-0.71)
Proton Pump Inhibitors	0.61 (0.57-0.64)	**0.69 (0.66-0.72)**	0.60 (0.56-0.64)
Phosphate	0.67 (0.61-0.75)	0.69 (0.62-0.77)	0.65 (0.58-0.72)
Iron	0.66 (0.62-0.71)	**0.68 (0.64-0.72)**	0.65 (0.61-0.69)
Anti-infective Misc. - Combinations	0.64 (0.60-0.70)	**0.68 (0.64-0.72)**	0.64 (0.59-0.68)
Parenteral Therapy Supplies	0.56 (0.50-0.62)	**0.68 (0.61-0.73)**	0.53 (0.47-0.59)
Gout Agents	0.66 (0.61-0.71)	0.67 (0.62-0.73)	0.63 (0.58-0.68)
Alternative Medicine - M's	0.57 (0.50-0.64)	**0.67 (0.61-0.73)**	0.58 (0.51-0.65)
Insulin	0.64 (0.61-0.67)	**0.67 (0.63-0.71)**	0.63 (0.59-0.67)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Alpha-Beta Blockers	0.54 (0.50-0.59)	**0.67 (0.63-0.71)**	0.51 (0.47-0.55)
Anti-infective Agents - Misc.	0.60 (0.55-0.64)	**0.67 (0.63-0.71)**	0.58 (0.53-0.62)
Loop Diuretics	0.62 (0.58-0.65)	**0.67 (0.62-0.70)**	0.61 (0.58-0.65)
Antiflatulents	0.61 (0.55-0.66)	**0.66 (0.61-0.72)**	0.58 (0.52-0.64)
Magnesium	0.59 (0.55-0.63)	**0.66 (0.62-0.71)**	0.57 (0.52-0.60)
Carbohydrates	0.63 (0.59-0.67)	**0.66 (0.61-0.70)**	0.61 (0.58-0.65)
Laxatives - Miscellaneous	0.60 (0.56-0.64)	**0.66 (0.62-0.70)**	0.59 (0.55-0.63)
5-HT3 Receptor Antagonists	0.57 (0.53-0.61)	**0.66 (0.62-0.70)**	0.56 (0.53-0.60)
Antihistamines - Ethanolamines	0.63 (0.59-0.68)	**0.66 (0.62-0.70)**	0.62 (0.59-0.66)
Nitrates	0.63 (0.58-0.67)	**0.66 (0.61-0.70)**	**0.61 (0.56-0.66)**
Cephalosporins - 3rd Generation	0.61 (0.57-0.65)	**0.66 (0.62-0.70)**	0.59 (0.55-0.63)
Opioid Antagonists	0.60 (0.55-0.65)	**0.66 (0.61-0.71)**	0.59 (0.53-0.64)
Thiazides and Thiazide-Like Diuretics	0.66 (0.61-0.71)	0.65 (0.60-0.71)	0.66 (0.60-0.70)
Glucocorticosteroids	0.61 (0.57-0.64)	**0.65 (0.62-0.69)**	0.60 (0.56-0.64)
Dibenzapines	0.60 (0.52-0.67)	0.65 (0.59-0.72)	0.54 (0.47-0.61)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Calcium	0.61 (0.56-0.66)	**0.65 (0.61-0.70)**	0.58 (0.54-0.62)
Calcium Channel Blockers	0.59 (0.55-0.63)	**0.65 (0.61-0.69)**	0.59 (0.55-0.64)
Fluoroquinolones	0.55 (0.50-0.58)	**0.65 (0.61-0.69)**	0.53 (0.49-0.56)
Biguanides	0.66 (0.61-0.70)	0.64 (0.59-0.69)	0.65 (0.61-0.71)
Thyroid Hormones	0.64 (0.59-0.68)	0.64 (0.59-0.69)	0.65 (0.61-0.70)
Local Anesthetic Combinations	0.57 (0.53-0.61)	**0.64 (0.60-0.68)**	0.54 (0.49-0.59)
Heparins And Heparinoid-Like Agents	0.61 (0.57-0.65)	**0.64 (0.60-0.68)**	0.60 (0.56-0.64)
Diabetic Supplies	0.65 (0.59-0.70)	0.64 (0.58-0.70)	0.63 (0.57-0.69)
Potassium	0.60 (0.55-0.64)	**0.64 (0.60-0.68)**	0.59 (0.55-0.62)
Antacid Combinations	0.63 (0.58-0.68)	0.64 (0.59-0.70)	0.61 (0.56-0.66)
Zinc	0.68 (0.60-0.77)	0.64 (0.56-0.72)	0.62 (0.53-0.70)
Liquid Vehicles	0.61 (0.53-0.67)	0.64 (0.58-0.71)	0.58 (0.51-0.66)
Electrolyte Mixtures	0.58 (0.54-0.62)	**0.64 (0.60-0.68)**	0.57 (0.53-0.61)
Antitussives	0.48 (0.42-0.55)	**0.64 (0.58-0.70)**	0.40 (0.33-0.47)
Sympathomimetics	0.60 (0.55-0.64)	**0.64 (0.60-0.68)**	0.60 (0.56-0.64)
Saline Laxatives	0.58 (0.54-0.62)	**0.64 (0.60-0.68)**	0.56 (0.52-0.61)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Antiarrhythmics Type III	0.56 (0.48-0.63)	**0.64 (0.56-0.71)**	**0.45 (0.38-0.53)**
Analgesics Other	0.60 (0.56-0.63)	**0.64 (0.59-0.67)**	**0.58 (0.54-0.61)**
Glycopeptides	0.61 (0.56-0.65)	**0.64 (0.59-0.67)**	0.60 (0.55-0.65)
Vasopressors	0.58 (0.54-0.62)	**0.64 (0.60-0.67)**	0.56 (0.52-0.60)
Stimulant Laxatives	0.62 (0.58-0.65)	0.63 (0.60-0.67)	0.61 (0.57-0.64)
Antihypertensive Combinations	0.68 (0.61-0.73)	0.63 (0.56-0.70)	0.61 (0.54-0.67)
Diagnostic Radiopharmaceuticals	0.58 (0.54-0.62)	**0.63 (0.59-0.68)**	0.58 (0.54-0.63)
Surfactant Laxatives	0.61 (0.56-0.64)	0.63 (0.59-0.67)	0.61 (0.57-0.64)
Antiperistaltic Agents	0.55 (0.48-0.62)	**0.63 (0.56-0.70)**	0.53 (0.46-0.60)
Benzodiazepines	0.61 (0.56-0.64)	**0.63 (0.59-0.67)**	0.60 (0.55-0.64)
Antidepressants - Misc.	0.69 (0.62-0.76)	0.63 (0.55-0.71)	0.56 (0.48-0.64)
Bicarbonates	0.62 (0.56-0.70)	0.63 (0.55-0.70)	0.57 (0.50-0.65)
Ophthalmic Steroids	0.67 (0.62-0.72)	0.63 (0.57-0.69)	0.63 (0.56-0.69)
Ophthalmics - Misc.	0.65 (0.59-0.70)	0.63 (0.57-0.68)	0.66 (0.61-0.72)
Sulfonylureas	0.61 (0.54-0.68)	0.63 (0.56-0.69)	0.57 (0.49-0.63)
Antibiotics - Topical	0.55 (0.50-0.59)	**0.62 (0.58-0.67)**	0.53 (0.49-0.59)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Cobalamins	0.61 (0.54-0.66)	0.62 (0.56-0.68)	0.58 (0.52-0.64)
Bacterial Vaccines	0.59 (0.54-0.64)	0.62 (0.58-0.67)	0.58 (0.54-0.62)
Platelet Aggregation Inhibitors	0.64 (0.60-0.68)	0.62 (0.57-0.67)	0.62 (0.58-0.66)
Penicillin Combinations	0.56 (0.52-0.61)	**0.62 (0.57-0.66)**	0.56 (0.51-0.60)
Steroid Inhalants	0.50 (0.41-0.59)	0.62 (0.53-0.71)	0.55 (0.46-0.64)
Tetracyclines	0.52 (0.46-0.58)	0.62 (0.56-0.68)	**0.41 (0.35-0.46)**
Lozenges	0.59 (0.53-0.67)	0.61 (0.54-0.69)	0.45 (0.39-0.52)
Alpha-2 Receptor Antagonists (Tetracyclics)	0.66 (0.57-0.74)	0.61 (0.51-0.71)	0.64 (0.55-0.73)
Beta Blockers Cardio-Selective	0.54 (0.51-0.58)	**0.61 (0.57-0.65)**	0.53 (0.49-0.57)
Phenothiazines	0.53 (0.49-0.58)	**0.61 (0.56-0.66)**	0.48 (0.44-0.53)
Angiotensin II Receptor Antagonists	0.62 (0.58-0.66)	0.61 (0.57-0.66)	0.60 (0.56-0.64)
Coumarin Anticoagulants	0.61 (0.56-0.68)	0.61 (0.55-0.67)	0.58 (0.53-0.63)
Irrigation Solutions	0.60 (0.52-0.68)	0.61 (0.53-0.68)	0.57 (0.50-0.65)
Gastrointestinal Stimulants	0.54 (0.49-0.59)	0.61 (0.56-0.66)	**0.51 (0.46-0.56)**

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Serotonin-Norepinephrine Reuptake Inhibitors (S...	0.65 (0.57-0.73)	0.60 (0.53-0.68)	0.56 (0.48-0.65)
Antidiarrheal/Probiotic Agents - Misc.	0.60 (0.52-0.66)	0.60 (0.53-0.68)	0.57 (0.51-0.65)
Laxative Combinations	0.59 (0.54-0.65)	0.60 (0.55-0.65)	0.56 (0.51-0.62)
Diagnostic Drugs	0.60 (0.56-0.66)	0.60 (0.56-0.65)	0.59 (0.55-0.63)
Urinary Anti-infectives	0.61 (0.54-0.68)	0.60 (0.53-0.68)	0.56 (0.48-0.63)
Misc. Nutritional Substances	0.56 (0.51-0.61)	0.60 (0.54-0.65)	0.55 (0.51-0.59)
Expectorants	0.53 (0.46-0.60)	0.60 (0.52-0.67)	0.39 (0.33-0.48)
Posterior Pituitary Hormones	0.52 (0.46-0.58)	**0.59 (0.52-0.66)**	**0.41 (0.35-0.47)**
Bronchodilators - Anticholinergics	0.60 (0.55-0.65)	0.59 (0.54-0.65)	0.57 (0.51-0.62)
Potassium Sparing Diuretics	0.61 (0.56-0.67)	0.59 (0.52-0.65)	0.59 (0.53-0.66)
Antihistamines-Topical	0.53 (0.45-0.62)	0.59 (0.50-0.66)	0.37 (0.29-0.43)
Water Soluble Vitamins	0.56 (0.52-0.61)	0.59 (0.55-0.63)	0.55 (0.49-0.60)
Antacids - Calcium Salts	0.56 (0.51-0.61)	0.59 (0.54-0.64)	0.54 (0.49-0.60)
Radiographic Contrast Media	0.60 (0.56-0.64)	0.59 (0.55-0.63)	0.60 (0.56-0.64)
Ophthalmic Local Anesthetics	0.66 (0.60-0.73)	0.59 (0.51-0.67)	0.60 (0.51-0.67)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Azithromycin	0.59 (0.54-0.64)	0.59 (0.53-0.63)	0.59 (0.53-0.63)
Beta Blockers Non-Selective	0.57 (0.48-0.66)	0.58 (0.50-0.67)	0.57 (0.48-0.66)
Antiadrenergic Antihypertensives	0.58 (0.51-0.65)	0.58 (0.52-0.66)	0.51 (0.44-0.57)
Ophthalmic Anti-infectives	0.60 (0.55-0.65)	0.58 (0.53-0.63)	0.54 (0.48-0.61)
Antianxiety Agents - Misc.	0.57 (0.50-0.65)	0.58 (0.49-0.66)	0.51 (0.44-0.59)
Opioid Agonists	0.58 (0.53-0.62)	0.58 (0.54-0.62)	0.56 (0.51-0.61)
Antiarrhythmics Type I-B	0.55 (0.51-0.58)	0.58 (0.54-0.61)	0.53 (0.49-0.56)
Antiemetics - Anticholinergic	0.55 (0.47-0.62)	0.58 (0.50-0.64)	0.50 (0.43-0.58)
Anesthetics - Misc.	0.51 (0.47-0.55)	0.58 (0.54-0.61)	0.47 (0.43-0.50)
Herpes Agents	0.51 (0.44-0.58)	0.58 (0.51-0.63)	0.50 (0.44-0.57)
Cephalosporins - 1st Generation	0.55 (0.51-0.58)	0.58 (0.54-0.62)	0.52 (0.48-0.56)
Serotonin Modulators	0.55 (0.49-0.61)	0.57 (0.52-0.63)	0.53 (0.48-0.58)
H-2 Antagonists	0.55 (0.51-0.59)	0.57 (0.54-0.62)	**0.49 (0.45-0.54)**
Artificial Tears and Lubricants	0.59 (0.52-0.65)	0.57 (0.49-0.64)	0.52 (0.46-0.58)
Local Anesthetics - Amides	0.57 (0.53-0.61)	0.57 (0.54-0.62)	0.55 (0.51-0.59)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Selective Serotonin Reuptake Inhibitors (SSRIs)	0.60 (0.56-0.66)	0.57 (0.52-0.62)	0.58 (0.54-0.63)
Oil Soluble Vitamins	0.58 (0.55-0.62)	0.57 (0.53-0.61)	0.59 (0.55-0.63)
Diagnostic Tests	0.62 (0.56-0.66)	0.57 (0.52-0.63)	**0.55 (0.49-0.60)**
Bulk Laxatives	0.55 (0.48-0.63)	0.57 (0.49-0.66)	0.44 (0.36-0.51)
Nondepolarizing Muscle Relaxants	0.57 (0.54-0.61)	0.57 (0.53-0.61)	0.56 (0.52-0.60)
Direct Factor Xa Inhibitors	0.60 (0.54-0.66)	0.57 (0.50-0.62)	0.56 (0.49-0.62)
Aminoglycosides	0.53 (0.45-0.61)	0.57 (0.50-0.64)	0.44 (0.36-0.52)
Antispasmodics	0.57 (0.52-0.61)	0.56 (0.52-0.61)	0.57 (0.53-0.61)
B-Complex Vitamins	0.49 (0.41-0.57)	0.56 (0.48-0.65)	0.44 (0.34-0.52)
Folic Acid/Folates	0.61 (0.56-0.67)	0.56 (0.51-0.62)	0.60 (0.54-0.65)
Protamine	0.51 (0.43-0.59)	0.56 (0.49-0.64)	0.46 (0.39-0.55)
Genitourinary Irrigants	0.46 (0.40-0.52)	0.56 (0.50-0.61)	0.49 (0.44-0.54)
Anesthetics Topical Oral	0.60 (0.54-0.66)	0.56 (0.50-0.63)	0.58 (0.53-0.65)
Non-Barbiturate Hypnotics	0.55 (0.51-0.59)	0.56 (0.52-0.60)	0.53 (0.49-0.58)
Hemostatics - Topical	0.57 (0.51-0.65)	0.56 (0.49-0.63)	0.53 (0.46-0.59)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Antidotes and Specific Antagonists	0.44 (0.38-0.49)	0.56 (0.50-0.61)	0.61 (0.55-0.67)
Influenza Agents	0.49 (0.40-0.57)	0.56 (0.47-0.64)	0.43 (0.36-0.51)
Multiple Vitamins w/ Minerals	0.59 (0.53-0.66)	0.55 (0.48-0.62)	0.56 (0.49-0.63)
Antihistamines - Non-Sedating	0.57 (0.51-0.62)	0.55 (0.49-0.60)	0.51 (0.44-0.57)
Local Anesthetics - Topical	0.56 (0.52-0.60)	0.55 (0.51-0.59)	0.56 (0.52-0.61)
Multivitamins	0.51 (0.47-0.55)	0.55 (0.51-0.59)	0.51 (0.47-0.55)
Antimyasthenic/Cholinergic Agents	0.55 (0.50-0.61)	0.55 (0.50-0.60)	0.54 (0.48-0.59)
Viral Vaccines	0.58 (0.52-0.63)	0.55 (0.50-0.59)	0.56 (0.50-0.61)
Aminopenicillins	0.53 (0.48-0.58)	0.55 (0.48-0.60)	0.54 (0.49-0.61)
Anti-inflammatory Agents - Topical	0.59 (0.52-0.64)	0.55 (0.49-0.61)	0.57 (0.50-0.63)
Acne Products	0.63 (0.54-0.74)	0.55 (0.45-0.64)	0.55 (0.45-0.65)
Urinary Antispasmodic - Antimuscarinics (Antich...	0.59 (0.51-0.66)	0.54 (0.47-0.61)	0.57 (0.49-0.64)
Sodium	0.51 (0.45-0.57)	0.53 (0.48-0.58)	**0.47 (0.42-0.52)**
Nasal Steroids	0.54 (0.49-0.59)	0.53 (0.48-0.57)	0.50 (0.45-0.55)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Opioid Combinations	0.54 (0.51-0.58)	0.53 (0.49-0.57)	0.52 (0.49-0.56)
ACE Inhibitors	0.51 (0.47-0.56)	0.52 (0.48-0.57)	**0.46 (0.42-0.50)**
Anticonvulsants - Misc.	0.54 (0.50-0.58)	0.52 (0.48-0.56)	0.51 (0.46-0.56)
Cephalosporins - 2nd Generation	0.49 (0.42-0.56)	0.52 (0.44-0.60)	0.47 (0.41-0.54)
Leukotriene Modulators	0.54 (0.46-0.61)	0.52 (0.43-0.61)	0.45 (0.37-0.53)
Hemostatics - Systemic	0.51 (0.44-0.59)	0.51 (0.43-0.59)	0.43 (0.35-0.51)
Cough/Cold/Allergy Combinations	0.57 (0.51-0.63)	**0.51 (0.44-0.57)**	0.55 (0.49-0.61)
Nonsteroidal Anti-inflammatory Agents (NSAIDs)	0.53 (0.49-0.57)	0.51 (0.47-0.55)	0.52 (0.48-0.56)
Corticosteroids - Topical	0.52 (0.47-0.56)	**0.50 (0.45-0.55)**	0.48 (0.43-0.52)
Toxoid Combinations	0.53 (0.45-0.61)	0.49 (0.43-0.56)	**0.41 (0.34-0.47)**
Depolarizing Muscle Relaxants	0.54 (0.48-0.60)	0.49 (0.43-0.55)	0.51 (0.47-0.56)
Miscellaneous Contrast Media	0.50 (0.47-0.54)	0.49 (0.45-0.53)	0.53 (0.49-0.58)
Central Muscle Relaxants	0.51 (0.46-0.55)	0.49 (0.44-0.53)	0.48 (0.43-0.53)
Antifungals - Topical	0.53 (0.48-0.59)	**0.46 (0.42-0.53)**	0.53 (0.48-0.58)

Continued on next page

Table A.2: Mean (95% confidence interval) area under the ROC curve for predicting medication usage, grouped by pharmaceutical subclass, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Pharmaceutical Subclass	Baseline	Methylation	Genotypes
Lincosamides	0.49 (0.43-0.56)	0.46 (0.40-0.53)	0.47 (0.41-0.54)
Dipeptidyl Peptidase-4 (DPP-4) Inhibitors	0.56 (0.45-0.65)	0.45 (0.36-0.53)	0.56 (0.48-0.64)
Alternative Medicine - C's	0.49 (0.42-0.55)	**0.44 (0.37-0.53)**	0.50 (0.43-0.58)
Tricyclic Agents	0.58 (0.51-0.66)	0.43 (0.36-0.50)	0.50 (0.43-0.58)
Cardiac Glycosides	0.53 (0.44-0.61)	0.40 (0.32-0.48)	0.53 (0.44-0.64)

Table A.3: Mean (95% confidence interval) R^2 for predicting the most recent lab result using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Lab Test	Baseline	Methylation	Genotypes
Troponin.	0.64 (0.52-0.73)	0.62 (0.49-0.74)	0.64 (0.54-0.74)
Troponin interpretation	0.47 (0.30-0.61)	0.47 (0.30-0.63)	0.38 (0.26-0.52)
Urea.nitrogen	0.02 (-0.03-0.06)	**0.46 (0.41-0.51)**	0.03 (-0.00-0.05)
Creatinine	0.07 (0.01-0.13)	**0.46 (0.42-0.50)**	0.08 (0.03-0.12)

Continued on next page

Table A.3: Mean (95% confidence interval) R^2 for predicting the most recent lab result using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Lab Test	Baseline	Methylation	Genotypes
Absolute eos count	-0.07 (-0.12–0.03)	**0.35 (0.27-0.42)**	0.00 (-0.02-0.02)
Hemoglobin	0.12 (0.06-0.18)	**0.30 (0.26-0.35)**	0.12 (0.07-0.17)
Neutrophil percent auto	0.23 (0.15-0.32)	**0.28 (0.21-0.35)**	0.24 (0.15-0.31)
Pft fev_1 pre	0.13 (-0.13-0.35)	0.26 (0.10-0.36)	**0.16 (-0.04-0.30)**
Hematocrit	0.07 (0.01-0.12)	**0.25 (0.20-0.29)**	**0.07 (0.02-0.11)**
Mean corpuscular hemoglobin	0.09 (0.03-0.15)	**0.22 (0.15-0.28)**	0.09 (0.05-0.15)
Mean corpuscular volume	0.09 (0.01-0.15)	**0.20 (0.13-0.26)**	0.08 (0.04-0.13)
Absolute lymphocyte count	0.07 (-0.01-0.18)	**0.18 (0.09-0.34)**	0.09 (0.03-0.22)
Absolute neut count	0.07 (0.00-0.12)	**0.17 (0.12-0.23)**	0.07 (0.03-0.12)
Platelet count auto	0.01 (-0.03-0.06)	**0.17 (0.13-0.20)**	0.02 (-0.00-0.05)
Chloride	0.01 (-0.03-0.04)	**0.15 (0.10-0.20)**	0.02 (-0.01-0.04)
White blood cell count	-0.01 (-0.07-0.06)	**0.14 (0.07-0.20)**	0.02 (-0.02-0.05)
Albumin	0.07 (0.02-0.14)	**0.14 (0.08-0.19)**	0.09 (0.04-0.13)

Continued on next page

Table A.3: Mean (95% confidence interval) R^2 for predicting the most recent lab result using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Lab Test	Baseline	Methylation	Genotypes
Absolute mono count	0.09 (0.02-0.15)	**0.13 (0.07-0.19)**	**0.08 (0.03-0.13)**
Neutrophils abs prelim.	0.05 (-0.01-0.11)	**0.13 (0.06-0.19)**	0.06 (0.01-0.09)
Sodium	-0.00 (-0.04-0.03)	**0.12 (0.08-0.16)**	0.02 (-0.01-0.03)
Hgb a1c hplc	-0.04 (-0.11-0.04)	**0.11 (0.06-0.17)**	-0.00 (-0.03-0.03)
Ferritin	-0.13 (-0.39-0.01)	**0.10 (0.03-0.15)**	0.00 (-0.03-0.03)
Total.protein	0.08 (0.01-0.14)	0.10 (0.04-0.15)	**0.08 (0.03-0.12)**
Absolute immature gran count	-0.05 (-0.24-0.04)	**0.10 (0.03-0.15)**	**0.00 (-0.03-0.03)**
Hematocrit osl	-0.37 (-0.99-0.06)	0.10 (-0.01-0.20)	-0.00 (-0.12-0.10)
Sedimentation rate erythrocyte	-0.25 (-0.70-0.06)	0.08 (-0.01-0.18)	0.04 (-0.05-0.10)
Absolute baso count	-0.08 (-0.22–0.01)	**0.08 (0.04-0.13)**	**-0.01 (-0.03–0.01)**
Iron binding capacity	-0.14 (-0.26–0.03)	**0.08 (0.01-0.14)**	-0.00 (-0.03-0.02)
Glucose	-0.02 (-0.07-0.02)	**0.06 (0.03-0.09)**	0.01 (-0.00-0.03)
Qrs duration	-0.02 (-0.10-0.05)	**0.05 (0.02-0.08)**	**0.01 (-0.02-0.03)**

Continued on next page

Table A.3: Mean (95% confidence interval) R^2 for predicting the most recent lab result using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Lab Test	Baseline	Methylation	Genotypes
Magnesium	-0.14 (-0.22–0.09)	0.05 (-0.00-0.10)	0.02 (-0.02-0.06)
Cholesterol hdl	-0.00 (-0.13-0.10)	**0.04 (-0.02-0.10)**	0.05 (-0.01-0.11)
Anion gap	-0.03 (-0.07-0.00)	**0.03 (0.00-0.06)**	-0.01 (-0.02-0.00)
Alanine aminotransferase	-0.03 (-0.07-0.01)	**0.03 (0.01-0.05)**	-0.00 (-0.02-0.01)
Ventricular rate	-0.06 (-0.14-0.01)	**0.03 (-0.00-0.06)**	0.01 (-0.02-0.04)
Cholesterol	-0.03 (-0.09-0.04)	0.03 (-0.03-0.08)	0.01 (-0.03-0.04)
Alkaline phosphatase	-0.01 (-0.06-0.03)	0.02 (-0.00-0.04)	0.01 (-0.01-0.02)
R axis	-0.01 (-0.08-0.05)	0.02 (-0.02-0.05)	0.02 (-0.02-0.05)
Aspartate aminotransferase	-0.03 (-0.10-0.02)	0.02 (0.00-0.03)	-0.00 (-0.02-0.01)
C reactive protein	-0.31 (-0.92–0.06)	0.02 (-0.05-0.08)	**-0.00 (-0.07-0.05)**
Qtc calculation bezet.	-0.06 (-0.13–0.01)	0.02 (-0.01-0.03)	-0.01 (-0.02-0.00)
Pft fev_1 pre percent ref	-0.34 (-0.59–0.14)	0.01 (-0.04-0.06)	**-0.04 (-0.08–0.02)**
Bnp	-0.16 (-0.56-0.06)	0.01 (-0.07-0.09)	0.03 (-0.07-0.11)
T axis	-0.08 (-0.19–0.01)	0.01 (-0.02-0.04)	-0.00 (-0.02-0.01)

Continued on next page

Table A.3: Mean (95% confidence interval) R^2 for predicting the most recent lab result using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Lab Test	Baseline	Methylation	Genotypes
Potassium	-0.02 (-0.05-0.01)	0.01 (-0.01-0.03)	0.01 (-0.01-0.02)
Cholesterol ldl calculated	-0.07 (-0.18-0.00)	0.01 (-0.03-0.05)	-0.01 (-0.04-0.01)
Glucose poc	-0.16 (-0.40-0.00)	0.01 (-0.02-0.04)	0.00 (-0.03-0.03)
Inr	-0.02 (-0.07-0.02)	0.01 (-0.01-0.02)	-0.00 (-0.02-0.01)
X saturation	-0.17 (-0.34–0.05)	0.01 (-0.03-0.04)	**-0.02 (-0.04–0.01)**
Prothrombin time	-0.04 (-0.10-0.01)	0.00 (-0.02-0.02)	0.00 (-0.01-0.02)
P axis	-0.04 (-0.10-0.01)	0.00 (-0.02-0.02)	0.01 (-0.02-0.03)
Q t interval	-0.07 (-0.15–0.01)	0.00 (-0.02-0.02)	0.00 (-0.01-0.02)
Bilirubin total	-0.04 (-0.15–0.01)	0.00 (-0.03-0.01)	-0.00 (-0.02-0.00)
Atrial rate	-0.02 (-0.09-0.03)	-0.00 (-0.03-0.02)	0.01 (-0.02-0.04)
P r interval	-0.02 (-0.10-0.06)	-0.00 (-0.04-0.03)	-0.00 (-0.04-0.03)
Aptt	-0.07 (-0.13–0.04)	-0.00 (-0.03-0.01)	-0.01 (-0.03–0.00)
Triglycerides	-0.16 (-0.25–0.07)	-0.00 (-0.03-0.02)	-0.01 (-0.03-0.01)
Tsh	-0.20 (-0.60–0.07)	-0.01 (-0.04-0.01)	**-0.01 (-0.05-0.00)**
Calcium	-0.04 (-0.08-0.01)	-0.01 (-0.03-0.00)	-0.01 (-0.02-0.01)
Glucose whole blood	-0.20 (-0.42–0.04)	-0.02 (-0.06-0.03)	-0.02 (-0.05–0.01)

Continued on next page

Table A.3: Mean (95% confidence interval) R^2 for predicting the most recent lab result using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Lab Test	Baseline	Methylation	Genotypes
Bilirubin conjugated	-0.52 (-1.18–0.27)	-0.02 (-0.09-0.03)	-0.02 (-0.16-0.06)
Iron	-0.13 (-0.28–0.02)	**-0.03 (-0.07-0.02)**	-0.01 (-0.04-0.01)
Urea nitrogen osl.	-0.60 (-1.75–0.07)	-0.03 (-0.12-0.03)	-0.02 (-0.09-0.03)
Blood lactate	-0.45 (-0.82–0.21)	-0.03 (-0.07–0.01)	-0.03 (-0.08-0.01)
Left ventricular ejection fraction	-2.95 (-7.20–0.84)	-0.10 (-0.31-0.04)	-0.04 (-0.29-0.08)
Hemoglobin osl.	-2.13 (-11.00–0.18)	**-0.10 (-0.46–0.01)**	-0.08 (-0.52-0.05)
Chloride osl.	-1.13 (-2.43–0.30)	-0.12 (-0.36-0.01)	-0.07 (-0.22-0.02)
Sodium osl.	-466.23 (-4036.73–0.08)	-2.12 (-17.00–0.02)	-2.40 (-17.82–0.02)
Potassium osl.	-160.57 (-921.21–0.12)	-2.45 (-7.77–0.02)	-2.53 (-7.72–0.02)

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
End stage renal disease	0.73 (0.69-0.77)	**0.91 (0.88-0.93)**	0.73 (0.69-0.78)
Renal dialysis	0.75 (0.71-0.79)	**0.89 (0.86-0.92)**	0.74 (0.69-0.79)
Morbid obesity	0.88 (0.82-0.93)	0.88 (0.83-0.93)	0.84 (0.75-0.91)
Cirrhosis of liver without mention of alcohol	0.71 (0.65-0.77)	**0.85 (0.81-0.90)**	0.66 (0.60-0.73)
Hyperplasia of prostate	0.86 (0.83-0.89)	0.85 (0.81-0.88)	0.84 (0.79-0.88)
Kidney replaced by transpant	0.75 (0.69-0.80)	**0.84 (0.80-0.87)**	0.72 (0.67-0.78)
Neutropenia	0.64 (0.56-0.72)	**0.83 (0.77-0.89)**	0.59 (0.52-0.67)
Tobacco use disorder	0.58 (0.49-0.67)	**0.83 (0.77-0.88)**	0.53 (0.45-0.62)
Obesity	0.84 (0.80-0.88)	0.83 (0.79-0.87)	0.83 (0.78-0.86)
Poisoning by primarily systemic agents	0.63 (0.55-0.72)	**0.83 (0.77-0.88)**	0.56 (0.47-0.64)
Disorders resulting from impaired renal function	0.69 (0.63-0.75)	**0.82 (0.77-0.86)**	0.65 (0.58-0.71)
Portal hypertension	0.71 (0.63-0.78)	**0.82 (0.76-0.87)**	0.66 (0.57-0.74)
Chronic renal failure [CKD]	0.64 (0.60-0.69)	**0.82 (0.79-0.85)**	0.64 (0.59-0.67)
Immunity deficiency	0.77 (0.73-0.82)	**0.82 (0.77-0.87)**	0.75 (0.70-0.81)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Noninflammatory female genital disorders	0.81 (0.77-0.85)	0.82 (0.78-0.85)	0.79 (0.74-0.83)
Menopausal and postmenopausal disorders	0.83 (0.78-0.87)	0.82 (0.77-0.85)	0.78 (0.72-0.85)
Hypertensive chronic kidney disease	0.61 (0.57-0.66)	**0.81 (0.78-0.85)**	0.59 (0.56-0.64)
Liver replaced by transplant	0.70 (0.61-0.77)	**0.81 (0.75-0.86)**	0.60 (0.51-0.69)
Renal failure	0.61 (0.58-0.65)	**0.81 (0.78-0.84)**	0.60 (0.57-0.64)
Decreased white blood cell count	0.67 (0.61-0.73)	**0.80 (0.75-0.86)**	0.62 (0.54-0.69)
Antineoplastic and immunosuppressive drugs caus...	0.64 (0.54-0.73)	**0.80 (0.73-0.88)**	0.60 (0.50-0.70)
Cancer of prostate	0.85 (0.80-0.89)	0.80 (0.72-0.86)	0.80 (0.74-0.86)
Disorders involving the immune mechanism	0.76 (0.71-0.80)	**0.80 (0.75-0.85)**	0.74 (0.69-0.79)
Anemia in chronic kidney disease	0.69 (0.63-0.75)	**0.80 (0.74-0.84)**	0.67 (0.61-0.72)
Erectile dysfunction [ED]	0.74 (0.69-0.79)	0.79 (0.73-0.85)	0.65 (0.56-0.73)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Family history	0.68 (0.61-0.74)	**0.79 (0.72-0.85)**	0.62 (0.52-0.72)
Anemia of chronic disease	0.70 (0.65-0.74)	**0.79 (0.75-0.82)**	0.69 (0.65-0.73)
Disorders of phosphorus metabolism	0.73 (0.66-0.80)	0.78 (0.73-0.84)	0.68 (0.59-0.76)
Overweight, obesity and other hyperalimentation	0.81 (0.77-0.85)	0.78 (0.73-0.82)	0.81 (0.76-0.85)
Osteoarthrosis NOS	0.78 (0.74-0.82)	0.77 (0.73-0.81)	0.76 (0.70-0.80)
Osteoarthrosis	0.78 (0.74-0.82)	0.77 (0.72-0.81)	0.77 (0.73-0.81)
Liver abscess and sequelae of chronic liver dis...	0.74 (0.68-0.81)	**0.77 (0.70-0.84)**	0.66 (0.57-0.75)
Osteoarthrosis, localized, primary	0.76 (0.71-0.82)	**0.76 (0.70-0.81)**	0.74 (0.67-0.80)
Fluid overload	0.67 (0.61-0.74)	**0.76 (0.71-0.80)**	0.64 (0.56-0.72)
Secondary hyperparathyroidism (of renal origin)	0.69 (0.62-0.75)	**0.76 (0.71-0.81)**	0.66 (0.59-0.72)
Respiratory failure, insufficiency, arrest	0.61 (0.55-0.67)	**0.75 (0.69-0.82)**	0.55 (0.47-0.63)
Splenomegaly	0.66 (0.58-0.73)	**0.75 (0.69-0.81)**	0.70 (0.63-0.77)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Ascites (non malignant)	0.74 (0.67-0.80)	0.75 (0.67-0.80)	0.63 (0.54-0.71)
Osteoarthritis; localized	0.77 (0.71-0.83)	0.75 (0.69-0.80)	0.77 (0.71-0.82)
Poisoning by hormones and synthetic substitutes	0.62 (0.53-0.70)	**0.74 (0.66-0.81)**	0.57 (0.49-0.65)
Hypertensive heart and/or renal disease	0.55 (0.50-0.60)	**0.74 (0.70-0.78)**	0.56 (0.51-0.61)
Altered mental status	0.64 (0.56-0.72)	**0.74 (0.67-0.82)**	0.59 (0.50-0.68)
Atherosclerosis	0.48 (0.41-0.55)	**0.73 (0.68-0.79)**	0.51 (0.44-0.58)
Nephritis and nephropathy in diseases classifie...	0.70 (0.62-0.76)	**0.73 (0.66-0.79)**	0.68 (0.61-0.75)
Thrombocytopenia	0.65 (0.59-0.71)	**0.73 (0.67-0.79)**	0.63 (0.57-0.69)
Arthropathy NOS	0.73 (0.66-0.80)	0.72 (0.65-0.77)	0.70 (0.64-0.76)
Degenerative skin conditions and other dermatoses	0.72 (0.66-0.76)	0.72 (0.67-0.76)	0.71 (0.66-0.75)
Cataract	0.73 (0.68-0.77)	0.71 (0.67-0.75)	0.72 (0.67-0.76)
Other and unspecified coagulation defects	0.53 (0.45-0.61)	**0.71 (0.62-0.79)**	0.50 (0.43-0.58)
Effects radiation NOS	0.64 (0.55-0.72)	0.71 (0.63-0.79)	0.57 (0.48-0.66)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Other disorders of the kidney and ureters	0.55 (0.50-0.60)	**0.71 (0.67-0.75)**	0.52 (0.47-0.57)
Emphysema	0.50 (0.43-0.58)	**0.71 (0.63-0.79)**	0.46 (0.38-0.53)
Spinal stenosis of lumbar region	0.68 (0.61-0.75)	0.71 (0.65-0.77)	0.57 (0.48-0.66)
Diverticulosis	0.74 (0.65-0.82)	0.71 (0.63-0.80)	0.66 (0.57-0.74)
Acid-base balance disorder	0.67 (0.62-0.73)	0.71 (0.65-0.76)	0.67 (0.61-0.72)
Complications of transplants and reattached limbs	0.68 (0.60-0.75)	0.71 (0.63-0.78)	0.62 (0.52-0.72)
Postinflammatory pulmonary fibrosis	0.69 (0.62-0.76)	0.71 (0.60-0.78)	0.59 (0.51-0.67)
Purpura and other hemorrhagic conditions	0.62 (0.55-0.69)	**0.71 (0.65-0.76)**	0.57 (0.50-0.64)
Hyperlipidemia	0.71 (0.67-0.74)	0.70 (0.67-0.74)	0.70 (0.67-0.74)
Coagulation defects	0.71 (0.65-0.77)	0.70 (0.64-0.78)	0.67 (0.60-0.74)
Acute renal failure	0.63 (0.59-0.67)	**0.70 (0.66-0.74)**	0.61 (0.57-0.66)
Ischemic Heart Disease	0.69 (0.65-0.72)	0.70 (0.67-0.73)	0.68 (0.64-0.72)
Hyperpotassemia	0.68 (0.63-0.72)	**0.70 (0.65-0.75)**	0.66 (0.61-0.71)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Other non-epithelial cancer of skin	0.72 (0.67-0.78)	0.70 (0.64-0.76)	0.68 (0.63-0.73)
Osteoporosis	0.72 (0.67-0.78)	0.70 (0.63-0.76)	0.66 (0.58-0.73)
Diverticulosis and diverticulitis	0.72 (0.66-0.78)	0.69 (0.63-0.76)	0.68 (0.60-0.75)
Actinic keratosis	0.72 (0.67-0.78)	**0.69 (0.64-0.75)**	**0.72 (0.67-0.76)**
Iron deficiency anemia secondary to blood loss ...	0.55 (0.47-0.63)	**0.69 (0.62-0.77)**	0.47 (0.40-0.55)
Type 2 diabetes	0.67 (0.63-0.71)	**0.69 (0.64-0.73)**	0.66 (0.62-0.70)
Secondary diabetes mellitus	0.64 (0.55-0.71)	0.69 (0.62-0.76)	0.64 (0.53-0.72)
Abnormal involuntary movements	0.40 (0.31-0.49)	**0.69 (0.61-0.77)**	0.56 (0.46-0.67)
Essential hypertension	0.67 (0.62-0.70)	0.69 (0.64-0.72)	0.67 (0.63-0.71)
Pneumonia	0.65 (0.59-0.70)	0.69 (0.63-0.74)	0.64 (0.59-0.70)
Viral infection	0.70 (0.63-0.76)	0.68 (0.61-0.76)	0.67 (0.58-0.75)
Respiratory failure	0.67 (0.61-0.74)	0.68 (0.61-0.75)	0.57 (0.50-0.65)
Cardiac pacemaker/device in situ	0.40 (0.32-0.48)	0.68 (0.61-0.75)	0.61 (0.53-0.69)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Other arthropathies	0.72 (0.66-0.78)	0.68 (0.59-0.77)	0.69 (0.61-0.76)
Other immunological findings	0.68 (0.60-0.76)	0.68 (0.60-0.76)	0.62 (0.52-0.71)
Coronary atherosclerosis	0.68 (0.65-0.72)	0.68 (0.65-0.72)	0.67 (0.64-0.71)
Sepsis	0.64 (0.59-0.70)	0.68 (0.62-0.73)	0.61 (0.54-0.66)
Acidosis	0.66 (0.61-0.72)	0.68 (0.63-0.73)	0.66 (0.61-0.73)
Skin cancer	0.73 (0.68-0.77)	0.68 (0.61-0.75)	0.70 (0.63-0.77)
Other venous embolism and thrombosis	0.59 (0.54-0.65)	**0.68 (0.63-0.74)**	0.57 (0.50-0.63)
Sensorineural hearing loss	0.67 (0.59-0.75)	0.68 (0.60-0.75)	0.62 (0.53-0.71)
Substance addiction and disorders	0.58 (0.49-0.66)	0.68 (0.59-0.77)	0.43 (0.35-0.51)
Nephritis; nephrosis; renal sclerosis	0.54 (0.47-0.61)	**0.68 (0.63-0.73)**	0.58 (0.52-0.64)
Protein-calorie malnutrition	0.66 (0.58-0.72)	0.68 (0.60-0.75)	0.61 (0.54-0.68)
Type 2 diabetes with renal manifestations	0.64 (0.57-0.69)	0.68 (0.61-0.74)	0.67 (0.61-0.71)
Osteoporosis NOS	0.71 (0.64-0.78)	0.68 (0.59-0.76)	0.64 (0.56-0.72)
Enthesopathy	0.67 (0.60-0.73)	0.68 (0.59-0.75)	0.60 (0.53-0.69)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Disorders of fluid, electrolyte, and acid-base ...	0.60 (0.56-0.65)	**0.67 (0.63-0.72)**	0.58 (0.54-0.63)
Hypertension	0.64 (0.60-0.68)	**0.67 (0.64-0.71)**	0.64 (0.59-0.68)
Diabetes mellitus	0.65 (0.61-0.70)	**0.67 (0.62-0.72)**	0.65 (0.60-0.69)
Myocardial infarction	0.64 (0.59-0.70)	0.67 (0.63-0.73)	0.58 (0.51-0.64)
Hemorrhage of gastrointestinal tract	0.65 (0.56-0.74)	0.67 (0.60-0.75)	0.61 (0.53-0.69)
Acute pain	0.57 (0.51-0.63)	**0.67 (0.62-0.73)**	0.60 (0.55-0.66)
Seborrheic keratosis	0.70 (0.64-0.75)	0.67 (0.60-0.74)	0.68 (0.62-0.74)
Septicemia	0.61 (0.55-0.67)	**0.67 (0.62-0.72)**	0.60 (0.55-0.67)
Disorders of lipoid metabolism	0.68 (0.64-0.71)	0.67 (0.63-0.71)	0.68 (0.64-0.71)
Carditis	0.59 (0.50-0.66)	0.67 (0.60-0.74)	0.45 (0.36-0.53)
Sleep apnea	0.66 (0.60-0.71)	0.67 (0.60-0.73)	0.67 (0.61-0.73)
Other anemias	0.61 (0.57-0.65)	**0.67 (0.63-0.71)**	0.61 (0.57-0.65)
Other forms of chronic heart disease	0.53 (0.45-0.61)	**0.67 (0.58-0.75)**	0.50 (0.43-0.58)
Disorders of diaphragm	0.70 (0.63-0.76)	0.67 (0.59-0.74)	0.67 (0.59-0.76)
Renal failure NOS	0.60 (0.52-0.68)	**0.67 (0.59-0.74)**	0.58 (0.51-0.67)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Disorders of vitreous body	0.72 (0.66-0.79)	0.67 (0.60-0.73)	0.70 (0.63-0.76)
Chronic liver disease and cirrhosis	0.62 (0.55-0.70)	0.67 (0.60-0.73)	0.61 (0.54-0.67)
Aortic valve disease	0.68 (0.63-0.72)	0.67 (0.61-0.72)	**0.65 (0.60-0.69)**
Hyposmolality and/or hyponatremia	0.63 (0.58-0.68)	0.66 (0.62-0.71)	0.59 (0.54-0.65)
Angina pectoris	0.66 (0.60-0.72)	0.66 (0.60-0.72)	0.60 (0.54-0.66)
Chronic Kidney Disease, Stage IV	0.57 (0.50-0.64)	0.66 (0.59-0.72)	0.54 (0.46-0.60)
Disorders of calcium/phosphorus metabolism	0.63 (0.57-0.70)	0.66 (0.59-0.74)	0.61 (0.54-0.68)
Other disorders of metabolism	0.58 (0.49-0.67)	**0.66 (0.58-0.74)**	0.55 (0.46-0.63)
Cardiac shunt/ heart septal defect	0.61 (0.53-0.68)	0.66 (0.57-0.73)	0.50 (0.39-0.62)
Nonrheumatic aortic valve disorders	0.68 (0.63-0.73)	0.66 (0.61-0.72)	0.66 (0.61-0.71)
Acute posthemorrhagic anemia	0.67 (0.62-0.72)	0.66 (0.59-0.72)	0.65 (0.60-0.71)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Fever of unknown origin	0.62 (0.56-0.69)	0.66 (0.60-0.72)	0.62 (0.55-0.68)
Sepsis and SIRS	0.63 (0.58-0.68)	0.66 (0.60-0.71)	0.61 (0.55-0.67)
Articular cartilage disorder	0.65 (0.58-0.71)	0.66 (0.57-0.72)	0.67 (0.61-0.73)
Cardiomegaly	0.52 (0.47-0.57)	**0.66 (0.61-0.70)**	0.47 (0.43-0.52)
Pain in joint	0.63 (0.59-0.67)	**0.66 (0.62-0.69)**	0.61 (0.57-0.66)
Bacteremia	0.60 (0.51-0.69)	0.65 (0.57-0.73)	0.59 (0.50-0.68)
Polyneuropathy in diabetes	0.61 (0.53-0.70)	0.65 (0.57-0.73)	0.56 (0.46-0.65)
Deep vein thrombosis [DVT]	0.66 (0.59-0.73)	0.65 (0.57-0.72)	0.63 (0.54-0.71)
Iron deficiency anemias, unspecified or not due...	0.68 (0.62-0.73)	0.65 (0.58-0.70)	0.66 (0.60-0.72)
Tachycardia NOS	0.65 (0.58-0.72)	0.65 (0.58-0.71)	0.62 (0.54-0.70)
Chronic sinusitis	0.38 (0.32-0.45)	**0.65 (0.59-0.72)**	0.62 (0.55-0.68)
Chronic airway obstruction	0.57 (0.51-0.62)	**0.65 (0.59-0.71)**	0.50 (0.44-0.57)
Nausea and vomiting	0.54 (0.49-0.60)	**0.65 (0.60-0.70)**	0.51 (0.46-0.57)
Senile cataract	0.66 (0.62-0.72)	0.65 (0.60-0.71)	0.64 (0.58-0.70)
Inflammation of the eye	0.64 (0.56-0.71)	0.65 (0.56-0.73)	0.55 (0.45-0.64)
Bacterial infection NOS	0.61 (0.55-0.66)	0.65 (0.59-0.70)	0.59 (0.53-0.64)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Osteoporosis, osteopenia and pathological fracture	0.62 (0.57-0.68)	0.65 (0.60-0.69)	0.58 (0.53-0.64)
Electrolyte imbalance	0.62 (0.58-0.67)	0.65 (0.60-0.69)	0.62 (0.57-0.66)
Congenital anomalies of great vessels	0.66 (0.61-0.72)	0.65 (0.59-0.70)	0.65 (0.58-0.70)
Primary/intrinsic cardiomyopathies	0.55 (0.49-0.62)	**0.65 (0.58-0.70)**	0.53 (0.47-0.59)
Intestinal infection	0.68 (0.60-0.76)	0.64 (0.55-0.74)	0.60 (0.52-0.70)
Chronic Kidney Disease, Stage III	0.59 (0.53-0.63)	**0.64 (0.59-0.70)**	0.52 (0.46-0.58)
Heart valve disorders	0.64 (0.60-0.68)	0.64 (0.60-0.69)	0.61 (0.57-0.66)
Atherosclerosis of aorta	0.66 (0.61-0.71)	0.64 (0.58-0.70)	0.61 (0.56-0.66)
Hypercholesterolemia	0.64 (0.60-0.70)	0.64 (0.60-0.69)	0.61 (0.56-0.66)
Infection/inflammation of internal prosthetic d...	0.58 (0.48-0.66)	0.64 (0.55-0.73)	0.59 (0.52-0.67)
Congestive heart failure; nonhypertensive	0.57 (0.51-0.63)	**0.64 (0.59-0.70)**	0.53 (0.46-0.59)
Voice disturbance	0.53 (0.44-0.62)	0.64 (0.55-0.73)	0.39 (0.32-0.46)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Abnormality of gait	0.70 (0.62-0.78)	0.64 (0.57-0.71)	0.63 (0.53-0.72)
Obstructive sleep apnea	0.65 (0.58-0.71)	0.64 (0.58-0.71)	0.67 (0.60-0.72)
Shortness of breath	0.63 (0.58-0.68)	0.64 (0.59-0.68)	0.61 (0.56-0.66)
Alcoholism	0.64 (0.53-0.76)	0.64 (0.54-0.74)	**0.63 (0.53-0.74)**
Nephritis and nephropathy without mention of gl...	0.55 (0.48-0.64)	**0.64 (0.57-0.71)**	0.58 (0.52-0.64)
Type 2 diabetes with neurological manifestations	0.66 (0.59-0.75)	0.64 (0.56-0.72)	0.62 (0.52-0.70)
Other disorders of intestine	0.58 (0.48-0.65)	0.64 (0.55-0.72)	0.54 (0.44-0.64)
Cerebrovascular disease	0.58 (0.53-0.64)	0.64 (0.58-0.69)	0.53 (0.47-0.58)
Congestive heart failure (CHF) NOS	0.59 (0.52-0.66)	0.63 (0.56-0.70)	0.53 (0.48-0.60)
Other disorders of eye	0.69 (0.64-0.75)	0.63 (0.57-0.70)	0.68 (0.61-0.75)
Frequency of urination and polyuria	0.61 (0.54-0.68)	0.63 (0.58-0.69)	0.55 (0.48-0.62)
Disorders of mineral metabolism	0.63 (0.58-0.68)	0.63 (0.57-0.69)	0.62 (0.57-0.68)
Atrial fibrillation	0.66 (0.62-0.71)	0.63 (0.58-0.68)	0.64 (0.58-0.70)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Chronic pulmonary heart disease	0.55 (0.49-0.62)	0.63 (0.57-0.69)	0.57 (0.52-0.64)
Other chronic nonalcoholic liver disease	0.61 (0.55-0.67)	0.63 (0.57-0.68)	0.59 (0.53-0.65)
Peripheral vascular disease, unspecified	0.56 (0.50-0.62)	0.63 (0.57-0.70)	0.53 (0.46-0.60)
Cystitis and urethritis	0.63 (0.57-0.68)	0.63 (0.54-0.70)	0.55 (0.46-0.61)
Candidiasis	0.66 (0.58-0.73)	0.63 (0.55-0.69)	0.60 (0.53-0.67)
Other local infections of skin and subcutaneous...	0.60 (0.50-0.69)	0.63 (0.55-0.69)	0.57 (0.48-0.66)
Heart valve replaced	0.55 (0.49-0.62)	0.62 (0.55-0.69)	0.47 (0.41-0.54)
Urinary tract infection	0.64 (0.59-0.68)	0.62 (0.57-0.67)	0.62 (0.57-0.66)
Degeneration of intervertebral disc	0.63 (0.56-0.70)	0.62 (0.55-0.69)	0.57 (0.50-0.64)
Hypotension	0.60 (0.53-0.67)	0.62 (0.57-0.67)	0.55 (0.49-0.62)
Atrial fibrillation and flutter	0.65 (0.60-0.69)	0.62 (0.57-0.67)	0.63 (0.58-0.68)
Benign neoplasm of colon	0.61 (0.55-0.65)	0.62 (0.56-0.68)	0.60 (0.54-0.66)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Pulmonary collapse; interstitial and compensato...	0.59 (0.53-0.65)	0.62 (0.55-0.69)	0.56 (0.48-0.63)
Disorder of skin and subcutaneous tissue NOS	0.59 (0.53-0.65)	0.62 (0.55-0.67)	0.55 (0.50-0.62)
Osteopenia or other disorder of bone and cartilage	0.62 (0.56-0.67)	0.62 (0.57-0.68)	0.60 (0.55-0.65)
Bacterial pneumonia	0.65 (0.57-0.74)	0.62 (0.55-0.70)	0.66 (0.56-0.75)
Urinary incontinence	0.66 (0.59-0.73)	0.62 (0.53-0.69)	0.63 (0.56-0.70)
Other diseases of lung	0.55 (0.50-0.60)	0.62 (0.57-0.67)	0.51 (0.46-0.57)
Lymphadenitis	0.56 (0.49-0.62)	0.62 (0.53-0.69)	0.49 (0.43-0.56)
Abnormal movement	0.64 (0.57-0.70)	0.61 (0.55-0.69)	0.61 (0.54-0.69)
Symptoms concerning nutrition, metabolism, and ...	0.63 (0.58-0.70)	0.61 (0.54-0.68)	0.60 (0.54-0.66)
Insulin pump user	0.61 (0.54-0.68)	0.61 (0.56-0.68)	0.59 (0.54-0.66)
Mixed hyperlipidemia	0.64 (0.58-0.70)	0.61 (0.55-0.67)	0.61 (0.55-0.66)
Other chronic ischemic heart disease, unspecified	0.68 (0.60-0.75)	0.61 (0.53-0.69)	0.63 (0.57-0.72)
Hypothyroidism	0.57 (0.49-0.63)	0.61 (0.55-0.67)	0.53 (0.46-0.60)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Iron deficiency anemias	0.64 (0.59-0.70)	0.61 (0.55-0.67)	0.63 (0.57-0.70)
Other disorders of synovium, tendon, and bursa	0.63 (0.55-0.71)	0.61 (0.54-0.68)	0.67 (0.60-0.73)
Abnormal findings examination of lungs	0.50 (0.45-0.54)	0.61 (0.56-0.67)	0.44 (0.39-0.50)
Heart failure NOS	0.57 (0.52-0.61)	0.61 (0.56-0.66)	0.50 (0.45-0.55)
Cystitis	0.60 (0.52-0.67)	0.61 (0.53-0.68)	0.51 (0.43-0.61)
Shock	0.61 (0.54-0.68)	0.61 (0.52-0.70)	**0.45 (0.37-0.55)**
Edema	0.55 (0.50-0.62)	0.61 (0.55-0.66)	0.55 (0.50-0.60)
Cerebral ischemia	0.46 (0.38-0.55)	0.61 (0.52-0.67)	0.52 (0.44-0.59)
Other aneurysm	0.61 (0.53-0.69)	0.61 (0.53-0.68)	0.57 (0.50-0.64)
Other hypertensive complications	0.61 (0.54-0.67)	0.61 (0.54-0.67)	0.60 (0.53-0.66)
Pain	0.58 (0.53-0.63)	0.60 (0.57-0.65)	0.57 (0.51-0.61)
Proteinuria	0.53 (0.44-0.61)	0.60 (0.54-0.67)	0.54 (0.47-0.61)
Other diseases of blood and blood-forming organs	0.58 (0.47-0.68)	0.60 (0.51-0.70)	0.46 (0.35-0.58)
Hypertensive heart disease	0.53 (0.47-0.60)	0.60 (0.52-0.67)	**0.67 (0.59-0.73)**

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Varicose veins	0.49 (0.39-0.58)	0.60 (0.51-0.70)	0.42 (0.34-0.51)
Circulatory disease NEC	0.44 (0.38-0.50)	0.60 (0.53-0.66)	0.56 (0.51-0.63)
Pleurisy; pleural effusion	0.61 (0.56-0.65)	0.60 (0.55-0.64)	0.58 (0.54-0.63)
Rheumatic disease of the heart valves	0.44 (0.37-0.52)	0.60 (0.54-0.67)	0.52 (0.43-0.60)
Other disorders of liver	0.60 (0.53-0.67)	0.60 (0.54-0.67)	0.56 (0.50-0.62)
Other ill-defined and unknown causes of morbidi...	0.68 (0.61-0.75)	0.60 (0.52-0.69)	0.65 (0.57-0.72)
Pulmonary heart disease	0.54 (0.47-0.60)	0.60 (0.53-0.65)	0.55 (0.48-0.61)
Alcohol-related disorders	0.63 (0.54-0.74)	0.60 (0.49-0.69)	0.56 (0.45-0.68)
Spinal stenosis	0.65 (0.58-0.72)	0.60 (0.52-0.67)	0.53 (0.44-0.61)
Dizziness and giddiness (Lightheadedness and v...	0.64 (0.58-0.69)	0.60 (0.54-0.66)	0.62 (0.55-0.67)
Other alveolar and parietoalveolar pneumonopathy	0.56 (0.47-0.65)	0.60 (0.49-0.71)	0.49 (0.40-0.60)
Nonrheumatic mitral valve disorders	0.56 (0.50-0.62)	0.60 (0.53-0.66)	0.53 (0.47-0.60)
Migraine	0.55 (0.47-0.64)	0.59 (0.51-0.68)	0.41 (0.33-0.50)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Hemorrhoids	0.59 (0.52-0.64)	0.59 (0.51-0.66)	0.54 (0.45-0.62)
Spondylosis without myelopathy	0.60 (0.54-0.66)	0.59 (0.53-0.65)	0.53 (0.45-0.59)
Other tests	0.56 (0.48-0.62)	0.59 (0.53-0.67)	0.60 (0.51-0.66)
Other infectious and parasitic diseases	0.46 (0.37-0.54)	0.59 (0.51-0.67)	0.50 (0.41-0.60)
Diseases of the larynx and vocal cords	0.52 (0.45-0.60)	0.59 (0.51-0.68)	0.45 (0.35-0.54)
Hypovolemia	0.56 (0.46-0.65)	0.59 (0.49-0.68)	0.51 (0.43-0.60)
Aortic aneurysm	0.63 (0.56-0.71)	0.59 (0.51-0.67)	0.60 (0.54-0.68)
Nevus, non-neoplastic	0.58 (0.50-0.67)	0.59 (0.51-0.67)	0.47 (0.41-0.55)
Peripheral enthesopathies and allied syndromes	0.60 (0.53-0.67)	0.59 (0.52-0.65)	0.60 (0.54-0.66)
Neoplasm of uncertain behavior	0.56 (0.48-0.64)	0.59 (0.51-0.69)	0.43 (0.34-0.51)
GERD	0.57 (0.53-0.61)	0.59 (0.54-0.63)	0.56 (0.52-0.60)
Other peripheral nerve disorders	0.58 (0.51-0.65)	0.59 (0.52-0.66)	0.52 (0.45-0.61)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Mitral valve disease	0.55 (0.50-0.62)	0.59 (0.53-0.65)	0.54 (0.49-0.60)
Mood disorders	0.57 (0.52-0.63)	0.59 (0.53-0.63)	0.52 (0.46-0.58)
Diseases of esophagus	0.58 (0.54-0.63)	0.59 (0.54-0.63)	0.55 (0.50-0.60)
Cardiac pacemaker in situ	0.45 (0.37-0.52)	0.59 (0.50-0.66)	0.47 (0.39-0.56)
Glaucoma	0.63 (0.57-0.70)	0.58 (0.50-0.65)	0.60 (0.54-0.68)
Other symptoms/disorders or the urinary system	0.60 (0.55-0.65)	0.58 (0.53-0.63)	0.58 (0.53-0.62)
Diseases of sebaceous glands	0.45 (0.36-0.53)	0.58 (0.52-0.67)	0.63 (0.55-0.71)
Inflammatory and toxic neuropathy	0.49 (0.42-0.56)	0.58 (0.51-0.65)	0.47 (0.39-0.54)
Benign neoplasm of skin	0.57 (0.52-0.62)	0.58 (0.51-0.64)	0.57 (0.51-0.62)
Intervertebral disc disorders	0.61 (0.54-0.68)	0.58 (0.51-0.65)	0.59 (0.53-0.65)
Other symptoms of respiratory system	0.58 (0.54-0.62)	0.58 (0.54-0.62)	0.58 (0.54-0.62)
Abnormal electrocardiogram [ECG] [EKG]	0.49 (0.43-0.54)	0.58 (0.53-0.63)	0.46 (0.41-0.51)
Peripheral vascular disease	0.62 (0.56-0.68)	0.58 (0.51-0.64)	0.54 (0.47-0.62)
Abnormal glucose	0.52 (0.48-0.56)	0.58 (0.53-0.63)	**0.40 (0.36-0.45)**

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Other disorders of bladder	0.66 (0.58-0.73)	0.58 (0.51-0.66)	0.59 (0.51-0.67)
Dysuria	0.54 (0.47-0.62)	0.58 (0.51-0.65)	0.47 (0.38-0.55)
Spondylosis and allied disorders	0.57 (0.51-0.62)	0.57 (0.51-0.63)	0.53 (0.46-0.60)
Cardiomyopathy	0.52 (0.46-0.58)	0.57 (0.51-0.63)	0.51 (0.45-0.58)
Abdominal pain	0.57 (0.53-0.62)	0.57 (0.53-0.62)	0.56 (0.52-0.60)
Blood in stool	0.55 (0.46-0.66)	0.57 (0.48-0.67)	0.49 (0.40-0.58)
Diaphragmatic hernia	0.56 (0.50-0.62)	0.57 (0.50-0.64)	0.50 (0.41-0.57)
Abnormal heart sounds	0.58 (0.51-0.66)	0.57 (0.52-0.63)	0.55 (0.48-0.61)
Back pain	0.56 (0.50-0.61)	0.57 (0.52-0.62)	0.55 (0.49-0.59)
Other retinal disorders	0.63 (0.53-0.72)	0.57 (0.47-0.66)	0.58 (0.50-0.67)
Chronic pain	0.58 (0.53-0.63)	0.57 (0.52-0.62)	0.56 (0.50-0.61)
Depression	0.55 (0.49-0.61)	0.57 (0.50-0.63)	0.56 (0.51-0.63)
Abnormal findings on exam of gastrointestinal t...	0.57 (0.50-0.64)	0.57 (0.47-0.65)	0.50 (0.43-0.58)
Other dyspnea	0.58 (0.53-0.64)	0.57 (0.51-0.61)	0.58 (0.53-0.64)
Heart failure with reduced EF [Systolic or comb...	0.59 (0.52-0.65)	0.57 (0.48-0.65)	0.51 (0.44-0.58)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Hearing loss	0.61 (0.55-0.67)	0.57 (0.51-0.62)	0.55 (0.48-0.62)
Cough	0.58 (0.54-0.63)	0.57 (0.52-0.61)	0.56 (0.51-0.60)
Cardiac dysrhythmias	0.58 (0.54-0.63)	0.57 (0.52-0.61)	**0.56 (0.51-0.60)**
Sleep disorders	0.57 (0.51-0.62)	0.56 (0.52-0.62)	0.53 (0.47-0.59)
Other biliary tract disease	0.60 (0.51-0.69)	0.56 (0.46-0.65)	0.46 (0.38-0.53)
Esophagitis, GERD and related diseases	0.58 (0.54-0.63)	0.56 (0.51-0.61)	**0.54 (0.50-0.59)**
Pulmonary congestion and hypostasis	0.61 (0.53-0.72)	0.56 (0.46-0.65)	0.56 (0.49-0.65)
Other abnormal blood chemistry	0.55 (0.50-0.61)	0.56 (0.50-0.61)	0.53 (0.47-0.58)
Palpitations	0.57 (0.50-0.65)	0.56 (0.48-0.63)	0.51 (0.43-0.57)
Thoracic or lumbosacral neuritis or radiculitis...	0.55 (0.47-0.62)	0.56 (0.47-0.64)	0.53 (0.46-0.60)
Hypothyroidism NOS	0.59 (0.53-0.64)	0.56 (0.50-0.61)	0.58 (0.53-0.63)
Abnormal results of function study of liver	0.58 (0.49-0.68)	0.56 (0.44-0.66)	0.47 (0.38-0.57)
Cardiac conduction disorders	0.52 (0.48-0.57)	0.56 (0.51-0.61)	**0.46 (0.41-0.51)**

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Neoplasm of uncertain behavior of skin	0.58 (0.52-0.65)	0.56 (0.49-0.62)	0.53 (0.47-0.60)
Dermatophytosis / Dermatomycosis	0.53 (0.45-0.59)	0.56 (0.49-0.62)	0.48 (0.41-0.54)
Other disorders of circulatory system	0.45 (0.38-0.52)	0.56 (0.50-0.61)	0.60 (0.55-0.67)
Allergic rhinitis	0.55 (0.47-0.63)	0.55 (0.48-0.64)	0.52 (0.45-0.59)
Diseases of white blood cells	0.49 (0.41-0.57)	0.55 (0.48-0.63)	0.41 (0.35-0.48)
Secondary malignant neoplasm	0.55 (0.49-0.62)	0.55 (0.45-0.64)	0.58 (0.50-0.66)
Dermatophytosis of nail	0.60 (0.49-0.72)	0.55 (0.46-0.64)	0.59 (0.49-0.69)
Cancer, suspected or other	0.57 (0.51-0.63)	0.55 (0.48-0.61)	0.60 (0.53-0.67)
Other abnormal glucose	0.54 (0.49-0.59)	0.55 (0.51-0.60)	**0.48 (0.44-0.53)**
Erythematous conditions	0.54 (0.45-0.63)	0.55 (0.45-0.64)	0.46 (0.35-0.56)
Major depressive disorder	0.56 (0.48-0.63)	0.55 (0.47-0.64)	0.46 (0.37-0.54)
Ill-defined descriptions and complications of h...	0.53 (0.46-0.59)	0.55 (0.48-0.62)	**0.41 (0.34-0.49)**

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Heart failure with preserved EF [Diastolic hear...	0.62 (0.56-0.69)	0.55 (0.47-0.64)	0.53 (0.45-0.63)
Transient cerebral ischemia	0.45 (0.36-0.54)	0.55 (0.47-0.63)	0.61 (0.53-0.69)
Occlusion and stenosis of precerebral arteries	0.67 (0.61-0.75)	0.55 (0.46-0.65)	0.58 (0.49-0.68)
Retention of urine	0.61 (0.53-0.68)	0.55 (0.48-0.62)	0.55 (0.46-0.63)
Disorders of protein plasma/amino-acid transpor...	0.66 (0.57-0.75)	0.55 (0.45-0.63)	0.60 (0.51-0.69)
Hypopotassemia	0.57 (0.49-0.65)	0.55 (0.48-0.63)	0.57 (0.50-0.64)
Asthma	0.58 (0.52-0.66)	0.55 (0.47-0.62)	0.51 (0.44-0.57)
Malaise and fatigue	0.58 (0.53-0.63)	0.55 (0.50-0.59)	0.56 (0.51-0.61)
Anxiety disorder	0.54 (0.50-0.58)	0.54 (0.49-0.60)	0.52 (0.45-0.57)
Other specified cardiac dysrhythmias	0.59 (0.53-0.63)	0.54 (0.49-0.60)	0.56 (0.51-0.62)
Dysphagia	0.54 (0.46-0.61)	0.54 (0.47-0.62)	0.50 (0.43-0.56)
Cardiac and circulatory congenital anomalies	0.49 (0.40-0.56)	0.54 (0.46-0.62)	0.49 (0.42-0.57)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Contusion	0.46 (0.38-0.56)	0.54 (0.46-0.63)	0.44 (0.35-0.51)
Myalgia and myositis unspecified	0.56 (0.50-0.63)	0.54 (0.48-0.61)	0.48 (0.42-0.55)
Neurological disorders	0.53 (0.46-0.60)	0.54 (0.47-0.61)	0.53 (0.46-0.59)
Solitary pulmonary nodule	0.48 (0.42-0.55)	0.54 (0.45-0.62)	0.44 (0.36-0.53)
Cyst of kidney, acquired	0.55 (0.49-0.62)	0.54 (0.46-0.62)	**0.48 (0.40-0.55)**
Hypercalcemia	0.61 (0.53-0.69)	0.54 (0.44-0.63)	0.52 (0.43-0.60)
Pericarditis	0.49 (0.41-0.58)	0.54 (0.46-0.61)	0.49 (0.41-0.58)
Chronic ulcer of skin	0.47 (0.38-0.54)	0.53 (0.45-0.62)	0.53 (0.45-0.61)
Superficial cellulitis and abscess	0.57 (0.51-0.63)	0.53 (0.47-0.61)	0.58 (0.51-0.65)
Chronic kidney disease, Stage I or II	0.52 (0.45-0.60)	0.53 (0.45-0.61)	0.45 (0.38-0.52)
Hypotension NOS	0.55 (0.50-0.62)	0.53 (0.48-0.60)	0.46 (0.40-0.52)
Nonrheumatic tricuspid valve disorders	0.53 (0.44-0.63)	0.53 (0.45-0.62)	0.47 (0.39-0.56)
Open wounds of head; neck; and trunk	0.48 (0.40-0.57)	0.53 (0.44-0.63)	0.42 (0.32-0.50)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Gastrointestinal hemorrhage	0.56 (0.49-0.64)	0.53 (0.46-0.61)	0.53 (0.45-0.60)
Vitamin deficiency	0.57 (0.51-0.62)	0.53 (0.48-0.58)	**0.49 (0.44-0.54)**
Other disorders of biliary tract	0.58 (0.51-0.65)	0.53 (0.43-0.62)	0.46 (0.38-0.56)
Gout	0.51 (0.42-0.59)	0.53 (0.43-0.64)	0.54 (0.44-0.62)
Secondary malignancy of lymph nodes	0.56 (0.48-0.65)	0.53 (0.43-0.62)	0.55 (0.46-0.62)
Other upper respiratory disease	0.46 (0.39-0.55)	0.53 (0.47-0.60)	0.51 (0.44-0.58)
Insomnia	0.55 (0.49-0.61)	0.53 (0.46-0.59)	0.51 (0.45-0.57)
Hematuria	0.52 (0.44-0.59)	0.53 (0.47-0.59)	0.45 (0.39-0.52)
Malignant neoplasm, other	0.52 (0.45-0.59)	0.53 (0.46-0.61)	0.47 (0.39-0.54)
Vitamin D deficiency	0.55 (0.50-0.61)	0.52 (0.48-0.57)	0.48 (0.42-0.53)
Anxiety disorders	0.57 (0.52-0.62)	0.52 (0.47-0.58)	0.54 (0.49-0.59)
Symptoms involving respiratory system and other...	0.58 (0.48-0.66)	0.52 (0.44-0.60)	0.62 (0.54-0.71)
Paroxysmal ventricular tachycardia	0.52 (0.45-0.59)	0.52 (0.44-0.62)	0.45 (0.37-0.55)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Other symptoms involving abdomen and pelvis	0.55 (0.46-0.63)	0.52 (0.45-0.60)	0.50 (0.42-0.59)
Elevated white blood cell count	0.48 (0.40-0.55)	0.52 (0.44-0.59)	**0.39 (0.32-0.46)**
Disease of tricuspid valve	0.45 (0.36-0.56)	0.52 (0.44-0.60)	0.57 (0.48-0.67)
Paroxysmal tachycardia, unspecified	0.45 (0.38-0.53)	0.52 (0.43-0.59)	0.47 (0.40-0.55)
Abdominal hernia	0.55 (0.50-0.60)	0.51 (0.44-0.57)	0.50 (0.44-0.56)
Rash and other nonspecific skin eruption	0.55 (0.49-0.61)	0.51 (0.45-0.57)	0.51 (0.44-0.59)
Nonspecific chest pain	0.50 (0.45-0.54)	0.51 (0.46-0.56)	0.49 (0.43-0.54)
Noninfectious gastroenteritis	0.48 (0.38-0.57)	0.51 (0.42-0.60)	0.46 (0.37-0.54)
Acute upper respiratory infections of multiple ...	0.53 (0.47-0.58)	0.51 (0.45-0.57)	0.50 (0.44-0.55)
Abnormal serum enzyme levels	0.49 (0.43-0.57)	0.51 (0.43-0.58)	0.46 (0.39-0.54)
Visual disturbances	0.52 (0.45-0.61)	0.51 (0.44-0.57)	0.56 (0.49-0.63)
Generalized anxiety disorder	0.56 (0.48-0.62)	0.51 (0.40-0.59)	0.55 (0.47-0.62)
Hepatomegaly	0.52 (0.44-0.59)	0.51 (0.44-0.58)	0.53 (0.45-0.59)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Other specified diseases of hair and hair folli...	0.42 (0.34-0.50)	0.51 (0.41-0.60)	0.60 (0.50-0.69)
Arrhythmia (cardiac) NOS	0.52 (0.44-0.59)	0.50 (0.43-0.58)	0.49 (0.41-0.56)
Cancer of urinary organs (incl. kidney and blad...	0.53 (0.43-0.62)	0.50 (0.41-0.59)	0.45 (0.36-0.55)
Nonspecific abnormal findings on radiological a...	0.57 (0.48-0.64)	0.50 (0.41-0.58)	0.51 (0.43-0.59)
Dermatophytosis	0.53 (0.47-0.60)	0.50 (0.43-0.57)	0.52 (0.44-0.59)
Poisoning by antibiotics	0.55 (0.46-0.65)	0.50 (0.41-0.58)	0.53 (0.45-0.61)
Diseases of hair and hair follicles	0.43 (0.35-0.54)	0.50 (0.42-0.59)	0.39 (0.27-0.48)
Functional digestive disorders	0.51 (0.43-0.60)	0.50 (0.41-0.59)	0.53 (0.41-0.64)
Paroxysmal supraventricular tachycardia	0.48 (0.41-0.56)	0.50 (0.42-0.58)	0.54 (0.45-0.63)
Empyema and pneumothorax	0.53 (0.46-0.61)	0.50 (0.41-0.58)	0.44 (0.37-0.51)
Premature beats	0.52 (0.44-0.60)	0.50 (0.41-0.57)	0.52 (0.45-0.60)
Hemangioma and lymphangioma, any site	0.58 (0.50-0.65)	0.50 (0.43-0.56)	0.47 (0.38-0.54)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Sciatica	0.51 (0.41-0.60)	0.50 (0.42-0.59)	0.50 (0.41-0.59)
Cardiac congenital anomalies	0.48 (0.40-0.56)	0.49 (0.42-0.57)	0.53 (0.45-0.63)
Sinoatrial node dysfunction (Bradycardia)	0.58 (0.48-0.67)	0.49 (0.39-0.58)	0.61 (0.51-0.70)
Encounter for long-term (current) use of antico...	0.52 (0.45-0.60)	0.49 (0.41-0.57)	0.49 (0.41-0.56)
Acute bronchitis and bronchiolitis	0.46 (0.37-0.56)	0.49 (0.41-0.59)	0.54 (0.45-0.64)
Swelling of limb	0.56 (0.50-0.63)	0.49 (0.42-0.56)	0.57 (0.51-0.64)
Other disorders of arteries and arterioles	0.43 (0.36-0.51)	0.49 (0.42-0.57)	0.49 (0.42-0.57)
Gout and other crystal arthropathies	0.49 (0.40-0.58)	0.48 (0.40-0.58)	0.61 (0.51-0.71)
Other headache syndromes	0.48 (0.43-0.53)	0.48 (0.42-0.55)	0.57 (0.50-0.63)
Disorders of refraction and accommodation; blin...	0.58 (0.46-0.67)	0.48 (0.40-0.58)	0.54 (0.44-0.65)
Atopic/contact dermatitis due to other or unspe...	0.53 (0.48-0.59)	0.48 (0.42-0.54)	0.55 (0.49-0.61)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Valvular heart disease/ heart chambers	0.54 (0.46-0.63)	0.48 (0.40-0.56)	0.50 (0.41-0.60)
Cervicalgia	0.54 (0.48-0.62)	0.48 (0.41-0.54)	0.50 (0.43-0.58)
Other diseases of respiratory system, not elsew...	0.50 (0.42-0.56)	0.47 (0.39-0.56)	0.49 (0.41-0.57)
Impacted cerumen	0.57 (0.50-0.65)	0.47 (0.39-0.55)	0.49 (0.41-0.57)
Opiates and related narcotics causing adverse e...	0.55 (0.46-0.65)	0.47 (0.37-0.56)	0.44 (0.36-0.52)
Other dyschromia	0.57 (0.50-0.65)	0.47 (0.38-0.55)	0.55 (0.48-0.61)
Disturbance of skin sensation	0.49 (0.41-0.58)	0.47 (0.38-0.55)	0.54 (0.45-0.62)
Diseases of pancreas	0.54 (0.47-0.61)	0.47 (0.39-0.55)	0.40 (0.34-0.49)
Sprains and strains	0.61 (0.54-0.68)	0.47 (0.39-0.54)	0.60 (0.53-0.67)
Symptoms and disorders of the joints	0.51 (0.43-0.60)	0.47 (0.39-0.57)	0.53 (0.44-0.60)
Nonspecific findings on examination of blood	0.48 (0.43-0.55)	0.46 (0.38-0.54)	0.48 (0.42-0.55)
Pruritus and related conditions	0.53 (0.44-0.62)	0.46 (0.38-0.54)	0.62 (0.54-0.70)

Continued on next page

Table A.4: Mean (95% confidence interval) area under the ROC curve for predicting patient diagnoses, grouped into Phecode phenotypes, using the baseline, methylation data, and genotype data. Confidence intervals determined using bootstrapping.

Phecode Phenotype	Baseline	Methylation	Genotypes
Complications of cardiac/vascular device, impla...	0.57 (0.48-0.64)	0.46 (0.39-0.53)	0.38 (0.30-0.47)
Swelling, mass, or lump in head and neck [Space...	0.46 (0.38-0.54)	0.46 (0.38-0.54)	0.61 (0.54-0.68)
Syncope and collapse	0.50 (0.42-0.57)	0.45 (0.38-0.53)	0.46 (0.39-0.53)
Other hypertrophic and atrophic conditions of skin	0.54 (0.46-0.60)	0.45 (0.39-0.52)	0.47 (0.41-0.55)
Calculus of kidney	0.52 (0.45-0.59)	0.44 (0.37-0.52)	0.51 (0.45-0.56)
Heart transplant/surgery	0.48 (0.40-0.58)	0.41 (0.33-0.49)	0.41 (0.31-0.49)
Dyschromia and Vitiligo	0.55 (0.49-0.63)	**0.41 (0.33-0.49)**	0.49 (0.42-0.55)
Acute pharyngitis	0.49 (0.40-0.58)	**0.32 (0.26-0.40)**	0.38 (0.29-0.48)

Table A.5: Number of samples with reported usage of medications in the pharmaceutical subclasses. Pharmaceutical subclasses are sorted by number of samples.

Pharmaceutical Subclass	**Number of Samples (Percent)**
Sodium	699 (80.9%)
Opioid Agonists	639 (74.0%)
Local Anesthetics - Amides	589 (68.2%)
Non-Barbiturate Hypnotics	584 (67.6%)
5-HT3 Receptor Antagonists	549 (63.5%)
Analgesics Other	544 (63.0%)
Radiographic Contrast Media	535 (61.9%)
Anesthetics - Misc.	507 (58.7%)
Glucocorticosteroids	499 (57.8%)
Salicylates	459 (53.1%)
Heparins And Heparinoid-Like Agents	459 (53.1%)
Opioid Combinations	458 (53.0%)
HMG CoA Reductase Inhibitors	456 (52.8%)
Proton Pump Inhibitors	443 (51.3%)
Oil Soluble Vitamins	434 (50.2%)
Vasopressors	421 (48.7%)
Surfactant Laxatives	398 (46.1%)
Electrolyte Mixtures	390 (45.1%)
Antiarrhythmics Type I-B	383 (44.3%)
Beta Blockers Cardio-Selective	383 (44.3%)
Cephalosporins - 1st Generation	369 (42.7%)
Calcium Channel Blockers	367 (42.5%)

Loop Diuretics	346 (40.0%)
Miscellaneous Contrast Media	341 (39.5%)
Nondepolarizing Muscle Relaxants	336 (38.9%)
Fluoroquinolones	327 (37.8%)
Stimulant Laxatives	326 (37.7%)
Nonsteroidal Anti-inflammatory Agents (NSAIDs)	313 (36.2%)
Sympathomimetics	308 (35.6%)
Antihistamines - Ethanolamines	301 (34.8%)
Laxatives - Miscellaneous	293 (33.9%)
Magnesium	290 (33.6%)
Local Anesthetics - Topical	280 (32.4%)
Potassium	277 (32.1%)
Insulin	269 (31.1%)
Benzodiazepines	265 (30.7%)
Diagnostic Radiopharmaceuticals	264 (30.6%)
Anticonvulsants - Misc.	260 (30.1%)
Carbohydrates	252 (29.2%)
Saline Laxatives	250 (28.9%)
Antispasmodics	250 (28.9%)
H-2 Antagonists	232 (26.9%)
Angiotensin II Receptor Antagonists	231 (26.7%)
ACE Inhibitors	225 (26.0%)
Penicillin Combinations	219 (25.3%)
Cephalosporins - 3rd Generation	217 (25.1%)
Nitrates	215 (24.9%)
Glycopeptides	213 (24.7%)
Alpha-Beta Blockers	210 (24.3%)

Calcium	207 (24.0%)
Multivitamins	207 (24.0%)
Local Anesthetic Combinations	200 (23.1%)
Anti-infective Misc. - Combinations	198 (22.9%)
Anti-infective Agents - Misc.	193 (22.3%)
Plasma Proteins	190 (22.0%)
Diagnostic Drugs	189 (21.9%)
Water Soluble Vitamins	188 (21.8%)
Phenothiazines	184 (21.3%)
Gastrointestinal Stimulants	182 (21.1%)
Corticosteroids - Topical	182 (21.1%)
Central Muscle Relaxants	181 (20.9%)
Viral Vaccines	175 (20.3%)
Iron	170 (19.7%)
Vasodilators	168 (19.4%)
Antibiotics - Topical	166 (19.2%)
Hematopoietic Growth Factors	165 (19.1%)
Azithromycin	164 (19.0%)
Antacids - Calcium Salts	164 (19.0%)
Antimyasthenic/Cholinergic Agents	158 (18.3%)
Nasal Steroids	158 (18.3%)
Selective Serotonin Reuptake Inhibitors (SSRIs)	157 (18.2%)
Thiazides and Thiazide-Like Diuretics	157 (18.2%)
Misc. Nutritional Substances	155 (17.9%)
Opioid Antagonists	155 (17.9%)
Platelet Aggregation Inhibitors	154 (17.8%)
Thyroid Hormones	149 (17.2%)

Antifungals - Topical	149 (17.2%)
Bacterial Vaccines	144 (16.7%)
Immunosuppressive Agents	142 (16.4%)
Phosphate Binder Agents	140 (16.2%)
Serotonin Modulators	136 (15.7%)
Laxative Combinations	136 (15.7%)
Biguanides	135 (15.6%)
Depolarizing Muscle Relaxants	135 (15.6%)
Genitourinary Irrigants	134 (15.5%)
Prostatic Hypertrophy Agents	134 (15.5%)
Bronchodilators - Anticholinergics	131 (15.2%)
Antiflatulents	130 (15.0%)
Antacid Combinations	127 (14.7%)
Aminopenicillins	126 (14.6%)
Imidazole-Related Antifungals	125 (14.5%)
Diagnostic Tests	120 (13.9%)
Cobalamins	118 (13.7%)
Folic Acid/Folates	116 (13.4%)
B-Complex w/ Folic Acid	116 (13.4%)
Antihistamines - Non-Sedating	113 (13.1%)
Anesthetics Topical Oral	108 (12.5%)
Diabetic Supplies	107 (12.4%)
Osmotic Diuretics	106 (12.3%)
Tetracyclines	105 (12.2%)
Multiple Vitamins w/ Minerals	105 (12.2%)
Ophthalmic Anti-infectives	104 (12.0%)
Metabolic Modifiers	102 (11.8%)

Potassium Removing Agents	102 (11.8%)
Potassium Sparing Diuretics	101 (11.7%)
Hemostatics - Topical	101 (11.7%)
Ophthalmics - Misc.	101 (11.7%)
Gout Agents	100 (11.6%)
Alternative Medicine - M's	99 (11.5%)
Parenteral Therapy Supplies	99 (11.5%)
Cough/Cold/Allergy Combinations	99 (11.5%)
Antiseptics - Mouth/Throat	98 (11.3%)
Direct Factor Xa Inhibitors	97 (11.2%)
Anti-infectives - Throat	94 (10.9%)
Anti-inflammatory Agents - Topical	93 (10.8%)
Coumarin Anticoagulants	92 (10.6%)
Posterior Pituitary Hormones	91 (10.5%)
Antidotes and Specific Antagonists	90 (10.4%)
Antiadrenergic Antihypertensives	90 (10.4%)
Ophthalmic Steroids	90 (10.4%)
Antitussives	88 (10.2%)
Lincosamides	84 (9.7%)
Dibenzapines	83 (9.6%)
Bone Density Regulators	81 (9.4%)
Antianxiety Agents - Misc.	80 (9.3%)
Phosphate	78 (9.0%)
Antiemetics - Anticholinergic	77 (8.9%)
Antiperistaltic Agents	76 (8.8%)
Herpes Agents	76 (8.8%)
Bicarbonates	75 (8.7%)

Liquid Vehicles	72 (8.3%)
Antiarrhythmics Type III	72 (8.3%)
Artificial Tears and Lubricants	71 (8.2%)
Antidiarrheal/Probiotic Agents - Misc.	71 (8.2%)
Toxoid Combinations	70 (8.1%)
Urinary Antispasmodic - Antimuscarinics (Antich...	67 (7.8%)
Lozenges	67 (7.8%)
CMV Agents	66 (7.6%)
Thrombolytic Enzymes	66 (7.6%)
Impotence Agents	65 (7.5%)
Alternative Medicine - C's	64 (7.4%)
Sulfonylureas	63 (7.3%)
Antihypertensive Combinations	63 (7.3%)
Specialty Vitamins Products	63 (7.3%)
Aminoglycosides	61 (7.1%)
Cephalosporins - 2nd Generation	60 (6.9%)
Alkalinizers	59 (6.8%)
Opioid Partial Agonists	73 (6.8%)
Urinary Anti-infectives	58 (6.7%)
Irrigation Solutions	58 (6.7%)
Influenza Agents	57 (6.6%)
Expectorants	57 (6.6%)
Beta Blockers Non-Selective	56 (6.5%)
Tricyclic Agents	56 (6.5%)
Serotonin-Norepinephrine Reuptake Inhibitors (S...	56 (6.5%)
Cephalosporins - 4th Generation	55 (6.4%)
Antihistamines-Topical	55 (6.4%)

Antacids - Bicarbonate	54 (6.2%)
Bulk Laxatives	53 (6.1%)
Alpha-2 Receptor Antagonists (Tetracyclics)	52 (6.0%)
Ophthalmic Local Anesthetics	49 (5.7%)
Hemostatics - Systemic	49 (5.7%)
Zinc	48 (5.6%)
Dipeptidyl Peptidase-4 (DPP-4) Inhibitors	47 (5.4%)
Gallstone Solubilizing Agents	47 (5.4%)
Cycloplegic Mydriatics	47 (5.4%)
Protamine	58 (5.4%)
Butyrophenones	46 (5.3%)
Antidepressants - Misc.	45 (5.2%)
Mucolytics	45 (5.2%)
Leukotriene Modulators	44 (5.1%)
B-Complex Vitamins	44 (5.1%)
Acne Products	44 (5.1%)

Table A.6: Medications used in each pharmaceutical subclass

Pharmaceutical Subclass	Drug Name
ALKALINIZERS	BICITRA
ALKALINIZERS	CITRIC
ALKALINIZERS	CYTRA-2
ALKALINIZERS	CYTRA-3
ALKALINIZERS	POT
ALKALINIZERS	POTASSIUM
ANTI-INFECTIVES - THROAT	CLOTRIMAZOLE
ANTI-INFECTIVES - THROAT	MICONAZOLE
ANTI-INFECTIVES - THROAT	NYSTATIN
B-COMPLEX W/ FOLIC ACID	B
B-COMPLEX W/ FOLIC ACID	B-COMPLEX
B-COMPLEX W/ FOLIC ACID	DIALYVITE
B-COMPLEX W/ FOLIC ACID	FULL
B-COMPLEX W/ FOLIC ACID	NEPHRO-VITE
B-COMPLEX W/ FOLIC ACID	NEPHROCAPS
B-COMPLEX W/ FOLIC ACID	RENA-VITE
B-COMPLEX W/ FOLIC ACID	RENAL
B-COMPLEX W/ FOLIC ACID	RENAL-VITE
B-COMPLEX W/ FOLIC ACID	VOL-CARE
B-COMPLEX W/ FOLIC ACID	VP-VITE
BIGUANIDES	METFORMIN
CALCIUM CHANNEL BLOCKERS	ADALAT
CALCIUM CHANNEL BLOCKERS	AFEDITAB

CALCIUM CHANNEL BLOCKERS	AMLODIPINE
CALCIUM CHANNEL BLOCKERS	CARTIA
CALCIUM CHANNEL BLOCKERS	DILT-XR
CALCIUM CHANNEL BLOCKERS	DILTIAZEM
CALCIUM CHANNEL BLOCKERS	FELODIPINE
CALCIUM CHANNEL BLOCKERS	ISRADIPINE
CALCIUM CHANNEL BLOCKERS	NICARDIPINE
CALCIUM CHANNEL BLOCKERS	NIFEDICAL
CALCIUM CHANNEL BLOCKERS	NIFEDIPINE
CALCIUM CHANNEL BLOCKERS	NIMODIPINE
CALCIUM CHANNEL BLOCKERS	NORVASC
CALCIUM CHANNEL BLOCKERS	VERAPAMIL
CMV AGENTS	VALCYTE
CMV AGENTS	VALGANCICLOVIR
DIBENZAPINES	OLANZAPINE
DIBENZAPINES	QUETIAPINE
DIBENZAPINES	ZYPREXA
HEMATOPOIETIC GROWTH FACTORS	ARANESP
HEMATOPOIETIC GROWTH FACTORS	DARBEPOETIN
HEMATOPOIETIC GROWTH FACTORS	EPOETIN
HEMATOPOIETIC GROWTH FACTORS	EPOGEN
HEMATOPOIETIC GROWTH FACTORS	FILGRASTIM
HEMATOPOIETIC GROWTH FACTORS	FILGRASTIM-SNDZ
HEMATOPOIETIC GROWTH FACTORS	MIRCERA
HEMATOPOIETIC GROWTH FACTORS	NEULASTA
HEMATOPOIETIC GROWTH FACTORS	NEUPOGEN
HEMATOPOIETIC GROWTH FACTORS	PEGFILGRASTIM

HEMATOPOIETIC GROWTH FACTORS	PROCRIT
HEMATOPOIETIC GROWTH FACTORS	ROMIPLOSTIM
HEMATOPOIETIC GROWTH FACTORS	ZARXIO
IMMUNOSUPPRESSIVE AGENTS	ANTI-THYMOCYTE
IMMUNOSUPPRESSIVE AGENTS	AZATHIOPRINE
IMMUNOSUPPRESSIVE AGENTS	BASILIXIMAB
IMMUNOSUPPRESSIVE AGENTS	BELATACEPT
IMMUNOSUPPRESSIVE AGENTS	CELLCEPT
IMMUNOSUPPRESSIVE AGENTS	CYCLOSPORINE
IMMUNOSUPPRESSIVE AGENTS	EVEROLIMUS
IMMUNOSUPPRESSIVE AGENTS	IDS
IMMUNOSUPPRESSIVE AGENTS	MYCOPHENOLATE
IMMUNOSUPPRESSIVE AGENTS	MYCOPHENOLIC
IMMUNOSUPPRESSIVE AGENTS	MYFORTIC
IMMUNOSUPPRESSIVE AGENTS	NEORAL
IMMUNOSUPPRESSIVE AGENTS	PROGRAF
IMMUNOSUPPRESSIVE AGENTS	RAPAMUNE
IMMUNOSUPPRESSIVE AGENTS	SIROLIMUS
IMMUNOSUPPRESSIVE AGENTS	TACROLIMUS
METABOLIC MODIFIERS	CALCITRIOL
METABOLIC MODIFIERS	CINACALCET
METABOLIC MODIFIERS	DOXERCALCIFEROL
METABOLIC MODIFIERS	HECTOROL
METABOLIC MODIFIERS	PARICALCITOL
METABOLIC MODIFIERS	ROCALTROL
METABOLIC MODIFIERS	SENSIPAR
METABOLIC MODIFIERS	ZEMPLAR

OSMOTIC DIURETICS	MANNITOL
PHOSPHATE BINDER AGENTS	AURYXIA
PHOSPHATE BINDER AGENTS	CALCIUM
PHOSPHATE BINDER AGENTS	FERRIC
PHOSPHATE BINDER AGENTS	FOSRENOL
PHOSPHATE BINDER AGENTS	LANTHANUM
PHOSPHATE BINDER AGENTS	PHOSLO
PHOSPHATE BINDER AGENTS	RENAGEL
PHOSPHATE BINDER AGENTS	RENVELA
PHOSPHATE BINDER AGENTS	SEVELAMER
PHOSPHATE BINDER AGENTS	SUCROFERRIC
PHOSPHATE BINDER AGENTS	VELPHORO
POTASSIUM REMOVING RESINS	KALEXATE
POTASSIUM REMOVING RESINS	KAYEXALATE
POTASSIUM REMOVING RESINS	KIONEX
POTASSIUM REMOVING RESINS	PATIROMER
POTASSIUM REMOVING RESINS	SODIUM
POTASSIUM REMOVING RESINS	VELTASSA
SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIS)	CITALOPRAM
SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIS)	ESCITALOPRAM
SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIS)	FLUOXETINE
SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIS)	FLUVOXAMINE
SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIS)	LEXAPRO
SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIS)	PAROXETINE
SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIS)	SERTRALINE
SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIS)	ZOLOFT
SPECIALTY VITAMINS PRODUCTS	MG-PLUS

SPECIALTY VITAMINS PRODUCTS	ONE-A-DAY
SPECIALTY VITAMINS PRODUCTS	PROSTATE
SULFONYLUREAS	GLIMEPIRIDE
SULFONYLUREAS	GLIPIZIDE
SULFONYLUREAS	GLYBURIDE
THROMBOLYTIC ENZYMES	ALTEPLASE
VASODILATORS	HYDRALAZINE
VASODILATORS	MINOXIDIL
VASODILATORS	NITROPRUSSIDE

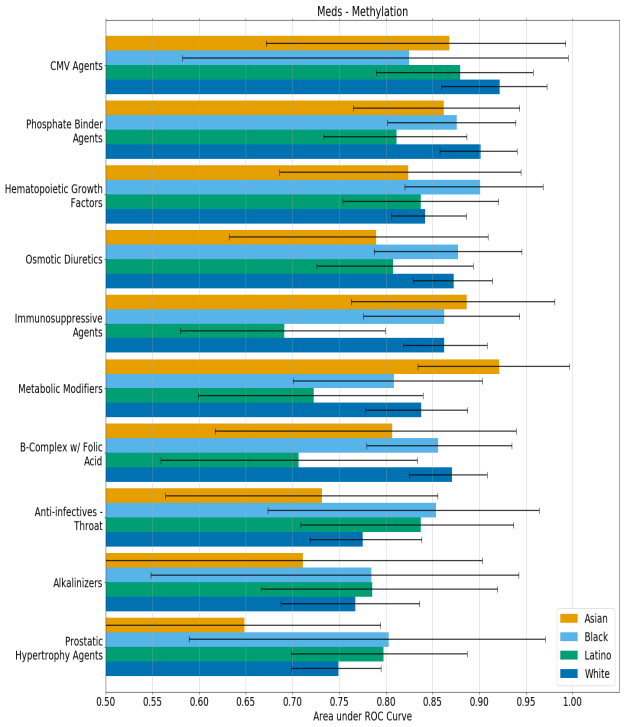

Figure A.3: Best methylation-predicted medications within ancestral populations. After training a model on the entire heterogeneous set of individuals, we evaluated the predictive performance within each population separately. We observed no significant differences within self-reported ancestral groupings.

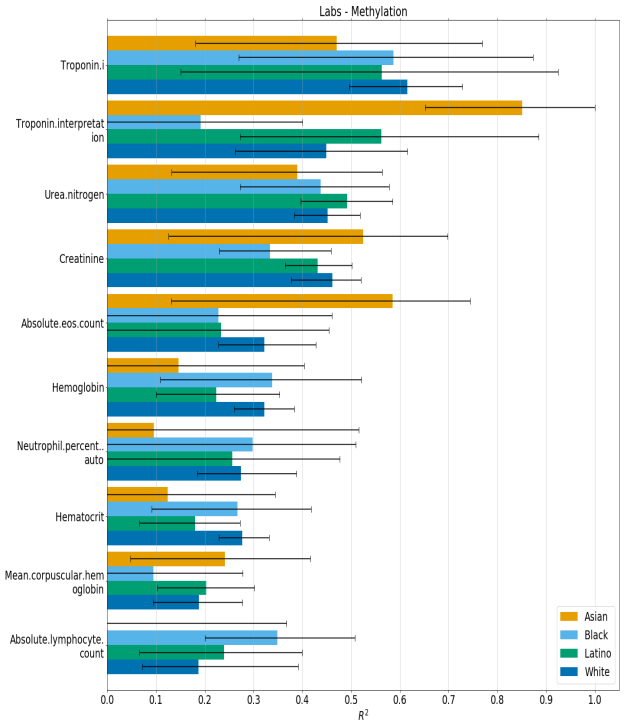

Figure A.4: Best methylation-predicted lab panels within ancestral populations. After training a model on the entire heterogeneous set of individuals, we evaluated the predictive performance within each population separately. We observed no significant differences within self-reported ancestral groupings.

APPENDIX B

Supplementary Material - Learning Higher-Order Dynamics in Video-Based Cardiac Measurement

B.1 Supplemental Methods

B.1.1 Example Video Frames

B.1.2 Model Architecture

The first two 3D convolutional layers in each branch each have 16 filters and the final two 3D convolutional layers in each branch each have 32 filters. Each convolutional layer has a filter size of 3x3x3 for all 3D convolutional layers in the network. All convolutional layers are padded such that they have the same height, width, and number of time steps in each consecutive layer. Convolutional layers use the hyperbolic tangent activation function, except for the convolutional layers used for the attention masks which use a sigmoid activation function for generating the soft masks. Attention masks (one per time step) are applied by applying an element-wise multiplication of the attention mask with each 3D convolutional feature map. Average pooling layers reduce the height and width of the frames by a factor of two, except for the final average pooling layer that pools over the entire frame (i.e. reduces each feature map to a single value per time step). Dropout (25% probability) is applied after every pooling layer to reduce overfitting.

After the final pooling layer, the learned features for each time step in a branch are concatenated together (i.e. combined across branches to share information). Each target

Figure B.1: Example video frames of different synthetic avatars generated for the training data set. The highly-parameterized avatar generation pipeline enables the creation of diverse subjects with varied demographics, lighting conditions, backgrounds, clothing/accessories, and movements.

signal uses its own set of (2) RNN layers to read the concatenated features over time and generate a target sequence. The first RNN layer is implemented as a bi-directional GRU (hyperbolic tangent activation function) with 64 total units (32 each direction). The second RNN layer is a GRU (linear activation function) layer with 1 output value per time step.

B.1.3 Metric Calculation

Heart Rate (HR) estimation To estimate the heart rate, we use an fast Fourier transform (FFT)-based method to calculate the dominant frequency in the signal, which corresponds to the heart rate. We first estimate power spectral density using the "periodogram" function

Figure B.2: Example video frames from four participants in the AFRL dataset used for model testing and evaluation.

from the scipy.signal [VGO20] library. Then we band-pass filter the PPG signal, with cutoff frequencies of 0.75-4.0 Hz (corresponding to a minimum HR of 45 BPM and maximum HR of 240 BPM). Finally, we select the frequency with the maximum power, and use this as our estimated HR.

Left Ventricle Ejection Time (LVET) estimation The LVET time is defined as the time interval between the diastolic peak and the dicrotic notch. To calculate this interval, we first identified the diastolic point in the second derivative (SD) of the PPG signal, which, because it is a "global" minima in the PPG heartbeat, appears as a "global" maxima (positive SD value) in the SD PPG. Then, in each predicted SD PPG waveform, we identified candidate dicrotic notch points. Since the dicrotic notch manifests as a "local" minima in the PPG

signal, it appears as a "local" maxima in the PPG SD signal (positive SD value). Using peak detection ("find_peaks" function in the scipy.signal library [VGO20]) we identify candiadate dicrotic notch points by finding local peaks that occur after a diastolic point, and use the dicrotic notch candidate point that is closest in time to the reference diastolic point.

Because both the ground truth PPG (and therefore its derivatives) and, in particular, the predicted PPG (and its derivatives), contain signal artifacts and noise, the peak detection process is not perfect. To reduce variability in the LVET interval estimates due to noise, we apply a smoothing operation. Specifically, we estimate the mean LVET interval within a 10-second non-overlapping window and use this as our estimate of true/predicted LVET. See Appendix Fig. B.3 for example LVET intervals over time, and the estimated LVET intervals after smoothing within windows.

B.2 Supplemental Results

Table B.1: Quantitative performance comparison between different architecture configurations. Values shown are (mean ± standard deviation). Beats-per-minute (BPM); First Derivative (FD); Heart Rate (HR); Mean Absolute Error (MAE); Second Derivative (SD); Left Ventricle Ejection Time (LVET).

Input Frames		Target Signals		HR MAE (BPM)	LVET MAE (ms)
FD	SD	FD	SD		
✗	✓	✓	✗	3.14 ± 7.07	83.09 ± 42.41
✗	✓	✓	✓	6.97 ± 16.30	65.10 ± 31.56
✗	✓	✗	✓	2.95 ± 6.57	57.67 ± 25.82

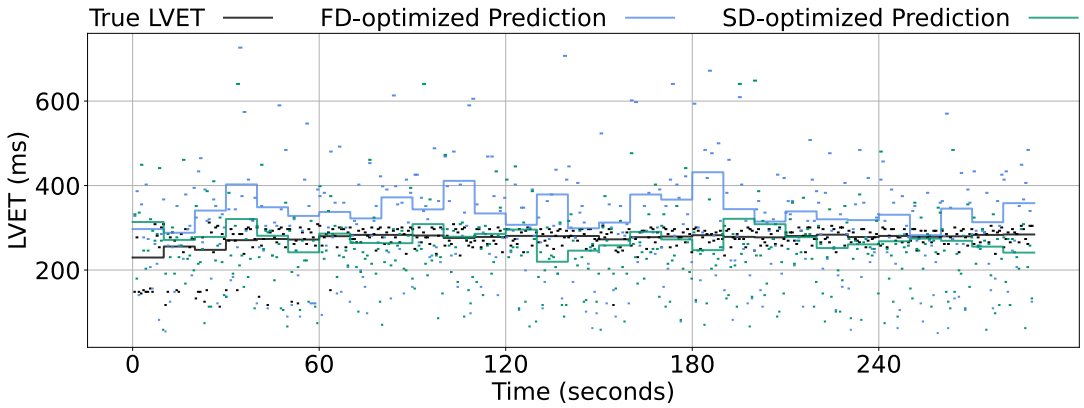

Figure B.3: Comparison of Left Ventricle Ejection Time (LVET) estimation over a 5-minute time period. Solid lines are computed as the mean LVET interval within non-overlapping 10-second windows.

REFERENCES

[AAB16] Martn Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems." *arXiv:1603.04467 [cs]*, March 2016. arXiv: 1603.04467.

[APH20] Tucker Annis, Susan Pleasants, Gretchen Hultman, Elizabeth Lindemann, Joshua A Thompson, Stephanie Billecke, Sameer Badlani, and Genevieve B Melton. "Rapid implementation of a COVID-19 remote patient monitoring program." *Journal of the American Medical Informatics Association*, **27**(8):1326–1330, August 2020.

[AS15] Gad Asher and Paolo Sassone-Corsi. "Time for food: the intimate interplay between nutrition, metabolism, and the circadian clock." *Cell*, **161**(1):84–92, Mar 2015.

[BAK14] Richard T. Barfield, Lynn M. Almli, Varun Kilaru, Alicia K. Smith, Kristina B. Mercer, Richard Duncan, Torsten Klengel, Divya Mehta, Elisabeth B. Binder, Michael P. Epstein, Kerry J. Ressler, and Karen N. Conneely. "Accounting for Population Stratification in DNA Methylation Studies." *Genetic Epidemiology*, **38**(3):231–241, 2014. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/gepi.21789.

[BB07] Sharon R. Browning and Brian L. Browning. "Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering." *The American Journal of Human Genetics*, **81**(5):1084–1097, November 2007. Publisher: Elsevier.

[BBH20] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M. Vardoulakis. "A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–12, New York, NY, USA, April 2020. Association for Computing Machinery.

[Bel17] Daniel W Belsky. "Translating Polygenic Analysis for Prevention: From Who to How." *Circulation. Cardiovascular genetics*, **10**(3):e001798, 06 2017.

[BHH18] Lisa Bastarache, Jacob J. Hughey, Scott Hebbring, Joy Marlo, Wanke Zhao, Wanting T. Ho, Sara L. Van Driest, Tracy L. McGregor, Jonathan D. Mosley, Quinn S. Wells, Michael Temple, Andrea H. Ramirez, Robert Carroll, Travis Osterman, Todd Edwards, Douglas Ruderfer, Digna R. Velez Edwards, Rizwan Hamid, Joy Cogan, Andrew Glazer, Wei-Qi Wei, QiPing Feng, Murray Brilliant, Zhizhuang J. Zhao, Nancy J. Cox, Dan M. Roden, and Joshua C. Denny. "Phenotype risk scores identify patients with unrecognized Mendelian disease patterns." *Science*, **359**(6381):1233–1239, 2018.

[BHV03] Andreas Bur, Harald Herkner, Marianne Vlcek, Christian Woisetschlger, Ulla Derhaschnig, Georg Delle Karth, Anton N. Laggner, and Michael M. Hirschl. "Factors influencing the accuracy of oscillometric blood pressure measurement in critically ill patients." *Critical Care Medicine*, **31**(3):793, March 2003.

[BLL09] Marek Brzezinski, Thomas Luisetti, and Martin J. London. "Radial Artery Cannulation: A Comprehensive Review of Recent Anatomic and Physiologic Investigations." *Anesthesia & Analgesia*, **109**(6):1763, December 2009.

[BLP13] Karl Y. Bilimoria, Yaoming Liu, Jennifer L. Paruch, Lynn Zhou, Thomas E. Kmiecik, Clifford Y. Ko, and Mark E. Cohen. "Development and evaluation of the universal ACS NSQIP surgical risk calculator: A decision aid and informed consent tool for patients and surgeons." *Journal of the American College of Surgeons*, 2013.

[BLS18] Brett K. Beaulieu-Jones, Daniel R. Lavage, John W. Snyder, Jason H. Moore, Sarah A. Pendergrass, and Christopher R. Bauer. "Characterizing and Managing Missing Structured Data in Electronic Health Records: Data Analysis." *JMIR Medical Informatics*, **6**(1):e8960, February 2018. Company: JMIR Medical Informatics Distributor: JMIR Medical Informatics Institution: JMIR Medical Informatics Label: JMIR Medical Informatics Publisher: JMIR Publications Inc., Toronto, Canada.

[BPP12] Jilles B. Bijker, Suzanne Persoon, Linda M. Peelen, Karel G. M. Moons, Cor J. Kalkman, L. Jaap Kappelle, and Wilton A. van Klei. "Intraoperative Hypotension and Perioperative Ischemic Stroke after General SurgeryA Nested Case-control Study." *Anesthesiology: The Journal of the American Society of Anesthesiologists*, **116**(3):658–664, March 2012.

[BQH18] Tor Biering-Srensen, Gabriela Querejeta Roca, Sheila M. Hegde, Amil M. Shah, Brian Claggett, Thomas H. Mosley, Kenneth R. Butler, and Scott D. Solomon. "Left Ventricular Ejection Time is an Independent Predictor of Incident Heart Failure in a Community based Cohort." *European journal of heart failure*, **20**(7):1106–1114, July 2018.

[BTR10] Christopher G. Bell, Andrew E. Teschendorff, Vardhman K. Rakyan, Alexander P. Maxwell, Stephan Beck, and David A. Savage. "Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus." *BMC Medical Genomics*, **3**(1):33, August 2010.

[CBH02] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. "SMOTE: Synthetic Minority Over-sampling Technique." *J. Artif. Int. Res.*, **16**(1):321–357, June 2002. Publisher: AI Access Foundation Place: USA.

[CG16] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, New York, NY, USA, 2016. ACM. Series Title: KDD '16.

[CGE18] Alexessander Couto Alves, Craig A. Glastonbury, Julia S. El-Sayed Moustafa, and Kerrin S. Small. "Fasting and time of day independently modulate circadian rhythm relevant gene expression in adipose and skin tissue." *BMC Genomics*, **19**(1):659, 2018.

[CHB19] Michelle M. Clark, Amber Hildreth, Sergey Batalov, Yan Ding, Shimul Chowdhury, Kelly Watkins, Katarzyna Ellsworth, Brandon Camp, Cyrielle I. Kint, Calum Yacoubian, Lauge Farnaes, Matthew N. Bainbridge, Curtis Beebe, Joshua J. A. Braun, Margaret Bray, Jeanne Carroll, Julie A. Cakici, Sara A. Caylor, Christina Clarke, Mitchell P. Creed, Jennifer Friedman, Alison Frith, Richard Gain, Mary Gaughran, Shauna George, Sheldon Gilmer, Joseph Gleeson, Jeremy Gore, Haiying Grunenwald, Raymond L. Hovey, Marie L. Janes, Kejia Lin, Paul D. McDonagh, Kyle McBride, Patrick Mulrooney, Shareef Nahas, Daeheon Oh, Albert Oriol, Laura Puckett, Zia Rady, Martin G. Reese, Julie Ryu, Lisa Salz, Erica Sanford, Lawrence Stewart, Nathaly Sweeney, Mari Tokita, Luca Van Der Kraan, Sarah White, Kristen Wigby, Brett Williams, Terence Wong, Meredith S. Wright, Catherine Yamada, Peter Schols, John Reynders, Kevin Hall, David Dimmock, Narayanan Veeraraghavan, Thomas Defay, and Stephen F. Kingsmore. "Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation." *Science Translational Medicine*, **11**(489), 2019.

[CKL18] Kristin M Corey, Sehj Kashyap, Elizabeth Lorenzi, Sandhya A Lagoo-Deenadayalan, Katherine Heller, Krista Whalen, Suresh Balu, Mitchell T Heflin, Shelley R McDonald, Madhav Swaminathan, and Mark Sendak. "Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): A retrospective, single-site study." *PLOS Medicine*, **15**(11):e1002701–e1002701, November 2018.

[CM18] Weixuan Chen and Daniel McDuff. "Deepphys: Video-based physiological measurement using convolutional attention networks." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 349–365, 2018.

[CM20] Weixuan Chen and Daniel McDuff. "DeepMag: Source-Specific Change Magnification Using Gradient Ascent." *ACM Transactions on Graphics*, **40**(1):2:1–2:14, September 2020.

[CMT16] Romanas Chaleckis, Itsuo Murakami, Junko Takada, Hiroshi Kondoh, and Mitsuhiro Yanagida. "Individual variability in human blood metabolites identifies age-related differences." *Proceedings of the National Academy of Sciences*, **113**(16):4252–4259, 2016.

[Co15] Franois Chollet and others. *Keras*. 2015.

[CSP94] Mary Charlson, Ted P Szatrowski, Janey Peterson, and Jeffrey Gold. "Validation of a combined comorbidity index." *Journal of Clinical Epidemiology*, **47**(11):1245–1251, November 1994. Publisher: Elsevier.

[CSV] Gari D Cliord, Daniel J Scott, and Mauricio Villarroel. "User Guide and Documentation for the MIMIC II Database." p. 76.

[CTS17] Audrey Y. Chu, Adrienne Tin, Pascal Schlosser, Yi-An Ko, Chengxiang Qiu, Chen Yao, Roby Joehanes, Morgan E. Grams, Liming Liang, Caroline A. Gluck, Chunyu Liu, Josef Coresh, Shih-Jen Hwang, Daniel Levy, Eric Boerwinkle, James S. Pankow, Qiong Yang, Myriam Fornage, Caroline S. Fox, Katalin Susztak, and Anna Köttgen. "Epigenome-wide association studies identify DNA methylation associated with kidney function." *Nature Communications*, **8**(1):1286, 2017.

[Daa11] Mohamed Daabiss. "American Society of Anaesthesiologists physical status classification." *Indian Journal of Anaesthesia*, **55**(2):111–115, 2011. Publisher: Medknow Publications Place: India.

[DBR13] Joshua C. Denny, Lisa Bastarache, Marylyn D. Ritchie, Robert J. Carroll, Raquel Zink, Jonathan D. Mosley, Julie R. Field, Jill M. Pulley, Andrea H. Ramirez, Erica Bowton, Melissa A. Basford, David S. Carrell, Peggy L. Peissig, Abel N. Kho, Jennifer A. Pacheco, Luke V. Rasmussen, David R. Crosslin, Paul K. Crane, Jyotishman Pathak, Suzette J. Bielinski, Sarah A. Pendergrass, Hua Xu, Lucia A. Hindorff, Rongling Li, Teri A. Manolio, Christopher G. Chute, Rex L. Chisholm, Eric B. Larson, Gail P. Jarvik, Murray H. Brilliant, Catherine A. McCarty, Iftikhar J. Kullo, Jonathan L. Haines, Dana C. Crawford, Daniel R. Masys, and Dan M. Roden. "Systematic comparison of phenome-wide association study

of electronic medical record data and genome-wide association study data." *Nature Biotechnology*, **31**(12):1102–1111, December 2013. Number: 12 Publisher: Nature Publishing Group.

[DKT11] M A Dalton Jarrod E., M D Kurz Andrea, M D Turan Alparslan, Ph.D. Mascha Edward J., M D Sessler Daniel I., and M D Saager Leif. "Development and Validation of a Risk Quantification Index for 30-Day Postoperative Mortality and Morbidity in Noncardiac Surgical Patients." *Anesthesiology*, **114**(6):1336–1344, June 2011.

[DNT14] Katherine J Dick, Christopher P Nelson, Loukia Tsaprouni, Johanna K Sandling, Dylan Aïssi, Simone Wahl, Eshwar Meduri, Pierre-Emmanuel Morange, France Gagnon, Harald Grallert, Melanie Waldenberger, Annette Peters, Jeanette Erdmann, Christian Hengstenberg, Francois Cambien, Alison H Goodall, Willem H Ouwehand, Heribert Schunkert, John R Thompson, Tim D Spector, Christian Gieger, David-Alexandre Trégouët, Panos Deloukas, and Nilesh J Samani. "DNA methylation and body-mass index: a genome-wide analysis." *The Lancet*, **383**(9933):1990–1998, 2014.

[Doz16] Timothy Dozat. "Incorporating Nesterov Momentum into Adam." February 2016.

[DRB10] Joshua C. Denny, Marylyn D. Ritchie, Melissa A. Basford, Jill M. Pulley, Lisa Bastarache, Kristin Brown-Gentry, Deede Wang, Dan R. Masys, Dan M. Roden, and Dana C. Crawford. "PheWAS: demonstrating the feasibility of a phenome-wide scan to discover genedisease associations." *Bioinformatics*, **26**(9):1205–1210, May 2010. Publisher: Oxford Academic.

[DSG19] L. Duncan, H. Shen, B. Gelaye, J. Meijsen, K. Ressler, M. Feldman, R. Peterson, and B. Domingue. "Analysis of polygenic risk score usage and performance in diverse human populations." *Nature Communications*, **10**(1):3328, July 2019. Number: 1 Publisher: Nature Publishing Group.

[DYZ17] Xiaorong Ding, Bryan P Yan, Yuan-Ting Zhang, Jing Liu, Ni Zhao, and Hon Ki Tsang. "Pulse Transit Time Based Continuous Cuffless Blood Pressure Estimation: A New Extension and A Comprehensive Evaluation." *Scientific Reports*, **7**(1):11554–11554, 2017.

[EBM14] Justin R. Estepp, Ethan B. Blackford, and Christopher M. Meier. "Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography." In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1462–1469, October 2014. ISSN: 1062-922X.

[EES13] Mohammed El Amine Lazouni, Mostafa El Habib Daho, Nesma Settouti, Mohammed Amine Chikh, and Sad Mahmoudi. "Machine Learning Tool for Automatic ASA Detection." pp. 9–16. Springer International Publishing, Cham, 2013.

[EFL19] Mohamed Elgendi, Richard Fletcher, Yongbo Liang, Newton Howard, Nigel H. Lovell, Derek Abbott, Kenneth Lim, and Rabab Ward. "The use of photoplethysmography for assessing hypertension." *npj Digital Medicine*, **2**(1):60, June 2019.

[EJ17] Anitha Edison and CV Jiji. "Optical acceleration for motion description in videos." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 39–47, 2017.

[EKN17] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature*, **542**:115–115, January 2017.

[FFM19] Christoph Feichtenhofer, Haoqi Fan, Jitendra Malik, and Kaiming He. "Slowfast networks for video recognition." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6202–6211, 2019.

[FHH21] Kara N Fitzgerald, Romilly Hodges, Douglas Hanes, Emily Stack, David Cheishvili, Moshe Szyf, Janine Henkel, Melissa W Twedt, Despina Giannopoulou, Josette Herdell, Sally Logan, and Ryan Bradley. "Potential reversal of epigenetic age using a diet and lifestyle intervention: a pilot randomized clinical trial." *Aging*, **13**(7):9419–9432, 04 2021.

[FHL04] Rupert Fincke, Judith S. Hochman, April M. Lowe, Venu Menon, James N. Slater, John G. Webb, Thierry H. LeJemtel, Gad Cotter, and Shock Investigators. "Cardiac power is the strongest hemodynamic correlate of mortality in cardiogenic shock: A report from the SHOCK trial registry." *Journal of the American College of Cardiology*, **44**(2):340–348, July 2004. Publisher: Journal of the American College of Cardiology Section: Clinical research: cardiogenic shock.

[FK18] Alexander L. Fogel and Joseph C. Kvedar. "Artificial intelligence powers digital medicine." *npj Digital Medicine*, 2018.

[GAG00] Goldberger Ary L., Amaral Luis A. N., Glass Leon, Hausdorff Jeffrey M., Ivanov Plamen Ch., Mark Roger G., Mietus Joseph E., Moody George B., Peng Chung-Kang, and Stanley H. Eugene. "PhysioBank, PhysioToolkit, and PhysioNet." *Circulation*, **101**(23):e215–e220, June 2000.

[GDM21] Monika Gawako, David Duncker, Martin Manninger, Rachel M.J. van der Velden, Astrid N.L. Hermans, Dominique V.M. Verhaert, Laurent Pison, Ron Pisters, Martin Hemels, Arian Sultan, Daniel Steven, Dhiraj Gupta, Hein Heidbuchel,

Afzal Sohaib, Petra Wijtvliet, Robert Tieleman, Henri Gruwez, Julian Chun, Boris Schmidt, John J. Keaney, Patrick Mller, Piotr Lodziski, Emma Svennberg, Olga Hoekstra, Ward P.J. Jansen, Lien Desteghe, Tom de Potter, David R. Tomlinson, Lis Neubeck, Harry J.G.M. Crijns, Nikki A.H.A. Pluymaekers, Jeroen M. Hendriks, Dominik Linz, and the TeleCheck-AF investigators. "The European TeleCheck-AF project on remote app-based management of atrial fibrillation during the COVID-19 pandemic: centre and patient experiences." *EP Europace*, **23**(7):1003–1015, July 2021.

[GGO17] Joshua M Galanter, Christopher R Gignoux, Sam S Oh, Dara Torgerson, Maria Pino-Yanes, Neeta Thakur, Celeste Eng, Donglei Hu, Scott Huntsman, Harold J Farber, Pedro C Avila, Emerita Brigino-Buenaventura, Michael A LeNoir, Kelly Meade, Denise Serebrisky, William Rodríguez-Cintrón, Rajesh Kumar, Jose R Rodríguez-Santana, Max A Seibold, Luisa N Borrell, Esteban G Burchard, and Noah Zaitlen. "Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures." *eLife*, **6**:e20532, jan 2017.

[GKI21] Trisha Greenhalgh, Matthew Knight, Matt Inada-Kim, Naomi J. Fulop, Jonathan Leach, and Cecilia Vindrola-Padros. "Remote management of covid-19 using home pulse oximetry and virtual ward support." *BMJ*, **372**:n677, March 2021. Publisher: British Medical Journal Publishing Group Section: Practice.

[GOA20] Amirata Ghorbani, David Ouyang, Abubakar Abid, Bryan He, Jonathan H. Chen, Robert A. Harrington, David H. Liang, Euan A. Ashley, and James Y. Zou. "Deep learning interpretation of echocardiograms." *npj Digital Medicine*, **3**(1):1–10, January 2020. Number: 1 Publisher: Nature Publishing Group.

[GPC16] Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip C. Nelson, Jessica L. Mega, and Dale R. Webster. "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." *JAMA*, **316**(22):2402–2410, December 2016.

[HAK12] Eugene Andres Houseman, William P. Accomando, Devin C. Koestler, Brock C. Christensen, Carmen J. Marsit, Heather H. Nelson, John K. Wiencke, and Karl T. Kelsey. "DNA methylation arrays as surrogate measures of cell mixture distribution." *BMC Bioinformatics*, **13**(1):86, May 2012.

[HBG18] Brian Hill, Robert P Brown, Eilon Gabel, Christine Lee, Maxime Cannesson, Loes Olde Loohuis, Ruth Johnson, Brandon Jew, Uri Maoz, Aman Mahajan, Sriram Sankararaman, Ira Hofer, and Eran Halperin. "Preoperative predictions of in-hospital mortality using electronic medical record data." *bioRxiv*, 2018.

[HBG19] Brian L. Hill, Robert Brown, Eilon Gabel, Nadav Rakocz, Christine Lee, Maxime Cannesson, Pierre Baldi, Loes Olde Loohuis, Ruth Johnson, Brandon Jew, Uri Maoz, Aman Mahajan, Sriram Sankararaman, Ira Hofer, and Eran Halperin. "An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data." *British Journal of Anaesthesia*, **123**(6):877–886, December 2019. Publisher: Elsevier.

[HDJ15] Nicholas J. Hackett, Gildasio S. De Oliveira, Umang K. Jain, and John Y.S. Kim. "ASA class is a reliable independent predictor of medical complications and mortality following surgery." *International Journal of Surgery*, **18**:184–190, June 2015. Publisher: Elsevier.

[HGP16] Ira S Hofer, Eilon Gabel, Michael Pfeffer, Mohammed Mahbouba, and Aman Mahajan. "A Systematic Approach to Creation of a Perioperative Data Warehouse." *Anesthesia & Analgesia*, **122**(6), 2016.

[HGT14] Jimmy L Huynh, Paras Garg, Tin Htwe Thin, Seungyeul Yoo, Ranjan Dutta, Bruce D Trapp, Vahram Haroutunian, Jun Zhu, Michael J Donovan, Andrew J Sharp, and Patrizia Casaccia. "Epigenome-wide differences in pathology-free regions of multiple sclerosis–affected brains." *Nature Neuroscience*, **17**(1):121–130, 2014.

[HHA19] Elizabeth Hibler, Lei Huang, Jorge Andrade, and Bonnie Spring. "Impact of a diet and activity health promotion intervention on regional patterns of DNA methylation." *Clinical Epigenetics*, **11**(1):133, 2019.

[HHK90] M. Hamada, K. Hiwada, and T. Kokubu. "Clinical significance of systolic time intervals in hypertensive patients." *European Heart Journal*, **11 Suppl I**:105–113, December 1990.

[HJB18] Feras Hatib, Zhongping Jian, Sai Buddi, Christine Lee, Jos Settels, Karen Sibert, Joseph Rinehart, and Maxime Cannesson. "Machine-learning Algorithm to Predict Hypotension Based on High-fidelity Arterial Pressure Waveform Analysis." *Anesthesiology*, **129**(4):663–674, October 2018.

[HJM19] Tim Hulsen, Saumya S. Jamuar, Alan R. Moody, Jason H. Karnes, Orsolya Varga, Stine Hedensted, Roberto Spreafico, David A. Hafler, and Eoin F. McKinney. "From Big Data to Precision Medicine." *Frontiers in Medicine*, **6**:34, 2019.

[HMW04] Raida I Harik-Khan, Denis C Muller, and Robert A Wise. "Racial difference in lung function in African-American and White children: effect of anthropometric, socioeconomic, nutritional, and environmental factors." *Am J Epidemiol*, **160**(9):893–900, Nov 2004.

[Hor13a] Steve Horvath. "DNA methylation age of human tissues and cell types." *Genome Biology*, **14**(10):3156, 2013.

[Hor13b] Steve Horvath. "DNA methylation age of human tissues and cell types." *Genome Biology*, **14**(10):3156, December 2013.

[HRH19] Awni Y Hannun, Pranav Rajpurkar, Masoumeh Haghpanahi, Geoffrey H Tison, Codie Bourn, Mintu P Turakhia, and Andrew Y Ng. "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network." *Nature Medicine*, **25**(1):65–69, 2019.

[HRR21] Brian L. Hill, Nadav Rakocz, kos Rudas, Jeffrey N. Chiang, Sidong Wang, Ira Hofer, Maxime Cannesson, and Eran Halperin. "Imputation of the continuous arterial line blood pressure waveform from non-invasive measurements using deep learning." *Scientific Reports*, **11**(1):15755, August 2021. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Machine learning;Medical research;Predictive medicine Subject_term_id: machine-learning;medical-research;predictive-medicine.

[IKY17] Noriko Inoue, Hideshi Kawakami, Hideya Yamamoto, Chikako Ito, Saeko Fujiwara, Hideo Sasaki, and Yasuki Kihara. "Second derivative of the finger photoplethysmogram and cardiovascular mortality in middle-aged and elderly Japanese women." *Hypertension Research*, **40**(2):207–211, February 2017.

[JLY20] Ali Jazayeri, Ou Stella Liang, and Christopher C. Yang. "Imputation of Missing Data in Electronic Health Records Based on Patients'Similarities." *Journal of Healthcare Informatics Research*, **4**(3):295–307, 2020.

[JPS16] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. "MIMIC-III, a freely accessible critical care database." *Scientific Data*, **3**:160035, May 2016.

[KB17] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization." *arXiv:1412.6980 [cs]*, January 2017. arXiv: 1412.6980.

[KCA18] Amit V Khera, Mark Chaffin, Krishna G Aragam, Mary E Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, Eric S Lander, Steven A Lubitz, Patrick T Ellinor, and Sekar Kathiresan. "Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations." *Nat Genet*, **50**(9):1219–1224, Sep 2018.

[KCB18] Matthieu Komorowski, Leo A Celi, Omar Badawi, Anthony C Gordon, and A Aldo Faisal. "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care." *Nature Medicine*, 2018.

[KKK12] Duk-Hyun Kang, Yong-Jin Kim, Sung-Han Kim, Byung Joo Sun, Dae-Hee Kim, Sung-Cheol Yun, Jong-Min Song, Suk Jung Choo, Cheol-Hyun Chung, Jae-Kwan Song, Jae-Won Lee, and Dae-Won Sohn. "Early Surgery versus Conventional Treatment for Infective Endocarditis." *New England Journal of Medicine*, **366**(26):2466–2473, June 2012.

[KLG13] Yuriy Kurylyak, Francesco Lamonaca, and Domenico Grimaldi. "A Neural Network-based method for continuous blood pressure estimation from a PPG signal." In *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 280–283. IEEE, May 2013.

[KLS14] Sang-Hyun Kim, Marc Lilot, Kulraj S Sidhu, Joseph Rinehart, Zhaoxia Yu, Cecilia Canales, and Maxime Cannesson. "Accuracy and Precision of Continuous Noninvasive Arterial Pressure Monitoring Compared with Invasive Arterial Pressure: A Systematic Review and Meta-analysis." *Anesthesiology: The Journal of the American Society of Anesthesiologists*, **120**(5):1080–1097, May 2014.

[KMH21] Miklos D. Kertai, Jonathan D. Mosley, Jing He, Abinaya Ramakrishnan, Mark J. Abdelmalak, Yurim Hong, M. Benjamin Shoemaker, Dan M. Roden, and Lisa Bastarache. "Predictive Accuracy of a Polygenic Risk Score for Postoperative Atrial Fibrillation After Cardiac Surgery." *Circulation: Genomic and Precision Medicine*, **14**(2):e003269, 2021/07/01 2021.

[KMK19] Sini Kerminen, Alicia R. Martin, Jukka Koskela, Sanni E. Ruotsalainen, Aki S. Havulinna, Ida Surakka, Aarno Palotie, Markus Perola, Veikko Salomaa, Mark J. Daly, Samuli Ripatti, and Matti Pirinen. "Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland." *The American Journal of Human Genetics*, **104**(6):1169–1181, June 2019.

[KNK19] Katarzyna Kamińska, Ewelina Nalejska, Marta Kubiak, Joanna Wojtysiak, Łukasz Żołna, Janusz Kowalewski, and Marzena Anna Lewandowska. "Prognostic and Predictive Epigenetic Biomarkers in Oncology." *Molecular Diagnosis & Therapy*, **23**(1):83–95, 2019.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[LAP13] Yun Liu, Martin J Aryee, Leonid Padyukov, M Daniele Fallin, Espen Hesselberg, Arni Runarsson, Lovisa Reinius, Nathalie Acevedo, Margaret Taub, Marcus Ronninger, Klementy Shchetynsky, Annika Scheynius, Juha Kere, Lars Alfredsson, Lars Klareskog, Tomas J Ekström, and Andrew P Feinberg. "Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis." *Nature Biotechnology*, **31**:142 EP –, 01 2013.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *Nature*, **521**(7553):436–444, May 2015. Publisher: Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

[LCR16] Yannick Le Manach, Gary Collins, Reitze Rodseth, Christine Le Bihan-Benjamin, Bruce Biccard, Bruno Riou, P.J. Devereaux, and Paul Landais. "Preoperative Score to Predict Postoperative Mortality (POSPOM)." *Anesthesiology*, **124**(3):570–579, March 2016.

[Lev10] Victor V Levenson. "DNA methylation as a universal biomarker." *Expert Review of Molecular Diagnostics*, **10**(4):481–488, 05 2010.

[LFP20] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. "Multi-task temporal shift attention networks for on-device contactless vitals measurement." *arXiv preprint arXiv:2006.03790*, 2020.

[LGH15] Jiajia Li, Gregory R. Grant, John B. Hogenesch, and Michael E. Hughes. "Chapter Sixteen - Considerations for RNA-seq Analysis of Circadian Rhythms." In Amita Sehgal, editor, *Circadian Rhythms and Biological Clocks, Part A*, volume 551 of *Methods in Enzymology*, pp. 349–367. Academic Press, 2015.

[LH19] Cathryn M Lewis and Saskia P Hagenaars. "Progressing Polygenic Medicine in Psychiatry Through Electronic Health Records." *JAMA Psychiatry*, **76**(5):470–472, May 2019.

[LHG18] Christine K Lee, Ira Hofer, Eilon Gabel, Pierre Baldi, and Maxime Cannesson. "Development and Validation of a Deep Neural Network Model for Prediction of Postoperative In-hospital Mortality." *Anesthesiology*, April 2018.

[LHZ21] Hao Lu, Hu Han, and S Kevin Zhou. "Dual-GAN: Joint BVP and Noise Modeling for Remote Physiological Measurement." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12404–12413, 2021.

[LLB01] Eric S. Lander, Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, Roel Funke, Diane Gage, Katrina Harris, Andrew Heaford, John Howland, Lisa Kann, Jessica Lehoczky, Rosie LeVine, Paul McEwan, Kevin McKernan, James Meldrim, Jill P. Mesirov, Cher Miranda, William Morris, Jerome Naylor, Christina Raymond, Mark Rosetti, Ralph Santos, Andrew Sheridan, Carrie Sougnez, Nicole Stange-Thomann, Nikola Stojanovic, Aravind Subramanian, Dudley Wyman, Jane Rogers, John Sulston, Rachael Ainscough, Stephan Beck, David Bentley, John Burton, Christopher Clee, Nigel Carter, Alan Coulson, Rebecca Deadman, Panos Deloukas, Andrew Dunham, Ian Dunham, Richard Durbin, Lisa French, Darren Grafham, Simon Gregory, Tim Hubbard, Sean Humphray, Adrienne Hunt, Matthew Jones, Christine Lloyd, Amanda

McMurray, Lucy Matthews, Simon Mercer, Sarah Milne, James C. Mullikin, Andrew Mungall, Robert Plumb, Mark Ross, Ratna Shownkeen, Sarah Sims, Robert H. Waterston, Richard K. Wilson, LaDeana W. Hillier, John D. McPherson, Marco A. Marra, Elaine R. Mardis, Lucinda A. Fulton, Asif T. Chinwalla, Kymberlie H. Pepin, Warren R. Gish, Stephanie L. Chissoe, Michael C. Wendl, Kim D. Delehaunty, Tracie L. Miner, Andrew Delehaunty, Jason B. Kramer, Lisa L. Cook, Robert S. Fulton, Douglas L. Johnson, Patrick J. Minx, Sandra W. Clifton, Trevor Hawkins, Elbert Branscomb, Paul Predki, Paul Richardson, Sarah Wenning, Tom Slezak, Norman Doggett, Jan-Fang Cheng, Anne Olsen, Susan Lucas, Christopher Elkin, Edward Uberbacher, Marvin Frazier, Richard A. Gibbs, Donna M. Muzny, Steven E. Scherer, John B. Bouck, Erica J. Sodergren, Kim C. Worley, Catherine M. Rives, James H. Gorrell, Michael L. Metzker, Susan L. Naylor, Raju S. Kucherlapati, David L. Nelson, George M. Weinstock, Yoshiyuki Sakaki, Asao Fujiyama, Masahira Hattori, Tetsushi Yada, Atsushi Toyoda, Takehiko Itoh, Chiharu Kawagoe, Hidemi Watanabe, Yasushi Totoki, Todd Taylor, Jean Weissenbach, Roland Heilig, William Saurin, Francois Artiguenave, Philippe Brottier, Thomas Bruls, Eric Pelletier, Catherine Robert, Patrick Wincker, Andr Rosenthal, Matthias Platzer, Gerald Nyakatura, Stefan Taudien, Andreas Rump, Douglas R. Smith, Lynn Doucette-Stamm, Marc Rubenfield, Keith Weinstock, Hong Mei Lee, JoAnn Dubois, Huanming Yang, Jun Yu, Jian Wang, Guyang Huang, Jun Gu, Leroy Hood, Lee Rowen, Anup Madan, Shizen Qin, Ronald W. Davis, Nancy A. Federspiel, A. Pia Abola, Michael J. Proctor, Bruce A. Roe, Feng Chen, Huaqin Pan, Juliane Ramser, Hans Lehrach, Richard Reinhardt, W. Richard McCombie, Melissa de la Bastide, Neilay Dedhia, Helmut Blcker, Klaus Hornischer, Gabriele Nordsiek, Richa Agarwala, L. Aravind, Jeffrey A. Bailey, Alex Bateman, Serafim Batzoglou, Ewan Birney, Peer Bork, Daniel G. Brown, Christopher B. Burge, Lorenzo Cerutti, Hsiu-Chuan Chen, Deanna Church, Michele Clamp, Richard R. Copley, Tobias Doerks, Sean R. Eddy, Evan E. Eichler, Terrence S. Furey, James Galagan, James G. R. Gilbert, Cyrus Harmon, Yoshihide Hayashizaki, David Haussler, Henning Hermjakob, Karsten Hokamp, Wonhee Jang, L. Steven Johnson, Thomas A. Jones, Simon Kasif, Arek Kaspryzk, Scot Kennedy, W. James Kent, Paul Kitts, Eugene V. Koonin, Ian Korf, David Kulp, Doron Lancet, Todd M. Lowe, Aoife McLysaght, Tarjei Mikkelsen, John V. Moran, Nicola Mulder, Victor J. Pollara, Chris P. Ponting, Greg Schuler, Jrg Schultz, Guy Slater, Arian F. A. Smit, Elia Stupka, Joseph Szustakowki, Danielle Thierry-Mieg, Jean Thierry-Mieg, Lukas Wagner, John Wallis, Raymond Wheeler, Alan Williams, Yuri I. Wolf, Kenneth H. Wolfe, Shiaw-Pyng Yang, Ru-Fang Yeh, Francis Collins, Mark S. Guyer, Jane Peterson, Adam Felsenfeld, Kris A. Wetterstrand, Richard M. Myers, Jeremy Schmutz, Mark Dickson, Jane Grimwood, David R. Cox, Maynard V. Olson, Rajinder Kaul, Christopher Raymond, Nobuyoshi Shimizu, Kazuhiko Kawasaki, Shinsei Minoshima, Glen A. Evans, Maria Athanasiou, Roger Schultz, Aristides Patri-

nos, Michael J. Morgan, International Human Genome Sequencing Consortium, Center for Genome Research: Whitehead Institute for Biomedical Research, The Sanger Centre:, Washington University Genome Sequencing Center, US DOE Joint Genome Institute:, Baylor College of Medicine Human Genome Sequencing Center:, RIKEN Genomic Sciences Center:, Genoscope and CNRS UMR-8030:, Institute of Molecular Biotechnology: Department of Genome Analysis, GTC Sequencing Center:, Beijing Genomics Institute/Human Genome Center:, The Institute for Systems Biology: Multimegabase Sequencing Center, Stanford Genome Technology Center:, University of Oklahoma's Advanced Center for Genome Technology:, Max Planck Institute for Molecular Genetics:, Lita Annenberg Hazen Genome Center: Cold Spring Harbor Laboratory, GBFGerman Research Centre for Biotechnology:, also includes individuals listed under other headings): *Genome Analysis Group (listed in alphabetical order, US National Institutes of Health: Scientific management: National Human Genome Research Institute, Stanford Human Genome Center:, University of Washington Genome Center:, Keio University School of Medicine: Department of Molecular Biology, University of Texas Southwestern Medical Center at Dallas:, US Department of Energy: Office of Science, and The Wellcome Trust:. "Initial sequencing and analysis of the human genome." *Nature*, **409**(6822):860–921, February 2001. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6822 Primary_atype: Research Publisher: Nature Publishing Group.

[LNA17] Guillaume Lemaitre, Fernando Nogueira, and Christos K Aridas. "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning." *Journal of Machine Learning Research*, **18**(17):1–5, 2017.

[LNC21] Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. "Video swin transformer." *arXiv preprint arXiv:2106.13230*, 2021.

[LP13] Ken Lee and Zdenka Pausova. "Cigarette smoking and DNA methylation." *Frontiers in Genetics*, **4**:132, 2013.

[LPT16] Wei Luo, Dinh Phung, Truyen Tran, Sunil Gupta, Santu Rana, Chandan Karmakar, Alistair Shilton, John Yearwood, Nevenka Dimitrova, Tu Bao Ho, Svetha Venkatesh, and Michael Berk. "Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View." *Journal of Medical Internet Research*, **18**(12):e323, 2016.

[LRK11] Magdalena Lewandowska, Jacek Rumiski, Tomasz Kocejko, and Jedrzej Nowak. "Measuring pulse rate with a webcam A non-contact method for evaluating cardiac activity." In *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 405–410, September 2011.

[LST13] Li-wei H. Lehman, Mohammed Saeed, Daniel Talmor, Roger Mark, and Atul Malhotra. "Methods of Blood Pressure Measurement in the ICU." *Critical care medicine*, **41**(1):34–40, January 2013.

[LTN19] Huiying Liang, Brian Y. Tsui, Hao Ni, Carolina C. S. Valentim, Sally L. Baxter, Guangjian Liu, Wenjia Cai, Daniel S. Kermany, Xin Sun, Jiancong Chen, Liya He, Jie Zhu, Pin Tian, Hua Shao, Lianghong Zheng, Rui Hou, Sierra Hewett, Gen Li, Ping Liang, Xuan Zang, Zhiqi Zhang, Liyan Pan, Huimin Cai, Rujuan Ling, Shuhua Li, Yongwang Cui, Shusheng Tang, Hong Ye, Xiaoyan Huang, Waner He, Wenqing Liang, Qing Zhang, Jianmin Jiang, Wei Yu, Jianqun Gao, Wanxing Ou, Yingmin Deng, Qiaozhen Hou, Bei Wang, Cuichan Yao, Yan Liang, Shu Zhang, Yaou Duan, Runze Zhang, Sarah Gibson, Charlotte L. Zhang, Oulan Li, Edward D. Zhang, Gabriel Karin, Nathan Nguyen, Xiaokang Wu, Cindy Wen, Jie Xu, Wenqin Xu, Bochu Wang, Winston Wang, Jing Li, Bianca Pizzato, Caroline Bao, Daoman Xiang, Wanting He, Suiqin He, Yugui Zhou, Weldon Haw, Michael Goldbaum, Adriana Tremoulet, Chun-Nan Hsu, Hannah Carter, Long Zhu, Kang Zhang, and Huimin Xia. "Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence." *Nature Medicine*, **25**(3):433–438, 2019.

[LTP17] Ira L. Leeds, Brindusa Truta, Alyssa M. Parian, Sophia Y. Chen, Jonathan E. Efron, Susan L. Gearhart, Bashar Safar, and Sandy H. Fang. "Early Surgical Intervention for Acute Ulcerative Colitis Is Associated with Improved Postoperative Outcomes." *Journal of Gastrointestinal Surgery*, 2017.

[LV20] Cathryn M. Lewis and Evangelos Vassos. "Polygenic risk scores: from research tools to clinical instruments." *Genome Medicine*, **12**(1):44, 2020.

[MA86] J Martin Bland and DouglasG. Altman. "STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT." *The Lancet*, **327**(8476):307–310, 1986.

[MBJ16] Paige Maas, Myrto Barrdahl, Amit D. Joshi, Paul L. Auer, Mia M. Gaudet, Roger L. Milne, Fredrick R. Schumacher, William F. Anderson, David Check, Subham Chattopadhyay, Laura Baglietto, Christine D. Berg, Stephen J. Chanock, David G. Cox, Jonine D. Figueroa, Mitchell H. Gail, Barry I. Graubard, Christopher A. Haiman, Susan E. Hankinson, Robert N. Hoover, Claudine Isaacs, Laurence N. Kolonel, Loic Le Marchand, I-Min Lee, Sara Lindström, Kim Overvad, Isabelle Romieu, Maria-Jose Sanchez, Melissa C. Southey, Daniel O. Stram, Rosario Tumino, Tyler J. VanderWeele, Walter C. Willett, Shumin Zhang, Julie E. Buring, Federico Canzian, Susan M. Gapstur, Brian E. Henderson, David J. Hunter, Graham G Giles, Ross L. Prentice, Regina G. Ziegler, Peter Kraft, Montse Garcia-Closas, and Nilanjan Chatterjee. "Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States." *JAMA Oncology*, **2**(10):1295–1302, 10 2016.

[MCP07] Dorinna D. Mendoza, Howard A. Cooper, and Julio A. Panza. "Cardiac power output predicts mortality across a broad spectrum of patients with acute cardiac disease." *American Heart Journal*, **153**(3):366–370, March 2007.

[MGP14] Daniel McDuff, Sarah Gontarek, and Rosalind W. Picard. "Remote Detection of Photoplethysmographic Systolic and Diastolic Peaks Using a Digital Camera." *IEEE Transactions on Biomedical Engineering*, **61**(12):2948–2954, December 2014.

[MHT10] Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. "Spectral Regularization Algorithms for Learning Large Incomplete Matrices." *Journal of machine learning research : JMLR*, **11**:2287–2322, March 2010.

[MHW20] Daniel McDuff, Javier Hernandez, Erroll Wood, Xin Liu, and Tadas Baltrusaitis. "Advancing Non-Contact Vital Sign Measurement using Synthetic Avatars." *arXiv preprint arXiv:2010.12949*, 2020.

[MKK19] Alicia R. Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. "Clinical use of current polygenic risk scores may exacerbate health disparities." *Nature Genetics*, **51**(4):584–591, April 2019. Number: 4 Publisher: Nature Publishing Group.

[MLK16] Riccardo Miotto, Li Li, Brian A. Kidd, and Joel T. Dudley. "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records." *Scientific Reports*, **6**(1):26094, 2016.

[MMD19] Nasim Mavaddat, Kyriaki Michailidou, Joe Dennis, Michael Lush, Laura Fachal, Andrew Lee, Jonathan P Tyrer, Ting-Huei Chen, Qin Wang, Manjeet K Bolla, Xin Yang, Muriel A Adank, Thomas Ahearn, Kristiina Aittomäki, Jamie Allen, Irene L Andrulis, Hoda Anton-Culver, Natalia N Antonenkova, Volker Arndt, Kristan J Aronson, Paul L Auer, Päivi Auvinen, Myrto Barrdahl, Laura E Beane Freeman, Matthias W Beckmann, Sabine Behrens, Javier Benitez, Marina Bermisheva, Leslie Bernstein, Carl Blomqvist, Natalia V Bogdanova, Stig E Bojesen, Bernardo Bonanni, Anne-Lise Børresen-Dale, Hiltrud Brauch, Michael Bremer, Hermann Brenner, Adam Brentnall, Ian W Brock, Angela Brooks-Wilson, Sara Y Brucker, Thomas Brüning, Barbara Burwinkel, Daniele Campa, Brian D Carter, Jose E Castelao, Stephen J Chanock, Rowan Chlebowski, Hans Christiansen, Christine L Clarke, J Margriet Collée, Emilie Cordina-Duverger, Sten Cornelissen, Fergus J Couch, Angela Cox, Simon S Cross, Kamila Czene, Mary B Daly, Peter Devilee, Thilo Dörk, Isabel Dos-Santos-Silva, Martine Dumont, Lorraine Durcan, Miriam Dwek, Diana M Eccles, Arif B Ekici, A Heather Eliassen, Carolina Ellberg, Christoph Engel, Mikael Eriksson, D Gareth Evans, Peter A Fasching, Jonine Figueroa, Olivia Fletcher, Henrik Flyger, Asta Försti, Lin Fritschi, Marike Gabrielson, Manuela Gago-Dominguez, Susan M Gapstur,

JoséA García-Sáenz, Mia M Gaudet, Vassilios Georgoulias, Graham G Giles, Irina R Gilyazova, Gord Glendon, Mark S Goldberg, David E Goldgar, Anna González-Neira, Grethe I Grenaker Alnæs, Mervi Grip, Jacek Gronwald, Anne Grundy, Pascal Guénel, Lothar Haeberle, Eric Hahnen, Christopher A Haiman, Niclas Håkansson, Ute Hamann, Susan E Hankinson, Elaine F Harkness, Steven N Hart, Wei He, Alexander Hein, Jane Heyworth, Peter Hillemanns, Antoinette Hollestelle, Maartje J Hooning, Robert N Hoover, John L Hopper, Anthony Howell, Guanmengqian Huang, Keith Humphreys, David J Hunter, Milena Jakimovska, Anna Jakubowska, Wolfgang Janni, Esther M John, Nichola Johnson, Michael E Jones, Arja Jukkola-Vuorinen, Audrey Jung, Rudolf Kaaks, Katarzyna Kaczmarek, Vesa Kataja, Renske Keeman, Michael J Kerin, Elza Khusnutdinova, Johanna I Kiiski, Julia A Knight, Yon-Dschun Ko, Veli-Matti Kosma, Stella Koutros, Vessela N Kristensen, Ute Krüger, Tabea Kühl, Diether Lambrechts, Loic Le Marchand, Eunjung Lee, Flavio Lejbkowicz, Jenna Lilyquist, Annika Lindblom, Sara Lindström, Jolanta Lissowska, Wing-Yee Lo, Sibylle Loibl, Jirong Long, Jan Lubiński, Michael P Lux, Robert J MacInnis, Tom Maishman, Enes Makalic, Ivana Maleva Kostovska, Arto Mannermaa, Siranoush Manoukian, Sara Margolin, John W M Martens, Maria Elena Martinez, Dimitrios Mavroudis, Catriona McLean, Alfons Meindl, Usha Menon, Pooja Middha, Nicola Miller, Fernando Moreno, Anna Marie Mulligan, Claire Mulot, Victor M Muñoz-Garzon, Susan L Neuhausen, Heli Nevanlinna, Patrick Neven, William G Newman, Sune F Nielsen, Børge G Nordestgaard, Aaron Norman, Kenneth Offit, Janet E Olson, Håkan Olsson, Nick Orr, V Shane Pankratz, Tjoung-Won Park-Simon, Jose I A Perez, Clara Pérez-Barrios, Paolo Peterlongo, Julian Peto, Mila Pinchev, Dijana Plaseska-Karanfilska, Eric C Polley, Ross Prentice, Nadege Presneau, Darya Prokofyeva, Kristen Purrington, Katri Pylkäs, Brigitte Rack, Paolo Radice, Rohini Rau-Murthy, Gad Rennert, Hedy S Rennert, Valerie Rhenius, Mark Robson, Atocha Romero, Kathryn J Ruddy, Matthias Ruebner, Emmanouil Saloustros, Dale P Sandler, Elinor J Sawyer, Daniel F Schmidt, Rita K Schmutzler, Andreas Schneeweiss, Minouk J Schoemaker, Fredrick Schumacher, Peter Schürmann, Lukas Schwentner, Christopher Scott, Rodney J Scott, Caroline Seynaeve, Mitul Shah, Mark E Sherman, Martha J Shrubsole, Xiao-Ou Shu, Susan Slager, Ann Smeets, Christof Sohn, Penny Soucy, Melissa C Southey, John J Spinelli, Christa Stegmaier, Jennifer Stone, Anthony J Swerdlow, Rulla M Tamimi, William J Tapper, Jack A Taylor, Mary Beth Terry, Kathrin Thöne, Rob A E M Tollenaar, Ian Tomlinson, Thérèse Truong, Maria Tzardi, Hans-Ulrich Ulmer, Michael Untch, Celine M Vachon, Elke M van Veen, Joseph Vijai, Clarice R Weinberg, Camilla Wendt, Alice S Whittemore, Hans Wildiers, Walter Willett, Robert Winqvist, Alicja Wolk, Xiaohong R Yang, Drakoulis Yannoukakos, Yan Zhang, Wei Zheng, Argyrios Ziogas, Alison M Dunning, Deborah J Thompson, Georgia Chenevix-Trench, Jenny Chang-Claude, Marjanka K Schmidt, Per Hall, Roger L Milne, Paul D P Pharoah, Antonis C Antoniou, Nilanjan Chatterjee,

Peter Kraft, Montserrat García-Closas, Jacques Simard, and Douglas F Easton. "Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes." *Am J Hum Genet*, **104**(1):21–34, Jan 2019.

[MNA16] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation." *arXiv:1606.04797 [cs]*, June 2016. arXiv: 1606.04797.

[MNM18] Kamal Maheshwari, Brian H. Nathanson, Sibyl H. Munson, Victor Khangulov, Mitali Stevens, Hussain Badani, Ashish K. Khanna, and Daniel I. Sessler. "The relationship between ICU hypotension and in-hospital mortality and morbidity in septic patients." *Intensive Care Medicine*, **44**(6):857–867, June 2018.

[MS18] Agnes S Meidert and Bernd Saugel. "Techniques for Non-Invasive Monitoring of Arterial Blood Pressure." *Frontiers in medicine*, **4**:231–231, January 2018.

[MSA01] David Moher, Kenneth F Schulz, and Douglas G Altman. "The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials." *The Lancet*, **357**(9263):1191–1194, 2001.

[MWG05] Catherine A McCarty, Russell A Wilke, Philip F Giampietro, Steve D Wesbrook, and Michael D Caldwell. "Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank." *Personalized Medicine*, **2**(1):49–79, 2021/01/04 2005.

[MZM13] Erika L. Moen, Xu Zhang, Wenbo Mu, Shannon M. Delaney, Claudia Wing, Jennifer McQuade, Jamie Myers, Lucy A. Godley, M. Eileen Dolan, and Wei Zhang. "Genome-Wide Variation of Cytosine Modifications Between European and African Populations and the Implications for Complex Traits." *Genetics*, **194**(4):987–996, August 2013. Publisher: Genetics Section: Investigations.

[NNB04] D. Nister, O. Naroditsky, and J. Bergen. "Visual odometry." In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pp. I–I, June 2004. ISSN: 1063-6919.

[PCA18] Ryan Poplin, Pi-Chuan Chang, David Alexander, Scott Schwartz, Thomas Colthurst, Alexander Ku, Dan Newburger, Jojo Dijamco, Nam Nguyen, Pegah T. Afshar, Sam S. Gross, Lizzie Dorfman, Cory Y. McLean, and Mark A. DePristo. "A universal SNP and small-indel variant caller using deep neural networks." *Nature Biotechnology*, **36**(10):983–987, November 2018. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 10 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Genome informatics;Genomics;Machine learning Subject_term_id: genome-informatics;genomics;machine-learning.

[PHJ06] Rupert M Pearse, David A Harrison, Philip James, David Watson, Charles Hinds, Andrew Rhodes, R Michael Grounds, and E David Bennett. "Identification and characterisation of the high-risk surgical population in the United Kingdom." *Critical Care*, **10**(3):R81, June 2006.

[PMP10] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. "Advancements in noncontact, multiparameter physiological measurements using a webcam." *IEEE transactions on biomedical engineering*, **58**(1):7–11, 2010.

[PNT07] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, ManuelA.R. Ferreira, David Bender, Julian Maller, Pamela Sklar, PaulI.W. deBakker, MarkJ. Daly, and PakC. Sham. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics*, **81**(3):559–575, September 2007.

[PTG20] Tania Pereira, Nate Tran, Kais Gadhoumi, Michele M. Pelter, Duc H. Do, Randall J. Lee, Rene Colorado, Karl Meisel, and Xiao Hu. "Photoplethysmography based atrial fibrillation detection: a review." *npj Digital Medicine*, **3**(1):1–12, January 2020.

[PVG11] Fabian Pedregosa, Gal Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and douard Duchesnay. "Scikit-learn: Machine Learning in Python." *J. Mach. Learn. Res.*, **12**:2825–2830, November 2011. Publisher: JMLR.org.

[QLC11] Hude Quan, Bing Li, Chantal M Couris, Kiyohide Fushimi, Patrick Graham, Phil Hider, Jean-Marie Januel, and Vijaya Sundararajan. "Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries." *American Journal of Epidemiology*, **173**(6):676–682, March 2011.

[RBD11] Vardhman K. Rakyan, Huriya Beyan, Thomas A. Down, Mohammed I. Hawa, Siarhei Maslau, Deeqo Aden, Antoine Daunay, Florence Busato, Charles A. Mein, Burkhard Manfras, Kerith-Rae M. Dias, Christopher G. Bell, Jörg Tost, Bernhard O. Boehm, Stephan Beck, and R. David Leslie. "Identification of Type 1 Diabetes–Associated DNA Methylation Variable Positions That Precede Disease Diagnosis." *PLOS Genetics*, **7**(9):1–9, 09 2011.

[RBS17] Tiffany C Randolph, Samuel Broderick, Linda K Shaw, Karen Chiswell, Robert J Mentz, Valentina Kutyifa, Eric J Velazquez, Francis R Gilliam, and Kevin L Thomas. "Race and Sex Differences in QRS Interval and Associated Outcome Among Patients with Left Ventricular Systolic Dysfunction." *Journal of the American Heart Association*, **6**(3):e004381, 03 2017.

[RD12] Caroline L Relton and George Davey Smith. "Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease." *Int J Epidemiol*, **41**(1):161–176, Feb 2012.

[RFO17] Alex Rubinsteyn, Sergey Feldman, Tim O'Donnell, and Brett Beaulieu-Jones. "hammerlab/fancyimpute: Version 0.2.0." September 2017.

[ROC18] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Peter J. Liu, Xiaobing Liu, Mimi Sun, Patrik Sundberg, Hector Yee, Kun Zhang, Gavin E. Duggan, Gerardo Flores, Michaela Hardt, Jamie Irvine, Quoc Le, Kurt Litsch, Jake Marcus, Alexander Mossin, Justin Tansuwan, De Wang, James Wexler, Jimbo Wilson, Dana Ludwig, Samuel L. Volchenboum, Katherine Chou, Michael Pearson, Srinivasan Madabushi, Nigam H. Shah, Atul J. Butte, Michael Howell, Claire Cui, Greg Corrado, and Jeff Dean. "Scalable and accurate deep learning for EHR - supplement." *npj Digital Medicine*, 2018.

[RPB08] D M Roden, J M Pulley, M A Basford, G R Bernard, E W Clayton, J R Balser, and D R Masys. "Development of a large-scale de-identified DNA biobank to enable personalized medicine." *Clin Pharmacol Ther*, **84**(3):362–369, Sep 2008.

[RRN21] Honnesh Rohmetra, Navaneeth Raghunath, Pratik Narang, Vinay Chamola, Mohsen Guizani, and Naga Rajiv Lakkaniga. "AI-enabled remote monitoring of vital signs for COVID-19: methods, prospects and challenges." *Computing*, March 2021.

[RRP20] Antnio H. Ribeiro, Manoel Horta Ribeiro, Gabriela M. M. Paixo, Derick M. Oliveira, Paulo R. Gomes, Jssica A. Canazart, Milton P. S. Ferreira, Carl R. Andersson, Peter W. Macfarlane, Wagner Meira Jr, Thomas B. Schn, and Antonio Luiz P. Ribeiro. "Automatic diagnosis of the 12-lead ECG using a deep neural network." *Nature Communications*, **11**(1):1760, April 2020. Number: 1 Publisher: Nature Publishing Group.

[RSM08] Andrew Reisner, PhillipA. Shaltis, Devin McCombie, HHarry Asada, DavidS. Warner, and MarkA. Warner. "Utility of the Photoplethysmogram in Circulatory Monitoring." *Anesthesiology*, **108**(5):950–958, May 2008.

[RSS17] Elior Rahmani, Liat Shenhav, Regev Schweiger, Paul Yousefi, Karen Huen, Brenda Eskenazi, Celeste Eng, Scott Huntsman, Donglei Hu, Joshua Galanter, Sam S. Oh, Melanie Waldenberger, Konstantin Strauch, Harald Grallert, Thomas Meitinger, Christian Gieger, Nina Holland, Esteban G. Burchard, Noah Zaitlen, and Eran Halperin. "Genome-wide methylation data mirror ancestry information." *Epigenetics & Chromatin*, **10**(1):1, January 2017.

[RZB16] Elior Rahmani, Noah Zaitlen, Yael Baran, Celeste Eng, Donglei Hu, Joshua Galanter, Sam Oh, Esteban G Burchard, Eleazar Eskin, James Zou, and Eran

Halperin. "Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies." *Nature Methods*, **13**:443, March 2016. Publisher: Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

[SAB18] Fredrick R Schumacher, Ali Amin Al Olama, Sonja I Berndt, Sara Benlloch, Mahbubl Ahmed, Edward J Saunders, Tokhir Dadaev, Daniel Leongamornlert, Ezequiel Anokian, Clara Cieza-Borrella, Chee Goh, Mark N Brook, Xin Sheng, Laura Fachal, Joe Dennis, Jonathan Tyrer, Kenneth Muir, Artitaya Lophatananon, Victoria L Stevens, Susan M Gapstur, Brian D Carter, Catherine M Tangen, Phyllis J Goodman, Ian M Jr Thompson, Jyotsna Batra, Suzanne Chambers, Leire Moya, Judith Clements, Lisa Horvath, Wayne Tilley, Gail P Risbridger, Henrik Gronberg, Markus Aly, Tobias Nordström, Paul Pharoah, Nora Pashayan, Johanna Schleutker, Teuvo L J Tammela, Csilla Sipeky, Anssi Auvinen, Demetrius Albanes, Stephanie Weinstein, Alicja Wolk, Niclas Håkansson, Catharine M L West, Alison M Dunning, Neil Burnet, Lorelei A Mucci, Edward Giovannucci, Gerald L Andriole, Olivier Cussenot, Géraldine Cancel-Tassin, Stella Koutros, Laura E Beane Freeman, Karina Dalsgaard Sorensen, Torben Falck Orntoft, Michael Borre, Lovise Maehle, Eli Marie Grindedal, David E Neal, Jenny L Donovan, Freddie C Hamdy, Richard M Martin, Ruth C Travis, Tim J Key, Robert J Hamilton, Neil E Fleshner, Antonio Finelli, Sue Ann Ingles, Mariana C Stern, Barry S Rosenstein, Sarah L Kerns, Harry Ostrer, Yong-Jie Lu, Hong-Wei Zhang, Ninghan Feng, Xueying Mao, Xin Guo, Guomin Wang, Zan Sun, Graham G Giles, Melissa C Southey, Robert J MacInnis, Liesel M FitzGerald, Adam S Kibel, Bettina F Drake, Ana Vega, Antonio Gómez-Caamaño, Robert Szulkin, Martin Eklund, Manolis Kogevinas, Javier Llorca, Gemma Castaño-Vinyals, Kathryn L Penney, Meir Stampfer, Jong Y Park, Thomas A Sellers, Hui-Yi Lin, Janet L Stanford, Cezary Cybulski, Dominika Wokolorczyk, Jan Lubinski, Elaine A Ostrander, Milan S Geybels, Børge G Nordestgaard, Sune F Nielsen, Maren Weischer, Rasmus Bisbjerg, Martin Andreas Røder, Peter Iversen, Hermann Brenner, Katarina Cuk, Bernd Holleczek, Christiane Maier, Manuel Luedeke, Thomas Schnoeller, Jeri Kim, Christopher J Logothetis, Esther M John, Manuel R Teixeira, Paula Paulo, Marta Cardoso, Susan L Neuhausen, Linda Steele, Yuan Chun Ding, Kim De Ruyck, Gert De Meerleer, Piet Ost, Azad Razack, Jasmine Lim, Soo-Hwang Teo, Daniel W Lin, Lisa F Newcomb, Davor Lessel, Marija Gamulin, Tomislav Kulis, Radka Kaneva, Nawaid Usmani, Sandeep Singhal, Chavdar Slavov, Vanio Mitev, Matthew Parliament, Frank Claessens, Steven Joniau, Thomas Van den Broeck, Samantha Larkin, Paul A Townsend, Claire Aukim-Hastie, Manuela Gago-Dominguez, Jose Esteban Castelao, Maria Elena Martinez, Monique J Roobol, Guido Jenster, Ron H N van Schaik, Florence Menegaux, Thérèse Truong, Yves Akoli Koudou, Jianfeng Xu, Kay-Tee Khaw, Lisa Cannon-Albright, Hardev Pandha, Agnieszka Michael, Stephen N Thibodeau,

Shannon K McDonnell, Daniel J Schaid, Sara Lindstrom, Constance Turman, Jing Ma, David J Hunter, Elio Riboli, Afshan Siddiq, Federico Canzian, Laurence N Kolonel, Loic Le Marchand, Robert N Hoover, Mitchell J Machiela, Zuxi Cui, Peter Kraft, Christopher I Amos, David V Conti, Douglas F Easton, Fredrik Wiklund, Stephen J Chanock, Brian E Henderson, Zsofia Kote-Jarai, Christopher A Haiman, and Rosalind A Eeles. "Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci." *Nat Genet*, **50**(7):928–936, Jul 2018.

[SCC21] Rencheng Song, Huan Chen, Juan Cheng, Chang Li, Yu Liu, and Xun Chen. "PulseGAN: Learning to generate realistic pulse waveforms in remote photoplethysmography." *IEEE Journal of Biomedical and Health Informatics*, **25**(5):1373–1384, 2021.

[SDZ17] Peng Su, Xiaorong Ding, Yuanting Zhang, Ye Li, and Ni Zhao. *Predicting Blood Pressure with Deep Bidirectional LSTM Network*. May 2017.

[SGA15a] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLOS Medicine*, **12**(3):e1001779–, 03 2015.

[SGA15b] Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, Bette Liu, Paul Matthews, Giok Ong, Jill Pell, Alan Silman, Alan Young, Tim Sprosen, Tim Peakman, and Rory Collins. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLOS Medicine*, **12**(3):e1001779, March 2015. Publisher: Public Library of Science.

[SKN16] C. Sideris, H. Kalantarian, E. Nemati, and M. Sarrafzadeh. "Building Continuous Arterial Blood Pressure Prediction Models Using Recurrent Networks." In *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 1–5, May 2016.

[SLM16] Matthew J.G. Sigakis, Lisa R. Leffert, Hooman Mirzakhani, Nadir Sharawi, Baskar Rajala, William M. Callaghan, Elena V. Kuklina, Andreea A. Creanga, Jill M. Mhyre, and Brian T. Bateman. "The Validity of Discharge Billing Codes Reflecting Severe Maternal Morbidity." *Anesthesia and Analgesia*, 2016.

[SR15] Takaya Saito and Marc Rehmsmeier. "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced

Datasets." *PLOS ONE*, **10**(3):e0118432, March 2015. Publisher: Public Library of Science.

[SSM10] M D Sessler Daniel I., Ph.D. Sigl Jeffrey C., Ph.D. Manberg Paul J., M D Kelley Scott D., M D Schubert M.B.A., Armin, and M S Chamoun Nassib G. "Broadly Applicable Risk Stratification System for Predicting Duration of Hospitalization and Mortality." *Anesthesiology*, **113**(5):1026–1037, November 2010.

[SSW15] Paula Singmann, Doron Shem-Tov, Simone Wahl, Harald Grallert, Giovanni Fiorito, So-Youn Shin, Katharina Schramm, Petra Wolf, Sonja Kunze, Yael Baran, Simonetta Guarrera, Paolo Vineis, Vittorio Krogh, Salvatore Panico, Rosario Tumino, Anja Kretschmer, Christian Gieger, Annette Peters, Holger Prokisch, Caroline L. Relton, Giuseppe Matullo, Thomas Illig, Melanie Waldenberger, and Eran Halperin. "Characterization of whole-genome autosomal differences of DNA methylation between men and women." *Epigenetics & Chromatin*, **8**(1):43, Oct 2015.

[STA21] Nasa Sinnott-Armstrong, Yosuke Tanigawa, David Amar, Nina Mars, Christian Benner, Matthew Aguirre, Guhan Ram Venkataraman, Michael Wainberg, Hanna M. Ollila, Tuomo Kiiskinen, Aki S. Havulinna, James P. Pirruccello, Junyang Qian, Anna Shcherbina, Fatima Rodriguez, Themistocles L. Assimes, Vineeta Agarwala, Robert Tibshirani, Trevor Hastie, Samuli Ripatti, Jonathan K. Pritchard, Mark J. Daly, Manuel A. Rivas, and FinnGen. "Genetics of 35 blood and urine biomarkers in the UK Biobank." *Nature Genetics*, **53**(2):185–194, 2021.

[Tak93] K Takazawa. "Clinical usefulness of the second derivative of a plethysmogram (acceleration plethysmogram)." *J Cardiol*, **23**:207–217, 1993.

[TMP20] Daniel Trejo Banos, Daniel L. McCartney, Marion Patxot, Lucas Anchieri, Thomas Battram, Colette Christiansen, Ricardo Costeira, Rosie M. Walker, Stewart W. Morris, Archie Campbell, Qian Zhang, David J. Porteous, Allan F. McRae, Naomi R. Wray, Peter M. Visscher, Chris S. Haley, Kathryn L. Evans, Ian J. Deary, Andrew M. McIntosh, Gibran Hemani, Jordana T. Bell, Riccardo E. Marioni, and Matthew R. Robinson. "Bayesian reassessment of the epigenetic architecture of complex traits." *Nature Communications*, **11**(1):2865, 2020.

[TO07] Chihiro Takano and Yuji Ohta. "Heart rate measurement based on a time-lapse image." *Medical engineering & physics*, **29**(8):853–857, 2007.

[TOM18] Shoichiro Takeda, Kazuki Okami, Dan Mikami, Megumi Isogai, and Hideaki Kimata. "Jerk-aware video acceleration magnification." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1769–1777, 2018.

[TTF98] Kenji Takazawa, Nobuhiro Tanaka, Masami Fujita, Osamu Matsuoka, Tokuyu Saiki, Masaru Aikawa, Sinobu Tamura, and Chiharu Ibukiyama. "Assessment of

Vasoactive Agents and Vascular Aging by the Second Derivative of Photoplethysmogram Waveform." *Hypertension*, **32**(2):365–370, August 1998.

[TZ03] X.F. Teng and Y.T. Zhang. "Continuous and noninvasive estimation of arterial blood pressure using a photoplethysmographic approach." In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*, volume 4, pp. 3153–3156. IEEE, 2003.

[VBA20] Dragana Vuckovic, Erik L. Bao, Parsa Akbari, Caleb A. Lareau, Abdou Mousas, Tao Jiang, Ming-Huei Chen, Laura M. Raffield, Manuel Tardaguila, Jennifer E. Huffman, Scott C. Ritchie, Karyn Megy, Hannes Ponstingl, Christopher J. Penkett, Patrick K. Albers, Emilie M. Wigdor, Saori Sakaue, Arden Moscati, Regina Manansala, Ken Sin Lo, Huijun Qian, Masato Akiyama, Traci M. Bartz, Yoav Ben-Shlomo, Andrew Beswick, Jette Bork-Jensen, Erwin P. Bottinger, Jennifer A. Brody, Frank J.A. van Rooij, Kumaraswamy N. Chitrala, Peter W.F. Wilson, Hélène Choquet, John Danesh, Emanuele Di Angelantonio, Niki Dimou, Jingzhong Ding, Paul Elliott, Tõnu Esko, Michele K. Evans, Stephan B. Felix, James S. Floyd, Linda Broer, Niels Grarup, Michael H. Guo, Qi Guo, Andreas Greinacher, Jeff Haessler, Torben Hansen, Joanna M.M. Howson, Wei Huang, Eric Jorgenson, Tim Kacprowski, Mika Kähönen, Yoichiro Kamatani, Masahiro Kanai, Savita Karthikeyan, Fotios Koskeridis, Leslie A. Lange, Terho Lehtimäki, Allan Linneberg, Yongmei Liu, Leo-Pekka Lyytikäinen, Ani Manichaikul, Koichi Matsuda, Karen L. Mohlke, Nina Mononen, Yoshinori Murakami, Girish N. Nadkarni, Kjell Nikus, Nathan Pankratz, Oluf Pedersen, Michael Preuss, Bruce M. Psaty, Olli T. Raitakari, Stephen S. Rich, Benjamin A.T. Rodriguez, Jonathan D. Rosen, Jerome I. Rotter, Petra Schubert, Cassandra N. Spracklen, Praveen Surendran, Hua Tang, Jean-Claude Tardif, Mohsen Ghanbari, Uwe Völker, Henry Völzke, Nicholas A. Watkins, Stefan Weiss, Na Cai, Kousik Kundu, Stephen B. Watt, Klaudia Walter, Alan B. Zonderman, Kelly Cho, Yun Li, Ruth J.F. Loos, Julian C. Knight, Michel Georges, Oliver Stegle, Evangelos Evangelou, Yukinori Okada, David J. Roberts, Michael Inouye, Andrew D. Johnson, Paul L. Auer, William J. Astle, Alexander P. Reiner, Adam S. Butterworth, Willem H. Ouwehand, Guillaume Lettre, Vijay G. Sankaran, and Nicole Soranzo. "The Polygenic and Monogenic Basis of Blood Traits and Diseases." *Cell*, **182**(5):1214–1231.e11, 2020.

[VGO20] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, lhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero,

Charles R. Harris, Anne M. Archibald, Antnio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. "SciPy 1.0: fundamental algorithms for scientific computing in Python." *Nature Methods*, **17**(3):261–272, March 2020. Number: 3 Publisher: Nature Publishing Group.

[VSN08] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. "Remote plethysmographic imaging using ambient light." *Optics express*, **16**(26):21434–21445, 2008.

[VVH70] Charles J Vacanti, Robert J Van Houten, and Robert C Hill. "A statistical analysis of the relationship of physical status to postoperative mortality in 68.368 cases." *Anesth Analg*, **49**:564, 1970.

[Was18] Jack O Wasey. "icd: Tools for Working with ICD-9 and ICD-10 Codes, and Finding Comorbidities." 2018.

[WBS00] Ting Wu, Vladimir Blazek, and Hans Juergen Schmitt. "Photoplethysmography imaging: a new noninvasive and noncontact method for mapping of the dermal perfusion changes." In *Optical Techniques and Instrumentation for the Measurement of Blood Composition, Structure, and Dynamics*, volume 4163, pp. 62–70. International Society for Optics and Photonics, 2000.

[WBS16] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard De Haan. "Algorithmic principles of remote PPG." *IEEE Transactions on Biomedical Engineering*, **64**(7):1479–1491, 2016.

[WDG13] Michael Walsh, Philip J. Devereaux, Amit X. Garg, Andrea Kurz, Alparslan Turan, Reitze N. Rodseth, Jacek Cywinski, Lehana Thabane, and Daniel I. Sessler. "Relationship between Intraoperative Mean Arterial Pressure and Clinical Outcomes after Noncardiac SurgeryToward an Empirical Definition of Hypotension." *Anesthesiology: The Journal of the American Society of Anesthesiologists*, **119**(3):507–515, September 2013.

[WGH20] Marije Wijnberge, Bart F. Geerts, Liselotte Hol, Nikki Lemmers, Marijn P. Mulder, Patrick Berge, Jimmy Schenk, Lotte E. Terwindt, Markus W. Hollmann, Alexander P. Vlaar, and Denise P. Veelo. "Effect of a Machine LearningDerived Early Warning System for Intraoperative Hypotension vs Standard Care on Depth and Duration of Intraoperative Hypotension During Elective Noncardiac Surgery: The HYPE Randomized Clinical Trial." *JAMA*, February 2020.

[WKW16] Judith A. R. van Waes, Wilton A. van Klei, Duminda N. Wijeysundera, Leo van Wolfswinkel, Thomas F. Lindsay, and W. Scott Beattie. "Association between Intraoperative Hypotension and Myocardial Injury after Vascular Surgery." *Anesthesiology*, **124**(1):35–44, January 2016.

[WRS12] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. "Eulerian video magnification for revealing subtle changes in the world." *ACM transactions on graphics (TOG)*, **31**(4):1–8, 2012.

[WSB13] Alexandra J White, Dale P Sandler, Sophia C E Bolick, Zongli Xu, Jack A Taylor, and Lisa A DeRoo. "Recreational and household physical activity at different time points and DNA global methylation." *Eur J Cancer*, **49**(9):2199–2206, Jun 2013.

[WWS96] U Wolters, T Wolf, H Sttzer, and T Schrder. "ASA classification and perioperative variables as predictors of postoperative outcome." *BJA: British Journal of Anaesthesia*, **77**(2):217–222, August 1996.

[XMZ19] Xiaoman Xing, Zhimin Ma, Mingyou Zhang, Ying Zhou, Wenfei Dong, and Mingxuan Song. "An Unobtrusive and Calibration-free Blood Pressure Estimation Method using Photoplethysmography and Biometrics." *Scientific Reports*, **9**(1):8611, June 2019.

[XNL16] Zongli Xu, Liang Niu, Leping Li, and Jack A. Taylor. "ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip." *Nucleic Acids Research*, **44**(3):e20, February 2016.

[XS16] Xiaoman Xing and Mingshan Sun. "Optical blood pressure estimation with photoplethysmography and FFT-based neural networks." *Biomedical Optics Express*, **7**(8):3007–3020, 2016.

[YLG11] Jian Yang, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. "GCTA: A Tool for Genome-wide Complex Trait Analysis." *American Journal of Human Genetics*, **88**(1):76–82, January 2011.

[YVG18] Takashige Yamada, Susana Vacas, Yann Gricourt, and Maxime Cannesson. "Improving Perioperative Outcomes Through Minimally Invasive and Non-invasive Hemodynamic Monitoring Techniques." *Frontiers in Medicine*, **5**, 2018.

[ZFW16] Linda Zhang, Daniel Fabbri, and Jonathan P Wanderer. "Data-Driven System for Perioperative Acuity Prediction." In *AMIA*, 2016.

[ZH05] Hui Zou and Trevor Hastie. "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**(2):301–320, 2005. Publisher: Wiley Online Library.

[ZMC11] Fang Fang Zhang, Alfredo Morabia, Joan Carroll, Karina Gonzalez, Kimberly Fulda, Manleen Kaur, Jamboor K. Vishwanatha, Regina M. Santella, and Roberto Cardarelli. "Dietary Patterns Are Associated with Levels of Global Genomic DNA Methylation in a Cancer-Free Population." *The Journal of Nutrition*, **141**(6):1165–1171, 04 2011.

[ZPV17] Yichao Zhang, Silvia L Pintea, and Jan C Van Gemert. "Video acceleration magnification." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 529–537, 2017.

[ZRH19] Bing Zhang, Huihui Ren, Guoyan Huang, Yongqiang Cheng, and Changzhen Hu. "Predicting blood pressure from physiological index data using the SVR algorithm." *BMC Bioinformatics*, **20**(1):109, February 2019.

[ZXX17] Tao Zheng, Wei Xie, Liling Xu, Xiaoying He, Ya Zhang, Mingrong You, Gong Yang, and You Chen. "A machine learning-based framework to identify type 2 diabetes through electronic health records." *International Journal of Medical Informatics*, **97**:120–127, 2017.