Using Machine Learning to Fill in Missing Values in Pulsative Data from Diverse Clinical Datasets

A Thesis

by

Nabin Kim

Approved by:

Dr. Rehg, Advisor
School of Interactive Computing
Georgia Institute of Technology

Date Approved: 05/04/2024

Dr. Ploetz
School of Interactive Computing
Georgia Institute of Technology

Date Approved: 05/02/2024

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Electrocardiogram (ECG) and Photoplethysmography (PPG) signals are physiological signals collected from body sensors, commonly used in diagnosing clinical conditions and monitoring patient conditions over time. However, a significant challenge arises from missing data in these signals, resulting from various causes such as unstable sensor attachment or data transmission loss (Bashar et al., 2019). This missing data severely impacts the clinical utility of ECG and PPG signals. The objective of this research project is to demonstrate the generalizability and improvement of the BDC Transformer, a machine learning model for imputing missing values in waveform data (Xu et al., 2022). Specifically, we aim to apply and enhance the BDC Transformer on the MODS (Multiple Organ Dysfunction Syndrome) dataset from Emory University, which provides a unique opportunity to evaluate machine learning algorithms on a more diverse set of patients and conditions compared to prior studies.

Previous machine learning research on patient health data, including both clinical and waveform data, predominantly relied on the MIMIC-III dataset for training (Johnson, Pollard, & Mark, 2016; Sarika & Elias, 2023; Bashar et al., 2019). While MIMIC-III is widely cited and offers extensive electronic health record and bedside monitoring data, it suffers from limitations in diversity and coverage of ICU care, primarily stemming from a single data source: the Beth Israel Deaconess Medical Center ("Commonwealth of Massachusetts", 2021; Beth Israel Deaconess Medical Center, 2022). The Emory MODS dataset, on the other hand, draws from two different Emory Healthcare systems, encompassing a much broader range of patient characteristics and demographics. With the goal of improving imputation methods, this project aims to assess the performance of the BDC Transformer on the MODS dataset and compare it with other imputation techniques. By addressing the issue of missing data in ECG waveform signals, we seek to enhance disease diagnosis accuracy, treatment planning, and remote disease detection capabilities.

Therefore the goal of this project is to apply and improve the BDC transformer on the MODS dataset and compare its performance to the MIMIC-III dataset. This research emphasizes the need to incorporate comprehensive clinical datasets and diverse patient populations in the development of algorithmic approaches to ensure their generalizability and effectiveness. Ultimately, our work contributes to advancing cardiovascular disease diagnosis and remote monitoring by providing valuable insights into the performance and applicability of machine learning imputation methods.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

Since 1975, cardiovascular diseases (CVD) have consistently been one of the top two causes of death in the United States. In fact, heart disease was the leading cause of mortality in 2015, accounting for 633,842 fatalities, or almost 1 in every 4 deaths. (Lopez, Ballard & Jan, 2022). The accurate diagnosis and monitoring of specific cardiovascular clinical conditions heavily rely on physiological signals collected from body sensors, such as electrocardiogram (ECG) and photoplethysmography (PPG). These signals provide crucial information about a patient's cardiovascular condition, but they are often corrupted for various reasons (Bashar et al., 2019). The presence of missing values in ECG waveform data can lead to inaccurate diagnoses and treatment plans. Therefore, the development of effective imputation methods to address missing data in time series analysis is of paramount importance.

2.1.1 Deep Learning Approaches for Missing Data Imputation

Numerous techniques have been devised to address the difficulties associated with missing data in time series analysis, and deep learning approaches have shown particular promise in this regard (Fang et al. 2020). However, there is a need to further explore and evaluate these methods to understand their strengths and limitations in accurately handling missing values in physiological signals.

2.1.2 Physiological Signals: ECG and PPG

Understanding and diagnosing specific cardiovascular clinical conditions often rely on physiological signals collected from body sensors. Two commonly utilized signals are ECG and PPG. ECG signals measure electrical activity of the heart and PPG signals measure changes in blood volume as the heart pumps blood, and as such, both provide information about a patient's cardiovascular condition such as atrial fibrillation, and monitor patient conditions over time, such as via heart rate variability. However, the data is often corrupted by motion, noise artifacts (MNA), poor electrode contact, and missing values (Bashar et al., 2019). Missing values in ECG waveform data can lead to inaccurate diagnoses and treatment plans. The need for imputation arises due to missing values in ECG and PPG waveform data, which will impact the accuracy of disease diagnosis and treatment plans. This is particularly relevant in the recent development of classification machine learning algorithms used for remote disease diagnosis (Aziz, Ahmed, & Alouini, 2021).

2.2 Previous Imputation methods

2.2.1 State-of-the-Art Imputation Methods

Various imputation methods have been developed to address the challenges of missing data in time series analysis. Fang et al. (2020) provides an overview of three types of time series data imputation methods: deletion methods, traditional methods, and learning-based methods. The study specifically highlights the effectiveness of deep learning methods within the learning-based approach, demonstrating superior performance in various time series data imputation tasks.

Liu et al. (2019) presented NAOMI (Non-Autoregressive Multiresolution Sequence Imputation), a deep-learning based time-series imputation method to overcome the cumulative error issue observed in autoregressive models. Cumulative error poses a significant challenge, particularly when accurately modeling long sequences. NAOMI successfully addresses this challenge by improving imputation accuracy and generalization for long-range sequences, with BRITS as a backbone. However, NAOMI falls short in accurately imputing the actual values of the signals being imputed.

The transformer model, first introduced by Vaswani et al. (2017), is a neural network architecture that relies solely on self-attention mechanisms. The authors demonstrated that the transformer model outperformed recurrent neural networks (RNNs) on multiple natural language processing tasks in terms of both accuracy and speed. By efficiently capturing long-range dependencies and modeling contextual relationships through self-attention mechanisms, positional encoding, and encoder-decoder architecture, the transformer model is well-suited for modeling sequential data such as time series. This makes it particularly relevant for reconstructing medical data in the context of healthcare. Due to these characteristics, the transformer model has higher accuracy of imputating time series by handling long-range dependencies even better than NAOMI.

Building upon the transformer model, Bansal et al. (2021) proposes DeepMVI as a novel model for imputing missing values in multidimensional time series datasets. DeepMVI combines coarse-grained and fine-grained patterns along a time series and utilizes a network architecture with a temporal transformer and kernel regression. Compared to previous studies, this model is designed specifically for multidimensional time series datasets, which is crucial for signals like ECG consisting of multiple channels. The utilization of multiple channels' signals allows for higher accuracy due to the correlations between channels. However, challenges arise when missing channels are present in certain circumstances, such as by transmission errors or limitations in wearable devices.

2.3 Investigating the BDC Transformer on the Emory MODS Dataset

2.3.1 PulseImpute and the BDC Transformer

However, many existing imputation methods do not effectively leverage the quasi-periodic nature of pulsative signals, leading to suboptimal results. PulseImpute introduced the BDC transformer, a model specifically designed to utilize temporal correlations in ECG and PPG waveform data for accurate imputation of missing values in a single channel. This paper demonstrates the BDC transformer's state-of-the-art performance in imputing physiological signals, highlighting its potential in accurately imputing pulsative signals (Xu et al., 2022).

Previous studies have demonstrated that the BDC transformer outperforms other SOTA imputation methods. Nearly all previous machine learning research on a hospital's clinical or waveform data, including BDC transformer, utilized the MIMIC-III dataset for training (Johnson, Pollard, & Mark, 2016; Sarika & Elias, 2023; Bashar et al., 2019). This is because MIMIC-III is currently the largest publicly-available source of electronic health record (EHR) and bed-side monitoring data, and as such has been cited extensively, with over 5,436 citations in various publications spanning from the healthcare space to the computer science space.

2.3.2 Limitations of Existing Datasets

However, it is important to note that the MIMIC dataset has limitations in terms of diversity and coverage of ICU care. The dataset has only a single data source: the Beth Israel Deaconess Medical Center in Boston, Massachusetts. This hospital is limited in terms of diversity and coverage of ICU care. Approximately 74.6% of the patients processed are white, with only 3.8% black, 1.6% hispanic and 5.4% asian. Additionally only 2% of patients surveyed in a 2022 community health needs assessment were found to be uninsured ("Commonwealth of Massachusetts", 2021). This demonstrates a higher proportion of wealthy patients and implies a narrower range of ICU care (Beth Israel Deaconess Medical Center, 2022).

This lack of diversity is problematic because there has been a clear correlation between racial/ethnic & socioeconomic patterns and cardiovascular disease (Singh et al., 2015). Since 1935, historical data show significantly higher cardiovascular disease (CVD) mortality rates among US blacks compared to whites. In 2012, compared to whites, the prevalence of CVD was at least 50% higher among blacks, Hispanics, and American Indians/Alaska Natives. Next, among individuals with lower education and incomes compared to those with higher education and income levels, 46-76% higher risks of CVDs were observed.

On the other hand, the Multiple Organ Dysfunction Syndrome (MODS) dataset from Emory university contains a much more diverse patient population. The dataset pulls from two different Healthcare systems, the Emory University Hospital and the Grady Memorial Hospital. The Emory University Hospital's patient population is 39.5% white, 38.6% black, 12.9% hispanic, 6.6% asian, and of the total patients, 14% are uninsured (Emory Healthcare, 2013). The Grady Memorial Hospital's patient population is 35% white, 31.2% black, 7.7% hispanic, 6.8% asian, and of the total patients, ~11% are uninsured (Grady Health System, 2022). These statistics exactly demonstrate how the Emory datasets cover a much wider range of patient characteristics and demographics.

2.3.3 Research Objective

To ensure the acquisition of a more comprehensive and representative dataset, the primary objective of this research project is to investigate the applicability and generalizability of the BDC transformer within the context of the Emory MODS dataset. Effective imputation of missing values in ECG data is crucial for accurate disease diagnosis and treatment planning, especially in remote disease detection and monitoring. Specifically, evaluating BDC transformers' performance on the MODS dataset and comparing it to the MIMIC-III dataset provides an opportunity to assess the generalizability of BDC transformers on a broader range of patients and conditions, enhancing our understanding of their strengths and limitations in real-world clinical scenarios.

2.4 Advancing Cardiovascular Disease Diagnosis and Remote Monitoring

This research project aims to investigate the generalizability of BDC transformer, an algorithm for imputing missing ECG waveform data, for the Emory MODS dataset. Effective imputation of missing values in ECG data is crucial for accurate disease diagnosis and treatment planning, especially in remote disease detection and monitoring. By evaluating PulseImpute's performance on the MODS dataset and comparing it to the MODS dataset, this project enhances our understanding of its strengths and limitations in real-world clinical scenarios, improving the utilization of imputed ECG data for remote cardiovascular disease detection. Ultimately, this research contributes to advancing cardiovascular disease diagnosis and remote monitoring by providing valuable insights into the performance and applicability of machine learning imputation methods.

CHAPTER 3. METHODS

3.1 Examination of the MODS Dataset

3.1.1 Dataset Volume and Composition

Preprocessing the MODS dataset was imperative to ensure the extraction of pertinent information for subsequent analysis. The dataset consists of waveform data from two distinct sources: Grady patients (totaling 15,463 patients with 12T data) and Emory patients (totaling 6,285 patients with 9.2T data). This dataset's substantial volume, when compared to the MIMIC-III dataset, which comprises 10,000 patients and 6.7 TB of data, offers extensive material for testing and validating research hypotheses.

Furthermore, within each dataset, multiple patients are included, and each patient is associated with several time sequences, each spanning 8 hours. These time sequences encompass a multitude of signal types, contributing to a comprehensive dataset for analysis. Each time sequence is an array of k x T where k is the number of modalities and T is the length. Notably, patient IDs starting with "A" denote Emory patients' data, while those starting with "C" represent Grady patients' data. This differentiation emphasizes the diverse sources of the dataset. The pat_id serves as a unique identifier within a specific health system. For example, for a file path Waveforms/A014-0495873493/A014-0495873493_0036, A014-0495873493 represents the patient ID, and A014-0495873493_0036 combines the patient ID with the time sequence ID.

3.1.2 Identification of Signal Types

Having examined the volume and compositions of the datasets, the focus then shifted to identifying the various signal types and analyzing missing data within the datasets. The discovery of various signal types within the dataset was facilitated by .hea files, revealing ECG signals, PA signals, and others. Among the ECG signals, 10 out of 12 leads were present in the dataset, with signals such as V2, V4, and AVL being completely missing. Some signals, unique to specific hospitals, necessitated consultation with medical professionals for accurate classification. By examining hea files, we'll gather information on the types of signals found in each of the time sequences.

3.1.3 Analysis of Missing Data

Additionally, evaluating the extent of missing data within the dataset is crucial to determine its suitability for imputation. Insufficient missingness may make it challenging to assess algorithm performance accurately, while excessive missingness can hinder accurate imputation, especially with prolonged missingness intervals. Therefore, scrutinizing both the percentage and consecutive duration of missingness blocks is imperative. While traditional methods rely on considering NaN values to identify missing data points, waveform visualization of multiple signals revealed anomalous negative values during data collection, indicative of missing data. Hence, along with identifying the types of signals present in the patients and time sequences, the percentage and consecutiveness of the threshold of -10,000 will be analyzed, with a prolonged consecutiveness defined as four times in a row below the threshold, to ensure validity.

3.1.4 Preprocessing Steps

After a thorough analysis of signal types and the extent of missing data, the preprocessing steps are initiated to ensure readiness for analysis. Irrelevant file formats like almth and 0d were omitted, and metadata from hea files provided essential information for understanding the waveforms. Hea files provided essential metadata, including sample rate, signal type, and units, necessary for understanding the context of the waveforms.

3.1.5 Extraction of Demographic Information

Following the meticulous cleaning procedures, attention turned towards extracting demographic information to enrich the dataset further. Extracting demographics from the MODS dataset involved accessing multiple DSV (Delimiter-Separated Values) files for Emory patients. Specifically, related demographics will be collected: pat_id, gender_code, gender, race_code, race, ethnicity_code, and ethnicity. For Grady hospitals, demographic columns such as pat_id, gender, race, ethnicity, and others were utilized to extract relevant demographic information.

The subsequent sections will provide detailed accounts of the specific preprocessing steps undertaken to input the data into the BDC transformer model.

3.2 Applying PulseImpute to the MODS dataset and Evaluation

3.2.1 Introduction to PulseImpute

PulseImpute represents a pivotal advancement in the domain of imputing missing values within waveform data, particularly in the context of pulsative signals. Renowned for its state-of-the-art (SOTA) performance, PulseImpute introduced the BDC transformer, specifically engineered to harness temporal correlations in ECG and PPG waveform data. This transformative model ensures accurate imputation of missing values, holding significant promise for enhancing data integrity in clinical settings. Notably, PulseImpute tackles three types of missingness: extended loss and transient loss for 10-second data, while for 5-minute data, it's trained using the MIMIC dataset.

3.2.2 *Detailed Explanation of PulseImpute's preprocessing algorithm*

The PulseImpute algorithm unfolds through a series of meticulous steps, commencing with preprocessing to ensure data quality. The preprocessing phase involves a comprehensive procedure, adhering to the MIMIC-III ECG curation protocol. This includes linear interpolation for filling missing values (NaNs), resampling the signal to 100 Hz, computing the power spectral density (PSD), and peak detection on the PSD using stringent criteria to identify clean ECG signals. Additionally, a segment of 5-minute data undergoes ECG peak detection to ensure physiological criteria are met before normalization. Efficient execution of preprocessing tasks is ensured through parallelization strategies, optimizing computational efficiency and reducing processing time.

3.2.3 *Testing Algorithms' Signal Identification and Visualization*

To validate the efficacy of the algorithms in identifying signals, a series of tests are conducted on segmented arrays with missingness percentages ranging from 10% to 50%. This rigorous testing involves simulating extended and transient missingness scenarios for 10-second data using the PTBXL dataset, while Longformer is employed for 5-minute data, trained using the MIMIC dataset. These inferences provide crucial insights into the algorithms' performance under varying degrees of missingness, informing their suitability for real-world clinical applications.

CHAPTER 4. RESULTS

Our investigation into the MODS dataset uncovers signal distribution, missingness patterns, demographic disparities, and imputed signal performance. Through rigorous analysis, we reveal key dataset aspects, guiding subsequent research in cardiovascular disease diagnosis and remote monitoring.

4.1 Distribution of Signal Types

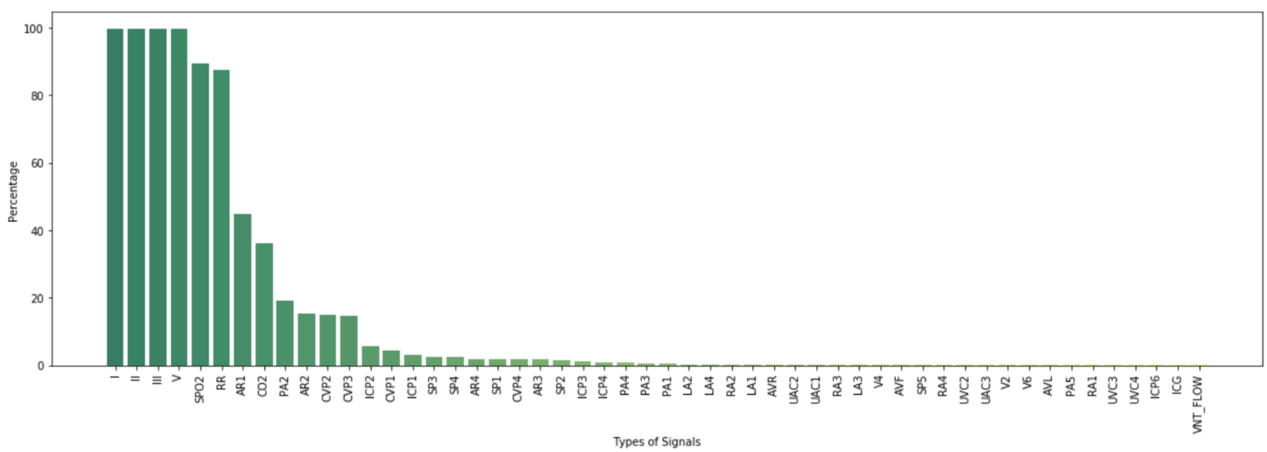

Figure 01: Percentage of Patients with Specific Signals

The visualization of the percentage of patients with specific signals highlighted notable trends. ECG signals, particularly I, II, III, and V, were prevalent across a majority of patients, comprising the top four signals. These findings are pivotal as the BDC Transformer is primarily trained on imputing ECG signals, facilitating direct comparisons between MODS and MIMIC-III datasets. Additionally, signals related to respiration, such as SPO2 (oxygen saturation) and RR (respiration rate), exhibited about 85% of presence, albeit not originally targeted by the model. The presence of these signals underscores the model's generability, warranting further exploration. Further examination revealed varying levels of presence for signals related to respiration and cardiovascular pressure measurements. While some signals achieved moderate presence, others exhibited minimal occurrence, posing challenges for performance testing and training. The investigation of hea files also provided essential contextual information regarding sample rate, signal types, and units, crucial for waveform analysis. A sample rate of 240 Hz, each signal having an array size of 6668160 (240 Hz* 60 seconds * 60 minutes * 8 hours). and varied signal units were identified, enhancing the understanding of waveform data parameters.

4.2 Analysis of Missingness Patterns

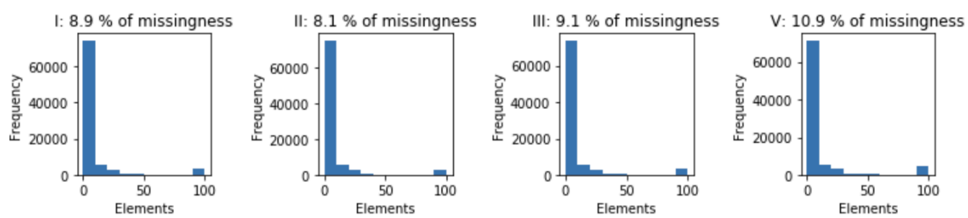

Figure 02: Missingness Percent Per Signal Type

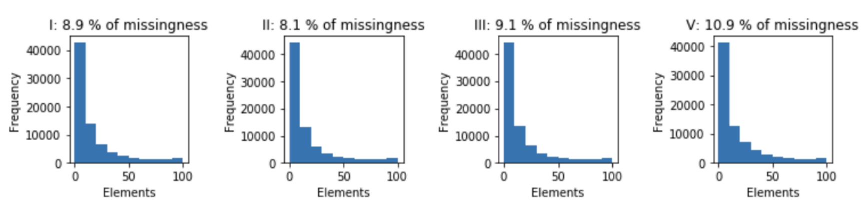

Figure 03: Amount of Missingness Blocks Per Signal Type

Moreover, the analysis of missingness patterns elucidated critical insights into data validity. High-presence ECG signals demonstrated minimal missingness about 9%, affirming their suitability for performance testing. Conversely, the visualization of missingness blocks revealed adequate variability in missingness levels from 10% to 50%, further validating the dataset's potential for model testing.

4.3 Demographic Comparisons

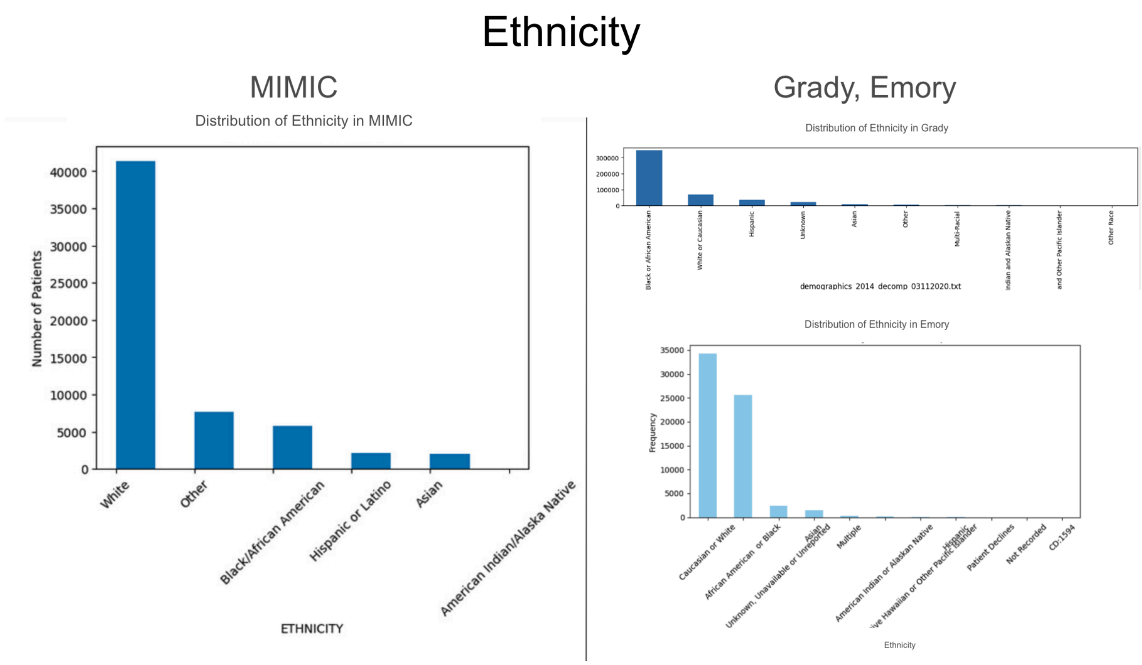

Figure 04: Comparison of Ethnicity Distribution in MIMIC-III vs MODS

Comparisons of demographics between MODS and MIMIC-III datasets unveiled notable disparities, particularly in demographic diversity. Grady and Emory's hospitals exhibited more diverse demographics compared to MIMIC-III, with higher proportions of Black individuals, singles, females, and unemployed individuals. These findings underscore the importance of considering demographic variations in dataset analyses.

4.4 Performance of Imputed Signals

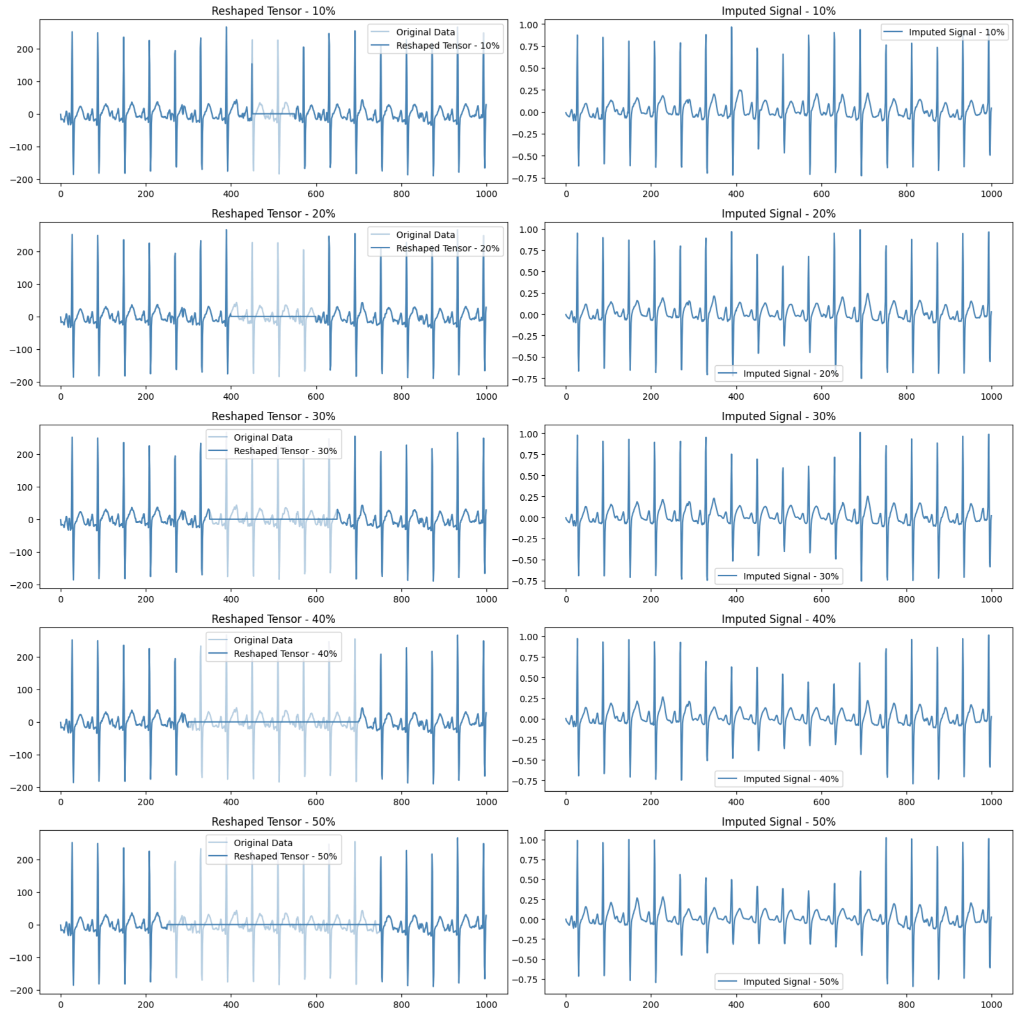

Figure 05: Imputation of MODS dataset using varied level of extended missingness

Finally, visualizations of constructed signals demonstrated the model's efficacy in identifying peak timings and mitigating complex patterns. However, imputed signals consistently exhibited smaller magnitudes compared to the originals, highlighting an area for potential improvement. This effect is more pronounced with consecutive missingness for a greater percentage. The model's tendency to consider extended missingness as part of the data patterns may contribute to this tendency.

In summary, our results provide comprehensive insights into the composition, validity, and performance of the MODS dataset, laying the groundwork for further research in cardiovascular disease diagnosis and remote monitoring.

CHAPTER 5. DISCUSSION

Our research endeavors to explore the generalizability and enhancement of the BDC Transformer for imputing missing ECG waveform data within the Emory MODS dataset. Through rigorous evaluation and comparison with existing methods, our study seeks to provide insights into the effectiveness and limitations of the BDC Transformer in real-world clinical scenarios.

5.1 Effectiveness and Limitations of the BDC Transformer

In our investigation, we found that the BDC Transformer effectively identifies the timing of peaks and mitigates complex patterns within the ECG waveform data. However, a notable limitation is the tendency for imputed signals to exhibit a smaller magnitude compared to the originals, especially in cases of consecutive missingness. This effect is more pronounced with consecutive missingness for a greater percentage. The model's tendency to consider extended missingness as part of the data patterns may contribute to this phenomenon.

Furthermore, it's crucial to highlight any assumptions made during the imputation process and their potential impact on the results. For instance, the BDC Transformer's preprocessing algorithm only accepts signals with typical ECG and PPG data. Although significant within the medical field, this limitation could be addressed by extending the model to impute a wider range of signal types through training on the MODS dataset.

5.2 Challenges Encountered During the Project

It's also essential to acknowledge the challenges encountered during the project that influenced the direction of our study. One significant challenge was related to data quality assurance measures, particularly the absence of data for certain signals, which limited the testing of the model's performance on a more varied set of signals. Additionally, computational resource limitations, despite access to GPU and parallelization, resulted in prolonged inference times, impacting the efficiency of the imputation process. Moreover, limitations in accessing demographic data, due to stricter rules around Protected Health Information (PHI), disrupted access to the dataset for a significant period, affecting the comparison of results based on demographics.

Despite these methodological constraints and challenges, our work holds significant importance in the realm of cardiovascular disease diagnosis and remote monitoring. By enhancing disease diagnosis accuracy and treatment planning, our findings have the potential to revolutionize clinical practices, particularly in remote healthcare settings. Furthermore, the generalizability of the BDC Transformer across diverse patient populations underscores its utility in addressing the complex challenges of missing data in clinical datasets.

5.3 Significance and Future Directions

Moving forward, our research opens avenues for future investigations in several key areas. Firstly, efforts to address the limitations of the BDC Transformer, such as the issue of signal magnitude, could lead to the development of more robust imputation methods. Additionally, the exploration of alternative algorithms and approaches may further enhance the performance of missing data imputation in clinical settings.

In conclusion, our study contributes valuable insights into the field of machine learning-based imputation methods for clinical waveform data. By highlighting both the strengths and limitations of the BDC Transformer, we pave the way for advancements in cardiovascular disease diagnosis and remote monitoring practices. Ultimately, our work underscores the importance of continued research and development in imputation techniques to address the complex challenges of missing data in clinical datasets.

CHAPTER 6. CONCLUSION

Our investigation into the MODS dataset provides comprehensive insights into signal distribution, missingness patterns, demographic disparities, and the performance of imputed signals. Through rigorous analysis, we illuminate key aspects of the dataset's composition, laying the groundwork for further research in cardiovascular disease diagnosis and remote monitoring.

In summary, our results offer valuable contributions to the field, guiding future endeavors in machine learning-based imputation methods for clinical waveform data. By addressing both the strengths and limitations of the BDC Transformer, we advance understanding and underscore the importance of continued research and development in addressing the complex challenges of missing data in clinical datasets.

Appendix

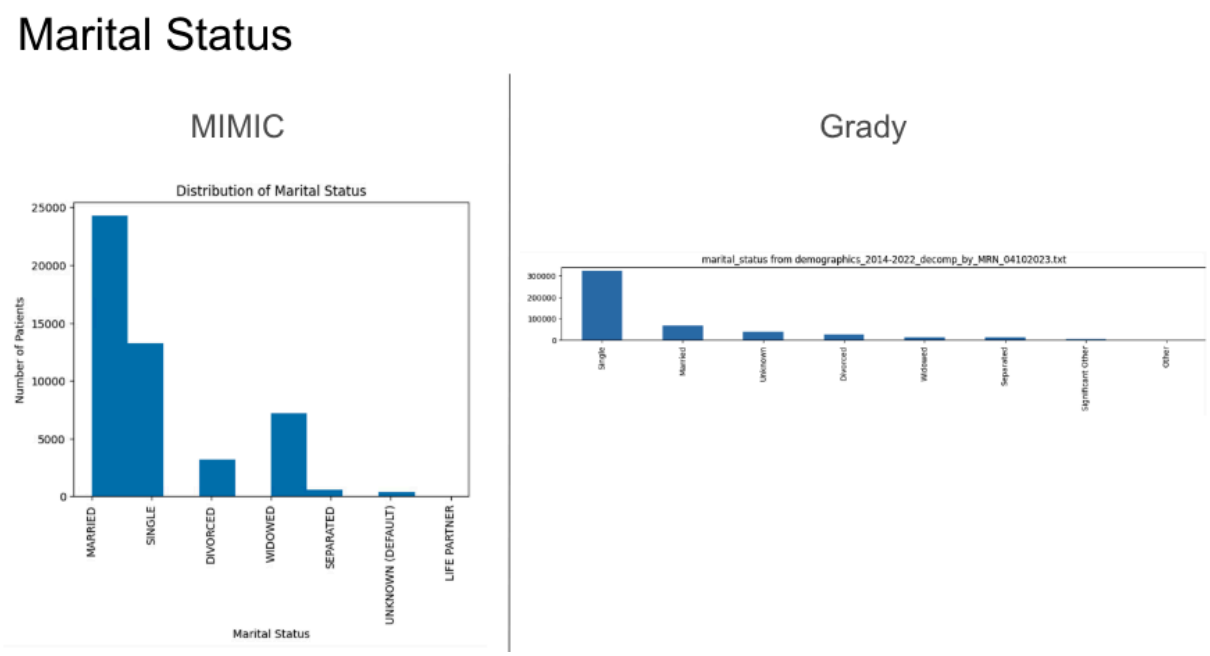

Figure 05: Comparison of Marital Status Distribution in MIMIC-III vs MODS

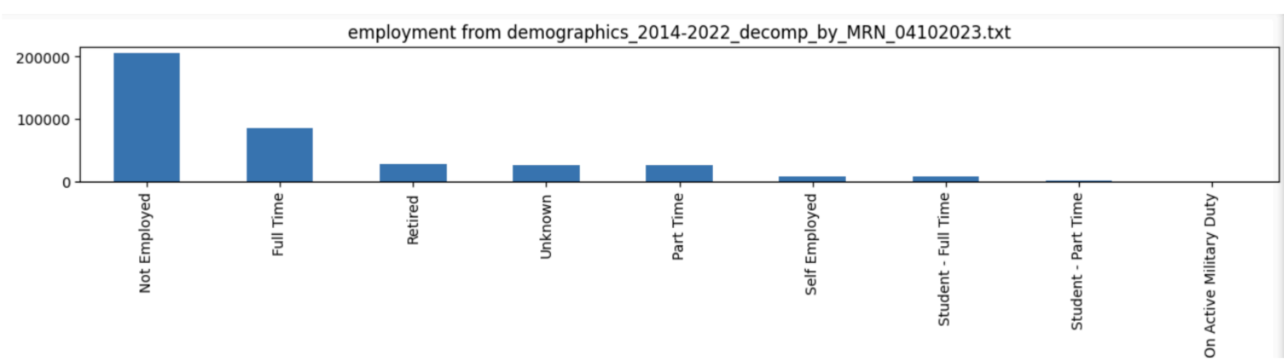

Figure 06: Employment Distribution in MODS (Grady)

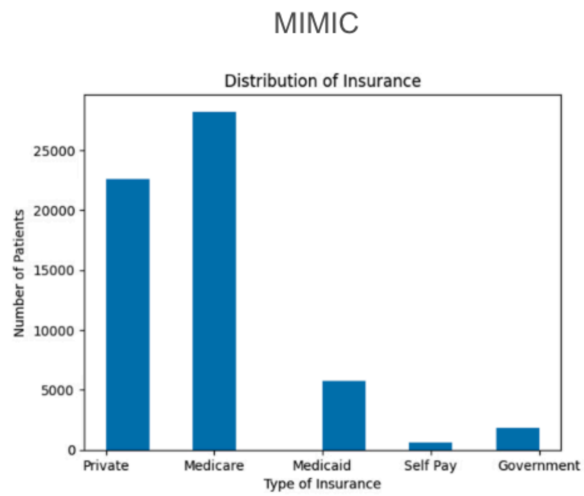

Figure 07: Comparison of Insurance Distribution in MIMIC-III

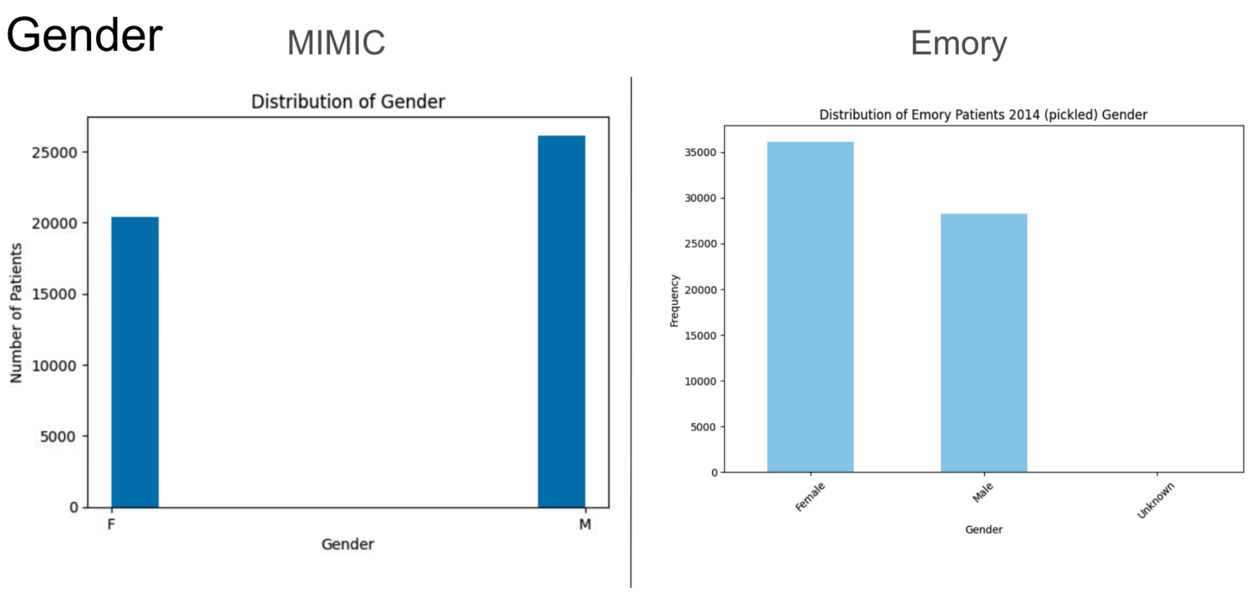

Figure 08: Comparison of Gender Distribution in MIMIC-III vs MODS

REFERENCES

Aziz, S., Ahmed, S., & Alouini, M.-S. (2021, September 21). ECG-based machine-learning algorithms for Heartbeat classification. Nature. https://www.nature.com/articles/s41598-021-97118-5

Bashar, S. K., Ding, E., Walkey, A. J., McManus, D. D., & Chon, K. H. (2019, July 1). Noise detection in electrocardiogram signals for Intensive Care Unit patients. IEEE access : practical innovations, open solutions. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7597656/

Beth Israel Deaconess Medical Center. (2022). Beth Israel Deaconess Medical Center 2022 Community Health Needs Assessment. bidmc. https://www.bidmc.org/-/media/files/beth-israel-org/about-bidmc/helping-our-community/community-initiatives/community-benefits/bidmc-2023-2025-implementation-strategy-093022.pdf

Commonwealth of Massachusetts. (2021, October 14). Beth Israel Lahey Health, Inc. – DoN Required Equipment – Staff Report. Mass.gov. https://www.mass.gov/doc/beth-israel-lahey-health-inc-don-required-equipment-staff-report/download

Emory Healthcare. (2013, July). Emory Healthcare Community Health Needs Asessment. Emory healthcare. https://www.emoryhealthcare.org/ui/pdfs/community/euh-chna071513.pdf

Grady Health System. (2022). Community health improvement plan - grady memorial hospital. Grady Health. https://www.gradyhealth.org/wp-content/uploads/Grady-Community-Health-Implementation-Strategy-2020-2022_FINAL_12.5.2019.pdf

HHS.gov. (2023, March 9). Office of Minority Health. Heart Disease and African Americans - The Office of Minority Health. https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=19

Johnson, A., Pollard, T., & Mark, R. (2016). MIMIC-III Clinical Database (version 1.4). PhysioNet. https://doi.org/10.13026/C2XW26.

Liang, Y., Abbott, D., Howard, N., Lim, K., Ward, R., & Elgendi, M. (2019, March 11). How effective is pulse arrival time for evaluating blood pressure? challenges and recommendations from a study using The mimic database. MDPI. https://www.mdpi.com/2077-0383/8/3/337

Sarika, K. R., & Elias, S. (2023, February 27). Strategies of predictive schemes and clinical diagnosis for prognosis using MIMIC-III: A systematic review. Healthcare (Basel, Switzerland). https://pubmed.ncbi.nlm.nih.gov/36900715/

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., & Yan, X. (2020). Enhancing the locality and breaking the memory bottleneck of ... - neurips. Neurips. https://proceedings.neurips.cc/paper/2019/file/6775a0635c302542da2c32aa19d86be0-Paper.pdf

Xu, M. A., Moreno, A., Nagesh, S., Aydemir, V. B., Wetter, D. W., Kumar, S., & Rehg, J. M. (2022, December 14). Pulseimpute: A novel benchmark task for pulsative physiological signal imputation. arXiv.org. https://arxiv.org/abs/2212.07514

Liu, Y., Yu, R., Zheng, S., Zhan, E., & Yue, Y. (2019, October 29). Naomi: Non-autoregressive multiresolution sequence imputation. arXiv.org. https://arxiv.org/abs/1901.10946