12-2022

A Machine Learning Approach for Predicting Patient Mortality with Heart Rate Variability Statistics

Matthew Thiele
matthewthiele@unomaha.edu

Dario Ghersi
University of Nebraska at Omaha

A Machine Learning Approach for Predicting Patient Mortality with Heart Rate Variability Statistics

Matthew Thiele

University of Nebraska at Omaha

December 2022

Dario Ghersi, M.D., Ph.D.

Faculty Mentor

School of Interdisciplinary Informatics

Table of Contents

Abstract

The prediction of patient mortality in the healthcare system provides a metric by which hospitals can better manage patient care and assess the needs of each individual patient. As such, the development of better predictive methods is vital for improving patient outcomes and overall quality of care. Heart rate variability (HRV) is a measure of the heart's complex beating patterns, giving medical professionals additional insight into patient health. Previous research has demonstrated the potential use of heart rate variability as a metric for patient mortality prediction for various conditions, however more work is necessary to validate HRV as a metric for a broader and more diverse set of patients. This study uses data from 2664 patients within the MIMIC-III clinical database matched with patient electrocardiogram (ECG) data to link HRV data with later patient mortality, examining the efficacy of HRV as a biomarker for predicting patient mortality and investigating possible avenues for future integration of HRV into patient mortality predictive algorithms.

Introduction

In a clinical setting, methods for predicting patient outcomes, and particularly patient mortality, are often used to enhance the decision-making of health-care professionals. These methods score patients based on a variety of physiological data points, attempting to provide a holistic picture of a patient's health status. This knowledge can then influence which treatments are considered and the level of urgency with which a particular patient is treated, and improve the quality of care provided. The development of better models for predicting patient mortality provides a promising path towards improving patient outcomes [1-3].

Heart rate variability (HRV) is a term describing the natural variation in the timing between subsequent heart beats. This gap is designated as an inter-beat interval or RR interval, named in relation to the QRS complex of a heartbeat as observed on an ECG, where the 'R' denotes the peak of electrical activity. When applied specifically to subsequent normal beats, filtering out beats that are classified as ectopic or otherwise abnormal, this is also called an NN interval [4]. Alongside time domain measurements that are derived from this interval, frequency domain measurements examine various frequency bands derived from the electrical signal of the heart. High variability is often associated with good cardiac health; a healthy heart follows complex behavior and more easily adapts to changes in conditions, with greater variability allowing for a better response. However, high HRV can also be a result of various pathologies, the most common of which being cardiac arrhythmias such as atrial fibrillation. As such, HRV analysis can provide insights on cardiovascular health. HRV measurements are also highly sensitive to context, and age, sex, and length of measurement must be considered for proper interpretation [5].

HRV has been well established as a component of heart health, particularly useful for patients suffering from cardiac conditions and those at risk for such diseases. A comprehensive systematic review of clinical studies examining HRV statistics in relation to patients recovering from an ST-elevation myocardial infarction, colloquially known as a heart attack, found a strong link between low HRV and later health concerns. Various studies examined in the review found a statistically significant relationship between measures such as SDNN, RMSSD, SDANN, mean RR interval, and frequency-domain measures, and later major clinical events, recurrent myocardial infarction, and later mortality, both from general causes and cardiac-related events [6]. Despite their value in assessing cardiovascular health, no commonplace methods of predicting patient mortality integrate HRV statistics. However, evidence shows that HRV is a potential predictor of patient mortality in cases of COVID-19, traumatic brain injury, and sepsis [7-9]. Further validation may open the door for use of HRV measurements for both targeted and general use patient outcome prediction models.

HRV statistics, as quantitative assessments of heart health, are well-suited to machine learning approaches, and can be easily integrated into existing models. Many patient mortality models function by assigning weights to several numerical medical measurements; one such application, the ALaRMS (Acute Laboratory Risk of Mortality Score) model, has found success in using a logistic regression approach to fit a wide variety of standard laboratory measurements into a prediction model [3]. Machine learning approaches are capable of recognizing and capitalizing on patterns within patient data that are not readily apparent to human observers. Such methods can also provide valuable information on the relevance and predictive capacity of various elements within the data, highlighting the most important factors in predicting outcomes. Machine learning methods have been shown to outperform other methods of evaluating patient mortality in sepsis patients, and may be applicable in a more general population [10]. However, machine learning methods are also highly reliant on data quantity and quality, and careful consideration should be given to data selection. Proper methodology and implementation of machine learning methods cannot compensate for poor quality data or data lacking in significant predictors. Alongside small sample size, these limiting factors can also lead to results that do not translate well to other similar datasets [11].

Finally, the growing body of publicly available data in the healthcare field presents an unparalleled opportunity for community-driven scientific progress. The sharing of data when possible, with regulations in place for deidentification of sensitive health data, allows data to be used for research in areas far beyond the initial investigation's direction and scope. The MIMIC-III database, along with its predecessors and newer addition MIMIC-IV, is one such resource, allowing credentialed access to deidentified electronic health records for over 40,000 patients of the Beth Israel Deaconess Medical Center critical care unit. The dataset supports research in multiple areas, but of specific note is the robust collection of ECG records matched to patients' health and demographic information, which provides a significant amount of data for examining HRV in relation to patient mortality in diverse populations [12-14].

This research aims to establish the efficacy of HRV statistic-based machine learning algorithms in predicting patient mortality in a diverse critical care environment, and illuminate future challenges and considerations in implementing such statistics into subsequent prediction models.

Related Work

Several studies have investigated the potential of HRV statistics in predicting patient outcomes for various conditions and situations. One such study examined the viability of the standard deviation of NN intervals (SDNN) metric, derived from ultra-short ECG recordings of 10 seconds, in predicting COVID-19 patient mortality. The study, conducted with 271 patients hospitalized with COVID-19, focused on 3-week survival and ICU readmission, and found that higher SDNN values (>8 ms) were associated with better outcomes in both areas, particularly for patients aged 70 year or older [7].

Another study, a systematic review on the relationship between HRV and patient mortality following traumatic brain injury, looked at data collected from 542 patients. Using frequency domain statistics, the study found that low frequency/high frequency (LF/HF) ratio, HF peak, and total power (TP) were each highly correlated with patient mortality [8]. A similar study focused instead on patient mortality due to sepsis, and proposed several novel measures of HRV that were strong predictors within their own dataset. In a cohort of 342 patients, a model based on these measures outperformed patient mortality models such as APACHE and SOFA in predicting patient mortality, based on the ROC curve [9].

Of studies relating to cardiovascular conditions, several characterized HRV statistics as highly correlated with patient outcomes. One paper used many 24-hour features related to SDNN, including SDNN itself, in a support vector machine-based approach to predict cardiac death following myocardial infarction [15]. Another study similarly found that a decrease in 24-hour SDNN and very low frequency band (VLF) values was associated with greater risk for developing cardiac arrhythmias in patients following myocardial infarction [16]. A third study followed 1757 patients diagnosed with coronary artery disease, and investigated SDNN In particular as well. The study noted that a decrease in SDNN was correlated with a large increase in risk of both sudden cardiac death and non-cardiac death [17].

Each of these studies further validates the notion that HRV is tied intimately to heart health and overall health, and that HRV has predictive potential when it comes to patient outcomes. Where each individual studies supports HRV as a predictor of outcomes for specific patient populations and conditions, this study aims to validate these findings with a diverse patient population, with varying medical conditions, for determining HRV efficacy in a general patient mortality prediction model.

Methods

Data Collection and Initial Processing

Electronic health record data and electrocardiogram data was collected from the MIMIC-III clinical database and MIMIC-III waveform database matched subset, respectively. For clinical data access, the CITI Data or Specimens Only Research course, as provided by the Massachusetts Institute of Technology, was completed and submitted to Physionet.org for review. The data use agreement for MIMIC-III was signed to indicate my agreement to data security and adherence to protect health data privacy, and my account and personal details verified to confirm my identity as a student researcher.

Data was downloaded and initially processed using the Physionet wfdb and pandas python packages, creating a directory of all patient ECG files containing more than ten minutes of data for each patient. Due to a combination of time constraints and processing constraints, this was later amended to a single file per hospital stay per patient, with the five to ten minute window specifically processed for HRV statistics for consistency between patients. During this initial processing, ECG data was converted to a text file containing RR interval data, which was then used for HRV calculations. Deidentified patient information, including mortality within ninety, thirty, seven, and one days of recording start, was organized with the relevant HRV data and stored.

HRV Analysis Approach

Six measures were calculated relating to patient ECG data. Average heart rate, in beats per minute (bpm), was calculated alongside average RR interval duration in milliseconds (ms). The minimum and maximum RR interval value, in ms, were also stored. The standard deviation of NN intervals, SDNN, was calculated for this 5 minute duration, along with the percentage of NN intervals that differ by more than 50 ms, the PNN50 value. Values for heart rate, SDNN, and PNN50 values were used for statistical analysis of the data as a whole. Statistical significance was calculated using a Mann-Whitney U rank test, using the scipy Python library.

To briefly examine how HRV trends change in the time approaching patient death, a small number of patients with multiple distinct ECG sessions within a short timeframe of patient death were found and their data collected. Patients were selected that had at least one recording within 10 days of death, at least one recording between 10 and 20 days of death, and at least one recording between 20 and 30 days of death. A full body of all ECG recordings exceeding 10 minutes was downloaded, and a series of 5 minute HRV statistics recorded for the length of each sample. HRV statistics were generated individually for each ECG session. This data was then used to plot the trajectory of patient heart rate, SDNN, and PNN50 in the 30 days prior to patient death. For estimating recording date, all samples had their respective date assigned as the waveform record date, as recorded in the master header files for each recording session.

Machine Learning Approach

A logistic regression model was created and trained on the data using the Python scikit-learn library. Heart rate, SDNN, and PNN50 were used to predict patient mortality within ninety days, thirty days, seven days, and one day. Following training, the model was tested on validation data and various metrics related to model accuracy were collected. A random test subset of 75% was used for model training, while a complementary testing subset of 25% was used for model validation, using the scikit-learn train_test_split() method.

Results

Of 4,753 patients whose ECG data was downloaded, 2,664 had data that was compatible with the wfdb peak detection algorithm without further processing. Signal noise within the raw ECG data could interfere with peak detection, causing the peak detection to fail and no list of RR-intervals to be generated. The statistical results derived from the remaining data are displayed in Table 1 below.

Table 1

Heart rate and HRV statistics for 4259 independent hospital admissions, from 2664 patients

		HR (bpm)*	SDNN (ms)*	PNN50 (%)*
Gender	Male (n = 2,366)	86.35 ± 22.02	34.00 ± 26.22	6.46 ± 10.50
	Female (n = 1,893)	88.49 ± 22.40	35.21 ± 27.50	7.57 ± 10.92
	Significance	$p < 0.001$	$p = 0.009$	$p < 0.001$
Mortality	Patient Deceased, 90 days (n = 803)	88.80 ± 22.80	34.65 ± 27.82	7.76 ± 22.79
	Patient Alive, 90 days (n = 3,456)	86.96 ± 22.06	34.51 ± 26.56	6.76 ± 10.62
	Significance	$p = 0.002$	$p = 0.766$	$p < 0.001$
Gender and Mortality	Male, Deceased (n = 434)	87.96 ± 21.75	33.47 ± 26.71	7.61 ± 10.69
	Male, Alive (n = 1,932)	85.99 ± 22.07	34.12 ± 26.11	6.19 ± 10.44
	Significance	$p = 0.003$	$p = 0.669$	$p < 0.001$

	Female, Deceased (n = 369)	89.80 ± 23.95	36.05 ± 29.01	7.93 ± 11.36
	Female, Alive (n = 1,524)	88.18 ± 21.99	35.01 ± 27.11	7.49 ± 10.81
	Significance	$p = 0.230$	$p = 0.990$	$p = 0.518$

* Measured across a five minute interval for each sample

Values for heart rate and PNN50 were significantly different ($p < 0.01$) across gender and mortality classes, with the exception of female mortality, for which heart rate and PNN50 were not significantly different between female patients that survived past 90 days and those that expired within 90 days. Values for SDNN were significantly different ($p = 0.009$) only between male and female patients, with no statistically significant difference between other classes. A slight, though significant ($p < 0.01$), increase is noted in all values in female patients. Consistent with previous research, patients that expire within 90 days have a somewhat higher heart rate compared to other patients, however HRV values are somewhat higher as well, contrary to previous research (Table 1).

A total of 31 patients within the full waveform database directory fulfilled the requirements for having an adequate range of recordings within 30 days of death. Of these patients, 18 failed the same requirements after processing, losing some recording data due to noisy ECG signals. From the 13 remaining patients, 4 were selected for deeper analysis based on both total range of recording days within the 30 day mark and proximity of latest recording to official date of death. The resulting graphs generated from these patients are shown below.

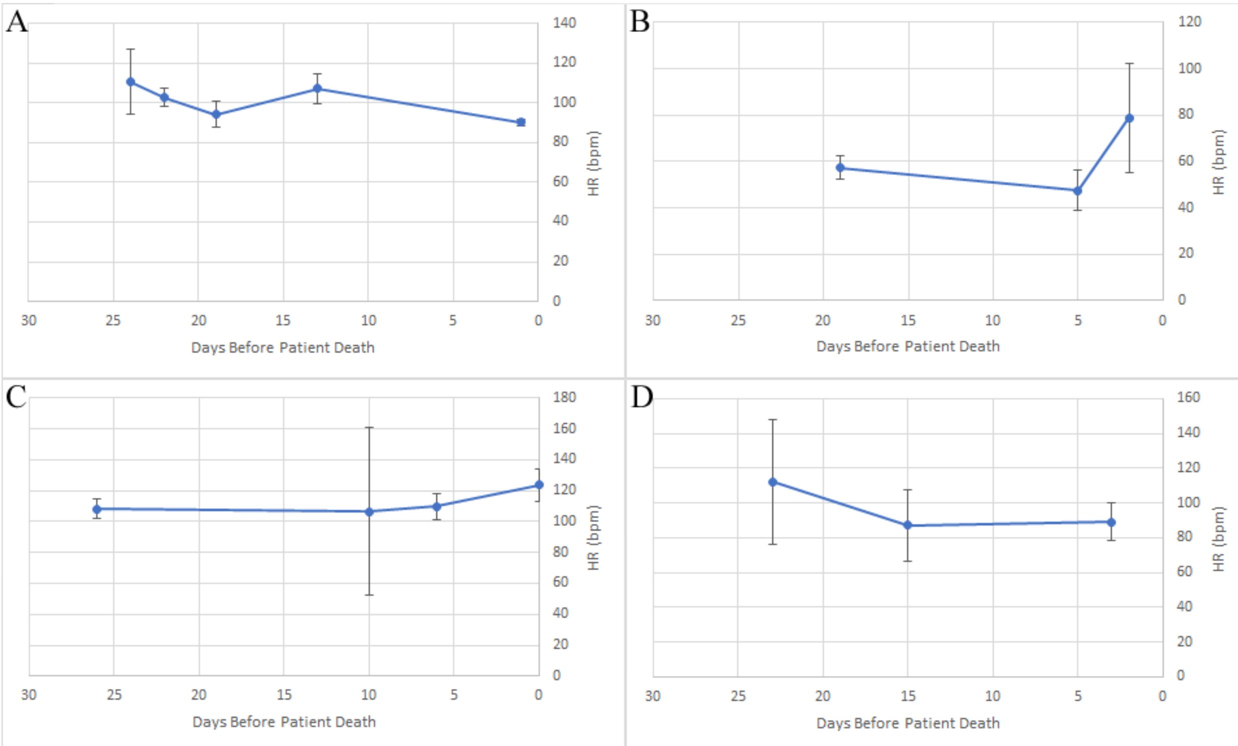

Figure 1 Average heart rate of four patients (A-D) from recordings taken within 30 days of eventual patient death.

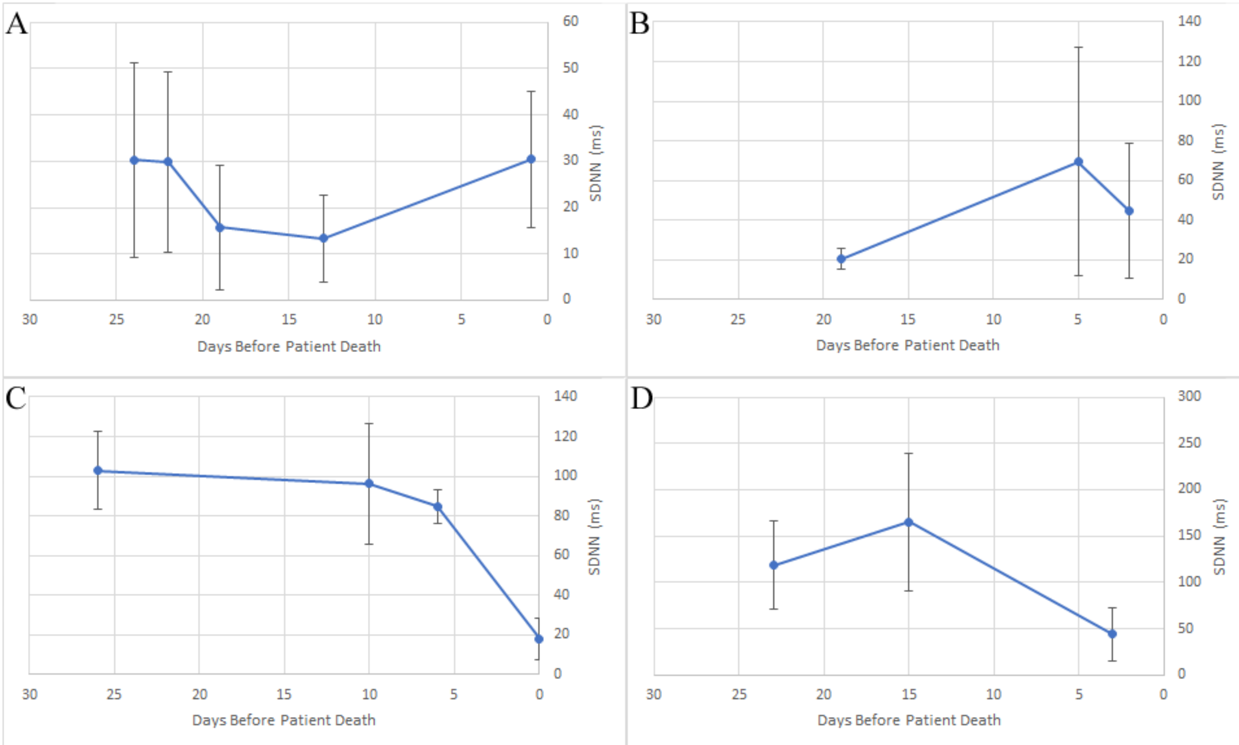

Figure 2. Average standard deviation of NN intervals (SDNN) of four patients (A-D) taken within 30 days of eventual patient death.

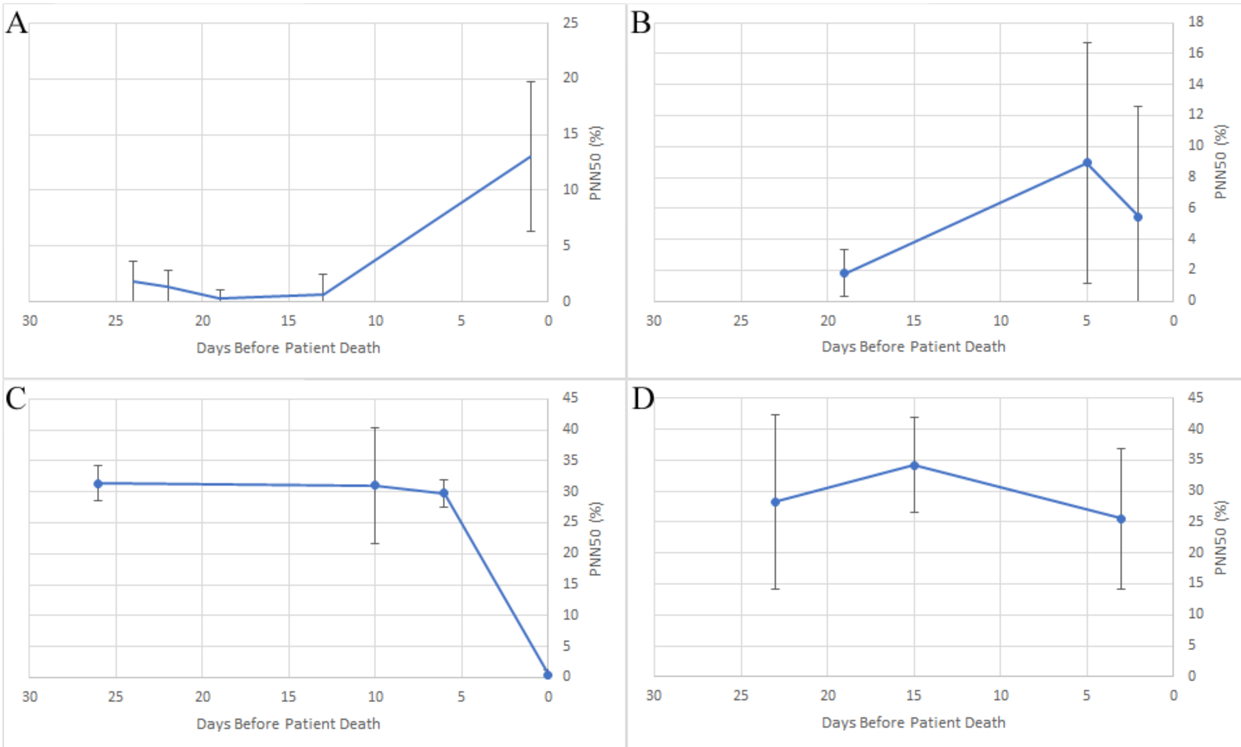

Figure 3. Average percentage of subsequent NN intervals that differ by more than 50 ms (PNN50) of four patients (A-D) taken within 30 days of eventual patient death.

Across all four patients, two showed a higher heart rate between their final and penultimate recordings, one showed a lower heart rate, and one showed a similar heart rate (Figure 1). Three showed a lower SDNN and PNN50 value between their final and penultimate recordings, however patient B's final value for both HRV measures was still higher than their first recorded value. Patient A showed a higher SDNN and PNN50, though their final value for SDNN was similar to earlier values in the examined timeframe (Figure 2-3). Of all patients, patient C showed the most drastic and consistent change in HRV values over time, with SDNN decreasing by 60 ms and PNN50 decreasing by 30 percentage points.

	precision	recall	f1-score	support
0	0.796244	1.000000	0.886566	848
1	0.000000	0.000000	0.000000	217
accuracy			0.796244	1065
macro avg	0.398122	0.500000	0.443283	1065
weighted avg	0.634005	0.796244	0.705923	1065

Figure 4. Logistic regression modeling results. Class 0 represents the living class, while class 1 represents the class of patients that passed away within 90 days of recording start.

Logistic regression analysis showed little to no success in differentiating data between patients within 90 days of death and other patients. Due to the large and almost entirely overlapping range of the data between both groups, the created model was heavily biased towards classification as the class alive after 90 days, as the larger portion of the dataset. Of validation samples from patients that passed away within 90 days, the model was unable to assign any as the correct class, assigning all samples to the living class (Figure 4).

Discussion

While results were somewhat inconclusive due to high variance in measured HRV statistics, the process of utilizing MIMIC-III data to investigate patient mortality in relation to such statistics has provided many insights into how HRV data fits into a patient's health record and what considerations should be taken into account in future research and structuring of publicly available datasets. Heart rate and PNN50 were found to be statistically significant between patients that passed away within 90 days and those that lived after 90 days, however a closer investigation into patient gender and mortality combined showed that this wasn't the case for examining female patient mortality. This may be due to the lower number of female patients in the sample set, with 2,366 male patient admissions and 1,893 female patient admissions. More data on female patients could clarify this difference, if unable to bring it more in line with other results. Nonetheless, SDNN was negligibly different between classes in all cases except in comparing both genders. For both HRV measures, mean values were unexpectedly higher when examining patients with 90 day mortality vs patients who lived past 90 days. This may be due to the relatively short 90 day timeframe in comparison to previous studies, which examined mortality on the scale of years more so than months. It may be that HRV values increase closer to the date of death, however more analysis is needed. Ultimately, more data and a greater focus on patient trends, as opposed to different hospital visits being treated independently, may be beneficial in future work.

This study was subject to a number of limitations. A notable constraint on the computational side of things was the amount of data available in the MIMIC-III waveform database, and future work should take into account the machine requirements for storing and processing multiple terabytes of data. While early statistical analysis on patients with greater amounts of data showed a similar difference between patients that passed away within 90 days of recording and living patients, more data would better illustrate each patient's health and better protect against possibly poor reliability of data, as it is currently gathered solely from a single five minute window of ECG data.

Context of recording is equally important to having high quality HRV data. HRV statistics are naturally tied to activity and other factors. As HRV tends to increase while an individual is asleep, a five minute recording of such an individual can reflect a higher HRV than would be expected from a typical five minute recording, taken at rest as is standard. Monitoring of patient activity and/or segmentation of ECG recordings according to patient status could counteract some of these uncertainties. This area in particular may contribute to the success of previous studies in predicting mortality with HRV statistics, as hands-on research lends itself naturally to strict control over sample context.

Another limitation of this study and/or the dataset involves the organization of data surrounding patient diagnoses. The MIMIC-III dataset provides two areas in which diagnosis information is present; a diagnosis table provides all official ICD-9 (International Classification of Diseases, version 9) codes of patient diagnoses, and the admissions table provides one or more plain-text diagnoses that are related to their hospital stay. When examining elements that have an impact on HRV statistics, the context of what ailments have contributed most to a hospital stay can be very beneficial. The ability to derive these diagnoses from the dataset would give further insight into how patient health may be affected by their HRV and vice versa. For future work, classification by ICD-9 code within this admissions dataset or usage of a text processing tool to derive these ICD-9 codes from plain text may allow for groups of patients to be better delineated along the lines of which particular pathologies are the most concerning. A comprehensive list of all ICD-9 diagnoses, which is stored in the diagnosis table, can clutter such data with less concerning data when it comes to evaluating HRV impacting ailments, though still important contextual information.

References

[1] Hashir, M., & Sawhney, R. (2020). Towards unstructured mortality prediction with free-text clinical notes. Journal of biomedical informatics, 108, 103489. https://doi.org/10.1016/j.jbi.2020.103489

[2] Fayed, M., Patel, N., Angappan, S., Nowak, K., Vasconcelos Torres, F., Penning, D. H., & Chhina, A. K. (2022). Sequential Organ Failure Assessment (SOFA) Score and Mortality Prediction in Patients With Severe Respiratory Distress Secondary to COVID-19. Cureus, 14(7), e26911. https://doi.org/10.7759/cureus.26911

[3] Tabak, Y. P., Sun, X., Nunez, C. M., & Johannes, R. S. (2014). Using electronic health record data to develop inpatient mortality predictive model: Acute Laboratory Risk of Mortality Score (ALaRMS). Journal of the American Medical Informatics Association : JAMIA, 21(3), 455–463. https://doi.org/10.1136/amiajnl-2013-001790

[4] Kleiger, R. E., Stein, P. K., & Bigger, J. T., Jr (2005). Heart rate variability: measurement and clinical utility. Annals of noninvasive electrocardiology : the official journal of the International Society for Holter and Noninvasive Electrocardiology, Inc, 10(1), 88–101. https://doi.org/10.1111/j.1542-474X.2005.10101.x

[5] Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. Frontiers in public health, 5, 258. https://doi.org/10.3389/fpubh.2017.00258

[6] Brinza, C., Floria, M., Covic, A., & Burlacu, A. (2021). Measuring Heart Rate Variability in Patients Admitted with ST-Elevation Myocardial Infarction for the Prediction of Subsequent Cardiovascular Events: A Systematic Review. Medicina (Kaunas, Lithuania), 57(10), 1021. https://doi.org/10.3390/medicina57101021

[7] Mol, M., Strous, M., van Osch, F., Vogelaar, F. J., Barten, D. G., Farchi, M., Foudraine, N. A., & Gidron, Y. (2021). Heart-rate-variability (HRV), predicts outcomes in COVID-19. PloS one, 16(10), e0258841. https://doi.org/10.1371/journal.pone.0258841

[8] Florez-Perdomo, W. A., García-Ballestas, E., Moscote-Salazar, L. R., Konar, S. K., Raj, S., Chouksey, P., Shrivastava, A., Mishra, R., & Agrawal, A. (2021). Heart Rate Variability as a Predictor of Mortality in Traumatic Brain Injury: A Systematic Review and Meta-Analysis. World neurosurgery, 148, 80–89. https://doi.org/10.1016/j.wneu.2020.12.132

[9] Liu, N., Chee, M. L., Foo, M., Pong, J. Z., Guo, D., Koh, Z. X., Ho, A., Niu, C., Chong, S. L., & Ong, M. (2021). Heart rate n-variability (HRnV) measures for prediction of mortality in sepsis patients presenting at the emergency department. PloS one, 16(8), e0249868. https://doi.org/10.1371/journal.pone.0249868

[10] van Doorn, W., Stassen, P. M., Borggreve, H. F., Schalkwijk, M. J., Stoffers, J., Bekers, O., & Meex, S. (2021). A comparison of machine learning models versus clinical evaluation for mortality prediction in patients with sepsis. PloS one, 16(1), e0245157. https://doi.org/10.1371/journal.pone.0245157

[11] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., Ashley, E., & Dudley, J. T. (2018). Artificial Intelligence in Cardiology. Journal of the American College of Cardiology, 71(23), 2668–2679. https://doi.org/10.1016/j.jacc.2018.03.521

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

[13] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific data, 3, 160035. https://doi.org/10.1038/sdata.2016.35

[14] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[15] Song, T., Qu, X. F., Zhang, Y. T., Cao, W., Han, B. H., Li, Y., Piao, J. Y., Yin, L. L., & Da Cheng, H. (2014). Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. BMC cardiovascular disorders, 14, 59. https://doi.org/10.1186/1471-2261-14-59

[16] Liu, X., Xiang, L., & Tong, G. (2020). Predictive values of heart rate variability, deceleration and acceleration capacity of heart rate in post-infarction patients with LVEF ≥35. Annals of noninvasive electrocardiology : the official journal of the International Society for Holter and Noninvasive Electrocardiology, Inc, 25(6), e12771. https://doi.org/10.1111/anec.12771

[17] Vuoti, A. O., Tulppo, M. P., Ukkola, O. H., Junttila, M. J., Huikuri, H. V., Kiviniemi, A. M., & Perkiömäki, J. S. (2021). Prognostic value of heart rate variability in patients with coronary artery disease in the current treatment era. PloS one, 16(7), e0254107. https://doi.org/10.1371/journal.pone.0254107