EARLY HOSPITAL MORTALITY PREDICTION USING ROUTINE VITAL SIGNS IN ICU PATIENTS

An Undergraduate Research Scholars Thesis

by

NAVEED KHIMANI

Submitted to the LAUNCH: Undergraduate Research office at
Texas A&M University
in partial fulfillment of requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Faculty Research Advisor: Dr. Bobak Mortazavi

May 2022

Major: Computing

RESEARCH COMPLIANCE CERTIFICATION

Research activities involving the use of human subjects, vertebrate animals, and/or biohazards must be reviewed and approved by the appropriate Texas A&M University regulatory research committee (i.e., IRB, IACUC, IBC) before the activity can commence. This requirement applies to activities conducted at Texas A&M and to activities conducted at non-Texas A&M facilities or institutions. In both cases, students are responsible for working with the relevant Texas A&M research compliance program to ensure and document that all Texas A&M compliance obligations are met before the study begins.

I, Naveed Khimani, certify that all research compliance requirements related to this Undergraduate Research Scholars thesis have been addressed with my Research Faculty Advisor prior to the collection of any data used in this final thesis submission.

This project did not require approval from the Texas A&M University Research Compliance & Biosafety office.

TABLE OF CONTENTS

Page

ABSTRACT 1

ACKNOWLEDGEMENTS 3

CHAPTERS

1. INTRODUCTION 4

    1.1 Purpose 4
    1.2 Related Works 5

2. METHODS 7

    2.1 Background 7
    2.2 Data Extraction 8

3. RESULTS 13

    3.1 Data Distribution 13
    3.2 Objective 1: Model Metrics for Single, Static Vital Set 14
    3.3 Objective 2: Model Metrics for Time-Series Optimization 17
    3.4 Objective 3: Optimized, Time-Series Model Metrics 20

4. CONCLUSION 25

REFERENCES 27

ABSTRACT

Early Hospital Mortality Prediction Using Routine Vital Signs in ICU Patients

Naveed Khimani
Department of Computer Science and Engineering
Texas A&M University

Research Faculty Advisor: Dr. Bobak Mortazavi
Department of Computer Science and Engineering
Texas A&M University

In a clinical setting, there are countless scenarios in which a statistical prognosis for patients can be extremely beneficial to medical professionals so that they may better allocate resources to provide the best patient care. The purpose of this paper is to identify when in a patient's stay a meaningful prediction of hospital mortality can be made to provide that prognosis.

In order to accomplish this, eight clinical variables were extracted from the MIMIC-III database for ICU patients and were supplied to a XGBoost model, an advanced Decision Tree Classifier that employs gradient boosting. Because of the imbalanced data, the positive values were weighted more heavily along with other optimized parameter values found from the use of GridSearchCV.

A static model demonstrated an average accuracy of 80.50% with an AUC-ROC of 0.800 and an AUC-PR of 0.429. However, a time-series analysis using extracted statistics from twelve-hours of compounded, time-varying data generated a model with an 83.28% accuracy with an AUC-ROC of 0.846 and an AUC-PR of 0.562. Additionally, the model demonstrated the

importance of GCS and airway management in the prediction of mortality indicating the need to focus more on these vitals in emergency situations.

The time-series model was shown to be most effective in predicting mortality, exemplifying the importance of providing time-series data that can detail the progress/decline of the patient. This implementation especially could be very impactful in clinical settings to provide healthcare professionals with the means to make quick and effective decisions.

ACKNOWLEDGEMENTS

Contributors

I would like to thank my faculty advisor, Dr. Bobak Mortazvi, and my lab peers, Lida Zhang, Hooman Sajjadi and Zhale Nowroozilarki, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my friends for their encouragement and to my family for their patience and love.

The data analyzed/used for this paper were provided by the MIMIC database.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

No funding was received from any entity for any part of this research.

1. INTRODUCTION

1.1 Purpose

In the medical field, there are several instances that have been explored where machine learning advances could prove fruitful. One such area, in-hospital mortality, has been investigated using a variety of methodologies. While there have been steps taken to classify patients using machine learning techniques, there can be improvements in the interpretability and extent of training data without sacrificing accuracy. By limiting the training data to basic vitals taken at admission, a more practical statistical prognosis can be provided to medical professionals in order to support the quick, efficient decision-making process that is necessary for intensive care situations. Additionally, this study will seek to identify when in the patient's stay this prediction should be made by weighing the impact of compounding time-varying data. The three objectives of this project are:

1. Determine the accuracy of a static model using a single set of vitals.
2. Compare the accuracy of models that use vitals taken across multiple time frames to demonstrate the effect of compounding data.
3. Interpret feature importance and benefit of time-variance in an optimized final model

This model could be employed in clinical settings during an ICU stay to identify how concerning a patient's condition is. Additionally, as the model computes the prognosis over time, the data will show whether the patient has been improving or declining. These are each very imporant pieces of information to ascertain as early as possible in an ICU stay while also maximizing the accuracy of the prediction.

This paper will seek to demonstrate that providing an XGBoost model with time-varying data compounded over an interval of up to twelve hours is more effective than using smaller time frames or using a static model while maintaining the interpretability and ease of training.

1.2 Related Works

The Acute Physiology and Chronic Health Evaluation (APACHE) II is a scoring mechanism used by medical professionals in an ICU setting to quantify severity based on demographic and physiologic measurements [1]. This is similar to other scores such as the Simplified Acute Physiology Score (SAPS) and Sequential Organ Failure Assessment (SOFA), among others. However, there is little known about the efficacy of the APACHE score especially and its implementation is not entirely accurate because of the variation across different professionals and inconsistencies in the data.

A quantified method to predict mortality will be better suited and provide improvements in consistency and efficiency in a clinical setting. Because the purpose of this paper is to predict early hospital mortality, there is an expanse of previous research to consider in the field of machine learning. The first is “Multitask learning and benchmarking with clinical time-series data” [2]. This paper by Harutyunyan, et al. has become a standard in the use of the MIMIC medical database because of its versatility in features, labels, and model types. For the purposes of this paper, the most pertinent results exist in the prediction of in-hospital mortality. This paper looks at seventeen clinical variables across the first 48 hours of a given patient’s stay in an effort to identify any patterns that may occur across that larger window. The dataset used contains approximately 21,000 ICU stays and contains an imbalance in the data where the mortality rate is only 13.23%. The best classification occurred using a multitask channel-wise LSTM (Long Short-Term Memory Network) for which the AUC-ROC was 0.870 and AUC-PR was 0.533.

Another paper that is relevant to predicting hospital mortality is “Early hospital mortality prediction using vital signals” [3]. This paper by Sadeghi, Banarjee, and Romine focuses on predicting mortality using data from the first hour of waveform ECG (echocardiogram) measurements on ICU patients. The best results for an interpretable model were from a decision tree with a precision of 0.90 and a recall of 0.92.

Based on these works, there exists a baseline that can be improved upon in some ways. Some of the MIMIC Benchmark models achieve impressive accuracy but make classifications through a black box. However, the alternative is a model that is understandable which would be much more suited to clinical purposes. While there are models from earlier papers that possess this interpretability, they generally sacrifice some accuracy to accomplish this. Additionally, the above works use large amounts of data to classify patients, be that several clinical variables or several features extracted from a waveform vital. A further objective of this paper is to use an optimal amount of data to identify the earliest point where a meaningful prediction can be made.

2. METHODS

2.1 Background

2.1.1 XGBoost

While gradient boosting trees have been implemented for quite some time, the XGBoost model was introduced in 2016 by Tianqi Chen, Carlos Guestrin. Gradient boosted trees are built similarly to decision tree ensembles, however, are more responsive to scaling. Specifically, the model “consists of a set of classification and regression trees (CART)” [4]. The XGBoost documentation visualizes a classification problem in which a decision tree is used to classify members of a family [5]. As opposed to a traditional decision tree where a leaf would be the classification, in a CART model, the leaves each have a raw score. As more trees are created, the sum of each of the raw scores of the leaves will equate to a decision. As defined by Jerome H. Friedman in his paper, “Greedy Function Approximation: A Gradient Boosting Machine”, gradient boosting simply put is a method to reduce error in each subsequent tree so that the final error is much lower than each individual tree [6]. Each additional tree focuses specifically on the errors that the previous tree made. Gradient boosted trees fill in the gaps of each individual tree creating an accurate and versatile model. An example of a XGBoost model is shown below in Figure 2.1.

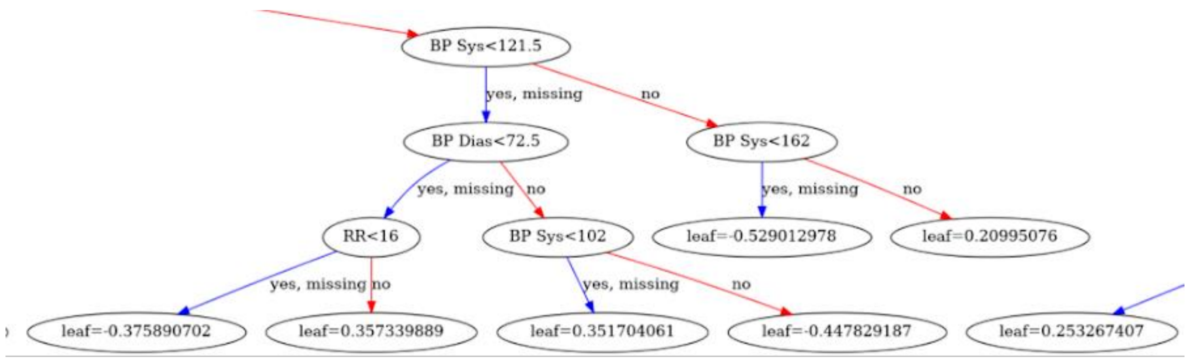

Figure 2.1: Example of Gradient Boosted Decision Tree from Static Model

2.1.2 MIMIC-III

The MIMIC-III database by PhysioNet "integrates de-identified, comprehensive clinical data of patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts. [7]" This database has been used quite prevalently in the field and has become a standard of data as seen in other papers referred to above. In this paper, the MIMIC database is used for several aspects of the data collection process. Of the twenty-six tables in the database, specifically the ADMISSIONS and CHARTEVENTS. CHARTEVENTS contains every vital, lab, or procedure completed on a patient codified by the column ITEMID in D_ITEMS which maps to the procedure name and information. While all of the data in the database is de-identified, the data is consistent between tables. For mortality prediction, the ADMISSIONS table contains the identifier HADM_ID and a column, DEATHTIME, that indicates whether the patient passed away during the care process.

2.2 Data Extraction

2.2.1 Feature Selection

There were several considerations that needed to be made when deciding upon which data set to employ as features for the model. The CHARTEVENTS table in the MIMIC-III database includes as many as 2,671,816 different vitals, labs, and other measurements and procedures that can be performed on a given patient. However, for the purposes of this paper, only routine vitals were selected in order to create a preliminary prognosis with as little of a data burden as possible. According to this 2021 publication, “Nursing Admission Assessment and Examination”, the most important vitals for a medical professional to take as a part of the admission process are “[t]emperature recorded in Celsius, heart rate, respiratory rate, blood pressure, pain level on admission, oxygen saturation” [10]. An additional paper from 2021 identified that the traditional vital signs “consist of temperature, pulse rate, blood pressure, and respiratory rate” and pulse oximetry has been shown to also be beneficial to assess [11]. The analysis done by Harutyunyan et al in the MIMIC Benchmark study employed seventeen clinical variables [2]. In order to maximize the abilities of the model, the features in this paper will employ the above five routine vital signs that are a part of the nursing assessment along with three additional features that were selected due to their effective use in the MIMIC Benchmark paper along with their frequency in the MIMIC database. These features are detailed below in Table 2.1. Column 2 displays the median value of these features in the CHARTEVENTS table.

Table 2.1: Features and Median Values

Feature Name	**Feature Median Value**
Heart Rate	92 bpm
Respiratory Rate	22 insp/min
Blood Pressure Systolic	118 mmHg
Blood Pressure Diastolic	59 mmHg
Temperature	97.20 °F
Oxygen Saturation	98%
Glasgow Coma Scale Total	11
Glucose	127 mg/dL
Fraction of Inspired Oxygen	40%

2.2.2 Time-Series Data

While the primary objective of this research is to identify the least amount of data and time that is necessary to make a worthwhile prediction, it is also important to identify the potential improvements that can be drawn from increasing that data to the standard length of prediction, twelve hours. In order to accomplish this, the single, static prediction based upon the vital data points at admission was compared to the datasets generated from one to twelve hours of time in the hospital in one-hour increments. However, XGBoost is not designed to handle time-series, two-dimensional data. To avoid this issue, instead of providing each data point to the model, statistical metrics were extracted from the available data in the given timeframe. The features that were extracted are listed in Table 2.2. These features were extracted for each of the eight vital signs that were discussed earlier to establish the progression of each vital over the course of the collection time frame. The two columns within the table show the disparity in features for the heart rate as an example for each extracted metric. The differences between the columns establishes the importance of this feature in predicting mortality. This is a similar technique to that which was demonstrated in Sadehgi's paper regarding early hospital mortality prediction using waveform ECG [3].

Table 2.2 Comparison of Features Across Labels

Feature	**Avg for Passed Away Patients**	**Avg for Alive Patients**
Maximum	98.03	107.59
Minimum	74.75	72.71
Mean	85.20	90.98
Median	84.77	90.88
Standard Deviation	7.23	9.67

2.2.3 XGBoost Optimization

There are several ways to optimize the XGBoost model in order to improve its ability to predict. The method that was employed in this process was hypertuning the parameters using the library GridSearchCV. This library performs an exhaustive search of all possible combinations of parameter values to ascertain which combination performs the best for the specified task. This method is taxing in terms of resources but is more proficient in finding the optimal parameters for the model because it can test all combinations, as opposed to a random search that may overlook a potential improvement [12]. Hypertuning the parameters in this study was accomplished by holding all other variables constant and varying a single parameter to find the ideal value for that one and repeating that procedure for the remainder of the variables.

3. RESULTS

3.1 Data Distribution

The theme in the distribution of each feature is that the median is around the generally accepted normal value for the vital, however, there are outliers in each distribution that describe more concerning situations for patients. For blood pressure, the data centers around fairly normal values of 118 mmHg for systolic and 63 mmHg for diastolic, however there are several outliers in the abnormal range that would contribute to the decline of a patient. The temperature distribution also follows a similar theme in that the median is just under the limit of febrility, but there are some realistic outliers both above and below the middle quartiles. This theme proves to be beneficial to a model so that it may use the outliers as an indicator for potential decline in the patient's health.

An accessory topic that is commonly debated in the medical field is the neglect of the respiratory rate vital. Often the measurement is simply estimated or downright ignored [13]. Respiratory rate is a very important vital in measuring how unwell a patient is but because of its lack of high-quality data, it is difficult to understand how impactful it is. An interesting plot, Figure 3.1, demonstrates the correlation between oxygen saturation and respiratory rate. In an ideal case, as respiratory rate exceeds normal boundaries (roughly twelve to twenty inspirations per minute) the oxygen saturation should drop. However, instead there is a fairly uniform band of oxygen saturation values across the range of most frequent respiratory rate values. The lack of correlation endorses the assertions made by Cretikos about the neglect of respiratory rate. When taken accurately this vital could play a large role in the mortality prediction, however, because of its sporadic nature, it may not be as beneficial.

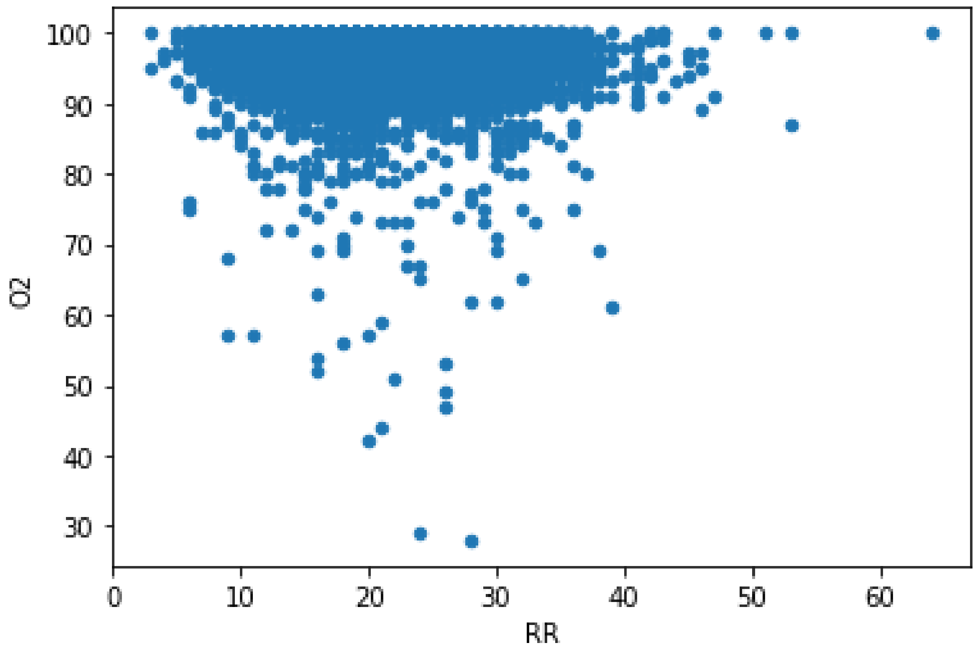

Figure 3.1: Airway Assessment Comparison

3.2 Objective 1: Model Metrics for Single, Static Vital Set

The XGBoost model created here has several quantitative means to measure its accuracy. The most standardized measure of accuracy is a stratified *k*-fold cross validation. The *k*-fold method describes when a model is trained with a portion of the data and then tested with the remainder, *k* times. A stratified version of this method is used specifically in instances of imbalanced data in order to ensure that each fold maintains a comparable distribution of positive and negative labels. A stratified 5-fold cross validation on the model trained with a static, single set of vitals resulted in an accuracy of 80.50% with a standard deviation of 2.14%.

Additionally, the receiver operating curve, Figure 3.2, and the precision recall curve, Figure 3.3, are both useful methods to understand the accuracy of the model. Below are the figures visualizing the area under the curve.

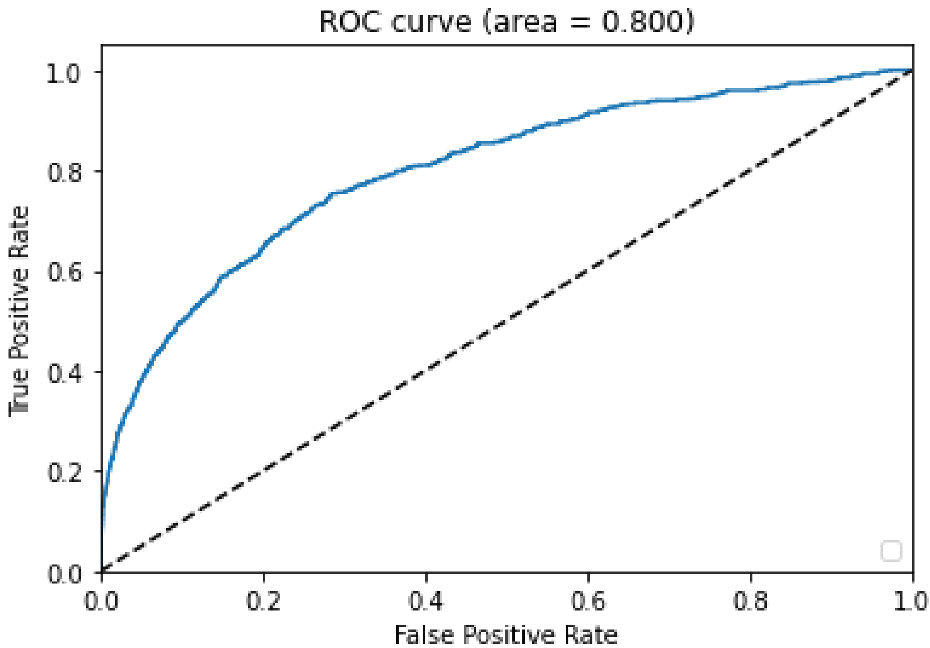

Figure 3.2: ROC Curve for Static Model

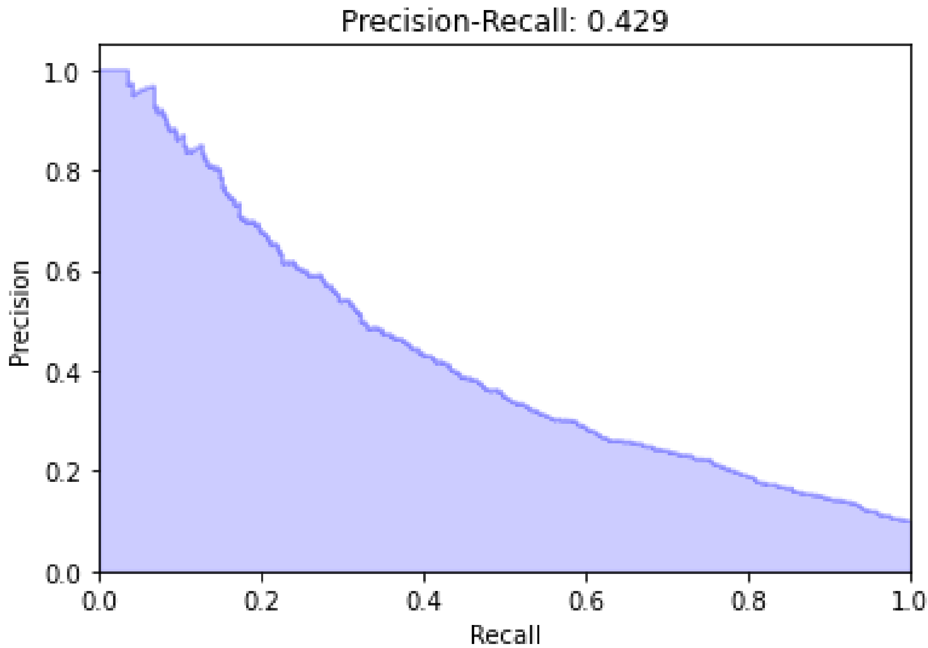

Figure 3.3: PR Curve for Static Model

This model has flaws that are evident in the confusion matrix below in Figure 3.4. For the purposes of hospital mortality prediction, the consequences of a false prediction are very dangerous. For false negatives, a death is predicted when the patient lived, the consequences are less severe in that a greater priority may be assigned to the patient when not necessary. However, false positives, shown in the upper right, indicate that a patient passed away when the model predicted the patient would live. If a lesser priority were assigned or some resources were allocated away from the patient, it could lead to even further deterioration of the patient and lead to fatality. While the key drawback of this model is the prevalence of false positives and negatives that are also seen by the lower AUC-PR, the standard deviation of the accuracy across the five folds is also much higher than other models. This is indicative of the model having to guess more causing this inconsistency. Both of these challenges can be solved for by providing the model with more data to consider in the form of time-series data.

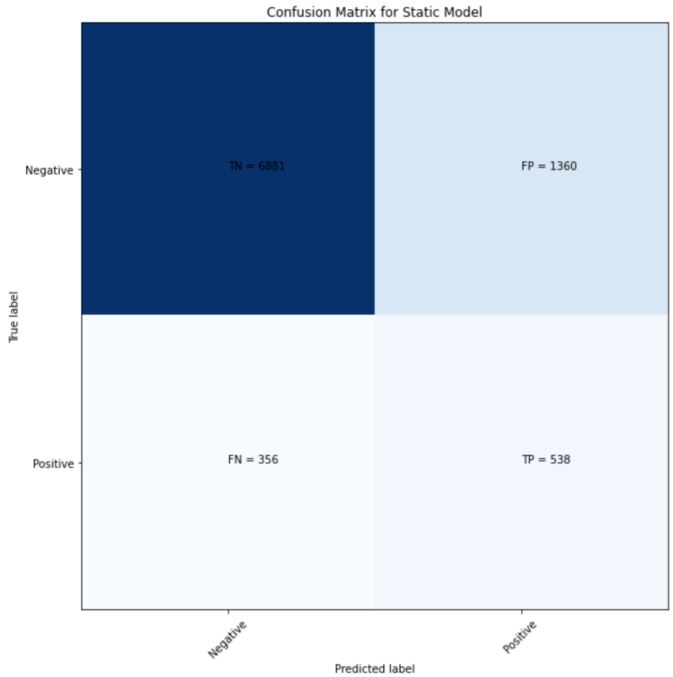

Figure 3.4: Confusion Matrix for Static Model

3.3 Objective 2: Model Metrics for Time-Series Optimization

Within the time-series tests, there were models run at one-hour intervals from one to twelve hours of ICU vitals data. This can then be compared with the static, single data point metrics that were detailed above. Table 3.1 details the accuracy metrics of each of the models trained with the accompanying time-series data. Figure 3.5 contains a graph that displays the change over time and the effects of compounding data over time.

Table 3.1: Time Optimization

Number of Hours	**Accuracy**	**AUC-PR**	**AUC-ROC**
1	79.32%	0.386	0.761
2	80.59%	0.386	0.771
3	80.77%	0.430	0.799
4	81.18%	0.451	0.798
5	81.55%	0.471	0.812
6	81.24%	0.462	0.809
7	82.03%	0.495	0.818
8	82.35%	0.494	0.826
9	82.25%	0.506	0.829
10	82.61%	0.503	0.814
11	82.76%	0.502	0.828
12	83.28%	0.562	0.846

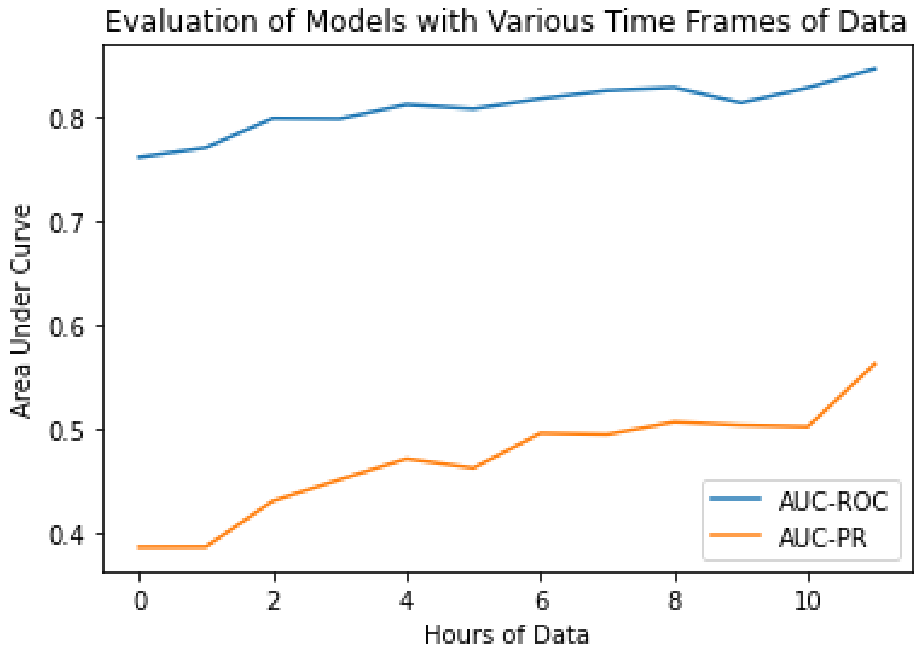

Figure 3.5: Evaluation across Compounded Time Frames

As shown above, there is a significant gain in accuracy, area under the precision-recall curve (AUC-PR), and area under the receiver operative curve (AUC-ROC) between using just the first hour of data and the first twelve hours of data. Across the various time points analyzed here, there is a moderate increase in each of those metrics. Thus, it would be most beneficial to employ a model based on the extracted metrics from the initial twelve hours of vitals available on admission. This method is more precise than using a single data point, however, would require waiting until twelve hours into the patient's stay to make the best prediction. Previous predictions can be made but will not be as accurate.

3.3.1 Test of Impact of Compounding

Using twelve hours of data is clearly the most effective way to make a prediction. However, the importance of this prediction is not in that specific hour mark, but rather in the methodology used. Providing the model more data over time allows it to understand the history of the patient rather than looking at a static point in time. To demonstrate that the gain in the model's ability to predict is based on this compounding of time-varying data and not just the final hour of data, a model was created only with the data between eleven and twelve hours for each ICU stay.

The model is the same XGBoost Classifier in every way including the below optimizations from section 3.4 but trained with and tested against different data. The results were not nearly as effective with an average accuracy of 74.77%, standard deviation of 0.61%, and an AUC-ROC of 0.659 and an AUC-PR 0.211.

The downfalls of this model highlight the successes of using compounding data. By looking at a twelve-hour time frame, there are fewer false positives and negatives, and the model becomes more attuned to predicting mortality for ICU patients over time.

3.4 Objective 3: Optimized, Time-Series Model Metrics

Using the optimization method discussed in section 2.2.3, the twelve-hour model from the time-series optimizations, was tested under multiple scenarios to identify the optimal parameter values. The Stratified-5-Fold cross validation resulted in an average accuracy of 83.28% with a standard deviation of 0.89% across the folds. The receiver-operating-characteristic curve and the precision-recall curve are displayed below in Figures 3.6 and 3.7, respectively. The exact parameter values optimized with GridSearchCV are detailed in Table 3.2.

Table 3.2: Optimized Parameter Values

Parameter Name	Optimized Value
min_child_weight	3
learning_rate	0.3
max_depth	5
max_delta_step	1
subsample	0.8

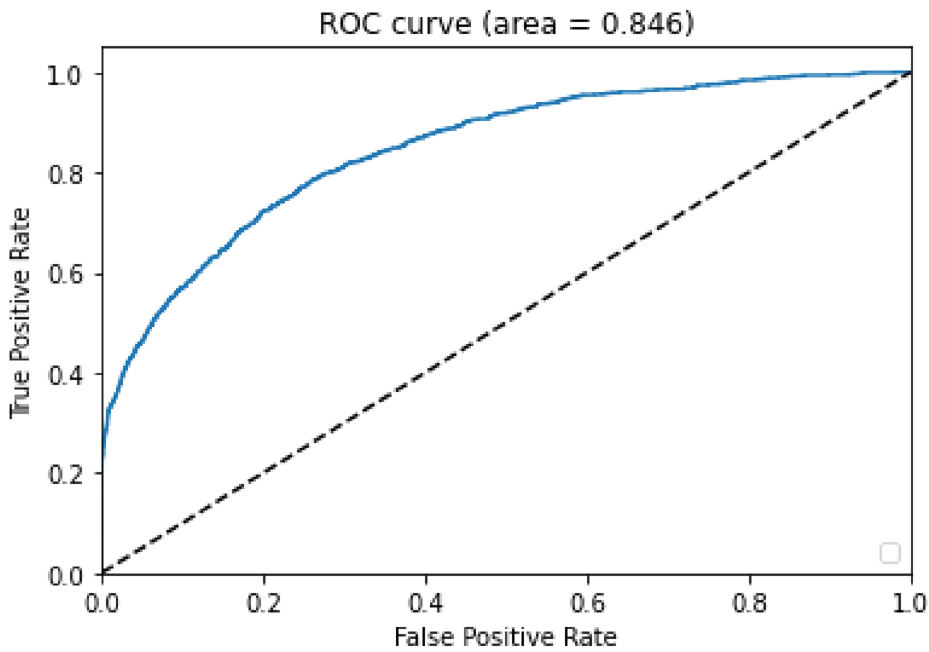

Figure 3.6: AUC-ROC for Optimized Model

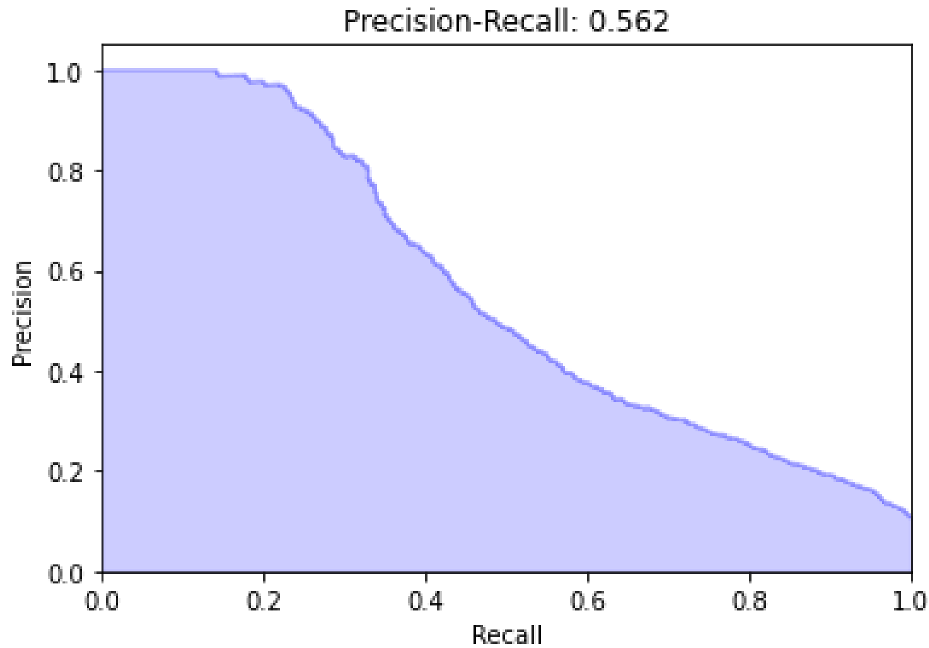

Figure 3.7: AUC-PR for Optimized Model

This model has shown to be very effective in the prediction of hospital mortality using the features explained above and could make large strides in the efficiency of an ICU. Quick and effective decisions in a hospital are necessary to save lives and with an accurate model like this, patients could be classified appropriately to provide healthcare professionals with the best information to make decisions regarding a patient's care.

3.4.1 Feature Importance

A final part of the third objective of this paper was to identify which features were most impactful in the model's ability to predict. The reason for this analysis is two-fold. First, retrospectively, it is valuable for health care administrators to identify which vitals are most impactful in the determination of early hospital mortality so that those vitals may be taken more often in a clinical setting in order to improve the accuracy of prediction. Additionally, the added weight to these vitals may also improve the accuracy in collection of the data as identified in section 3.1 specifically regarding the discussions about the validity of respiratory rate. Secondly, if this model were to be implemented in a hospital setting, an instantaneous metric of the vital that is most contributing to a prediction of mortality may assist the medical professional in prioritizing care in an emergency situation. For example, if a patient were brought in with moderately low temperature and a plummeting blood pressure, the model may indicate mortality with a feature importance attributing the blood pressure as the more likely cause as opposed to the temperature. The medical staff may then focus on administering pressors to bring up the blood pressure prior to worrying about the temperature or other extraneous signs or symptoms. This process mimics the critical analysis used by emergency medical services and would supplement a physician’s decision-making skills.

For the most optimized model, the below plot, Figure 3.8, visualizes the impact of each of the top 5 most impactful features by gain or the amount of accuracy added that can be attributed to the given feature. Clinically, this plot is relevant because the most important features are descriptors of airway management and Glasgow Coma Scale, both of which are commonly used in most if not all emergency cases. The minimum of oxygen saturation and fraction of inspired oxygen are valued highly because they evaluate the patient’s ability to breath on their own, a valuable indicator in the ICU. The GCS total is also highly impactful becase it indicates the patient’s level of consciousness and potential for brain injury. Closer monitoring of these vitals in particular could provide valuable insight into a patient’s prognosis.

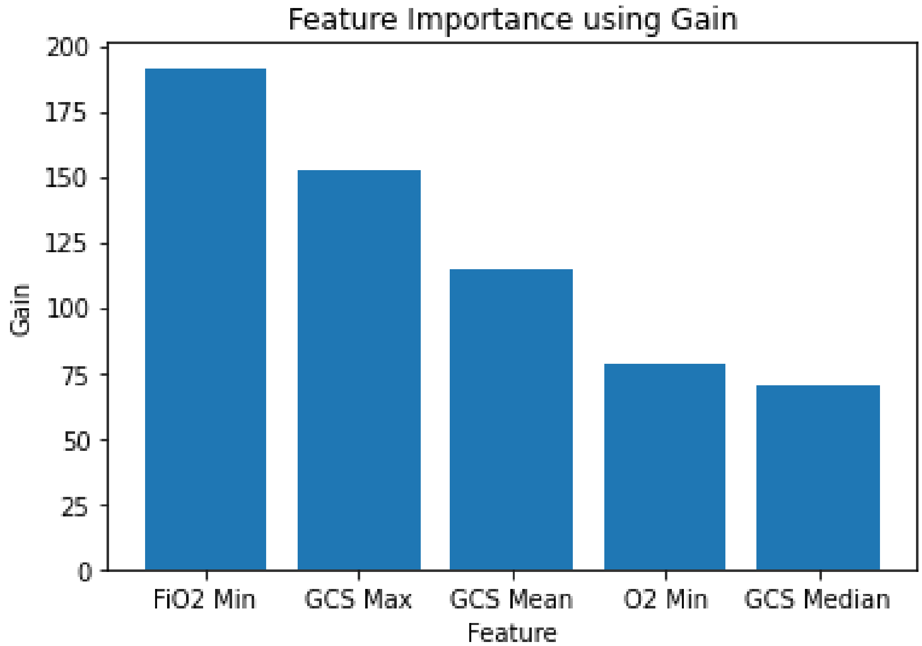

Figure 3.8: Feature Importance for 5 Highest Features

4. CONCLUSION

The purpose of the paper was to identify a statistical prognosis early into a patient's stay in the hospital in order to provide more efficient, prioritized care. The three objectives as a result were:

1. Determine the accuracy of a static model using a single set of vitals.
2. Compare the accuracy of models that use vitals taken across multiple time frames to demonstrate the effect of compounding data.
3. Interpret feature importance and benefit of time-variance in an optimized final model

The static model was created with a single set of features based upon the most important vitals taken during nursing assessments. With this static data the model failed to have high levels of precision likely because of the imbalance in the little data that it had and its inconsistent nature.

The second objective was accomplished by comparing twelve models each with one to twelve hours of vitals data. Across the twelve-period, the AUC-PR and AUC-ROC both increased consistently resulting in a more effective prediction with the cumulative data of all twelve hours. This mortality prediction would be most beneficial to medical professionals who could then make accommodations and adjust their treatment plan as more data becomes available later in the patient's stay.

The optimal model was thus the twelve-hour model which proved to best minimize the incorrect predictions, especially false negatives which are most clinically dangerous. From this model, feature importance was also extracted which identified that airway management and GCS are likely the most statistically significant vitals in a hospital admission. In addition to the direct benefits of this model, medical professionals could also monitor these vitals more closely because of their relevance in indicating the potential for loss of life.

Although this research is a strong step towards improving the field of early hospital mortality prediction, there is still much work to be done from this point. First is the way that the imbalance in mortality prediction data is handled. In this study, the imbalance was dealt with by weighing the effects of positively labeled (deceased) hospital stays greater than the more frequent negatively labeled (survived) hospital stays. The alternative is to improve the data in the pre-processing phase with an algorithm like Synthetic Minority Over-sampling Technique (SMOTE). In its developmental paper, Chawla, Bowyer, Hall and Kegelmeyer detailed this technique's ability to improve the AUC-ROC by creating a synthetic minority class from the oversampled minority data [14]. Other options for improvements moving forward are in the model, its tuning, and the time-series aspect of the data. XGBoost has been a very robust model that has won several competitions and awards recently, however it does not support time-series data and switching to models like long short-term memory networks that are more suited to such a task. However, if maintained, the XGBoost model could also be tuned differently with Bayesian Optimization as opposed to the GridSearchCV library to further optimize the parameters. Finally, this study identified the trend from the first hour in the ICU to the twelfth. Across that period, there was no consistent periods of stagnancy or decrease in the model's ability to predict. A valuable future study could look across the first twenty-four hours of ICU data to identify the extent of improvement and when waiting for more data becomes negligible or even detrimental.

REFERENCES

[1] K. H. Polderman, A. R. Girbes, L. G. Thijs, and R. J. Strack van Schijndel, “Accuracy and reliability of Apache II scoring in two Intensive Care Units,” *Anaesthesia*, vol. 56, no. 1, pp. 47–50, 2001.

[2] H. Harutyunyan, H. Khachatrian, D. C. Kale, G. Ver Steeg, and A. Galstyan, “Multitask learning and benchmarking with clinical time series data,” *Scientific Data*, vol. 6, no. 1, 2019.

[3] R. Sadeghi, T. Banerjee, and W. Romine, “Early hospital mortality prediction using vital signals,” *Smart Health*, vol. 9-10, pp. 265–274, 2018.

[4] “Introduction to boosted trees¶,” *Introduction to Boosted Trees - xgboost 1.5.2 documentation*. [Online]. Available: https://xgboost.readthedocs.io/en/stable/tutorials/model.html. [Accessed: 06-Mar-2022].

[5] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *CoRR*, vol. abs/1603.02754, Aug. 2016.

[6] J. Friedman and B. Popescu, “Importance sampled learning ensembles,” 09-Sep-2003. [Online]. Available: https://jerryfriedman.su.domains/ftp/isle.pdf. [Accessed: 07-Mar-2022].

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

[8] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., 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.

[9] 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.

[10] Toney-Butler TJ, Unison-Pace WJ. Nursing Admission Assessment and Examination. [Updated 2021 Aug 30]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK493211/

[11] Sapra A, Malik A, Bhandari P. Vital Sign Assessment. [Updated 2021 May 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK553213/

[12] "3.2. tuning the hyper-parameters of an estimator," scikit. [Online]. Available: https://scikit-learn.org/stable/modules/grid_search.html. [Accessed: 06-Mar-2022].

[13] O. Steichen, G. Grateau, and E. Bouvard, "Respiratory rate: The neglected vital sign," *Medical Journal of Australia*, vol. 189, no. 9, pp. 531–532, 2008.

[14] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1613/jair.953