Using Artificial Intelligence in Predicting Early-Onset Hypotension

An Tran, M.S

7887 Dunbrook Rd, Ste D, San Diego, CA, 92126, United States, Email: antran@darrochmedical.com, Phone: (858) 610-6384.

Robert Topp, RN, PhD

415 Kingswood, Lexington, KY. 40502, Email: toppcounsulting@yahoo.com, Phone: (502) 593-1865

Ebrahim Tarshizi, PhD, MBA

University of San Diego, Shiley-Marcos School of Engineering, Belanich Engineering Center, Room 328, 5998 Alcala Park Way, San Diego, CA, 92110, United States Email: etarshizi@sandiego.edu, Phone: (775) 771-3559.

Anthony Shao, B.S

7887 Dunbrook Rd, Ste D, San Diego, CA, 92126, United States, Email: anthonyshao@darrochmedical.com, Phone: (858) 357-8181.

Author contribution/biographies:

An Tran is a data scientist for Darroch Medical Solutions, Inc. An obtained his Master's degree in Computer Engineering where he specialized in biomedical signals processing and another Master's degree in Data Science. He contributed to this study by conducting data processing and analysis, developing and validating the machine learning models, and contributing to the writing of the manuscript.

Robert Topp, is President of Topp Consulting. He contributed to this study by developing the protocol, conducting data analysis and contributing to the writing of the manuscript.

Ebrahim Tarshizi is a Professor of Practice and Academic Director of the Applied Data Science (MS-ADS) & Applied Artificial Intelligence (MS-AAI) programs at the University of San Diego's Shiley-Marcos School of Engineering. Ebrahim provided technical expertise in the development and refinement of the machine learning models used in this study, particularly in the design and application of AI methodologies. He also contributed to editing and revising the manuscript to enhance its technical accuracy and clarity.

Anthony Shao is the CEO of Darroch Medical Solutions, Inc. He contributed to the writing of the manuscript.

Conflicts of Interest and Source of Funding: None declared.

Abstract

Hypotension, defined as a drop in blood pressure below acceptable levels, is associated with increases in morbidity and mortality among patients admitted to an intensive care unit. Current methods of predicting if a patient will experience hypotension in the clinical environment are limited by delayed analysis and a lack of accurate predictive algorithms to anticipate an evolving hypotensive event. The purpose of this research was to employ Artificial Intelligence methodologies using vital signs data to predict the onset of hypotension in Intensive Care Unit patients from the Medical Information Mart for Intensive Care-IV dataset. By comparing an Explainable Boosting Machine, Long Short-Term Memory with attention mechanism, and Logistic Regression models, the research utilized minute-by-minute vital signs from various historical windows to predict hypotension one hour before onset. After analyzing 1340 hypotension and 2027 non-hypotension cases, the Explainable Boosting Machine model achieved 90% sensitivity and 85% specificity using a 3-hour lookback window, outperforming Logistic Regression and matching the Deep Learning model's performance. These findings indicate that the onset of hypotension can be accurately predicted up to one hour in advance of the event. This early identification of patients who have a high risk of developing hypotension has the potential to support early interventions and improve patient outcomes.

Keywords: hypotension; artificial intelligence; continuous monitoring; early prediction; vital signs

Introduction

Hypotension, or a drop in blood pressure, experienced by patients in the Intensive Care Unit (ICU) is often a harbinger of critical events including Acute Kidney Injury (AKI), Myocardial Infarction (MI), and other life-threatening conditions (Haase et al., 2009). The economic burden of treating hypotension has escalated quickly to $870 million in 2023, and this amount is estimated to double by 2032 (Market Study Report, 2023). Traditional methods of identifying patients who are at risk for developing hypotension often involve time-consuming chart reviews and data that have been manual entered into electronic medical records (EMRs), making the process inefficient and susceptible to errors which may delay clinical interventions (Despins, 2017).

The development of Artificial Intelligence (AI) algorithms capable of analyzing real-time vital sign data presents a paradigm shift in managing critically ill patients. By enabling the early detection of hypotensive trends, AI technologies hold the promise of facilitating intervention strategies that can preempt the onset of severe complications that may result from a hypotensive event and other associated negative health consequences.

However, despite recent technological advances, the integration of AI into the early identification of hypotension is still in its infancy, often lacking near real-time analytics and accurate predictive capabilities. This study seeks to bridge this gap, offering a novel AI approach that aims to improve on the early detection and management of hypotension. By analyzing continuous near real-time data from bedside monitors and centralized data systems, AI models have the potential to interpret complex data streams and synthesize early warning alerts for healthcare professionals, thereby enhancing the clinical decision-making process and improving patient outcomes.

Objectives

The purpose of this research was to employ Artificial Intelligence methodologies using vital signs data to predict the onset of hypotension in Intensive Care Unit patients from the Medical Information Mart for Intensive Care-IV dataset.

To achieve this purpose, the study pursued two primary objectives. The first objective was to develop and compare three AI-driven hypotension prediction models - Logistic Regression, Long Short-Term Memory (LSTM) Network with an attention mechanism, and Explainable Boosting Machine (EBM). These models were then evaluated using key performance metrics including sensitivity, specificity, area under the receiver operating characteristics curve (AUC), and an F-1 score. The optimal historical data window for accurate hypotension prediction was determined by balancing prediction accuracy with practical clinical utility.

The second objective was to enhance the robustness of the prediction process by implementing an autoencoder-based anomaly detection system, following approaches demonstrated by Malhotra et al. (2016) and Zhang et al. (2019) in clinical time series analysis. This system identified irregular patterns in vital signs data before prediction, allowing the models to express uncertainty, rather than forcing potentially unreliable binary predictions. Similar to the approach validated by Zhou et al. (2021) in ICU settings, this created a three-category prediction system including hypotension likely, hypotension unlikely, and anomalous patterns requiring further clinical review.

Methodology

Dataset

Data in this study were extracted from the Medical Information Mart for Intensive Care (MIMIC)-IV database; this is an open-access resource developed by the MIT Lab for Computational

Physiology (Johnson et al., 2023). and contains de-identified patient demographics, vital signs, laboratory tests, and medications for Intensive Care Unit (ICU) patients at Beth Israel Deaconess Medical Center between 2008 and 2019. Patient privacy was maintained by removing all Protected Health Information (PHI) in compliance with Health Insurance Portability and Accountability Act (HIPAA) standards; therefore, IRB oversight of this study was not required.

Inclusion criteria were adult individuals aged 18 and over who were admitted to the ICU. Eligible individuals had vital signs recorded at least once every hour, with no gaps in data exceeding one hour for any vital sign measures. This criterion ensured a balance between data completeness and sample size, allowing for the inclusion of patients with reasonably continuous monitoring while not excluding those with occasional gaps in their recorded vital signs. Vital signs extracted from the MIMIC-IV database included heart rate, temperature, respiratory rate, SpO2, systolic, diastolic blood pressure, and mean arterial pressure. For instances where multiple recordings of a vital sign occurred within the same minute, these values were averaged to provide a minute-averaged measure. In situations where a vital sign reading was absent for a given minute, imputation was performed using the most recent data available for the patient. The imputation techniques employed in this study were the last observation carried forward (LOCF) and the next observation carried backward (NOCB) methods. These techniques are widely recognized in healthcare research for their efficacy in managing missing follow-up observations, as they respect the sequential nature of clinical data and are less likely to introduce bias than simpler mean or median imputation methods (Hamer and Simpson, 2009). LOCF was first applied to fill forward any missing values, followed by NOCB to fill any remaining gaps backward. This approach was chosen due to its respect for data sequence in a time series data and suitability for real-time applications. LOCF and NOCB maintain temporal integrity,

avoid introducing artificial patterns, and are less likely to introduce bias compared to mean or median imputation methods.

A hypotensive event was defined when at least five consecutive readings within any 10-minute window exceeded a systolic blood pressure (SBP) of 90 mmHg or lower and a mean arterial pressure (MAP) of 60 mmHg or lower (Yoon et al., 2020). To accurately identify these events, a minute-by-minute rolling window analysis over a 10-minute duration was employed for each patient stay, enabling dynamic tracking of each patient's conditions. This rolling window approach continuously evaluated each 10-minute segment of a patient's stay, advancing one minute at a time, and for each window, the algorithm checked whether the least five readings met the hypotension criteria. If this condition was satisfied at any point during the stay, the patient was classified as having a hypotensive event, and the first such occurrence was marked as the onset of hypotension. If no 10-minute window throughout the entire stay met this criterion, the patient was classified as non-hypotensive. This method allowed for a comprehensive examination of each patient's vital signs, ensuring that no hypotensive events were missed, regardless of when they occurred during the ICU stay. Based on this definition and analysis, the study classified 1340 patient stays as hypotensive and 2027 as non-hypotensive.

To ensure the integrity of the predictive analysis, this study aligned the dataset by standardizing the timeframe from admission to the onset of hypotension across all patient stays. Specifically, the average time from admission to hypotension onset for patients who experienced hypotensive events was calculated. This average time was then used as a reference to align the timeline of non-hypotensive patient stays, ensuring each non-hypotensive stay was analyzed over a comparable duration (see Figure 1). This alignment was important for several reasons. First, it addressed the potential bias that could arise from analyzing patients at different stages of their ICU stay, for without alignment, patients who neared discharge could show more stable

vital signs due to their improved condition, potentially skewing the analysis. Similarly, patients in the early stages of their stay could exhibit more volatile vital signs as they were still being stabilized. This standardization allowed for a more fair and accurate comparison between hypotensive and non-hypotensive stays, with an emphasis on the analysis of the critical period leading up to potential a hypotension event.

Analysis Plan

An extensive feature engineering process was undertaken to capture a wide range of physiological patterns and trends based on the vital signs being studied. Derived clinical features included Shock Index, Pulse Pressure, and Cardiac Output were calculated as follows:

$$Shock\ Index = \frac{Heart\ Rate}{Systolic} \quad (1)$$

$$Pulse\ Pressure = Systolic - Diastolic \quad (2)$$

$$Cardiac\ Output = Pulse\ Pressure * Heart\ Rate \quad (3)$$

These derived features provided additional insights into cardiovascular function and hemodynamic status that were not directly captured by individual vital signs alone as they had been previously recognized as potential predictors for hypotension onset (Berger et al., 2013; Corrêa et al., 2013). Statistical features were computed, including rolling statistics such as min, max, mean, standard deviation, and median, for each vital sign over the specified look-back windows (see Figure 2). These rolling statistics captured the central tendency, variability, and extreme values of vital signs over time, potentially revealing patterns of stability or instability in a patient's condition.

To complement these statistical measures, frequency domain analysis of vital signs was employed to detect subtle patterns that might precede hypotension. These metrics included

maximum magnitudes and dominant frequencies, which measure the strength and prevalence of specific rhythmic patterns in vital signs (Clifford et al., 2015). The spectral centroid, defined as the weighted average of the frequencies present in the signal, indicated where the center of mass of the frequency spectrum lied (Smith, 2011). The spectral spread measured how far the frequencies deviated from this center, while spectral entropy quantified the randomness or unpredictability in the signal (Pincus, 1991). For example, a sudden increase in spectral entropy of blood pressure measurements could have indicated the deterioration of normal cardiovascular regulatory mechanisms before a hypotensive event. These frequency-based measurements could also reveal recurring patterns in vital signs that may not be apparent through traditional time-based analysis.

The analysis also incorporated trend analysis features developed by Darroch Medical Solutions that quantify changes in vital signs stability over time. These specialized features were designed to capture nuanced patterns in physiological measurements that could escape detection through conventional statistical or frequency-based analysis. At their core, these features tracked two key aspects of vital sign behavior, including the frequency of directional changes and the magnitude of these variations.

The directional change count represented how often a vital sign's trajectory reversed direction within specified time windows and served as an indicator of signal volatility. Complementing this measure with the mean, minimum, and maximum values of these changes both captured the scale of the variations and provided context about the intensity of the fluctuations. For instance, if heart rate measurements showed frequent directional changes of large magnitude, this could have indicated the cardiovascular system was instable which may have preceded a hypotensive event. These trend measurements provided additional insight into blood pressure stability beyond what could be observed through standard statistical measures, particularly in identifying patterns of physiological compensation that could occur before acute deterioration. By

combining clinical measurements, statistical analysis, frequency domain analysis, and trend analysis, the feature engineering process created a detailed representation of patients' vital signs. This comprehensive approach allowed the AI models to detect both obvious and subtle indicators of impending hypotension, potentially improving prediction accuracy.

The original MIMIC-IV dataset was partitioned into three subsets, including training (70%), validation (15%), and testing (15%) sets, with care taken to maintain the original distribution of hypotensive and non-hypotensive stays across all sets. The training set was used to teach the models of the various patterns and relationships within the data. The validation set served two primary purposes: First, this data set supported the tuning of the model's hyperparameters and provided an unbiased evaluation of the model's performance during the training process. This process primarily addressed overfitting, a phenomenon in which a model learns training data in a manner that internalizes noise and peculiarities (Goodfellow et al., 2016). In many cases, overfitting causes a model to internalize random fluctuations as meaningful patterns, which can result in poorer performance on unseen data. Second, the validation set aided model selection by allowing the comparison of different model architectures or parameters on data not used in training. The testing set, which is not engaged during the training and tuning phases, is used for the final evaluation of the model's performance, providing an unbiased assessment of the model's psychometric performance with novel data. This separate analysis of the three subsets of the original data ensured that the developed models were robust and capable of performing well on future novel data.

Model performance was evaluated based on sensitivity, specificity, the area under the receiver operating characteristic curve (AUC), and the F1-score in predicting hypotension 60 minutes prior to clinical onset. Sensitivity measures the model's ability to correctly identify impending hypotensive events, while specificity assesses its ability to correctly identify non-hypotensive cases. The F1-score is the harmonic mean of precision and recall, and it provides a balanced

measure of the model's performance, particularly useful in scenarios with imbalanced datasets. AUC provides an overall measure of the model's discriminative power across various classification thresholds. Together, these metrics offer a comprehensive assessment of the model's predictive capabilities, balancing its ability to identify true positives while minimizing false alarms.

An innovative approach of this study was the integration of anomaly detection to enhance the robustness of predictions. An autoencoder was trained on non-hypotensive patient data to learn typical vital sign patterns. During the prediction phase, input sequences were passed through this autoencoder to calculate a reconstruction error that served as the anomaly score. Cases that exceeded the threshold were flagged as potential anomalies, which allowed for special clinical consideration and improved the model's performance in unusual or complex cases (see Figure 3). A comparative analysis of the AI models, included Logistic Regression, Explainable Boosting Machine (EBM), and Long Short-Term Memory (LSTM) Network with attention mechanism, was conducted across different look-back windows to determine the optimal configuration. This comparison aimed to identify the most effective combination of model architecture and historical data duration for accurate hypotension prediction which balanced predictive power with computational efficiency.

Model Training

The model training phase involved three distinct approaches: Explainable Boosting Machine (EBM), Long Short-Term Memory (LSTM) with Attention Mechanism, and Logistic Regression. Each model was trained using the same preprocessed data to ensure a fair comparison of their performance in predicting hypotension onset.

Explainable Boosting Machine

The first machine learning model employed for hypotension onset prediction was the Explainable Boosting Machine (EBM), a decision tree-based model trained on derived features

from patients' vital signs (Caruana et al., 2015). EBM is typically selected for its powerful interpretability, making it highly suitable for high-stake decision fields such as healthcare settings. EBM provides clear explanations of how each vital sign contributes to its predictions, allowing healthcare providers to understand and validate the model's decision-making process.

For work in this project, EBM examined each vital sign independently and then considered how pairs of vital signs worked together to predict hypotension. For example, it might have learned that a moderate drop in systolic blood pressure alone may not have indicated impending hypotension, but when combined with an elevated heart rate, it became a significant predictor. This approach was designed to mirror how clinicians often evaluate multiple vital signs together to assess patient status.

The EBM model generated a probability value, ranging from 0 to 1, that represented the likelihood of the look-back window belonging to a particular hypotensive or non-hypotensive event; this was based on observed trends in patients' vital signs in that look-back window. To make predictions more accurate, a decision threshold was used to convert these probabilities into discrete class assignments that maximized both sensitivity and specificity. The optimal decision threshold was determined using the validation set which provided a basis for evaluating the model's performance at various thresholds in order to ensure the best balance between sensitivity and specificity before assessment on the test set. When the generated probability was compared against the decision threshold by the model, patterns within a specific look-back window were classified as either hypotensive or non-hypotensive. If the probability was higher than the threshold, the patterns were classified as hypotensive positive. Conversely, if the probability was lower than the threshold, the patterns were classified as non-hypotensive, which suggested a lower likelihood of hypotensive development within the given time frame.

LSTM with Attention Mechanism

Long Short-Term Memory (LSTM) models are a unique kind of deep learning capable of learning long-term dependencies. The LSTM architecture is particularly well-suited for time series data, as it can learn to retain important information over long sequences while forgetting irrelevant details. This is particularly useful in the context of predicting hypotension, where patterns leading to the event may develop over extended periods.

The attention mechanism, on the other hand, allows a model to focus on different parts of the input sequence when making predictions. In the context of vital signs, this means the model can learn to pay more attention to specific time periods or particular vital sign patterns that are most indicative of impending hypotension. For instance, it might learn to focus more on recent sharp drops in blood pressure or on specific combinations of heart rate and blood pressure changes that often precede hypotensive events.

For work on this project, this combination of LSTM and attention mechanism enabled the model to effectively process the time series data of vital signs so as to capture the subtle temporal patterns that might be indicative of impending hypotension. The model could then learn to recognize complex patterns such as gradual trends, sudden changes, or specific sequences of vital sign fluctuations that are associated with increased risk of hypotension.

Logistic Regression

In addition to the EBM and LSTM with Attention Mechanism, a Logistic Regression (LR) model was utilized as a baseline for comparative analysis. LR has been widely used for classification tasks and is known for its inherent interpretability as it transforms input features into a linear regression (Fan et al., 2008). The general form of the LR model is given as follows:

$$p(y) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}} \quad (5)$$

In this equation, $p(y)$ represents the probability of the output variable y being equal to hypotensive (1), given the input variables x_1 to x_n and the associated weights β_1 to β_n. The logistic regression model allows for a global interpretation of feature importance by analyzing the values of the β feature weights. However, due to its linear nature, the LR model is not able to provide direct assessment of local explanations, as the feature importances are adjusted as being the same for the entire features distribution. Due to its simplicity, logistic regression often serves as an important baseline model. Its performance, relative to more complex models like EBM and LSTM, can provide insights into the complexity of the relationship between the input features and hypotension onset

Anomaly Detection

An innovative aspect of this study was the integration of an autoencoder that preceded the prediction models. Essentially a pattern-learning system, the autoencoder worked by first compressing (encoding) incoming vital signs data into a simplified form and then by attempting to reconstruct (decode) the original data. The purpose was to create a system which could learn the difference between normal patterns and nonnormal patterns; in the case of the latter, the case was flagged as potentially ambiguous.

The autoencoder was first trained on both hypotensive (cases) and non-hypotensive (controls) patient data to learn the characteristic vital sign patterns associated with each group. When vital signs data was processed by the autoencoder, a reconstruction error was calculated which represented how different the reconstructed data was from the original input. A higher reconstruction error indicated that the pattern being observed was less similar to any of the patterns the autoencoder learned during training from either group. A threshold at the 95th percentile of reconstruction errors in the training set was established; anything above the 95th percentile was considered anomalous. Such cases would then be flagged as “uncertain” rather than a binary hypotensive or hypotensive event.

The integration of anomaly detection before the prediction models created a more comprehensive and robust system for hypotension prediction. By acknowledging when patterns appeared unusual compared to both typical hypotensive and non-hypotensive trajectories, the system could indicate when additional clinical scrutiny would be warranted. After all, one of the chief deterrents of any warning system is the unreliability in false positives or negatives particularly in fringe cases which contain unfamiliar patterns, and the inclusion of a third criteria yields a more comprehensive model which addresses unusual cases which can reduce model predictions accuracy and clinical utility in practice.

Results

The performance of the three machine learning models - Logistic Regression, Explainable Boosting Machine (EBM), and Long Short-Term Memory (LSTM) - was evaluated on the test set for predicting hypotension onset, both with and without an anomaly detector. The models were tested with different look-back windows (1, 2, 3, and 6 hours) while maintaining a consistent 1-hour prediction window.

Without the Anomaly Detector, the Logistic Regression showed the lowest performance, with sensitivity ranging from 71% to 76% and specificity from 69% to 73% across different look-back windows (see Table 1). The EBM model demonstrated improved performance, with sensitivity between 78% and 80.7% and specificity between 75% and 78%. While EBM outperformed Logistic Regression due to its ability to capture non-linear relationships, it fell short compared to LSTM in capturing complex temporal dependencies. As a tree-based model, EBM treated each time point independently, potentially missing important sequential patterns in the data.

The LSTM model consistently outperformed both previous models, achieving the highest sensitivity of 85% (83-87% CI) and specificity of 82% (80-84% CI) with a 3-hour look-back window. LSTM's superior performance can be attributed to its ability to capture long-term

dependencies in sequential data. Its architecture, featuring memory cells and gating mechanisms, allowed it to selectively remember or forget information over long sequences, making it particularly suited for analyzing time-series vital sign data. The attention mechanism further enhanced its performance by allowing the model to focus on the most relevant parts of the input sequence when making predictions.

The integration of the anomaly detector led to performance improvements across all models. Logistic Regression's sensitivity increased to 74-79% and specificity to 72-76% (see Table 2); the EBM model showed further enhancement, with sensitivity rising to 81-83.7% and specificity to 78-81%; the LSTM model maintained its superior performance, reaching a peak sensitivity of 88% (86-90% CI) and specificity of 85% (83-87% CI), again with a 3-hour look-back window.

To illustrate the statistical significance of these findings, Figure X presents the Receiver Operating Characteristic (ROC) curves for all models, clearly demonstrating the superior performance of the LSTM model, particularly when combined with the anomaly detector. The area under the ROC curve (AUC) for LSTM with anomaly detection was 0.92 (0.90-0.94 CI), significantly higher than EBM at 0.88 (0.86-0.90 CI) and Logistic Regression at 0.83 (0.81-0.85 CI).

Discussion

The results demonstrate the effectiveness of AI models in predicting hypotension onset in ICU patients, with the LSTM model consistently outperforming other approaches. This finding can be attributed to LSTM's ability to capture long-term dependencies and complex temporal patterns in time-series data, which is crucial for predicting physiological events like hypotension. The consistent improvement in performance with the 3-hour look-back window across all models suggests that this timeframe provides an optimal balance of historical information. Shorter

windows may not capture enough context, while longer windows might introduce noise or irrelevant data.

The integration of the anomaly detector resulted in notable performance improvements across all models. This enhancement underscores the importance of identifying and handling atypical cases in clinical prediction tasks. The anomaly detector helps by filtering out unusual cases that might confuse the main predictive models, and by potentially identifying subtle, atypical patterns that precede hypotension but might be missed by standard approaches.

The Explainable Boosting Machine, while not matching the LSTM's performance, showed significant improvements over logistic regression. Its strong performance, coupled with its computational efficiency, makes it a viable option for clinical settings where responsive prediction is required.

Logistic Regression, despite its simplicity, showed meaningful improvements with the anomaly detector, highlighting the value of this approach even with less complex models. However, its lower performance compared to EBM and LSTM suggests that the relationship between vital signs and hypotension onset is likely non-linear and complex, requiring more sophisticated modeling approaches. The high sensitivity and specificity achieved by the LSTM model with anomaly detection (88% and 85%, respectively) represented a significant advancement in hypotension prediction. This level of performance could translate to earlier interventions and potentially improved patient outcomes in clinical settings.

However, it is important to note that while the anomaly detector improved performance across all models, it also adds complexity to the system. The trade-off between improved accuracy and increased system complexity should be carefully considered in real-world implementations. These results demonstrate the potential of advanced machine learning techniques, particularly

LSTM with anomaly detection, in improving the prediction of hypotension onset in ICU patients. The consistent superior performance of the 3-hour look-back window provides valuable insights for future model designs and clinical protocols.

However, the success of these models in research settings relies heavily on the quality and frequency of the EMR's vital signs measurements. Translating this performance to real-world clinical environments presents significant challenges. The current and most pressing challenge is that EMR systems often rely on manual data entry by healthcare providers, introducing potential errors and irregular timing between measurements. Even in institutions with automated vital signs recording, data is typically batch-uploaded into EMR systems every few hours rather than streaming in real-time. As a result, any deployed hypotension model may have difficulty in compensating for the differences in time; in addition, without an anomaly detection system, incorrect values may be interpreted by the model which would yield potentially spurious outputs. Other limitations in EMRs include data availability, as data from other traditionally non-interoperable medical devices is not available. Thus, adverse events that have precursors primarily in the form of non-vital signs data cannot be addressed. Successfully deploying any model in clinical practice would require modernizing current infrastructure to support automated, high-frequency vital signs monitoring with real-time data streaming capabilities.

One such system is Darroch Medical Solutions', Beat Analytics System (BAS). The BAS is a monitoring platform that alerts healthcare providers on a mobile application when data from various bedside devices indicate an increased likelihood of a patient experiencing an adverse event in the near future. Healthcare providers can then provide early intervention in an attempt to prevent adverse event precursors from developing into more serious complications. This system addresses several of the aforementioned data consistency and availability issues. The BAS functions by autonomously aggregating high-frequency data directly from IV pumps,

hospital beds, and vital signs monitors and providing a vertically integrated data platform that feeds real time data into various early warning algorithms. In events where either providers or the BAS determines a patient may require medical attention, intelligent alerts are then provided to the healthcare practitioner via a mobile application, Lumori.

This hypotension prediction model could be implemented through a phone-based mobile application user interface such as Lumori (see Figure 4). This interface provides real-time visualization of a patient's hypotension risk over a 24-hour period, with predictions made every 15 minutes. The graph shows the progression of risk throughout the day, allowing healthcare providers to identify patients who are at an increased risk for developing hypotension. The current A.I. prediction, displayed on the user interface, provides the user with an understanding of the patient's status, showing both a range (52-56%) and timeframe (1 hour) for potential hypotension onset. This is complemented by an at-a-glance overview of key vital signs, enabling quick assessment of the patient's current physiological state (see Figure 4).

Furthermore, the mobile application allows for customization of the prediction model parameters (see Figure 5). Healthcare providers can adjust the parameters including data collection frequency, decision threshold, prediction window, and look-back window; this is the benefit of having a vertically integrated data aggregation, analysis, and presentation platform. As parameters are adjusted, the system updates sensitivity and specificity values, allowing providers to see how each configuration affects prediction performance. For instance, a high-acuity unit might benefit from more frequent predictions while a step-down unit might prefer a longer prediction window. The decision threshold can be adjusted based on whether the unit prioritizes catching more potential cases (higher sensitivity) or reducing false alarms (higher specificity). While the 3-hour look-back window was found to be optimal in this study, the ability to adjust this setting allows units to better align the predictions with their patients' physiological patterns and monitoring protocols. The real-time display of sensitivity (82%) and specificity

(75%) for each configuration enables healthcare providers to optimize settings for their specific clinical environment. By providing interpretable predictions and the ability to customize model parameters, this application has the potential to enhance clinical decision-making and facilitate early interventions for hypotension in ICU settings.

Limitations

Despite the promising findings of this study, several limitations should be acknowledged to provide a comprehensive understanding of the results and their implications. First, the study utilized a single dataset from a specific population, which may limit the generalizability of the findings to other populations and clinical settings. The performance of the trained models in this study may vary across different populations with diverse demographic characteristics, underlying health conditions, and risk factors. Future research should aim to validate the models used in this study on multiple datasets from various populations and clinical environments to ensure their applicability and reliability across diverse settings.

Additionally, the study did not consider the impact of treatments, such as vasopressor administration or fluid resuscitation, on the development of hypotension. The absence of these variables may affect prediction accuracy, particularly in cases where interventions alter the course of physiological decline. Furthermore, the dataset did not include information on comorbidities or other patient-specific factors that could influence hypotension risk. Future studies should aim to incorporate these treatment-related and patient-specific variables to provide a more comprehensive picture of hypotension risk and potentially improve model performance.

Lastly, the study's retrospective nature limits any ability to assess the real-time performance and clinical impact of the models in a prospective setting. Prospective validation studies are

necessary to evaluate the model's performance in real-world clinical scenarios and to assess its impact on patient outcomes and clinical decision-making.

In light of these limitations, it is essential to interpret the findings of this study with caution. While the results demonstrate the potential of the trained AI models for hypotension prediction, further research is required to address the limitations in deployment and validate the models' performance in diverse, real-world clinical settings. This includes prospective studies, multi-center validations, and investigations into the impact of these predictive models on clinical decision-making and patient outcomes.

Conclusion

This study presents a novel feature engineering and artificial intelligence approach for predicting hypotension in ICU patients using near real-time vital signs data. The EBM and LSTM model with attention mechanism demonstrated strong predictive performance compared to traditional logistic regression model, offering the potential for earlier intervention and improved patient outcomes. The integration of the anomaly detection system enhanced the robustness of the prediction system, allowing for the identification of atypical cases that may require special clinical consideration.

The comparative analysis of different model architectures and look-back windows provided valuable insights into the temporal dynamics of hypotension development and the most effective approaches for prediction. This knowledge can inform future research and development of clinical decision support systems. The ability to predict hypotension up to one hour before its onset could significantly impact patient management strategies, reducing the incidence of hypotension-related complications and improving overall patient outcomes.

In conclusion, this study represents a significant step forward in the application of AI for critical care management, specifically in the prediction of hypotension. It lays the groundwork for future developments in personalized patient monitoring and proactive critical care interventions, potentially transforming the landscape of intensive care medicine.

Figure 1. Alignment of Hypotensive and Non-hypotensive Patient Stays

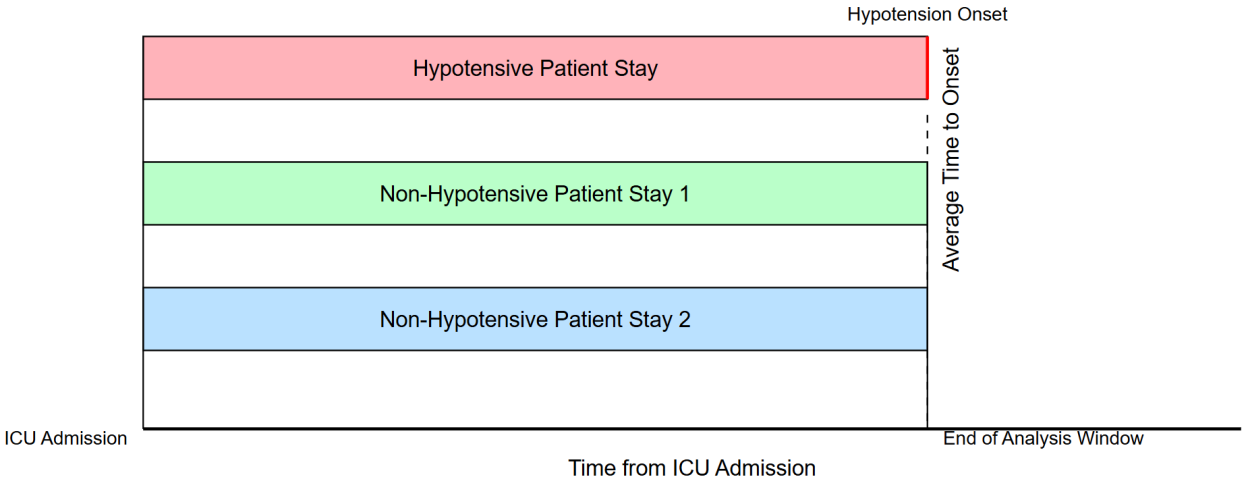

Figure 2. Rolling Window Statistics Across Vital Signs

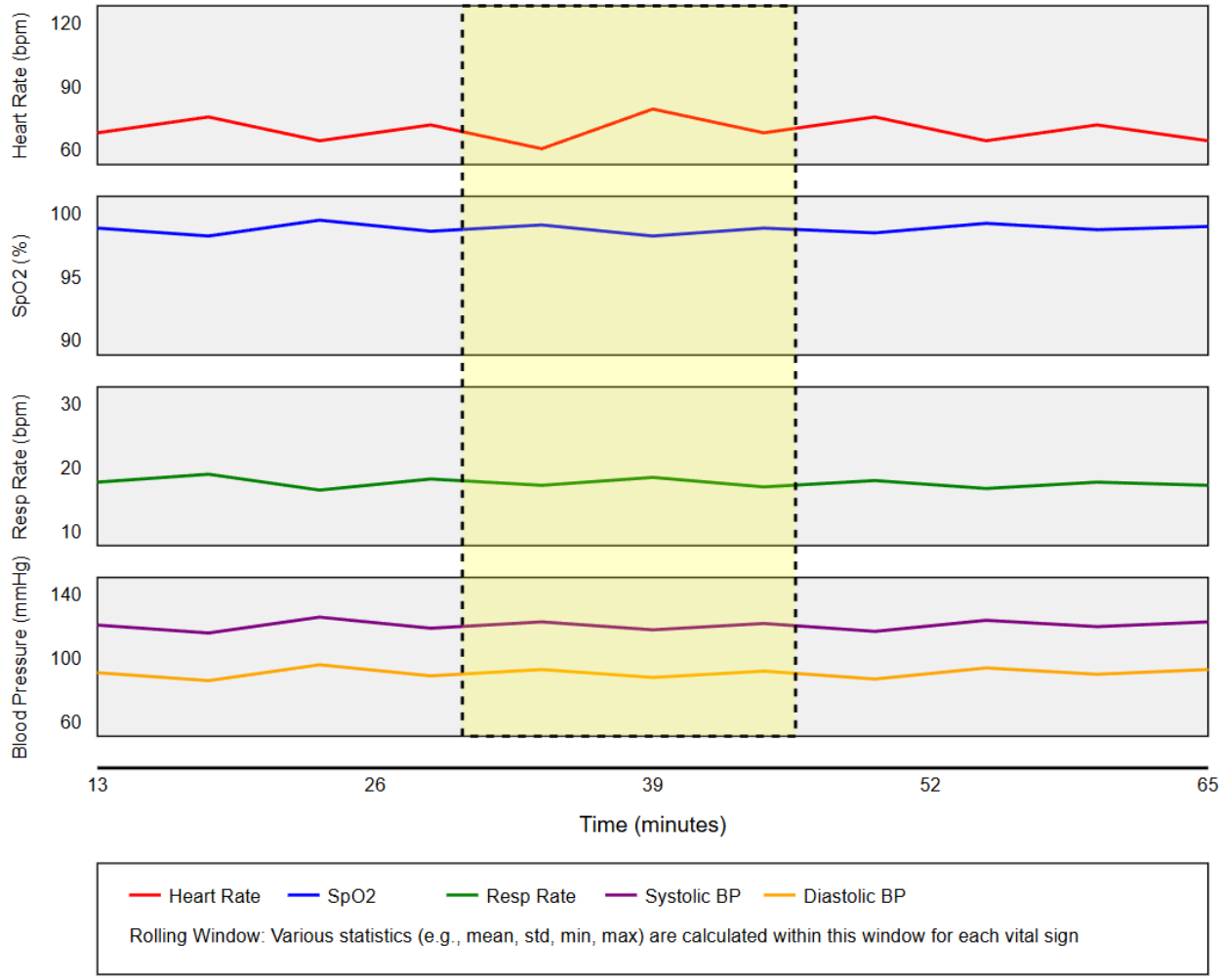

Figure 3. Project Flow Diagram

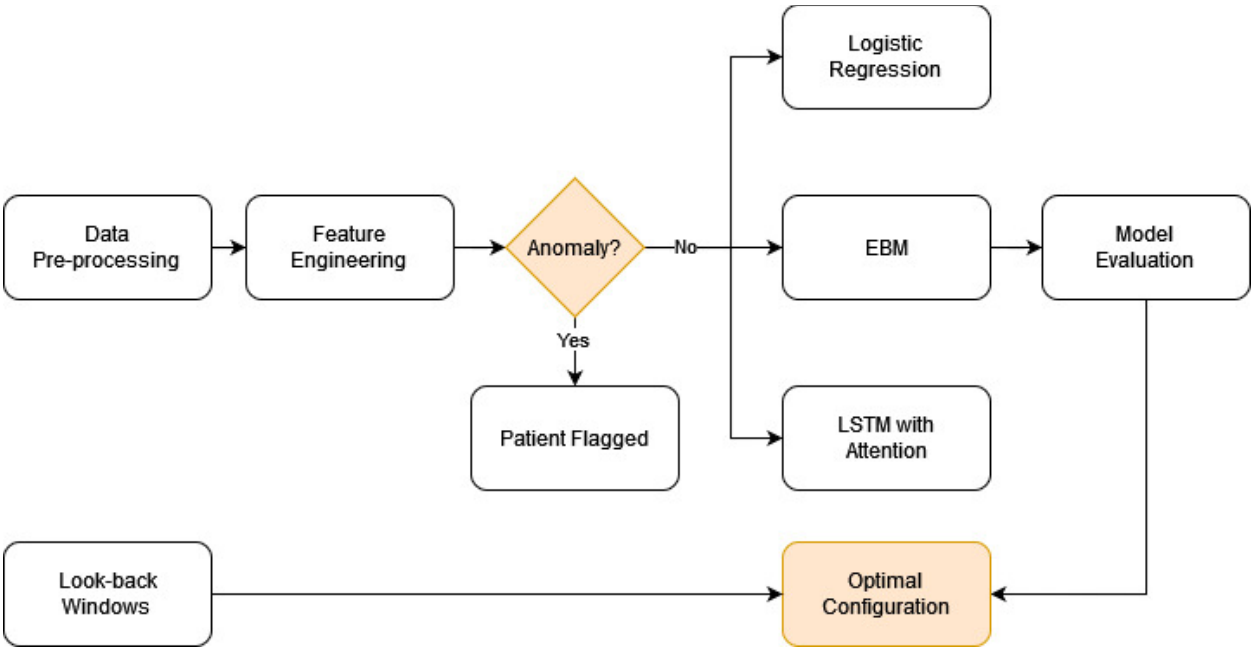
Logistic Regression
Data Pre-processing
Feature Engineering
Anomaly?
No
EBM
Model Evaluation
Yes
Patient Flagged
LSTM with Attention
Look-back Windows
Optimal Configuration

Figure 4. Hypotension User Interface

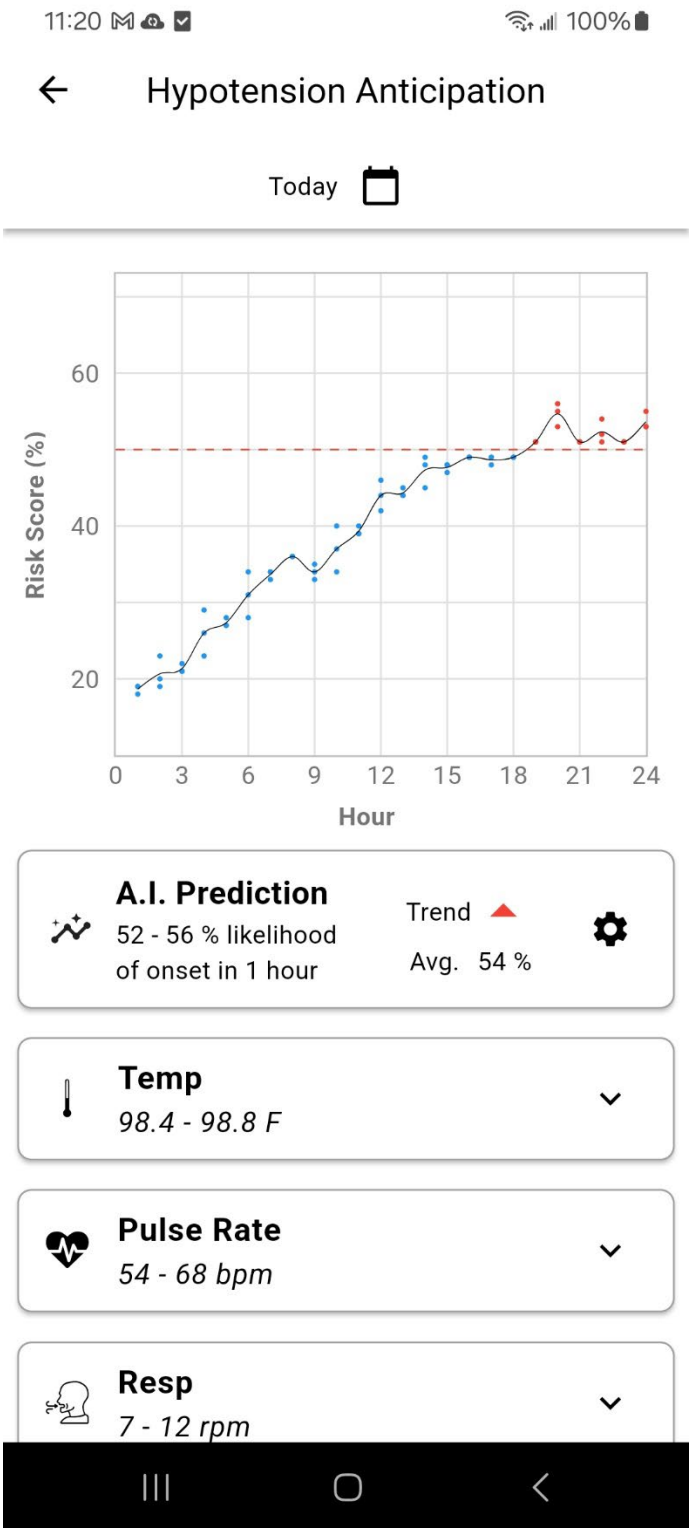

Figure 5. Hypotension Model Configurations

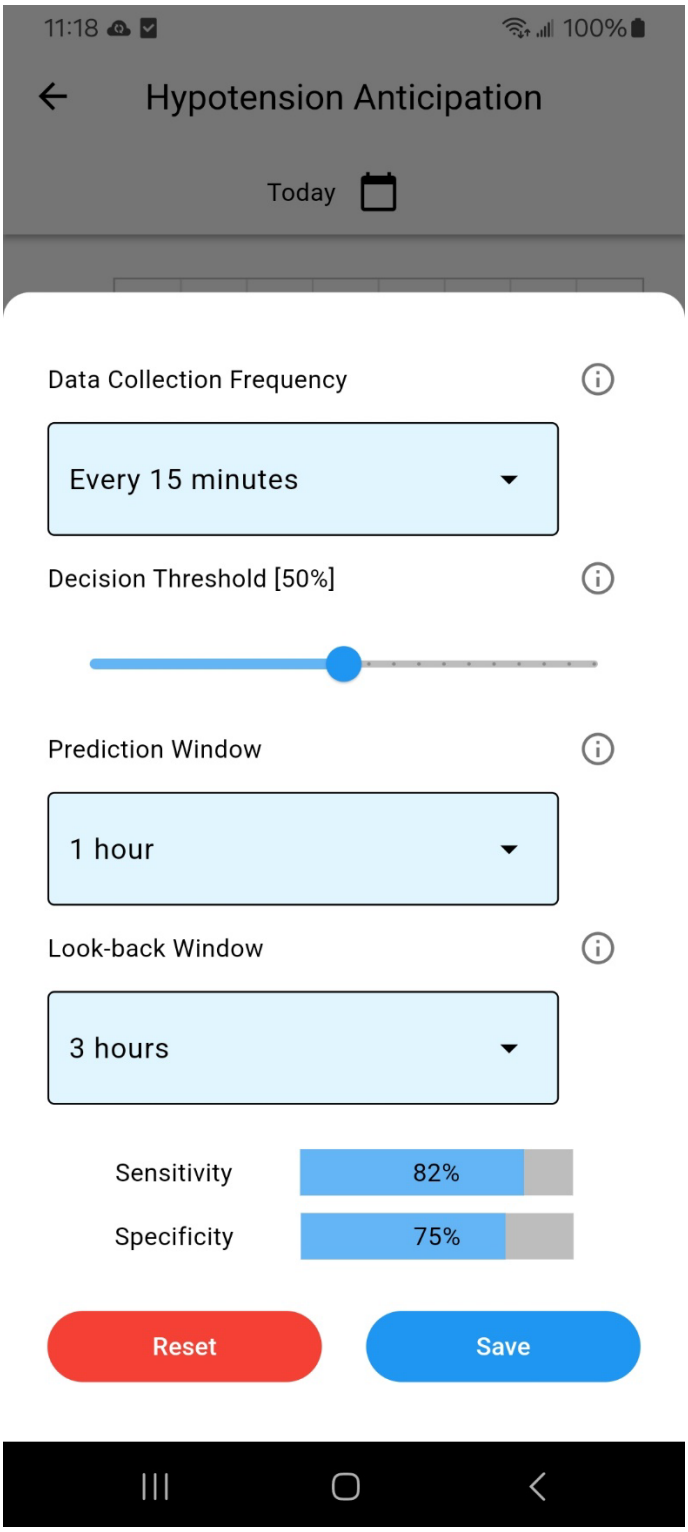

Table 1. Performance Results on Test Set without Anomaly Detector

Model	Look-back Window	Sensitivity (%)	Specificity (%)	F1-Score
Logistic Regression	1 hour	71 (69-73)	69 (67-71)	0.75
	2 hours	73 (71-75)	71 (69-73)	0.77
	3 hours	76 (74-78)	73 (71-75)	0.80
	6 hours	74 (72-76)	70.5 (68.5-72.5)	0.78
EBM	1 hour	78 (76-80)	75 (73-77)	0.82
	2 hours	79 (77-81)	76 (74-78)	0.83
	3 hours	80.7 (78.7-82.7)	78 (76-80)	0.85
	6 hours	79 (77-81)	76 (74-78)	0.83
LSTM with Attention	1 hour	82 (80-84)	79 (77-81)	0.86
	2 hours	83 (81-85)	80 (78-82)	0.87
	3 hours	**85 (83-87)**	**82 (80-84)**	**0.89**
	6 hours	83 (81-85)	80 (78-82)	0.87

Table 2. Performance Results on Test Set with Anomaly Detector

Model	Look-back Window	Sensitivity (%)	Specificity (%)	F1-Score
Logistic Regression	1 hour	74 (72-76)	72 (70-74)	0.73
	2 hours	76 (74-78)	74 (72-76)	0.75
	3 hours	79 (77-81)	76 (74-78)	0.78
	6 hours	77 (75-79)	73.5 (71.5-75.5)	0.76
	1 hour	81 (79-83)	78 (76-80)	0.80

EBM	2 hours	82 (80-84)	79 (77-81)	0.81
	3 hours	83.7 (81.7-85.7)	81 (79-83)	0.83
	6 hours	82 (80-84)	79 (77-81)	0.81
LSTM with Attention	1 hour	85 (83-87)	82 (80-84)	0.84
	2 hours	86 (84-88)	83 (81-85)	0.85
	3 hours	**88 (86-90)**	**85 (83-87)**	**0.87**
	6 hours	86 (84-88)	83 (81-85)	0.85