ICU Patients' Pattern Recognition and Correlation Identification of Vital Parameters Using Optimized Machine Learning Models

Original Scientific Paper

Ganesh Yallabandi
Department of Nephrology,
Apollo Hospitals, Jubilee Hills,
Hyderabad, 500033, Telangana, India

Veena Mayya*
Department of Information & Communication Technology, Manipal Institute of Technology,
Manipal, Manipal Academy of Higher Education,
Manipal, 576104, Karnataka, India.
veena.mayya@manipal.edu *Corresponding author

Jayakumar Jeganathan
Department of General Medicine,
Kasturba Medical College,
Manipal Academy of Higher Education,
Mangalore - 575001, Karnataka, India

Sowmya Kamath S.
Healthcare Analytics and Language Engineering (HALE) Lab, Department of Information Technology,
National Institute of Technology Karnataka,
Surathkal, Mangalore - 575025, Karnataka, India

Abstract *– Early detection of patient deterioration in the Intensive Care Unit (ICU) can play a crucial role in improving patient outcomes. Conventional severity scales currently used to predict patient deterioration are based on a number of factors, the majority of which consist of multiple investigations. Recent advancements in machine learning (ML) within the healthcare domain offer the potential to alleviate the burden of continuous patient monitoring. In this study, we propose an optimized ML model designed to leverage variations in vital signs observed during the final 24 hours of an ICU stay for outcome predictions. Further, we elucidate the relative contributions of distinct vital parameters to these outcomes The dataset compiled in real-time encompasses six pivotal vital parameters: systolic (0) and diastolic (1) blood pressure, pulse rate (2), respiratory rate (3), oxygen saturation (SpO2) (4), and temperature (5). Of these vital parameters, systolic blood pressure emerges as the most significant predictor associated with mortality prediction. Using a fivefold cross-validation method, several ML classifiers are used to categorize the last 24 hours of time series data after ICU admission into three groups: recovery, death, and intubation. Notably, the optimized Gradient Boosting classifier exhibited the highest performance in detecting mortality, achieving an area under the receiver-operator curve (AUC) of 0.95. Through the integration of electronic health records with this ML software, there is the promise of early notifications regarding adverse outcomes, potentially several hours before the onset of hemodynamic instability.*

Keywords*: Mortality Prediction, Clinical Decision Support Systems, Healthcare Informatics*

---

1. INTRODUCTION

In critical care applications, the process of taking practical decisions on managing the care of intensive care patients can help augment the efficiency of caregivers, through the use of predictive data analysis on the large amounts of data generated while monitoring these patients. The most important aspect of a clinical decision support system (CDSS) in the ICU is, undoubtedly, its ability to accurately predict in advance the mortality or severity risk of a patient so that doctors and other healthcare personnel can be prepared to intervene in time with the resources available in the ICU. Apart from measuring the severity of illness, mortality prediction can also play a crucial role in the assessment of treatment and critical care policies in a hospital. Hence, ICU mortality prediction has remained a well-researched problem over the years. Detecting the deterioration of patients in the ICU at an early stage has the potential to enhance patient outcomes. In ICUs, conventional severity scores, including the Acute Physiology and Chronic Health Evaluation (APACHE) score and the Simplified Acute Physiology Score [1], have become essential tools for assessing mortality risk. Globally, APACHE-II, SAPS-II, SOFA [2-6] remain the most widely utilized techniques for gauging mortality risk. However, the factors considered and the severity level assigned can vary significantly based on the chosen severity scale. The computation of severity scores relies on laboratory findings and a patient's medical history, and this process is both time-consuming and complex.

Given the limitations of traditional scoring systems, there is a growing interest among researchers in leveraging machine learning (ML) techniques to predict mortality [7]. Various studies, such as those conducted by Wong et al. [8], Johnson et al. [9], and Schuetz et al. [10], have demonstrated the superior performance of ML models compared to conventional severity scores. The mortality prediction algorithm put forth by Pirracchio et al. [11] utilized a set of 17 variables that are present in the SAPS- II score. In their study, Nemati et al. [12] utilized a set of 65 variables computed on an hourly basis and subsequently given to a ML algorithm to forecast the initiation of sepsis. In a different study, Zahid et al. [13] employed a self-normalizing neural network, leveraging over 20 parameters, to foresee the mortality outcomes for patients within the ICU. Another recent investigation by Camacho-Cogollo et al. [14] adopted a distinct approach by employing 31 medically relevant features (MRF) to predict sepsis. These features were meticulously chosen from a pool of 145 potential features, guided by the expert medical insights of a proficient physician. Subsequently, a variety of ML models were tested using this refined set of features. Weissman et al. [15] found that the inclusion of clinical notes along with structured clinical data dramatically improved the ability of ML models to predict ICU mortality. Payrovnaziri et al. [16] used both unstructured (discharge summaries) and structured patient data for performing myocardial infarction based mortality prediction. In a recent study by Huang et al. [17], a novel stacking ensemble model was devised to address the challenge of mortality risk assessment in patients with cerebrovascular conditions. This innovative model made use of multimodal data, integrating various sources, including laboratory test data, structured information, and textual radiology reports. However, the incorporation of these predictive systems into the healthcare realm encounters noteworthy challenges. This is primarily due to the need for a substantial number of features, including intricate laboratory findings, measurements of urine output, evaluations based on the Glasgow Coma Scale (GCS), and even clinical notes. These intricacies create impediments for the practical implementation of these systems, as they necessitate medical personnel to manually input a multitude of parameters to ensure the precision of predictions. Often, this process demands repeated investigations, further contributing to the intricacy and potentially hindering the effective application of these predictive models.

Conducting statistical analysis through bivariable trend models, Churpek et al. [18] determined that vital sign trends play a pivotal role in the detection of critical illness. In the study conducted by Bloch et al. [19], the authors manually selected four important features. These features were determined by analyzing their significance across a range of tested models. The selected features include the median change in heart rate, the number of trend changes in respiratory rate, the minimal change in respiratory rate, and arterial pressure. Recent studies by Baker et al. [20] emphasize the significance of vital signs as influential factors. In their work, they fused convolutional (CNN) layers with bidirectional long short-term memory (BiLSTM) networks to anticipate mortality using statistics characterizing variations in heart rate, blood pressure, respiratory rate, blood oxygen levels, and temperature. They derived a total of 49 statistical features for each of the seven vital signs. It's apparent from the study that prior to employing ML models, it's imperative to compute the statistical properties of the vital signs. The need to perform these computations prior to implementing the ML models introduces complexities that could hinder the seamless application of this approach in a clinical setting.

A substantial amount of time and effort is dedicated to recording vital signs in the ICU. However, there is a scarcity of studies focused solely on recognizing trends derived from these fundamental and straightforward parameters. The importance of vial sign trends in relation to patient outcomes remains relatively underexplored. The primary objective of this study is to identify discernible patterns within essential vital signs, namely blood pressure, respiratory rate, pulse rate, and SpO2. These vital signs have been observed to display correlations with the eventual outcomes of patients in the ICU. This investigation employs optimized ML techniques to achieve this objective and further examines the individual contributions of each vital parameter to these outcomes.

The rest of the paper is organized as follows: Section 2 provides a comprehensive review of the most relevant and effective mortality prediction systems reported in the literature. Section 3 details the data collation process and the methodology used for patients' pattern recognition and correlation identification of vital parameters. Section 4 documents the evaluation of the ML models and details the extensive experiments conducted using ML models trained under different timelines and hyperparameter optimization. Section 5 concludes the proposed experimental study and presents future work.

2. LITERATURE REVIEW

Mortality prediction plays a crucial role in assessing the severity of an illness and aiding in the enhancement of patients' prognoses. In recent years, researchers have focused on designing non-parametric CDSSs built using data mining, ML, and deep learning (DL) techniques to enable higher accuracy for ICU mortality prediction. The majority of these studies have made use of publicly accessible datasets such as the multiparameter intelligent monitoring in intensive care (MIMIC) dataset [21-23], the PhysioNet computing in cardiology challenge dataset [24], the high time-resolution ICU dataset (HiRID) [25], and the women in data science (WiDS) challenge dataset [26]. Through these studies, the versatility and effectiveness of ML in the critical care domain have been convincingly demonstrated. Nevertheless, a notable challenge arises from the complexity of these datasets, each containing more than 20 parameters. This complexity poses a hindrance to the practical implementation of these ML systems, as it requires medical personnel to manually input a multitude of parameters to ensure the accuracy and precision of the predictions. This manual input process can be time-consuming and prone to errors, potentially undermining the overall utility of these predictive systems.

As the field of ML continues to evolve, addressing this challenge is pivotal for achieving seamless integration of these predictive models into real-world clinical settings. Streamlining the data input process, reducing the number of required parameters, or developing automated methods for data extraction could all contribute to enhancing the feasibility and effectiveness of ML applications in critical care scenarios. This ongoing effort to bridge the gap between complex datasets and practical implementation holds the potential to revolutionize the way critical care is managed and optimized. Recent studies utilized automated feature selection [27-30] and reduction [31, 32] techniques to select the important parameters from the recorded ICU datasets. As the number of ICU patients increases, accumulating a large number of parameters becomes increasingly difficult. In this investigation, we conducted a pilot study with only six vital signs that are routinely monitored during ICU stays. According to recent studies, ML models were found to have superior predictive capabilities with structured data input than DL models [17]. Therefore, this study focuses on optimizing ML models for predicting mortality using crucial vital signs.

3. MATERIALS AND METHODS

3.1. PARTICIPANTS AND DATA

The study was conducted in ICUs attached to Kasturba Medical College, Mangalore, and Manipal Academy of Higher Education, Manipal, India. Patients with an age greater than 18 years who are admitted to the ICU from August 2019 to November 2020 and who provide their consent in a written informed form are included in this study. Patients who stayed less than 24 hours in the ICU or who were admitted before August 2019 and those who denied consent are excluded from the study. A total of 285 patients' data were considered for the study. Each patient record includes age, gender, length of ICU stay, outcome (recovered, death, intubated), and the last 24 hours of time series data. Time series data includes six vital parameters: systolic (0), diastolic (1), pulse rate (2), respiratory rate (3), SpO2 (4), and temperature (5). Fig. 1 illustrates the comprehensive procedure employed during data collection, with the utilization of the Philips mp20 monitor for recording vital signs. The distribution of the ICU patient data is depicted in Fig. 2. It can be observed that the data is highly imbalanced in terms of recorded patient outcomes.

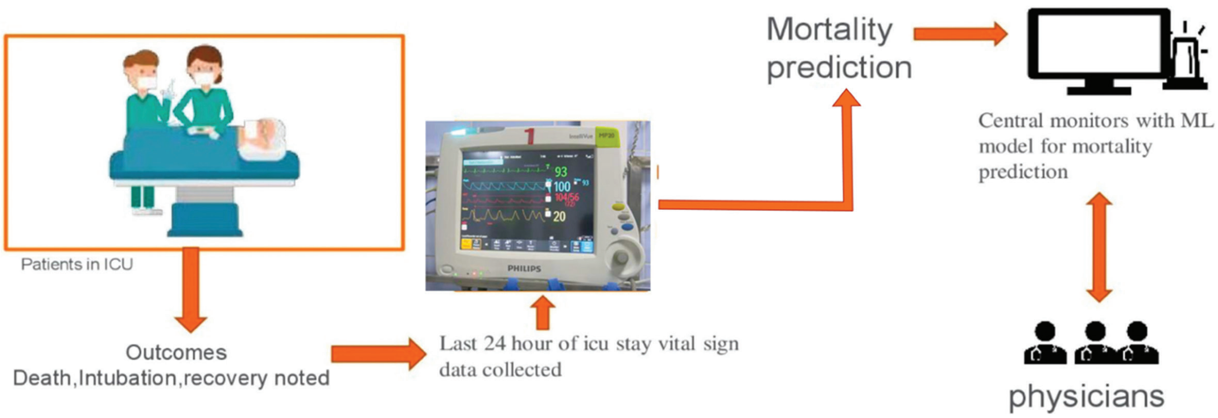


Fig 1. Structure of the Proposed CVD Prediction System

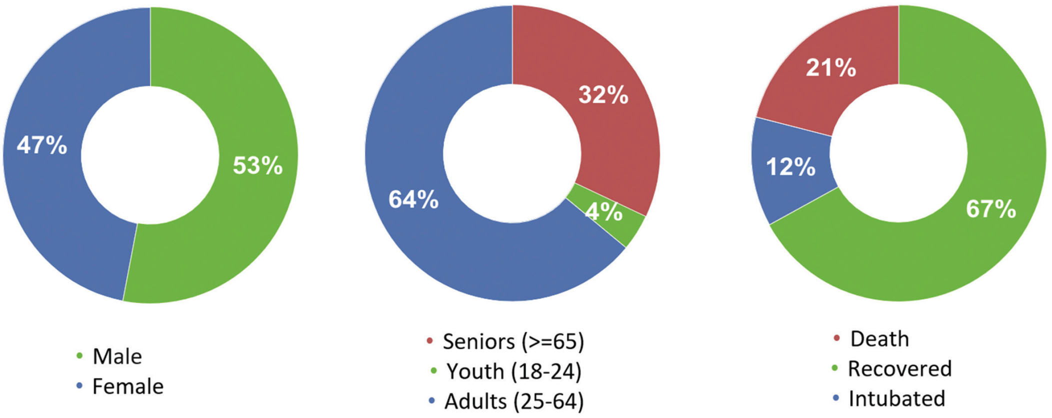


Fig 2. Patient Data Analysis: (a)Gender-wise distribution (b)Age-wise distribution (c)Outcome

3.2. METHODOLOGY

Due to the fact that a subset of rows (about 250) were marked as ”not recorded” (NR), these cases were imputed using the ”backward filling” method, which used the set of recorded patient data that came before it. The dataset encompassed a total of 285 patient records, each encompassing six vital parameters, and spanned a continuous 24-hour period. This data was organized in a structured format (285, 24, 6). Because ML models work best with one-dimensional (1D) data, the data for each patient record, which was a sequence of 24 values for each of six vitals, was put into a linear format with a row major layout, which led to 144 features. As a result of this reformatting, a distinct 1D dataset was generated for each individual patient, facilitating compatibility with the ML algorithms.

Ten different ML classifiers, including K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), AdaBoost, RUSBoost, Random Forest (RF), Decision Tree (DT), Gradient Boosting (GB), XGBoost classifier, and textRNN, were put through a thorough evaluation and fine-tuning process. KNN is a simple and intuitive algorithm that classifies data records based on the majority class among their k-nearest neighbours in the feature space. It measures distances between data points and assigns labels based on the neighbours' labels. SVM seeks to find a hyperplane that best separates different classes of data. SVM aims to maximize the margin between classes and can handle both linear and non-linear separation. MLP consists of multiple layers of interconnected nodes (neurons). It is capable of learning complex relationships in data through forward and backward propagation of signals, making it suitable for a wide range of tasks. AdaBoost (Adaptive Boosting) is an ensemble learning technique that combines the outputs of multiple weak classifiers to create a stronger overall classifier. It assigns higher weights to misclassified instances in each iteration to improve classification performance. RUSBoost (Random Under- Sampling Boosting) is a variant of AdaBoost that incorporates random undersampling of the majority class. This helps address the class imbalance in the dataset, making it particularly useful for imbalanced data scenarios. RF is an ensemble method that constructs multiple DTs during training and combines their outputs to make predictions. It improves accuracy and reduces overfitting by introducing randomness in the tree-building process. A DT is a hierarchical structure that recursively splits data into subsets based on the values of input features. It makes decisions by traversing the tree from the root node to a leaf node, where the final classification is determined. Gradient Boosting is another ensemble method that builds a strong model by sequentially adding weak learners (usually DTs) and focusing on instances that were previously misclassified. It aims to minimize the prediction error iteratively. XG Boost (Extreme GB) is an optimized and highly efficient implementation of GB. It includes regularization techniques, handling missing values, and parallel processing to enhance performance and predictive accuracy. In order to configure the TextRNN model’s hyperparameters, a combination of empirical observations and systematic experimentation was utilized. Through careful calibration of multiple variables, it was determined that implementing GRU (Gated Recurrent Unit) units in a two-layer configuration produced significantly enhanced outcomes.

The primary objective was to effectively categorize the input ICU time series dataset into 3 discrete classes: recovery, mortality, and intubation. Initially, each individual vital parameter is used to predict the outcomes. Further, all possible combinations of vital signs were experimented to determine the most crucial vital parameters. Since the data is highly imbalanced, combining the weak classifiers would improve the performance of the model. So the experiment was conducted using boosting classifiers.

3.3. EXPERIMENTS AND RESULTS

We evaluated the ML classifiers using a five-fold cross-validation method. Within this approach, a single fold was dedicated to testing, while the remaining folds were employed for training the classification model. This process was reiterated across all folds to ensure a uniform and stable performance evaluation. The Python open-source ML packages [33] were used for carrying out the experiments. Initially, the default values as set by the Python packages were set for network parameters. The results obtained are listed in Table 1. In our efforts to enhance performance, we endeavored to ensemble the top three performing ML models (RF, AdaBoost, and Gradient Boost) using voting and stacking algorithms. However, the process of ensembling did not yield a substantial improvement in overall performance. This could potentially indicate that the inherent ensembling nature of GB already integrated the advantages offered by ensembling with different classifiers.

Table 1. Mortality prediction results

Classifier	Precision	Recall	F1-score
TextCNN	0.51	0.57	0.54
RUSBoost	0.61	0.64	0.62
MLP	0.68	0.67	0.67
KNN	0.68	0.73	0.66
DT	0.68	0.69	0.69
AdaBoost	0.69	0.72	0.70
RF	0.73	0.75	0.73
XGBoost	0.74	0.76	0.74
SVM	0.65	0.74	0.68
GB	0.75	0.77	0.75

Given the superior performance of tree-based algorithms in comparison to other ML models, we meticulously refined their hyperparameters through the utilization of the particle swarm optimizaiton (PSO) algorithm [34–36]. Notably, we observed that certain hyperparameters—namely, n_estimators, max_depth, min_samples_leaf, max_features, and min_samples_split held significant importance across the spectrum of tree-based ML models. After achieving refined parameters using the PSO algorithm for the RF model, these fine-tuned hyperparameters were applied as initializations for the remaining ML models. Remarkably, it was discerned that the optimized hyperparameters from the initial RF tuning yielded the best performance across the other ML models as well. The hyperparameter ranges and the corresponding optimal values determined by the PSO approach have been detailed in Table 2. The outcomes achieved through this optimization process, encompassing the refined hyperparameters, have been documented in Table 3. Notably, to ensure a fair comparison, the same random state was upheld throughout the experiments. Evidently, the proposed optimization methodology led to a significant performance enhancement, with improvements of up to 10%.

Table 2. Hyper-parameters Range

Hyperparameter	Lower bound	Upper bound	PSO chosen value
n_estimators	10	200	168
max_features	1	20	8
max_depth	2	20	10
min_samples_split	2	20	10
min_samples_leaf	1	20	1

Table 3. Optimized classifiers adopted for mortality prediction

Classifier	Precision	Recall	F1-score
RUSBoost	0.68	0.68	0.68
DT	0.69	0.71	0.70
AdaBoost	0.76	0.78	0.77
RF	0.80	0.81	0.79
XGBoost	0.75	0.77	0.76
GB	0.81	0.82	0.80

Due to the minimal variance observed in body temperature among the studied cases within the last 24 hours after admission, its contribution to predictive modelling was limited. Consequently, the focus shifted to the remaining vital parameters for subsequent experimentation, aimed at comprehending the individual significance of each parameter on performance. In this context, a minimum of three vital parameters were selected at a time for classification using the GB classifier. A comprehensive set of 16 combinations was tested, and the corresponding accuracies for each combination are detailed in Table 4. Notably, the optimized models showcased enhancements in the classifier's performance. While the combination of systolic blood pressure, pulse rate, and SpO2 demonstrated potential for achieving higher accuracy on its own, it was observed that incorporating all vital parameters led to improvements not only in accuracy but also in precision.

Table 4. Mortality prediction with Adaboost classifier using combinations of vitals

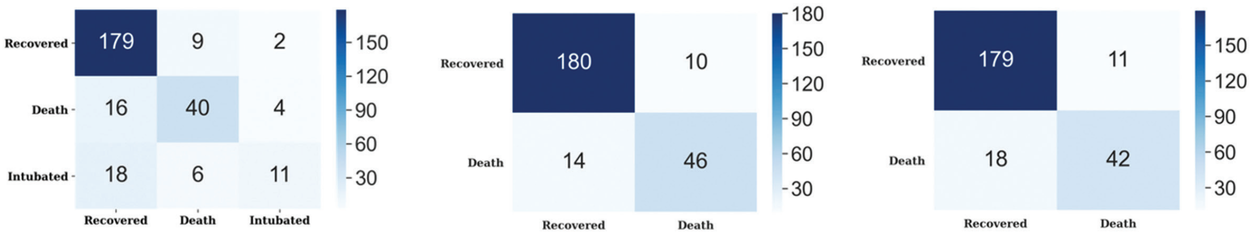
Combination of vitals	Accuracy	Precision	Recall	F1-score
systolic(0)	0.72	0.67	0.72	0.68
diastolic(1)	0.68	0.61	0.68	0.63
pulse rate(2)	0.71	0.66	0.71	0.68
respiratory rate(3)	0.69	0.64	0.69	0.65
SpO2(4)	0.71	0.67	0.71	0.68
(0,1,2)	0.74	0.67	0.74	0.69
(0,1,3)	0.74	0.69	0.74	0.70
(0,1,4)	0.76	0.73	0.76	0.74
(0,2,3)	0.75	0.70	0.75	0.71
(0,2,4)	0.82	0.80	0.80	0.80
(0,3,4)	0.78	0.75	0.78	0.76
(1,2,3)	0.74	0.69	0.74	0.70
(1,2,4)	0.78	0.76	0.78	0.77
(1,3,4)	0.75	0.72	0.75	0.73
(2,3,4)	0.79	0.77	0.79	0.77
(0,1,2,3)	0.75	0.71	0.75	0.71
(0,1,2,4)	0.79	0.77	0.79	0.78
(0,1,3,4)	0.76	0.73	0.76	0.74
(0,2,3,4)	0.82	0.80	0.81	0.80
(1,2,3,4)	0.79	0.77	0.79	0.77
(0,1,2,3,4)	0.82	0.81	0.82	0.80

Fig. 3(a) illustrates the confusion matrix derived from the predictions generated by employing the combination of all vital parameters with the optimized GB classifier. Notably, the model demonstrates a substantial capacity to accurately distinguish between recovered patients and those at risk of mortality. However, due to the study's limited inclusion of only 13% of intubated patients, the model struggled to discern the intricate patterns necessary for precise classification. To address this limitation, further analysis was undertaken with a focus on binary classification into two distinct categories: the possibility of death or survival. In pursuit of this, all data pertaining to intubated patients was excluded. The outcome of this refined classification strategy is depicted in Fig. 3(b), showcasing an impressive 90% accuracy achieved through the utilization of the proposed optimized GB classifier.

The study achieved remarkable results in binary classification using an optimized GB classifier. Precision, recall, accuracy, and F1 score were all recorded at 0.90, indicating a high level of performance. In the context of tree-based ML models, the process of identifying mortality instances can be understood by examining the DTs employed. The visualization of these DTs can be seen in Fig. 4. The data for each individual record, denoted as X in Fig. 4, is organized in a specific order: systolic blood pressure (0), diastolic blood pressure (1), pulse rate (2), respiratory rate (3), and SpO2 (4). Each parameter has associated data collected over a 24-hour period. By analyzing the DTs, it is possible to pinpoint the exact hour and parameter that contributed to a specific decision. This provides valuable insights into the factors influencing the classification outcome.

The ROC curve, presented in Fig. 5(a), showcases the performance of the optimized GB binary classifier. This curve is constructed by plotting the true positive rate (TPR) against the false positive rate (FPR), with TPR represented on the y-axis and FPR on the x-axis. Remarkably, the area under the ROC curve, which amounts to 0.95, signifies the model's ability to make accurate predictions approximately 95% of the time based on the last 24 hours of vital sign data.

For a more detailed analysis, we conducted an assessment by excluding vital sign data from the 12 hours immediately preceding the outcome within the last 24 hours. Fig. 3(c) represents the confusion matrix derived from predictions made by the optimized GB binary classifier utilizing combinations of all vital parameters. Impressively, the model continues to exhibit substantial accuracy in distinguishing between recovered patients and those potentially facing mortality. Illustrated in Fig. 5(b), the area under the ROC curve is presented, plotting the TPR against the FPR.



(a) (b) (c)

Fig. 3. Confusion matrix obtained for (a)all classes (b)two classes (c)two classes for 12 hours data

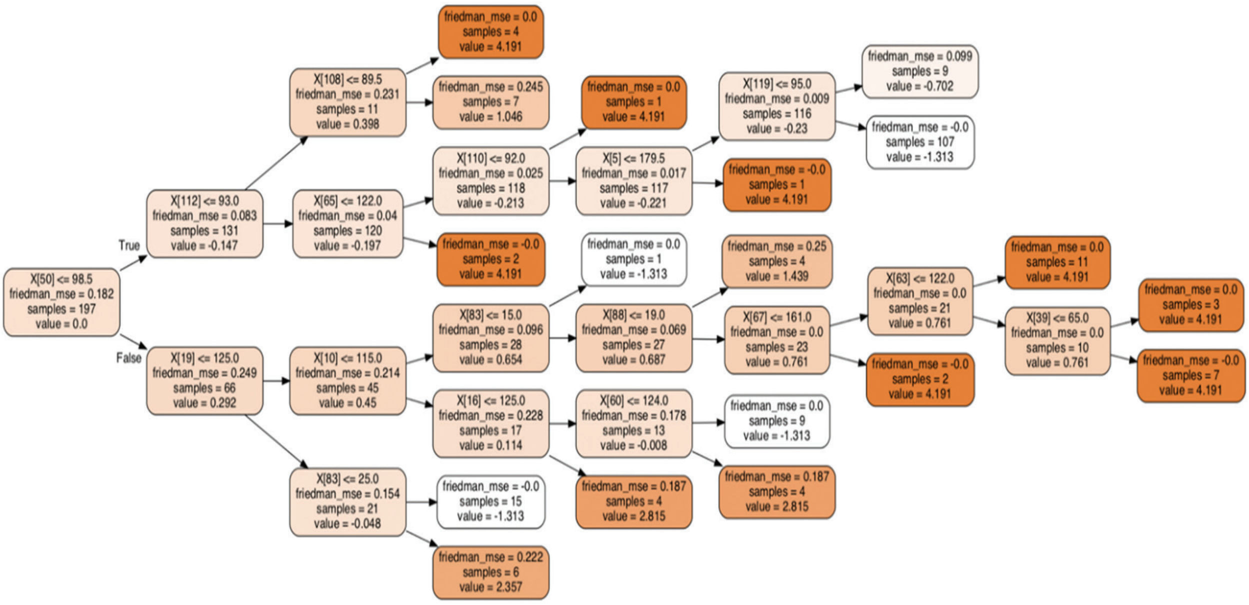
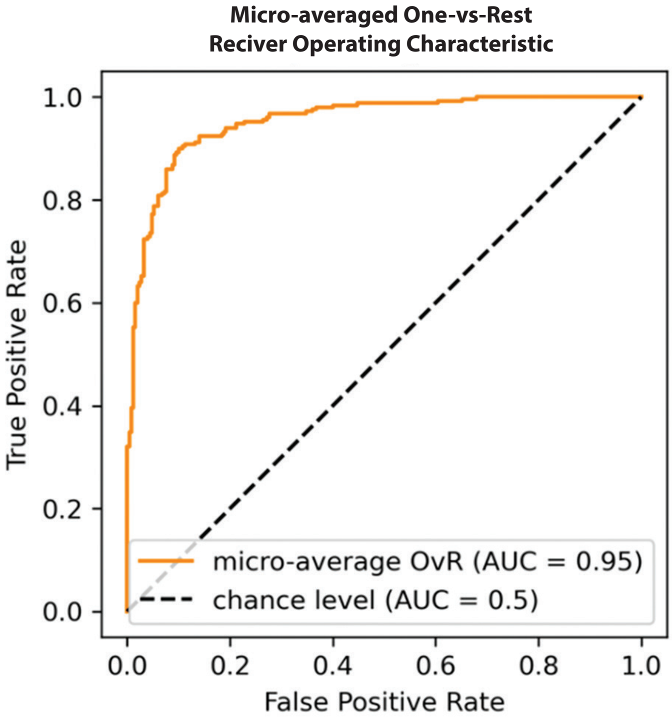
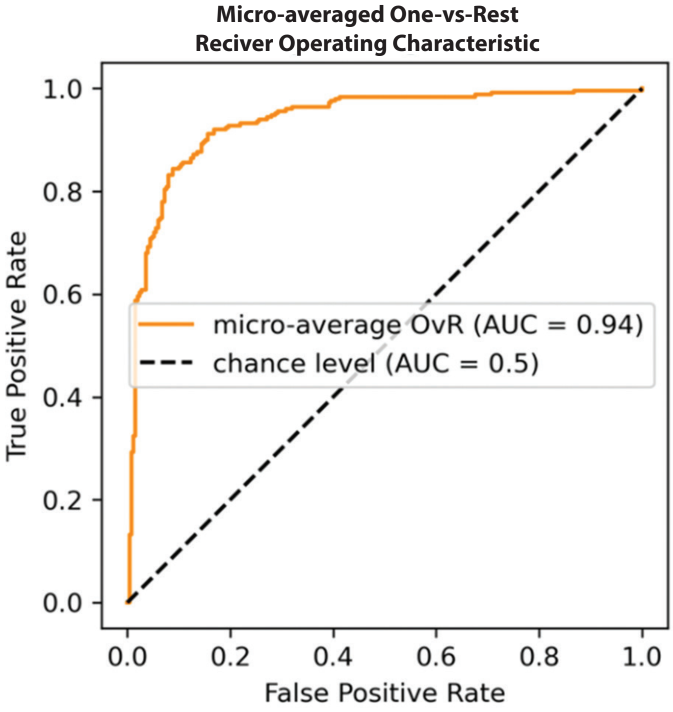


Fig. 4. The decision tree for mortality detection



(a)



(b)

Fig. 5. ROC curve obtained for mortality detection (a) for 24 hours input (b) for 12 hours input

The performance mirrors the outcomes achieved with the 24-hour data window, showcasing the classifier's consistency and ability to predict mortality effectively. In Table 5, a comprehensive summary of precision, recall, F1 score, and accuracy for the initial 12-hour window preceding the actual outcome is provided. The average precision, recall, and F1 score are detailed in the same table, reflecting the model's performance. Notably, the optimized GB classifier achieves 87% accuracy in predicting mortality, even when utilizing data collected during the first 12 hours before the actual outcome. This result underscores the classifier's robust performance across this critical early time frame. In essence, the proposed optimized GB classifier showcases commendable performance across both 12-hour and 24-hour time windows, attesting to its proficiency in forecasting outcomes.

Table 5. Performance obtained for 12 hours data

Metric	Recovered	Death	Weighted average
Precision	0.90	0.76	0.87
Recall	0.93	0.68	0.87
F1 Score	0.92	0.72	0.87
Accuracy	-	-	0.87

4. DISCUSSION

In this work, we have proposed an optimized model for predicting hospital mortality in ICU patients, designed to be especially applicable and beneficial in low- and middle-income countries. Our approach centers on utilizing a streamlined set of variables that are both readily accessible and straightforward to collect. Notably, the model that exhibited the highest performance in our investigation was a GB classifier, which harnesses the strength of an ensemble of weak classifiers, requiring solely vital sign data. These are routinely measured and effortlessly acquired within an ICU setting. Significantly, these variables do not necessitate knowledge of the patient's diagnosis or laboratory results.

Our findings closely align with those of a study conducted by Alistair et al. [37]. In their work, Alistair et al. [37] developed a model relying on a staggering 148 distinct variables, the majority of which emanate from intricate laboratory results, urine output, and Glasgow Coma Scale (GCS) measurements—a collection of vital parameters. This approach, however, introduces complexities that can hinder the model's practical utility. Medical practitioners would need to measure and input an extensive array of factors for an accurate prognosis, which not only consumes time but also necessitates repeated investigations. In contrast, our study offers a user-friendly solution devoid of labour-intensive data entry or cumbersome investigations during the ICU admission process. The optimized GB classifier showcased notable performance even when using vital sign data gathered 12 and 24 hours before the outcome. While our findings are encouraging, further research is required to explore the impact of more frequent intervals for vital sign data collection, aiming to enhance the accuracy of ML models in predicting outcomes.

4.1. LIMITATIONS

The performance of the ML model in identifying intubation outcomes among patients in our study was limited, possibly due to inadequate training on a large dataset. Additionally, our study did not encompass data from post-intubated patients. Furthermore, patients who were recovered and discharged from the ICU were not monitored until their discharge, potentially impacting the model's predictive capability. The low variance in temperature observed within the last 24 hours may have hindered its contribution to prediction, and a more frequent temperature monitoring interval would be necessary for a comprehensive generalization of findings. Notably, within our study, systolic blood pressure emerged as the most influential vital parameter for mortality prediction. However, it's important to acknowledge that blood pressure fluctuations are often evident in patients facing terminal cardiovascular collapse. The unique attribution of critical parameters by the ML model to each feature might be specific to our institution. Despite accounting for patient variability in our hospital and study, external and independent evaluations are imperative to validate the findings. It's crucial to underscore that our results are reflective of a single hospital population, whereas other studies have drawn insights from patient data collected across multiple hospitals. Therefore, external validation using data from a distinct institution is pivotal before broad conclusions can be drawn. Our research employed a smaller database compared to previous studies that aimed to make extensive population-level generalizations. Being retrospective in nature, our data gathering process could benefit from oversight post-results. While our dataset primarily captures the last 24 hours of a patient's ICU stay, an approach involving data collected throughout the entire ICU stay and analyzed across different time windows could enhance model accuracy. The exploration of DL techniques [38, 39], often deemed superior to supervised learning techniques, is an avenue worth exploring in future studies. However, despite experimenting with state-of-the-art DL models, the limitations posed by our data's scope led to relatively lower performance with DL models. To realize practical applicability, the developed ML model must undergo testing in a real-world environment.

5. CONCLUSIONS

Among the vital parameters studied, systolic blood pressure emerged as the most significant predictor linked to mortality prediction, underscoring its pivotal role. Additionally, SpO2 and pulse rate exhibited notable associations with predictive outcomes. Conversely, temperature variance exhibited a limited contribution to predicting outcomes in this study, potentially due to its low variability. An intriguing observation is the potential of a combination of systolic blood pressure, pulse rate, and SpO2 to yield enhanced accuracy. Moreover, incorporating all vital parameters not only enhances accuracy but also improves precision. To further refine the model's accuracy in predicting intubated patients, additional training with larger databases is essential.

Remarkably, the proposed optimized GB classifier achieves 90% accuracy in predicting recovery or mortality, utilizing a mere six recorded vital parameters. These findings suggest the viability of employing ML techniques for routine monitoring of ICU patients. This presents an opportunity to evolve beyond traditional prognostic scores like APACHE, SAPS, and SOFA and integrate more accessible and less intricate ML techniques that rely solely on essential vital parameters. Regular implementation of such techniques can serve as a valuable supplement to conventional ICU scoring systems. As ICU technologies progress towards greater automation, integrating central monitors with ML software could provide an early warning system, preemptively alerting healthcare providers to potential adverse outcomes hours before hemodynamic instability manifests.

ETHICAL APPROVAL

Ethical approval for this study was obtained from Institutional Ethics Committee Kasturba Medical College, Mangaluru (ID IEC KMC MLR 0919/445).

DATA AVAILABILITY

The data are not publicly available due to privacy and ethical concerns. The data that support the findings of this study are available from the first author, Ganesh Y., upon reasonable request.

6. REFERENCES

[1] J.-L. Vincent et al. "The SOFA score to describe organ dysfunction/failure", Intensive Care Medicine, Vol. 22, No. 7, 1996, pp. 707-710.

[2] W. Knaus, E. Draper, D. Wagner, J. Zimmerman, "Apache ii: A severity of disease classification system", Critical Care Medicine, Vol. 13, No. 10, 1985, p. 818-829.

[3] W. Knaus et al. "The APACHE III prognostic system: Risk prediction of hospital mortality for critically ill hospitalized adults", Chest, Vol. 100, No. 6, 1991.

[4] J. E. Zimmerman, A. A. Kramer, D. S. McNair, F. M. Malila, "Acute physiology and chronic health evaluation (Apache) iv: Hospital mortality assessment for today's critically ill patients", Critical Care Medicine, Vol. 34, No. 5, 2006.

[5] J.-R. Le Gall, P. Loirat, A. Alperovitch, P. Glaser, C. Granthil, D. Mathieu, P. Mercier, R. Thomas, D. Villers, "A simplified acute physiology score for ICU patients", Critical Care Medicine, Vol. 12, No. 11, 1984, pp. 975-977.

[6] J.-R. Gall, S. Lemeshow, F. Saulnier, "A new simplified acute physiology score (SAPS II) based on a European/North American multi-center study", Journal of the American Medical Association, Vol. 270, No. 24, 1993.

[7] A. Naemi et al. "Machine learning techniques for mortality prediction in emergency departments: a systematic review", BMJ Open, Vol. 11, No. 11, 2021.

[8] L. Wong, J. Young, "A comparison of ICU mortality prediction using the Apache II scoring system and artificial neural networks", Anaesthesia, Vol. 54, No. 11, 1999, pp. 1048-1054.

[9] A. Johnson, A. Kramer, G. Clifford, "A new severity of illness scale using a subset of acute physiology and chronic health evaluation data elements shows comparable predictive accuracy", Critical Care Medicine, Vol. 41, 2013.

[10] P. Schuetz et al. "Effect of procalcitonin-guided antibiotic treatment on mortality in acute respiratory infections: a patient level meta-analysis", The Lancet Infectious Diseases, Vol. 18, 2018.

[11] R. Pirracchio, M. L. Petersen, M. Carone, M. R. Rigon, S. Chevret, M. J. van der Laan, "Mortality prediction in intensive care units with the super ICU learner algorithm (sicula): a population-based study." The Lancet. Respiratory medicine, Vol. 31, 2015, pp. 42-52.

[12] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, T. G. Buchman, "An interpretable machine learning model for accurate prediction of sepsis in the ICU", Critical Care Medicine, Vol. 46, No. 4, 2018, pp. 547-553.

[13] M. Zahid, J. Lee, "Mortality prediction with self normalizing neural networks in intensive care unit patients", Proceedings of the IEEE EMBS International Conference on Biomedical and Health Informatics, Las Vegas, NV, USA, 4-7 March 2018, pp. 226-229.

[14] J. E. Camacho-Cogollo, I. Bonet, B. Gil, E. Iadanza, "Machine learning models for early prediction of sepsis on large healthcare datasets", Electronics, Vol. 11, No. 9, 2022.

[15] G. Weissman, R. Hubbard, L. Ungar, M. Harhay, C. Greene, B. Himes, S. Halpern, "Inclusion of

unstructured clinical text improves early prediction of death or prolonged ICU stay", Critical Care Medicine, Vol. 46, No. 7, 2018, pp. 1125–1132.

[16] S. N. Payrovnaziri, L. A. Barrett, D. Bis, J. Bian, Z. He, "Enhancing prediction models for one-year mortality in patients with acute myocardial infarction and post myocardial infarction syndrome", Studies in health technology and informatics, Vol. 264, 2019, pp. 273–277.

[17] R. Huang et al. "Stroke mortality prediction based on ensemble learning and the combination of structured and textual data", Computers in Biology and Medicine, Vol. 155, 2023.

[18] M. M. Churpek, R. Adhikari, D. P. Edelson, "The value of vital sign trends for detecting clinical deterioration on the wards." Resuscitation, Vol. 102, 2016, pp. 1-5.

[19] E. Bloch, T. Rotem, J. Cohen, P. Singer, Y. Aperstein, "Machine learning models analysis of vital signs dynamics: A case for sepsis onset prediction", Journal of Healthcare Engineering, Vol. 2019, 2019.

[20] S. B. Baker, W. Xiang, I. M. Atkinson, "Continuous and automatic mortality risk prediction using vital signs in the intensive care unit: a hybrid neural network approach", Scientific Reports, Vol. 10, No. 1, 2020.

[21] M. Saeed et al. "Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database", Critical care medicine, Vol. 39, 2011.

[22] A. E. Johnson et al. "Mimic-iii, a freely accessible critical care database", Scientific Data, Vol. 3, 2016.

[23] A. E. Johnson et al. "Mimic-iv, a freely accessible electronic health record dataset", Scientific Data, Vol. 10, 2023.

[24] I. Silva, G. Moody, D. J. Scott, L. A. Celi, R. G. Mark, "Predicting in-hospital mortality of ICU patients: The physionet/computing in cardiology challenge 2012", in Computing in Cardiology, Vol. 39, 2012, pp. 245-248.

[25] S. L. Hyland et al. "Early prediction of circulatory failure in the intensive care unit using machine learning", Nature Medicine, Vol. 26, 2020.

[26] M. Lee et al. "WIDS datathon 2020: ICU mortality prediction", PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals, 2020.

[27] C. Steinmeyer, L. Wiese, "Sampling methods and feature selection for mortality prediction with neural networks", Journal of Biomedical Informatics, Vol. 111, 2020.

[28] G. S. Krishnan, S. Kamath, "A novel GA-ELM model for patient-specific mortality prediction over large-scale lab event data", Applied Soft Computing, Vol. 80, 2019, pp. 525-533.

[29] Y.-C. Huang, K.-Y. Chen, S.-J. Li, C.-K. Liu, Y.-C. Lin, M. Chen, "Implementing an ensemble learning model with feature selection to predict mortality among patients who underwent three-vessel percutaneous coronary intervention", Applied Sciences, Vol. 12, No. 16, 2022.

[30] Q. Wang, G. Chen, X. Jin, S. Ren, G. Wang, L. Cao, Y. Xia, "Bit-mac: Mortality prediction by bidirectional time and multi-feature attention coupled network on multivariate irregular time series", Computers in Biology and Medicine, Vol. 155, 2023, p. 106586.

[31] A. Kline, T. Kline, Z. S. Hossein Abad, J. Lee, "Novel feature selection for artificial intelligence using item response theory for mortality prediction", Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, Montreal, QC, Canada, 20-24 July 2020, pp. 5729-5732.

[32] B. Srimedha, R. Naveen Raj, V. Mayya, "A comprehensive machine learning based pipeline for an accurate early prediction of sepsis in ICU", IEEE Access, Vol. 10, 2022, pp. 105120-105132.

[33] F. Pedregosa et al. "Scikit-learn: Machine learning in python", Journal of Machine Learning Research, Vol. 12, 2011, pp. 2825-2830.

[34] J. Kennedy, "Particle Swarm Optimization", Springer US, 2010, pp. 760-766.

[35] M. Clerc, J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multi-dimensional complex space", IEEE Transactions on Evolutionary Computation, Vol. 6, 2002, pp. 58-73.

[36] P. R. Lorenzo, J. Nalepa, M. Kawulok, L. S. Ramos, J. R. Pastor, “Particle swarm optimization for hyper-parameter selection in deep neural networks”, Proceedings of the Genetic and Evolutionary Computation Conference, July 2017, p. 481-488.

[37] A. E. W. Johnson, R. G. Mark, “Real-time mortality prediction in the intensive care unit”, AMIA Annual Symposium Proceedings, 2017, pp. 994-1003.

[38] S. Purushotham, C. Meng, Z. Che, Y. Liu, “Benchmarking deep learning models on large healthcare datasets”, Journal of Biomedical Informatics, Vol. 83, 2018, pp. 112-134.

[39] Z. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu, “Recurrent neural networks for multivariate time series with missing values”, Scientific Reports, Vol. 8, No. 1, 2018, p. 6085.