Grenze International Journal of Engineering and Technology, June Issue

An Accurate Early Mortality Risk Prediction Model Based on ICU over MIMIC III using Machine Learning

Shilpa Sharma[1], Mandeep Kaur[2] and Savita Gupta[3]
[1-3]UIET Panjab University Chandigarh, India
Email: shilpasharma5978@gmail.com, {mandeep@pu.ac.in, savita2K8@yahoo.com}

***Abstract*— Mortality risk prediction of ICU patients is a valuable and challenging task due to limited clinical data. Accurate mortality risk prediction can improve the utilization of resources. In this work, we explore the use of feature engineering followed by applying a machine learning approach for predicting mortality risk in ICU patients using the MIMIC III dataset. The results of this study show that our prediction accuracy is 99.99% using a neural network.**

***Index Terms*— Mortality, MIMIC, Machine learning, Neural Network, ICU, Feature engineering, Prediction**

I. Introduction

In an effort to provide patients with higher-quality care, the healthcare sector is embracing more and more technology innovations. The use of different technologies has led to an increase in the amount of information that is retained, including non-clinical (such as billing information, age, gender, ethnicity, background, and medical reports) and clinical (such as radiological reports, physician reports, and reports about treatment and medicine) information. Both electronic medical records (EMR) and electronic health records (EHR) may be used to store this data. Such a vast amount of medical data is kept on cloud servers or physical servers, where different analytical techniques are used to improve communication. In this study, an investigation of one such set of data—the MIMIC-III dataset is discussed. This standardized dataset provides data on hospital archives, the Intensive Care Unit (ICU), mortality, physiological data, and other information. The patient's records from the time of their ICU admission until their formal discharge are included in the MIMIC-III dataset. These raise the standard of care that its patients receive. The MIMIC-III dataset provides highly standardized information that has been carefully selected to ensure that any research conducted using these data will produce broadly applicable, definitive results. Since there isn't much research being done in this area right now, it is crucial to understand how crucial it is to analyze the MIMIC-III dataset to provide an accurate diagnosis. The first crucial element in comprehending the difficulties in analyzing the MIMIC-III dataset is ICD-9 codes, which assign a unique set of codes to each disease and procedure, and are used in its indexing method. While this is done to make the analysis process simpler, it also generates redundant data, which leads to outliers. The second issue is that current approaches do not prioritize the quality of the dataset and instead simply apply the learning strategy to the medical data. Therefore, the research challenge of the existing system is to opt for a technique that could enrich the medical data quality.
The quality of care and clinical outcomes for patients in intensive care units (ICUs) depends on how well the risk of death can be estimated and how new interventions can be evaluated [1]. Therefore, it is important to have

Grenze ID: 01.GIJET.10.2.211_1

reliable methods of predicting mortality for ICU patients. Risk prediction models have been instrumental in finding disparities in disease mortality across global populations, particularly in the context of the transition to modernity, which has increased the risk of non-communicable diseases [2]. Multiple scoring systems have been invented over time to estimate ICU patients' death probability. Some of the most widely used tools are the Acute Physiology and Chronic Health Evaluation (APACHE), the MPM, and the Simplified Acute Physiology Score (SAPS). In the literature below, we have discussed SAPS 2 and APACHE 2; [3] scoring systems that have been designed for adults, and PRIM 2; a scoring system developed specifically for children has been proven to be effective in predicting mortality in adult ICU patients. These models have stratified individuals into low-risk or high-risk groups based on their projected mortality rates, aiding clinical decision-making and resource allocation. Additionally, risk prediction models have extended to assessing the accuracy and predictive power of established scoring systems [4]. According to projections, under the baseline scenario, the average global lifespan will rise from 65.3 years in 2002 to 73 years in 2030 75.5 years under the optimistic scenario, and 71.1 [5] years under the pessimistic scenario. Therefore, mortality prediction needs improvement to account for the uncertainties and complexities of the factors that affect human longevity.

One feasible way to improve mortality prediction is to use machine learning (ML) techniques. These methods [6] have several advantages over traditional scoring systems. They can handle copious amounts of big data and can recognize complex relationships between variables that may not be apparent using traditional methods. Additionally, ML models like logistic regression (LR), K-nearest neighbors (K-NN), and random forest (RF) can estimate mortality probabilities, classify patients based on similarity, and measure the importance of different variables. Hence, the proposed system introduces a feature engineering technique that is meant to enrich the data quality of MIMIC-III and accurately predict the mortality risk. This study is organized as follows: Section II describes the dataset and machine learning tools used in this study Section III describes the experiment and results of this work Section IV describes the discussion of this work Section V describes the conclusion and future work.

II LITERATURE REVIEW

This section gives a literature review of several techniques for mortality risk prediction (MRP) which is based on ML and different datasets. Several ML methods are used for MRP. The mortality risk prediction is divided into two categories, such as adult scoring system and the Child scoring system. there are different subtypes of risk-predicting systems, such as APACHE III, SAPS II, APACHE II, and SAPS III. These different versions use different factors and different weights for the factors, which can lead to different predictions of mortality risk.

[7] This study uses the MIMIC-III database of ICU patients to develop and compare mortality prediction models. APACHE II is an established scoring system that sums age, chronic health, and acute physiology variables during the first day of ICU admission. While widely used, it has complex drawbacks, deteriorating calibration over time, and bias from ICU care delivery. The authors compare APACHE II to ML methods including XGBoost on a dataset of 24,777 adults. Features mirror APACHE II variables: demographics, vitals, labs, etc. With grid search hyperparameter tuning, XGBoost shows higher accuracy (0.858 vs. 0.742), precision, recall, and AUC than APACHE II in 5-fold cross-validation. XGBoost maintains the highest performance metrics among ML algorithms, improving to 0.867 accuracy and 0.811 AUC. The model is externally validated on an eICU dataset, achieving 0.79 AUC. XGBoost proves superior in discrimination, calibration, and generalizability. The tree SHAP method provides insight into feature impacts, with urine, WBC, and bilirubin most predictive. The scalable tree-boosting XGBoost model outperforms APACHE II thanks to flexible representation, robustness, and automation. This shows the potential to augment established scoring systems with modern ML. [8] implemented the XGBoost algorithm, a robust ML technique, to define a supervised model for in-hospital all-cause deaths among ICU patients. The MIMIC-IV database served as the training set, while the eICU was used for external validation. The study concentrated on assessing the model's efficacy, using metrics to gain insights into discrimination and calibration. The model showed a positive benefit over a threshold probability range of 0% to 90%, emphasizing its distinct competitiveness compared to the other two models. SAPS II is a mortality prediction model in intensive care units (ICUs) [9]. It uses simple physiological and laboratory variables to calculate a score that correlates with the risk of hospital mortality. Higher SAPS II scores show a higher risk of mortality. [9] conducted a standardized review of several studies that validated the use of SAPS II for risk assessment and mortality prediction in ICU patients, particularly the elderly. Authors [10] discuss the use of ANN to improve the SAPS 3, a logistic regression model for early ICU mortality prediction. The authors used data from the Swedish Intensive Care Registry, which included 217,289 first-time adult ICU admissions during 2009–2017. The authors conducted training and validation of various ANN models featuring hidden layers and hyper-parameters, aligning with the parameters of

the SAPS 3 model. Performance evaluation of the ANN models was based on the area under the receiver operating AUC and the Brier score, with a comparative analysis against the SAPS3 model. The best ANN model (AUC 0.88, Brier score 0.097) showed superior predictive capabilities compared to the SAPS 3 model (Brier score 0.110, AUC 0.86) for forecasting 30-day mortality in ICU patients. The ANN model also showed enhanced proficiency in adjusting for age, a pivotal prognostic factor. [11] explored the SVM technique for a SOFA predictive model. The scope of the author was to develop ML models using SOFA scores to analyze in-hospital mortality risk. These component scores were extracted for patients in the first 24 hours along with outcomes. The SVM classifier finds the optimal hyperplane that separates the feature space into classes with maximum margin. The SVM model with a Gaussian kernel was refined by employing grid search and cross-validation on the training data, leading to improved performance. The predictive result was evaluated on the test set using metrics like accuracy, AUC-ROC, calibration plots, etc. The final SVM model achieved a test accuracy of 0.862 and AUC-ROC of 0.67 using only 3 features - renal, cardiovascular, and central nervous system SOFA score.

III DATASET-TOOLS

In this section, we will discuss all components of our experiment, and describe the dataset and the techniques used

A. MIMIC III

The model was developed from a retrospective analysis of a cohort of patients' medical information Mart for intensive care (mimic iii) a large public dataset that includes information on 46,520 patients who were admitted to ICU from 2001 to 2012 at Beth Israel. Deaconess Medical Centre in Boston, MA, USA. The database holds records of demographics, hourly vital signs from bedside monitors, laboratory tests, International Classification of Diseases and Ninth Revision (ICD-9) codes diagnoses, and other clinical characteristics. The users had to pass a test to qualify to register for the database and to be approved by the MIMIC-III database administration staff [12]. We selected the features from the combinations of the below-mentioned tables.

- Admissions: holds information related to patients' admission
- ICU Stays: holds the patient's stay information
- Diagnosis ICD: holds the disease code information
- CPT Events: holds patients' medical chart information
- Services: holds the information services given to the patients during the ICU stay
- Patients: holds: Patients demographic information
- Diagnosis: holds the disease information of patients
- Procedures ICD: holds the procedure code information

1. *Data Extraction* In this research, firstly the above-mentioned tables are merged based on subject ID and HADM ID then we drop the irrelevant features like edge time, row ID, Has chart events data, etc. The obtained dataset contains 19 attributes like subject ID, HADM ID, gender, age, admission type, address location, discharge location, hospital expires flag, procedure sequence no, diagnosis sequence number, previous care unit, and current care unit. we evaluate the number of procedures, number of diagnoses, and number of CPT events per day and days difference as shown in the table1.
2. *Data Pre-processing.* Data preprocessing is an essential step in machine learning and data mining. The quality of data affects the learning ability of a machine learning model. Therefore, it is important to prepare the data before feeding it to the model [14]. Data preprocessing is a technique that is used to change the raw data into a clean dataset [13]. The feature engineering process supplies refined input and significantly contributes towards reducing the computational burden of the learning technique [14]. This study applies feature engineering to improve the performance of the model all the steps of feature engineering are given below:

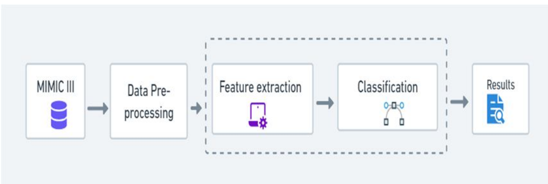

Figure 1. Proposed Methodology

2.1 Correlation analysis for ICD Data: Using feature engineering method study carried out on selected data. If the current illness is not carried out properly and does not follow the procedure accurately then it causes the secondary illness. To overcome this problem, firstly filter out the dataset by setting diagnosis sequence no equal to 1 it gives the first diagnosis data at admission time of the patient then compare the frequencies of such repetitive diagnosis ICD-9 code with threshold score. The frequencies within the threshold are considered as a common illness, while others are considered as risky illnesses. Then concatenate the transformed resultant data with extracted data. Finally, we get information about common and risky diseases and what kind of procedures are followed for diseases. Fig 2 represents the correlation between variables of a resultant dataset of this step. Inclusion of the first procedure ICD9 code, that is the ICD9 code of the first procedure carried out during an admission, may also prove beneficial. The nature of the procedure(s) carried out would be expected to provide further information about mortality risk. renal, cardiovascular, and central nervous system SOFA score.

2.2 Outliers Eliminations: The MIMIC III dataset holds a large amount of data with an iterative form of data. Handling this type of data is a challenging task. So, this work drops the outliers of repeated patient hospital visits by considering only long-stay patients with setting LOS >=1 on a meta-dataset obtained from an earlier step.

2.3 Encoding and Feature Extraction: The resultant dataset of the above step has missing and infinite values. To handle this problem replace the missing values and large values with zero. In the encoding step, this study finds categorical columns and then maps unique categorical values in each column to numerical indices. The resultant data frame with categorical columns is replaced by numerical indices. After applying these steps, the size of the sample (346113, 21) was obtained. After that we split the dataset into 70% training and 30% testing data and the hospital expire flag variable is used as the target variable then with Random Forest, we reduced the features from 20 to 10 most prominent features. The extracted features are given below in Table 2 and Fig 3 shows the features score of features selected by the random forest.

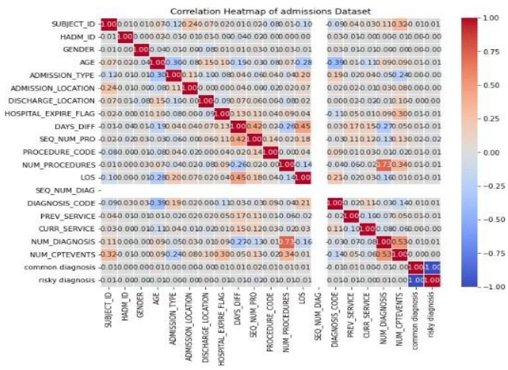

Figure 2. Stands for the correlations of attributes of meta-dataset

B. Classifiers

1) Logistic Regression: This statistical technique simulates the relationship between a binary dependent variable and one or more independent variables and is used for binary classification. The logistic function is utilized to convert a linear combination of input data into a probability, which makes it very helpful for activities with binary outcomes (yes/no, 1/0, etc.).

2) k-Nearest Neighbors (KNN): This instance-based, non-parametric machine learning technique is used for regression and classification tasks. A feature space observation is classified using KNN based on the majority vote of its k nearest neighbors. It bases conclusions on the properties of neighboring data points, assuming that similar cases exist nearby.

3) Multi-Layer Perceptron (MLP): This is a kind of artificial neural network intended for use in supervised learning. It is made up of an input layer, one or more hidden layers, and an output layer, which are all layers of nodes, or neurons. During training, the network learns by modifying the weights assigned to each connection between nodes. MLPs are frequently employed in a variety of machine-learning tasks, such as regression and classification, due to their well-known capacity to represent complex relationships.

IV. EXPERIMENT & RESULTS

A. Overview of the Experiment

The goal was to precisely forecast the MIMIC III dataset's early mortality risk for ICU patients. The MIMIC III database records several diagnoses for every admission. The diagnoses are arranged in order of sequence number, starting from 1 and going up to n, which represents the maximum number of diagnoses that can be recorded for that patient and that hospitalization. ICD9 diagnosis codes, such as 1197, offer an encoding for three, four, or five-digit diagnoses. We have utilized the ICD9 code for the diagnosis when Num equals 1, as this typically indicates the principal diagnosis or the diagnosis that is most important to the patient during that hospital stay. Additionally, four variables were made for our table: Num CPT indicates the total number of charted events, while" Num Procedures" reflects the total number of procedures performed on the patient during the ICU stay, without focusing on the specific type of those procedures. Days difference indicates the total LOS based on the period between discharge and admittance time, and age determines the patient's age based on admittance time and dob.

TABLE II. REPRESENTS FEATURES EXTRACTED BY USING RANDOM FOREST

S.no	Attributes	Description
1	Discharge Location	Represents the patient's location at discharge time
2	No of CPT EVENTS	Total number of charted events for that admission
3	Days Difference	Difference between admit time and discharge time
4	LOS	Length of stay
5	Age	Age in years
6	Diagnosis ICD 9 code	Disease code represented in a code
7	No of Procedures	Total number of procedures followed on patients for that admission
8	No of diagnosis	Total number of disease diagnoses for that admission
9	Subject ID	Each patient has a unique id
10	Curr services	Current service is given to patients after admission

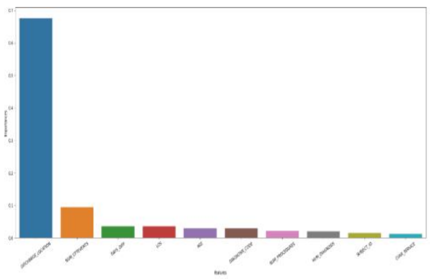

Figure 3. Represents the feature Importance score by random forest model

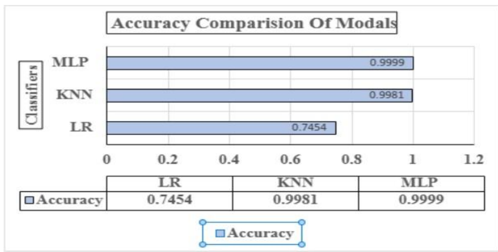

Figure 4. Represents the accuracy score of the proposed

B. Results

The machine learning techniques of LR, KNN, and MLP were employed in this work. As seen in Figs. 4 to 9, the model's performance is evaluated using accuracy, precision, recall, F1 score, and AUC precision-recall. The results showed that the LR achieved an accuracy of 0.7454, a precision of 0.364, a recall of 0.7408, an F1 score of 0.4720, an AUC of 0.998, a KNN accuracy of 0.99812, a precision of 0.98906, recall of 0.9988, F1 score of 0.9939, and an AUC 0.998 whereas MLP gradient descent achieved the highest accuracy of 0.9999, precision of 0.9999, F1score 1.0, recall of 0.999 and an AUC 1.00, therefore, the MLP NN model correctly identifies 99.99% of all survival patients in ICU and when predicts the status of a patient as survival is correct as 0.999 of the time by setting parameter values hidden layer sizes= (100,), max iter=1000, and learning rate init=0.0001 of neural network. In addition, Table 6 compares the suggested scheme with three relevant literature that have used the MIMIC-III dataset and have as their common objective the early prediction of mortality. Upon closer inspection of Table 3, it is evident that the existing methods described by [15] have employed backpropagation in a manner like the suggested scheme. Nevertheless, their dataset processing fails to consider comprehensive feature management, leading to a decline in AUC. However, the inclusion of multiple learning schemes results in a significant increase in processing time. This contrasts with the work given by [16], which used multiple learning schemes to raise the AUC score. Like this, [17] employed the extreme gradient boosting strategy (XGBoost) to achieve an AUC of 0.857. However, as compared to the current method, the accuracy has enhanced due to the effective processing steps that the proposed model applied before putting it through learning.

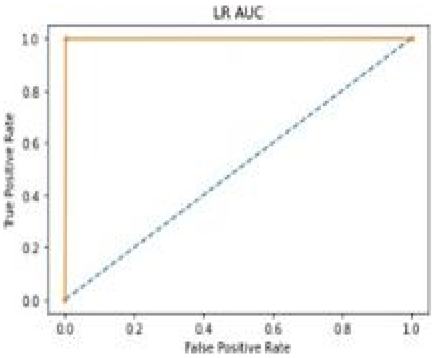

Figure 5. Represents the auc of lr model

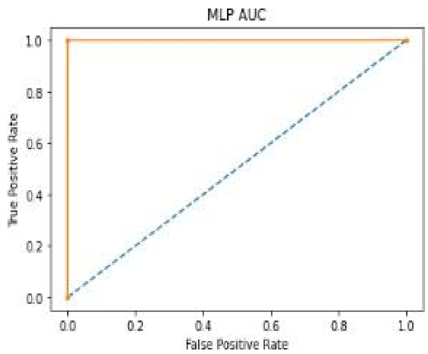

Figure 6. Represents the auc of the mlp model

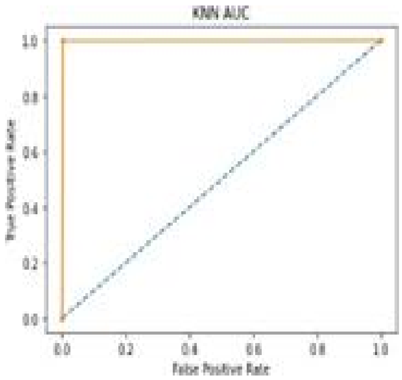

Figure 7. Represents the auc of knn mode

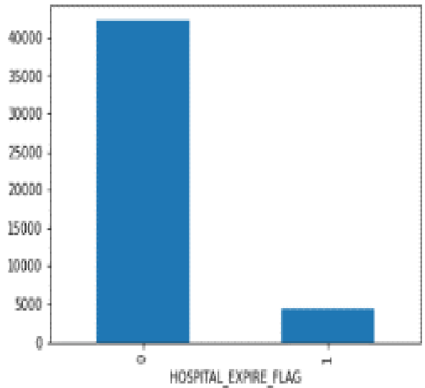

Figure 8. Represents the total number of survivors and Non- survivors inthe mimic iii dataset

V. Discussion

Given that it is not limited to any disease, the created model demonstrated remarkable accuracy in forecasting the mortality risk of ICU patients. Time series, pharmaceutical, laboratory, clinical, microbiological, imaging outcomes, etc. are not used in this investigation. Its main advantage is this model's simplicity in terms of inputs that must be applied to every patient. Furthermore, rather than using tests or clinical data collection, these inputs are known for every patient and are obtained through administrative procedures. For instance, a patient's age, gender, marital status, race, and religion are among the various inputs that are known at the time of admission. Similar to this, the ICU-related inputs are obtained from transfer records, such as the ICU care units in which a patient stayed. The set of attributes inputted to the classifiers does not vary per the health condition of the patient. This has implications beyond the model's general and condition-agnostic nature, as the model is simple enough to be applied by hospital administrative support without requiring domain-specific clinician input to calculate. One method of designating a patient's condition, but again without using physiological data or tests specific to that condition, is using the attribute with diagnosis ICD9 code; Although it might be expanded to accept the ICD9 codes of other concurrent conditions for the patient, the classifiers only use the principal diagnosis' ICD9 code at this time. These co-occurring diagnoses would provide more details about the patient's health, but as the sequence number increases, the priority or relevance of the diagnosis to the admission normally decreases. It should be noted that the classifiers can accurately predict short stays versus lengthy stays even when they are only given the primary diagnosis and other minimum diagnostic information. In addition to the ICD9 code of the major diagnosis for each patient, the total number of diagnoses for a patient at that admission is also used as input. This corresponds to the maximum diagnosis sequence number that can be assigned to that patient for that admission. The greater this figure is for a particular patient, the more diagnoses or health issues the patient is experiencing. This may be connected in some way to their current health status and, consequently, to their risk of dying. Similar to this, information about procedures is used sparingly. Procedure ICD9 codes identify the medical procedures carried out on a patient in a hospital. Sequence numbers designate the order in which these procedures are carried out.

TABLE I. REPRESENTS THE FEATURES SELECTED FOR THIS STUDY

S.no	Attributes	Description
1	Discharge Location	Represents the patient's location at discharge time
2	No of CPT EVENTS	Total number of charted events for that admission
3	Days Difference	Difference between admit time and discharge time
4	LOS	Length of stay
5	Age	Age in years
6	Diagnosis ICD 9 code	Disease code represented in a code
7	No of Procedures	Total number of procedures followed on patients for that admission
8	Subject ID	Each patient has a unique id
9	No of diagnosis	Total number of disease diagnoses for that admission
10	Curr services	Current service is given to patients after admission
11	HADM ID	unique admission ID of each patient
12	Admission location	Patient location at admission type
13	Gender	male, female
14	Admission Type	elective, urgent, emergency
15	Hospital expire flag	survive=1, non-survive=0
16	seq num pro	The procedure followed on the patient for that admission
17	Prev service	Earlier service was given to the patients
18	common diagnosis	yes=1, no=0
19	Risky diagnosis	yes=1, no=0
20	seq num Diag	The diagnosis sequence number for that admission

TABLE II. COMPARISON OF PROPOSED WORK WITH EXISTING WORK

References	Methods	AUC
[15]	Back Propagation ANN	0.769
[16]	Machine learning techniques	0.821
[17]	XGBoost	0.857
Proposed	Neural network	1.000

TABLE III. PERFORMANCE METRICS OF THE PROPOSED MODEL

Models	Precision	F1 score	Recall
LR	0.364	0.4720	0.7408
KNN	0.98906	0.9939	0.9988
MLP	0.9999	1.0	0.999

With a sequence number of 1, they are indicating that this is the initial treatment performed on the patient for the admission being entered. It is reasonable to assume that the ICD9 code of the initial procedure, as well as the codes of every procedure that takes place throughout the patient's hospitalization, will reveal more details about the patient's health and may even influence the chance of death. The classifiers use demographic inputs such as age, gender, and marital status; these putative characteristics have also been demonstrated to influence mortality risk in previous condition-specific investigations [17], [18]. The cost center input variable taken from the MIMIC III CPT EVENTS table provides limited information on the nature of the patient's condition, having just two possible values ICU or RESP (respiratory). The inputs from the first and last care units serve as a stand-in for details regarding the patient's general health state and level of care. A few examples of potential values for care units are the following: trauma/surgical intensive care unit (TSICU), medical intensive care unit (MICU), neonatal intensive care unit (NICU), cardiac surgery recovery unit (CSRU), and coronary care unit 1199 (CCU). Thus, mortality may be impacted by these care unit levels. In a similar vein, the input current service acts as a stand-in for providing data on a wide range of medical issues. The care team that is caring for the patient under which the patient was admitted is referred to as the service, and they may possess certain values, such as dental (DENT), gynecological (GYN), general care for internal medicine (MED), psychiatric (PSYCH), and heart surgery (CSURG). The ward the patient is staying in may not always be a good indicator of the type of care or general state of health; sometimes, patients are assigned to a specific physical ward based only on bed availability, even when that arrangement does not align with the medical team's recommendations. The Prev service for an entry in the MIMIC III SERVICES table frequently has no value. As a result, its value is diminished. In this instance, the record records the patient's initial hospital admission, as is the case with most of the entries in the SERVICES table. Neural networks have the advantage of allowing for prediction in situations where the relationships between the input variables are intricate and uncertain.

Firstly, the subset dataset was randomly divided into 70% train data and 30% test data then on the training dataset applied 5-fold cross-validation. The dataset has an imbalanced data problem which means that no of survivor patients' is more than the no survivors which affects the results of classifiers as shown in Fig 6. To overcome this problem this study used the Python Ram over sampler package which made the dataset balanced and then normalized the resultant dataset using a standard scaler method. There are various machine learning models; this paper used the most well-known and prominent machine learning model to compare with a new model.

VI CONCLUSION & FUTURE WORK

Unlike any previous approaches where datasets are directly given to learning schemes, the suggested scheme has established a novel process of extracting and maintaining features from the standard MIMIC- III dataset. The application of this methodology improves accuracy. In contrast to all other illness prediction methods now in use, which involve directly learning from the dataset, the suggested methodology emphasizes enhancing data quality through feature engineering. Because of this feature, the suggested model is more affordable than conventional learning techniques and existing literature in the same domain. The application of machine learning is also investigated in this work. The system was trained by the MIMIC III database's sizable data set. The precision of the predictive model's performance is demonstrated by the about 99.99% accuracy rate of mortality prediction from the test set. This offers a precise prediction technique based on previously published models.

Future research will incorporate measurements of physiological data, observations from caregivers, results from lab and imaging tests, and information about medications. These additions could potentially enhance predicting abilities in both general and condition-specific scenarios.

REFERENCES

[1] M. T. Keegan, O. Gajic, and B. Afessa, "Severity of illness scoring systems in the intensive care unit," Critical care medicine, vol. 39, no. 1, pp. 163–169, 2011.

[2] T. L. Bothe, A. Patzak, and N. Pilz, "The b-score is a novel metric for measuring the true performance of blood pressure estimation models," Scientific Reports, vol. 12, no. 1, p. 12173, 2022.

[3] J. I. Salluh and M. Soares, "Icu severity of illness scores: Apache, saps and mpm," Current opinion in critical care, vol. 20, no. 5, pp. 557– 565, 2014.

[4] K. Janikowski, R. Morawiec, B. Jegier, R. Jaszewski, and M. Lelonek, "Euroscore ii does not show better accuracy nor predictive power in comparison to original euroscore: a single-center study," Kardiologia Polska (Polish Heart Journal), vol. 74, no. 5, pp. 469–475, 2016.

[5] C. D. Mathers and D. Loncar, "Projections of global mortality and burden of disease from 2002 to 2030," PLoS Medicine, vol. 3, no. 11, p. e442, 2006.

[6] ANaemi, T. Schmidt, M. Mansourvar, M. Naghavi-Behzad, A. Ebrahimi, and U. K. Wiil, "Machine learning techniques for mortality prediction in emergency departments: a systematic review," BMJ open, vol. 11, no. 11, p. e052663, 2021.

[7] Y. Luo, Z. Wang, and C. Wang, "Improvement of apache ii score system for disease severity based on xgboost algorithm," BMC Medical Informatics and Decision Making, vol. 21, no. 1, pp. 1–12, 2021.

[8] Z. Chen, T. Li, S. Guo, D. Zeng, and K. Wang, "Machine learning- based in-hospital mortality risk prediction tool for intensive care unit patients with heart failure," Frontiers in Cardiovascular Medicine, vol. 10, p. 1119699, 2023.

[9] R. Puchades, B. Gonza´lez, M Contreras, A. Gullo´n, R. de Miguel, D. Mart´ın, C. Gutierrez, and R. Navarro, "Cardiovascular profile in critically ill elderly medical patients: Prevalence, mortality and length of stay," European journal of internal medicine, vol. 26, no. 1, pp. 49– 55, 2015.

[10] H.-C. Thorsen-Meyer, A. B. Nielsen, A. P. Nielsen, B. S. Kaas- Hansen, P. Toft, J. Schierbeck, T. Strøm, P. J. Chmura, M. Heimann, L. Dybdahl, et al., "Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records," The Lancet Digital Health, vol. 2, no. 4, pp. e179–e191, 2020.

[11] X. Pan, J. Xie, L. Zhang, X. Wang, S. Zhang, Y. Zhuang, X. Lin, S. Shi, S. Shi, and W. Lin, “Evaluate prognostic accuracy of sofa component score for mortality among adults with sepsis by machine learning method,” BMC Infectious Diseases, vol. 23, no. 1, pp. 1–8, 2023.
[12] F. Viton, M. Elbattah, J.-L. Gue´rin, and G. Dequen, “Heatmaps for visual explainability of cnn-based predictions for multivariate time series with application to healthcare,” in 2020 IEEE International Conference on Healthcare Informatics (ICHI), pp. 1–8, IEEE, 2020.
[13] S. A. Alasadi and W. S. Bhaya, “Review of data preprocessing techniques in data mining,” Journal of Engineering and Applied Sciences, vol. 12, no. 16, pp. 4102–4107, 2017.
[14] R. Ghorbani, R. Ghousi, A. Makui, and A. Atashi, “A new hybrid predictive model to predict the early mortality risk in intensive care units on a highly imbalanced dataset,” IEEE Access, vol. 8, pp. 141066–141079, 2020.
[15] N. Ding, C. Guo, C. Li, Y. Zhou, X. Chai, et al., “An artificial neural networks model for early predicting in-hospital mortality in acute pancreatitis in mimic-iii,” BioMed research international, vol. 2021, 2021
[16] Y. Zhu, J. Zhang, G. Wang, R. Yao, C. Ren, G. Chen, X. Jin, J. Guo, S. Liu, H. Zheng, et al., “Machine learning prediction models for mechanically ventilated patients: analyses of the mimic-iii database,” Frontiers in medicine, vol. 8, p. 662340, 2021.
[17] P. R. Hachesu, M. Ahmadi, S. Alizadeh, and F. Sadoughi, “Use of data mining techniques to determine and predict length of stay of cardiac patients,” Healthcare informatics research, vol. 19, no. 2, pp. 121– 129, 2013.