

Research



Predicting ICU Stay Duration for Hemodialysis Patients with Poisoning: A Study Comparing Deep Learning with Machine Learning Models

Zehirlenme Nedeniyle Hemodiyaliz Hastalarında Yoğun Bakımda Kalış Süresinin Tahmini: Makine Öğrenimi Modellerini Derin Öğrenme Modelleriyle Karşılaştıran Bir Çalışma

Khadijeh Moulaei¹, Shahin Shadnia², Mohammad Reza Afrash³, Babak Mostafazadeh², Heliya Rafsanjani⁴, Peyman Erfan Talab Evini², Mohanna Sharifi⁵, Mohadeseh Sarbaz Bardsiri^{2,6}, Hadi Kazemi-Arpanahi⁷, Babak Sabet⁸, Mohammad Parvin⁹, Vahideh Montazeri³, Amir Hossein Daeechini¹⁰, Mitra Rahimi²

¹Iran University of Medical Sciences, Health Management and Economics Research Center, Health Management Research Institute, Tehran, Iran

²Shahid Beheshti University of Medical Sciences, Lohman Hakim Hospital, Toxicological Research Center, Excellence Center of Clinical Toxicology, Department of Clinical Toxicology, Tehran, Iran

³Smart University of Medical Sciences, Artificial Intelligence in Medical Sciences Research Center, Tehran, Iran

⁴Network of Interdisciplinarity in Neonates and Infants (NINI), Universal Scientific Education and Research Network (USERN), Tehran, Iran

⁵Tor Vergata University, Department of Pharmacy, Rome, Italy

⁶Iran University of Medical Sciences, Firouzgar Hospital, Department of Clinical Toxicology, Tehran, Iran

⁷Abadan University of Medical Sciences, Department of Health Information Technology, Abadan, Iran

⁸Shahid Beheshti University of Medical Sciences Faculty of Medicine, Department of Surgery, Tehran, Iran

⁹Auburn University, Department of Industrial and Systems Engineering, Auburn, AL, USA

¹⁰Shahid Beheshti University of Medical Sciences, School of Allied Medical Sciences, Student Research Committee, Department of Health Information Technology and Management, Tehran, Iran

ABSTRACT

Objective: This study compares the effectiveness of machine learning and deep learning models in predicting hemodialysis patients' length of stay in the intensive care unit (ICU).

Methods: This retrospective cohort study used data from 980 poisoned patients undergoing hemodialysis. A variety of eight well-known machine learning [support vector machine, extreme gradient boosting, random forest (RF), decision tree] and deep learning (deep neural network, feedforward neural network, long short-term memory, convolutional neural network) models were employed.

Results: Feature importance analyses using Shapley Additive exPlanation and local interpretable model-agnostic explanation methodologies identified Glasgow Coma Scale (GCS <8), intubation, acute kidney injury, PO₂, blood urea nitrogen, metabolic acidosis, and number of hemodialysis sessions as key predictors of ICU stay duration in poisoned hemodialysis patients, with intubation score, GCS score, and ICU admission type being the most significant predictors. Overall, the RF model displayed exceptional performance across various metrics.

Conclusion: Our findings emphasize the importance of neurological status, respiratory function, and renal injury in predicting ICU duration, offering valuable insights for clinical decision-making and resource allocation in this high-risk population.

Keywords: Machine learning, deep learning, length of stay, intensive care, poisoning

Address for Correspondence: Mitra Rahimi MD, Shahid Beheshti University of Medical Sciences, School of Allied Medical Sciences, Student Research Committee, Department of Health Information Technology and Management, Tehran, Iran

E-mail: mrahimi744@gmail.com **ORCID ID:** orcid.org/0000-0002-8785-8207

Cite as: Moulaei K, Shadnia S, Afrash MR, Mostafazadeh B, Rafsanjani H, Evini PET, et al. Predicting ICU stay duration for hemodialysis patients with poisoning: a study comparing deep learning with machine learning models. Med J Bakirkoy. 2025;21(3):274-287

Received: 11.11.2024

Accepted: 24.03.2025

Publication Date: 03.09.2025



ÖZ

Amaç: Bu çalışma, hemodiyaliz hastalarının yoğun bakım ünitesinde (YBÜ) kalış sürelerini tahmin etmede makine öğrenimi ve derin öğrenme modellerinin etkinliğini karşılaştırır.

Gereç ve Yöntem: Bu retrospektif kohort çalışmasında, hemodiyalize giren 980 zehirlenmiş hastadan alınan veriler kullanılmıştır. Çeşitli sekiz iyi bilinen makine öğrenmesi (destek vektör makinesi, aşırı gradyan artırma, rastgele orman, karar ağacı) ve derin öğrenme (derin sinir ağı, ileri beslemeli sinir ağı, uzun kısa süreli bellek, evrişimli sinir ağı) modeli kullanılmıştır.

Bulgular: Shapley Katkı Açıklaması ve yerel yorumlanabilir modelden bağımsız açıklama metodolojileri kullanılarak yapılan özellik önem analizleri, zehirlenmiş hemodiyaliz hastalarında YBÜ'de kalış süresinin temel belirleyicileri olarak Glasgow Koma Skalası (GCS <8), entübasyon, akut böbrek hasarı, PO₂, kan üre azotu, metabolik asidoz ve hemodiyaliz seansı sayısını belirlemiştir. En önemlileri entübasyon puanı, GCS puanı ve YBÜ'ye kabul türü olmuştur. Genel olarak, rastgele orman modeli çeşitli ölçütlerde olağanüstü bir performans göstermiştir.

Sonuç: Bulgularımız, YBÜ'de kalış süresini tahmin etmede nörolojik durum, solunum fonksiyonu ve böbrek hasarının önemini vurgulamakta ve bu yüksek riskli popülasyonda klinik karar alma ve kaynak dağıtımı için değerli bilgiler sunmaktadır.

Anahtar Kelimeler: Makine öğrenimi, derin öğrenme, kalış süresi, yoğun bakım, zehirlenme

INTRODUCTION

Poisoning, whether intentional or accidental, is a significant health issue worldwide, imposing substantial financial, physical, and mental burdens on patients, families, and society (1,2). According to the World Health Organization (WHO), over 3 million people are poisoned annually, resulting in 220,000 deaths, mostly in developing countries due to easy access to toxic substances, lack of awareness, and limited hospital resources (3,4). Poisoning accounts for over 2.4% of emergency department visits and 3-6% of intensive care unit (ICU) admissions (5,6). While most poisoned patients recover with supportive treatment, critically ill patients with severe symptoms are admitted to ICUs for intensive care and monitoring. Studies show that 4.1-30.8% of life-threatening poisoning cases require ICU admission. In developing countries, including Iran, poisoning incidents, especially from pesticides and rice tablets, have doubled in recent decades despite improved ICU facilities (7,8).

Hemodialysis patients are frequently admitted to the ICU due to poisoning, with about 2% of chronic dialysis patients requiring ICU care annually. They are at increased risk of infections due to factors like immune deficiency from uremia, impaired phagocytic function, older age, and comorbidities such as diabetes. Frequent use of vascular access for hemodialysis heightens the risk of bloodstream infections (9). Managing poisoned hemodialysis patients in the ICU is challenging due to their altered pharmacokinetics and pharmacodynamics, which affect toxin clearance. Renal failure complicates toxin elimination, often necessitating hemodialysis. Additionally, these patients are prone to fluid and electrolyte imbalances, requiring close monitoring. Infections, especially bloodstream infections from vascular access, are a major concern for immunocompromised patients (10). Comprehensive ICU management for these

patients involves toxin removal, careful renal replacement therapy, fluid balance, electrolyte management, and infection control.

Moreover, healthcare systems face constant pressure to improve patient outcomes and reduce costs. ICUs, which provide critical care, are expensive and resource-intensive (11,12). The increasing number of poisoned patients has heightened the demand for ICU beds. The WHO highlights the importance of monitoring the length of stay (LOS) as a measure of care quality and resource use (13,14). Patients with prolonged LOS consume a significant portion of resources, so reducing LOS can enhance bed turnover, optimize resource allocation, improve patient safety, and lower costs. Identifying patients with long LOS, particularly in overwhelmed hospitals, can alleviate pressure and boost ICU efficiency. Policymakers are adopting evidence-based solutions to optimize ICU resources like beds, staff, and mechanical ventilation (15-17).

Recently, artificial intelligence-based solutions such as machine learning (ML) and deep learning (DL) have gained further attention for their ability to predict the outcome of interest based on a great amount of data, especially when the relationships between variables are complex and non-linear (18-20). DL and ML models can be used for predicting the patients' ICU LOS (21,22). These models leverage large datasets comprising patient demographics, clinical variables, and possibly real-time monitoring data such as vital signs and laboratory results (21). By analyzing patterns and correlations within these data, DL and ML algorithms can generate accurate predictions of how long a patient is likely to stay in the ICU (22,23). This predictive capability not only helps in optimizing resource allocation and bed management but also assists healthcare providers in identifying patients at risk of prolonged ICU stays early on, enabling proactive interventions and potentially improving patient outcomes (21,24).

To the best of our knowledge, no study has compared ML models with DL models for predicting ICU stay duration for hemodialysis patients with poisoning. Therefore, the purpose of this study is to develop and evaluate the effectiveness of both ML and DL models in predicting the ICU LOS for hemodialysis patients suffering from poisoning. By comparing the performance of these models, we aim to identify the most accurate and reliable approach, ultimately improving resource allocation and patient outcomes in critical care settings.

METHODS

Study Design and Setting

This retrospective cohort study was performed among all poisoned patients who were admitted to the ICU at the Loghman Hakim Hospital (LHH) between January 1, 2016, and December 31, 2020. The hospital, which is known as the Iran's largest poison center, is crucial for managing patients who need specialized care as a result of various poisonings and toxic exposures. The ICU of LHH is essential to the management of critical situations, providing cutting-edge medical techniques and specialized care to guarantee the best outcomes for patients. In this study, the LOS of patients who were admitted to the ICU was analyzed. Patients were categorized into two groups based on the LOS: short (lasting 4 days or less), and long (exceeding 4 days). We utilized several data-driven ML and DL models to develop an accurate prediction model the ICU LOS in

poisoned patients undergoing hemodialysis. The key steps taken were as follows (Figure 1):

- establishing the study roadmap and experiment environment,
- preprocessing the data,
- using feature selection algorithms,
- selecting appropriate classification algorithms and their hyperparameters,
- splitting the data into training and testing sets, and
- evaluating model performance.

Data Collection and Preprocessing

The data set was collected by reviewing the electronic medical records (EMRs) of patients undergoing hemodialysis who were poisoned admitted to the ICU at LHH between January 2016 and December 2020. Relevant variables were extracted from poisoned patient records and entered into a database. Age, sex, kind of poisoning, history of underlying conditions, medication usage and habits, laboratory test results, and vital signs, number of extracorporeal technique uses, type of extracorporeal method, and patient outcome were among the variables under investigation (Table 1).

Statistical Analysis

In this study, before further analysis and feeding data into ML methods, the rows of datasets collected from EMRs laboratory tests, underwent a series of preprocessing steps.

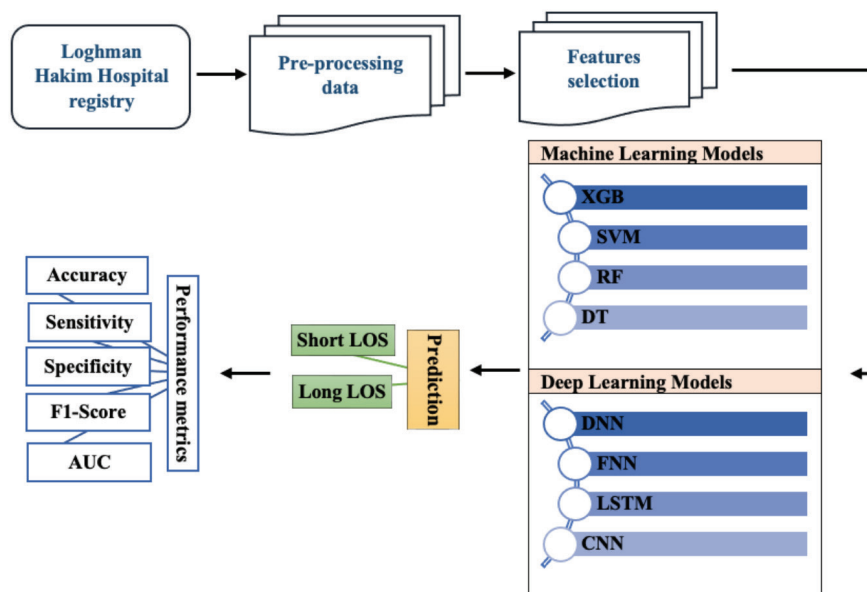


Figure 1. The road map of the proposed system for the prognosis of poisoned patients who are candidates for dialysis

AUC: Area under the curve, LOS: Length of stay, XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, FNN: Feedforward neural network, CNN: Convolutional neural network

Table 1. The descriptive statistics of variables

Demographic data			
Variables		Frequency (%)	Values
Gender	Female	187 (19.1%)	Female/male
	Male	793 (80.9%)	
Age (year)	17-20	100 (10/4%)	Numerical
	21-40	604 (61.6%)	
	41-60	211 (21.5%)	
	>61	65 (6.6%)	
Co-ingestion		62 (6.3%)	Yes/no
Smoking		46 (4.7%)	Yes/no
Alcohol consumption		627 (64%)	Yes/no
Opium abuse		75 (7.7%)	Yes/no
Stimulants abuse		13 (1.3%)	Yes/no
History of the previous disease		150 (15.3%)	Yes/no
History of taking medication		77 (7.9%)	Yes/no
Hemoperfusion		9 (0/9%)	Yes/no
Intubation		203 (20.7%)	Yes/no
Antidote therapy		904 (92.2%)	Yes/no
Duration of hospitalization (day) median (min-max)		8 (1-116)	Numerical
Laboratory and clinical data			
13≤GCS<15		659 (67.2%)	Numerical
8≤GCS<13		62 (6.3%)	Numerical
Coma (GCS <8)		108 (11%)	Yes/no
Bradypnea		46 (4.7%)	Numerical
Temperature (mean±SD)		36.9±0.56	Numerical
Bradycardia		14 (1.4%)	Numerical
Tachycardia		145 (14.8%)	Numerical
Hypotension		38 (3.9%)	Numerical
Hypertension		293 (29.9%)	Numerical
Metabolic acidosis		823 (84%)	Numerical
Acute kidney injury (AKI)		536 (54.7%)	Numerical
Renal disease		641 (65%)	Yes/no
Serum HCO ₃ (mean±SD)		14.4±23.8	Numerical
BUN (meq/L) (mean±SD)		36.7±27.2	Numerical
Creatinine (meq/L) (mean±SD)		1.7±4.1	Numerical
Blood glucose (mg/dL) (mean±SD)		132.5±66.6	Numerical
Blood pH (mean±SD)		7.2±0.4	Numerical
Sodium (meq/L) (mean±SD)		137.8±10.8	Numerical
Potassium (meq/L) (mean±SD)		4.7±3.4	Numerical
Partial pressure of carbon dioxide (PCO ₂), partial pressure of oxygen (PO ₂), bicarbonate (HCO ₃), base excess (BE), blood sugar (Bs), white blood cell count (WBC), hemoglobin (Hb), hematocrit (Hct), platelet count (PLT), aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), lactate dehydrogenase (LDH), creatine phosphokinase (CPK), creatine phosphokinase-MB (CPK-MB), prothrombin time (PT), partial thromboplastin time (PTT), international normalized ratio (INR).			Numerical
Outcome variable: ICU LOS	Short LOS ≤4 days	640	Yes/no
	Long LOS >4 days	340	Yes/no
ICU: Intensive care unit, LOS: Length of stay, SD: Standard deviation, GCS: Glasgow Coma Scale			

These steps include; any rows with over 70% missing values were removed. Patient data were excluded if they were missing critical demographic or clinical information (e.g., comorbidities) necessary for analysis. Additionally, duplicate records and entries with inconsistent or implausible values (e.g., negative age or erroneous laboratory results) were excluded to ensure data quality. Missing values for remaining variables were handled using imputation techniques where feasible, and rows with irreparable missingness for key variables were removed. Next, minimum-maximum scaling was applied to normalize all values between 0 and 1. Standard scaler scaling was then used to standardize the data distribution. Data validation checked the integrity and accuracy of the dataset. To handle class imbalance, under-sampling, balanced the classes by keeping all samples from the minority class.

After preprocessing, the final dataset contained 980 patients. This dataset was randomly split so that 70% of the data (686 patients) was assigned to the training set, and the remaining 30% (294 patients) was assigned to the test set. The training set was used to develop the feature selection and ML models, while the test set was held out for model evaluation. The descriptive statistics of the variables in the dataset are shown in Table 1. This includes variable names, the frequency of each variables, and their values.

Model Development and Evaluations

The LOS in the ICU for hemodialysis patients with poisoning was predicted using eight well-known models from the domains of DL and ML. Convolutional neural networks (CNN), feedforward neural networks, long short-term memory (LSTM), and deep neural networks (DNN) were among the DL models. The ML models included random forest (RF), decision tree (DT), support vector machine, and extreme gradient boosting (XGB).

Cross-validation and Tweaking of Hyperparameters

We trained all suggested models using 10-fold cross-validation to reduce overfitting. The dataset is divided into ten equal segments using this method. The model is trained on nine segments and validated on the remaining segments in each iteration. To ensure complete validation, this procedure is carried out ten times. A robust assessment of the model's overall performance is provided by the final performance metric, which is produced by averaging the outcomes from each iteration (25).

In addition to 10-fold cross-validation, we implemented other regularization techniques such as L2 regularization and dropout to further reduce the risk of overfitting. L2 regularization adds a penalty to the loss function based on

the magnitude of the model's coefficients, discouraging overly complex models. Furthermore, we monitored training and validation loss during the training process to detect signs of overfitting early. Early stopping was employed to halt training if the validation loss did not improve for a pre-defined number of epochs, thus preventing the model from overfitting to the training data.

We also performed hyperparameter tuning to improve the performance of each method. Using a grid search approach, we methodically assessed a large number of hyperparameter values. Finding the parameter configurations that maximize each model's accuracy and efficiency was the goal. We successfully adjusted the models because of this exhaustive and repetitive analysis of the hyperparameter space. Our models' capacity to assess and forecast outcomes using the provided dataset is greatly improved by this meticulous calibration (26).

Justification and Explanation of the Machine Learning and Deep Learning Models' Output

Because of their intricate and opaque internal workings, ML and DL techniques are commonly referred to as "black box" models (27,28). This intricacy frequently leads to a lack of interpretability, which can be especially troublesome in crucial domains, like healthcare, where comprehending the reasoning behind forecasts is essential. Researchers have been creating methods to improve the interpretability of these models in order to address this problem. Shapley Additive exPlanations (SHAP), first presented by Lundberg and Lee (29), is a well-known technique that has gained popularity recently. By utilizing the idea of Shapley values from cooperative game theory, SHAP seeks to clarify the predictions of ML models. Because of its ability to yield insightful information on model predictions, this method has become widely accepted and used in a variety of fields, including clinical research (30,31).

In our study, we incorporated SHAP to interpret the outputs of our ML and DL models. By applying SHAP, we were able to break down the predictions into contributions from each feature, offering a clear and detailed understanding of how each variable influenced the model's decisions. This transparency is particularly valuable in healthcare applications, as it allows clinicians to trust and verify the predictions made by the models. Furthermore, we complemented SHAP with other interpretability techniques, such as local interpretable model-agnostic explanations (LIME) and feature importance analysis, to provide a multifaceted view of model behavior. These methods together helped us ensure that our models were not only accurate but also interpretable and trustworthy.

By integrating these interpretability methods, we aimed to bridge the gap between model complexity and usability, ensuring that our ML and DL models can be effectively deployed in real-world healthcare settings, where understanding and trust are paramount. This comprehensive approach to interpretability underscores our commitment to developing reliable and transparent predictive models that can aid in critical decision-making processes.

Model Performance Evaluation

Performance measures obtained from the confusion matrix were used to thoroughly assess the efficacy of both ML and DL models, as shown below. Key performance indicators such as accuracy, specificity, sensitivity, F1-score, and the receiver operating characteristic (ROC) curve were used in a thorough evaluation of the predictive models.

$$1) \text{ Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

$$2) \text{ Sensitivity} = \frac{TP}{TP+FN} * 100$$

$$3) \text{ Specificity} = \frac{TN}{TN+FP} * 100$$

$$5) \text{ f - measure} = 2 \frac{\text{precision} * \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

Ethical Considerations

For this study, ethical approval was obtained from the ethics committee at Shahid Beheshti University of Medical Sciences (approval no: IR.SBMU.RETECH.REC.1401.767, date: 12.02.2023). Due to the study's non-invasive methodology and strict adherence to patient anonymity and data confidentiality, the ethics committee granted a waiver for written informed consent. This waiver ensured the collection of confidential data without any identifying information. Access to the data was limited to the research team, thereby mitigating any potential risk to patients in accordance with the study's protocols.

RESULTS

Patients' Characteristics

This retrospective observational study examined data from 68,181 patients hospitalized for poisoning over a specified time. After applying exclusion criteria, 980 (1.4%) underwent hemodialysis. The cohort included 793 (80.9%) males and 187 (19.1%) females with a mean age of 36.5 ± 14 years. The age distribution significantly differed, with most patients (604, 61.6%) aged 21-40 years, ($p < 0.001$). Intentional poisoning accounted for 117 (11.9%) cases. Methanol was the most common poisoning agent (858, 87.6%), followed by

multidrug ingestions. The majority (830, 84.7%) had no prior history of kidney disease, and 903 (92.1%) had no prior drug use. However, 627 (64%) admitted alcohol use. Hemodialysis was the most widely used extracorporeal method (980 cases, 99.1%). Hemoperfusion was additionally used to treat 9 patients poisoned by methanol, multidrug ingestions, or methadone. This study characterized the demographics, toxins, and extracorporeal treatment approaches for a large cohort of poisoned patients requiring hemodialysis. Figure 2 shows the cause of intoxication in the studied patients.

Hyperparameters Tuning

Table 2 presents the tuned hyperparameters of four ML algorithms.

Performance Evaluation of Selected Models

Deep Learning Models

Table 3 presents the performance assessment of selected models. According to the findings in this table, the best DL model identified is the CNN with a sensitivity of 82%, specificity of 83%, accuracy of 95%, and F1-score of 82%. The best ROC value was associated with the DNN model.

Machine Learning Models

Overall, the performance of the RF model was superior to all other models, with a sensitivity of 92%, an accuracy of 98%, an F1-score of 98%, and an ROC score of 98%. However, the XGB model also achieved a sensitivity of 92%, an accuracy of 98%, and an ROC score of 98%. Additionally, the DT model achieved a sensitivity of 92% and an accuracy of 98%.

Deep Learning vs. Machine Learning Models

In terms of sensitivity, DL models performed from 80.0% to 82.0%, whereas ML models achieved sensitivities between

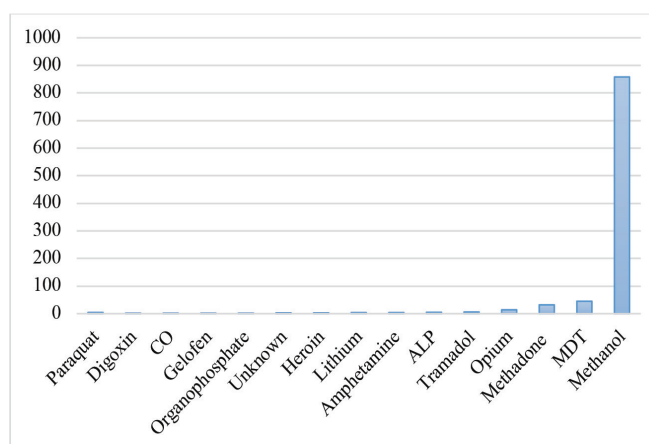


Figure 2. Cause of intoxication in the studied patients
MDT: Multiple drug toxicity

80.0% and 92.0%. The specificity of DL models ranged from 78.0% to 83.0%, whereas ML models achieved specificities between 84.0% and 94.0%. DL models had an accuracy ranging from 94.0% to 95.0%, compared to ML models, which ranged from 95.0% to 98.0%. The F1-scores for DL models were between 79.0% and 82.0%, while ML models ranged from 82.0% to 95.0%. Additionally, DL models achieved ROC scores between 97.0% and 98.0%, whereas ML models had ROC scores ranging from 95.0% to 98.0%.

Overall, while the RF model demonstrated exceptional performance across various metrics, other ML models showed superior performance compared to DL models across most measures.

Figure 3 shows the performance of various ML and DL models in predicting ICU stay duration for hemodialysis patients with poisoning, while Figure 4 compares the ROC curves of these models.

Table 2. Hyperparameters selected for ML and DL models

ML and DP models	Hyper parameters
CNN	Conv1D: filters: 64, kernel size: 3, activation: 'relu', MaxPooling1D: pool size: 2 dense (layer 1): units: 64 activation: 'relu', dense (output layer): units: 1, activation: 'sigmoid', optimizer: 'adam', loss function: 'binary_crossentropy', epochs: 10, batch size: 32
DT	Criterion: 'gini', max depth: none, min samples split: 2, min samples leaf: 1, max features: none, splitter: 'best', random state: 42
DNN	Dense (layer 1): units: 64, activation: 'relu', input shape: (X.shape[1,]), dense (layer) units: 32, activation: 'relu', dense (output layer): units: 1, activation: 'sigmoid' optimizer: 'adam', loss function: 'binary_crossentropy', epochs: 10, batch size: 32
FNN	Hidden layer sizes: (100,) max iterations: 1000, random state: 42, n_splits: 10, shuffle: true, random_state: 42
LSTM	Units: 100, input_shape: (X_train.shape[1], X_train.shape[2]) optimizer: 'adam' loss: 'binary_crossentropy' metrics: ['accuracy'] Training hyperparameters epochs: 100, batch_size: 32, verbose: 0 StratifiedKFold parameters, n_splits: 10, huffle: true, random_state: 42
RF	n_estimators: 100, random_state: 42
SVM	Kernel: 'linear', robability: true, random_state: 42
XGB	Base_score: 0.5, booster: gbtrees, colsample_bylevel: 1, colsample_bynode: 1 colsample_bytrees: 1, gamma: 0, gpu_id: -1, importance_type: gain, learning_rate: 0.300000012, max_delta_step: 0, max_depth: 6, min_child_weight: 1, missing: nan, monotone_constraints: () n_estimators: 100, random_state: 42, validate_parameters: 1

XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, FNN: Feedforward neural network, CNN: Convolutional neural network, LSTM: Long short-term memory, ML: Machine learning, DL: Deep learning

Table 3. Performance evaluation of selected models

Models	Performance of each deep learning models				
	Sensitivity (%)	Specificity (%)	Accuracy (%)	F1-score (%)	ROC (%)
CNN	82.0	83.0	95.0	82.0	97.0
LSTM	82.0	80.0	94.0	81.0	97.0
DNN	80.0	82.0	94.0	81.0	98.0
FNN	80.0	78.0	94.0	79.0	97.0
Models	Performance of each machine learning models				
	Sensitivity (%)	Specificity (%)	Accuracy (%)	F1-score (%)	ROC (%)
RF	92.0	92.0	98.0	95.0	98.0
DT	92.0	92.0	98.0	92.0	95.0
SVM	80.0	84.0	95.0	82.0	96.0
XGB	90.0	94.0	98.0	90.0	98.0

XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, FNN: Feedforward neural network, CNN: Convolutional neural network, LSTM: Long short-term memory, ROC: Receiver operating characteristic

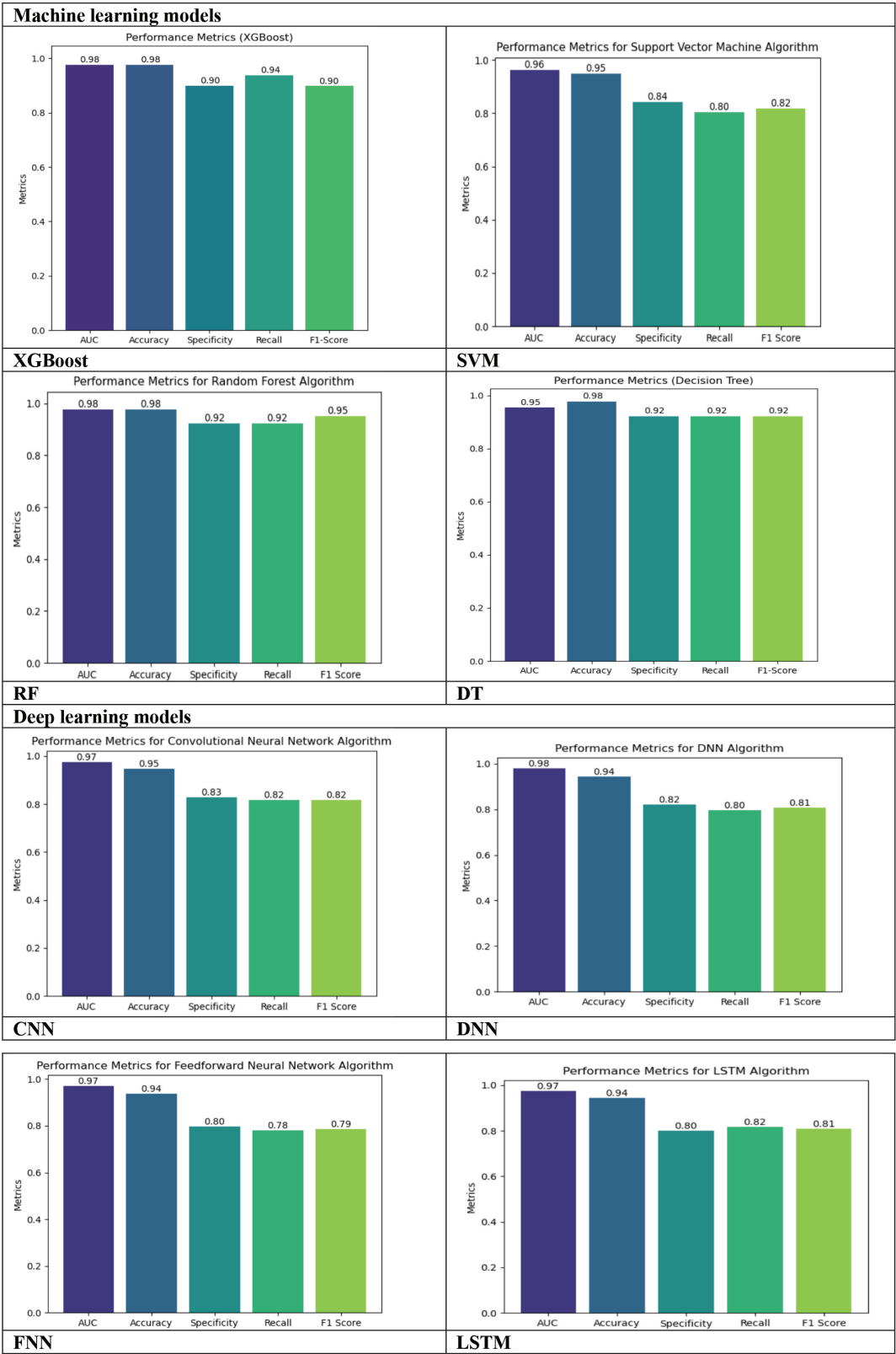


Figure 3. Performance evaluation of ML and DL models in predicting ICU stay duration for hemodialysis patients with poisoning
XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, CNN: Convolutional neural network, FNN: Feedforward neural network, LSTM: Long short-term memory, ML: Machine learning, DL: Deep learning, ICU: Intensive care unit, AUC: Area under the curve

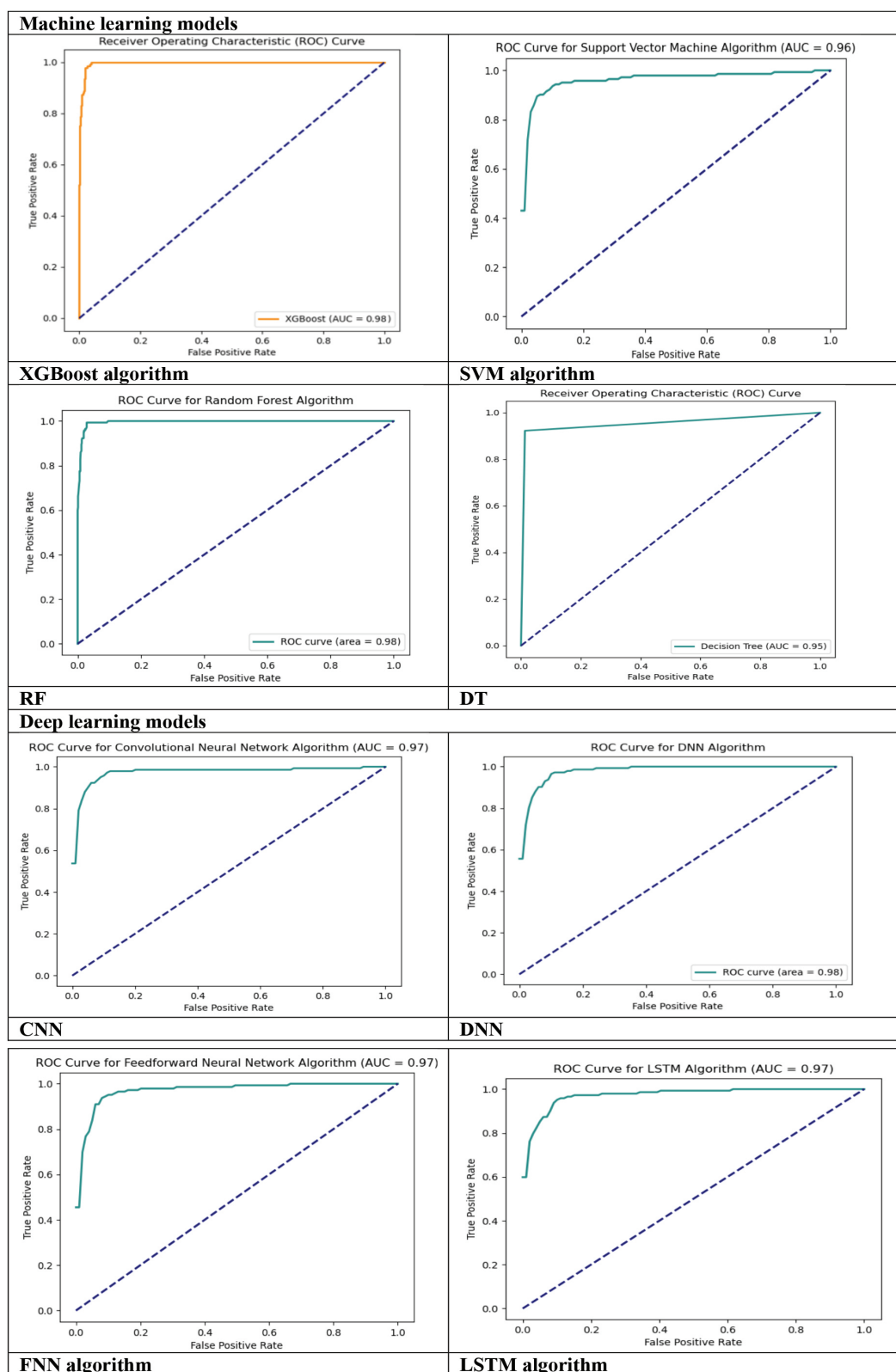


Figure 4. ROC curves comparing ML and DL models in predicting ICU stay duration for hemodialysis patients with poisoning
 XGB: Extreme gradient boosting, SVM: Support vector machine, RF: Random forest, DT: Decision tree, DNN: Deep neural network, CNN: Convolutional neural network, FNN: Feedforward neural network, LSTM: Long short-term memory, ROC: Receiver operating characteristic, ML: Machine learning, DL: Deep learning, ICU: Intensive care unit

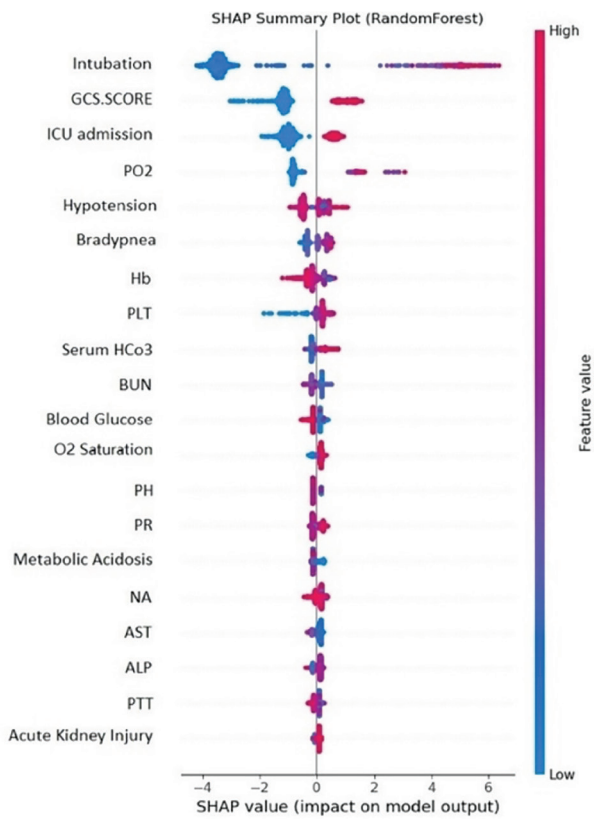


Figure 5. SHAP summary plot
SHAP: Shapley Additive exPlanations, GCS: Glasgow Coma Scale, ICU: Intensive care unit

Explanation and Justification the Output of Machine Learning and Deep Learning Models

Shapley Additive exPlanations (SHAP)

The SHAP summary plot shows the impact of each feature on the model's output, with higher values indicating a greater positive impact and lower values indicating a greater negative impact (Figure 5). For example, the feature "intubation" has a positive impact on the model's output, while the feature "blood glucose" has a negative impact. The most important features for predicting the duration of ICU stay among hemodialysis patients suffering from poisoning were intubation score, GCS score, and ICU admission type.

Local Interpretable Model-agnostic Explanations (LIME)

Figure 6 illustrates LIME plots depicting the cumulative influence of essential features on the model's prediction of ICU stay duration for hemodialysis patients with poisoning. The most influential features were GCS <8, intubation, acute kidney injury (AKI), PO₂, blood urea nitrogen (BUN), metabolic acidosis, and number of hemodialysis sessions.

DISCUSSION

In this study, we aimed to predict the ICU LOS for hemodialysis patients with poisoning by comparing the performance of various ML and DL models. The importance of timely prediction and intervention in poisoning cases cannot be overstated, as delayed treatment can lead

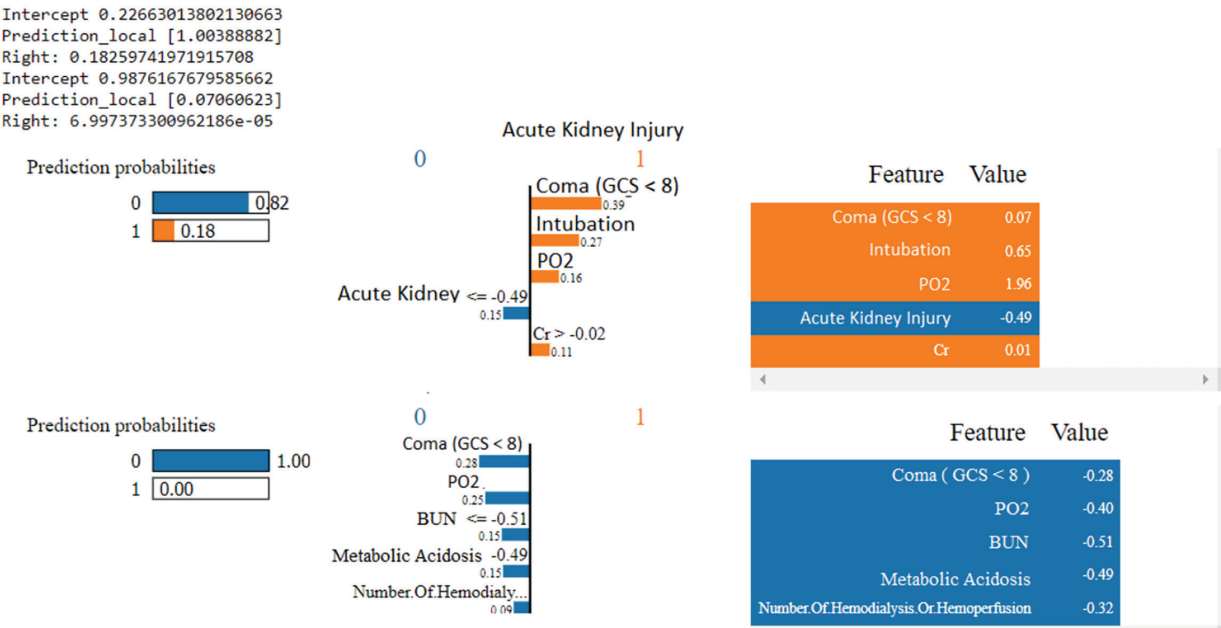


Figure 6. LIME plot
LIME: Local interpretable model-agnostic explanations, GCS: Glasgow Coma Scale

to severe complications or even death. Our findings demonstrated that the RF model outperformed all other models, achieving sensitivity, accuracy, F1-score, and an ROC score. In comparison, the best-performing DL model, the CNN, achieved a sensitivity, specificity, accuracy, and an F1-score. Overall, ML models exhibited higher sensitivities, specificities, accuracies, and F1-scores than DL models, indicating that, especially the RF model, they are more effective for predicting ICU stay duration for this patient population.

As mentioned above, our results showed that the RF model surpassed all ML and DL models, achieving superior sensitivity, accuracy, F1-score, and ROC score. In contrast, the best-performing DL model, the CNN, attained notable sensitivity, specificity, accuracy, and F1-score. The RF model and the CNN are pivotal in our study due to their exemplary performance and complementary strengths, which highlight the potential of both ML and DL in medical predictions. The RF model's superior sensitivity, accuracy, F1-score, and ROC score underscore its robustness and reliability in predicting ICU stay duration for hemodialysis patients with poisoning. This high performance can be attributed to the RF model's ability to handle diverse data inputs and reduce overfitting through its ensemble learning approach, making it a powerful tool for clinical decision-making. On the other hand, the CNN's impressive sensitivity, specificity, accuracy, and F1-score demonstrate the advanced capabilities of DL in capturing complex patterns within data. The CNN's performance highlights its potential for applications where intricate data structures and high-dimensional features are present. Together, these models illustrate the importance of leveraging both traditional ML techniques and advanced DL methods to achieve optimal predictive accuracy and clinical relevance, ultimately improving patient outcomes through precise and timely interventions. Huang et al. (32) study revealed that out of all the algorithms examined, the RF and ensemble methods exhibited superior predictive performance. The study indicated that RF is particularly effective for predictive modeling of blood pressure during hemodialysis (32). Other studies (33,34) have also shown that CNN can be effectively used to predict hemodialysis.

Our study findings indicate that ML models exhibited higher sensitivities, specificities, accuracies, and F1-scores than DL models, suggesting that particularly the RF model is more effective for predicting ICU stay duration for this patient population. These results underscore the potential of ML techniques in clinical decision-making processes for poisoned hemodialysis patients. The importance of comparing ML and DL models in our study lies in their distinct strengths and applications within

predictive modeling for medical outcomes. ML models have shown superior sensitivities, specificities, accuracies, and F1-scores compared to DL models, highlighting their effectiveness in handling structured data and generating precise predictions. This superiority is crucial in clinical settings where accurate predictions can significantly impact patient care and outcomes. ML models, such as RF, excel in interpreting relationships between input variables and outcomes, making them valuable tools for predicting ICU stay durations and guiding timely interventions for poisoned hemodialysis patients (32). Conversely, DL models like CNN and LSTM offer advantages in capturing intricate patterns from complex, unstructured data, although in our study, their performance metrics were comparatively lower. By understanding and leveraging the strengths of both ML and DL approaches, healthcare practitioners can enhance their predictive capabilities and ultimately improve patient care and treatment outcomes in critical medical scenarios.

Jordan and Mitchell (35) mentioned that ML addresses these limitations by enhancing a computer program's performance through experience with specific tasks and performance measures. Essentially, ML aims to automate analytical model building for cognitive tasks like object detection or natural language translation. This is accomplished by applying algorithms that iteratively learn from training data, enabling computers to uncover hidden insights and complex patterns without explicit programming (36). By learning from past computations and identifying regularities in large datasets, ML can facilitate reliable and repeatable decision-making. Consequently, ML algorithms have been successfully applied in numerous domains, including fraud detection, credit scoring, next-best offer analysis, speech and image recognition, and natural language processing (36).

In our study, we have identified several crucial prognostic factors that warrant further discussion. Our findings represent a significant advancement in the field, as previous research has predominantly focused on overall hospital LOS, rather than specifically addressing ICU duration in poisoned patients requiring hemodialysis. The most influential features identified in our study were GCS <8, intubation, AKI, PO₂, BUN, metabolic acidosis, and number of hemodialysis sessions. Among these, Intubation score, GCS score, and ICU admission type emerged as the paramount predictors of ICU stay duration. These results align with and expand upon the findings of Rahimi et al. (37), who reported that intubation, GCS, and ICU admission were significant prognostic factors in poisoned patients undergoing hemodialysis. Furthermore, the study by Brenner et al. (38) on arteriovenous access failure highlights

the impact of complications on hospitalization duration and costs. This emphasizes the need for meticulous vascular access management in poisoned hemodialysis patients to potentially reduce the duration of ICU stays. Additionally, the study by Yan et al. (39) on racial and ethnic disparities in hospitalization rates among hemodialysis patients introduces an important dimension worthy of exploration in future research on ICU stay duration. While our current study did not specifically address these demographic factors, their potential influence on ICU outcomes in poisoned hemodialysis patients merits further investigation. Overall, this section of our study represents a significant advancement in understanding the determinants of ICU stay duration for poisoned patients requiring hemodialysis. By focusing on specific ICU-related outcomes, we provide valuable insights that can inform clinical practices and improve patient management strategies in this critical patient population. Future research should aim to incorporate demographic factors and explore their potential impact on ICU outcomes to build a more comprehensive understanding of the determinants influencing ICU stay duration in poisoned hemodialysis patients.

Study Limitations

The study has several limitations, including the single-center data from Loghman Hakim Hospital, which may affect generalizability, and the sample size of 980 patients, which may not capture all variability in ICU stay duration for hemodialysis patients with poisoning. Future research should aim to include multicenter data to enhance generalizability and increase sample size to capture a broader range of variability.

Additionally, this study acknowledges that the models were developed based on static data, which may limit their applicability to real-time clinical decision-making. Future research should focus on testing and validating these models in dynamic, real-time clinical environments to assess their practical utility and performance in such settings.

Moreover, this study acknowledges potential biases introduced by the class imbalance in the dataset, particularly the higher prevalence of short LOS cases compared to long LOS cases. Although under-sampling techniques were applied to address this issue, such methods might not fully eliminate the inherent bias, potentially impacting the generalizability and robustness of the model's predictions for underrepresented classes. To further address class imbalance, methods such as SMOTE, class weighting, hybrid

sampling, or ensemble approaches could be employed, alongside emphasizing evaluation metrics like F1-score and precision-recall-area under the curve.

Lastly, this study acknowledges that temporal trends, such as changes in hospital practices or treatments between 2016 and 2020, were not explicitly accounted for in the analysis. Such trends could impact the generalizability of the findings, as variations in clinical protocols, resource availability, or treatment methods over time might influence the outcomes. Future studies should consider incorporating temporal stratification or modeling to assess the potential impact of these trends and minimize their confounding effects on the results.

CONCLUSION

This study aimed to predict the ICU LOS for hemodialysis patients suffering from poisoning, by comparing ML models with DL models. The results showed that among DL models, the CNN performed best with sensitivity, specificity, accuracy, and F1-score values, while the DNN achieved the best ROC value. However, the RF model, an ML model, outperformed all other DL and ML models, achieving higher scores across sensitivity, accuracy, F1-score, and ROC metrics.

In general, ML models demonstrated superior performance compared to DL models, with higher sensitivities, specificities, accuracies, and F1-scores. These findings suggest that ML models, particularly the RF model, are more effective for predicting ICU stay duration for hemodialysis patients with poisoning. Therefore, incorporating ML models into clinical practice could enhance the prediction and management of ICU stay durations in this patient population, potentially improving outcomes and resource allocation.

ETHICS

Ethics Committee Approval: Ethical approval was obtained from the ethics committee at Shahid Beheshti University of Medical Sciences (approval no: IR.SBMU.RETECH.REC.1401.767, date: 12.02.2023).

Informed Consent: Retrospective study.

Acknowledgment

We thank the Research Deputy of the Shahid Beheshti University of Medical Sciences for financially supporting this project. We also would like to extend our gratitude to all the medical experts at Loghman Hakim Hospital who participated in this study.

FOOTNOTES

Authorship Contributions

Surgical and Medical Practices: M.R., Concept: K.M., M.R.A., H.R., P.E.T.E., M.S., H.K.A., Design: K.M., S.S., M.R.A., B.M., Data Collection or Processing: K.M., S.S., H.R., B.S., M.P., M.S.B., Analysis or Interpretation: K.M., M.R., Literature Search: K.M., B.M., P.E.T.E., H.K.A., V.M., A.H.D., M.R., M.S.B., Writing: K.M., B.M., V.M., M.R.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declare that this study received no financial support.

REFERENCES

- Yaylacı S, Genc A, Demir M, Cinemre H, Tamer A. Retrospective evaluation of patients at follow-up with acute poisoning in intensive care unit. *Niger J Clin Pract.* 2016;19:223-6.
- Kabiri M, Hosseini SM, Veisi F, Mortezaei S, Ahmadi A, Agashteh Z, et al. Investigation of poisoning prevalence and its related factors in patients referred to Farsan's Hospital during 2018-2019, Iran. *Epidemiol Health Syst J.* 2022;9:75-9.
- Boedeker W, Watts M, Clausing P, Marquez E. The global distribution of acute unintentional pesticide poisoning: estimations based on a systematic review. *BMC public health.* 2020;20:1-19.
- Elgazzar FM, Afifi AM, Shama MAE, Askary AE, El-Sarnagawy GN. Development of a risk prediction nomogram for disposition of acute toxic exposure patients to intensive care unit. *Basic Clin Pharmacol Toxicol.* 2021;129:256-67.
- Lindqvist E, Edman G, Hollenberg J, Nordberg P, Forsberg S. Long-term mortality and cause of death for patients treated in intensive care units due to poisoning. *Acta Anaesthesiol Scand.* 2019;63:500-5.
- Lindqvist E, Edman G, Hollenberg J, Nordberg P, Ösby U, Forsberg S. Intensive care admissions due to poisoning. *Acta Anaesthesiol Scand.* 2017;61:1296-304. Erratum in: *Acta Anaesthesiol Scand.* 2020;64:139.
- Resiere D, Kallel H, Oxybel O, Chabartier C, Florentin J, Brouste Y, et al. Clinical and epidemiological characteristics of severe acute adult poisoning cases in Martinique: implicated toxic exposures and their outcomes. *Toxics.* 2020;8:28.
- Ng F. Ten-year profile of acute poisoning patients presenting to an accident and emergency department requiring intensive care in a regional hospital of Hong Kong. *Hong Kong J Emerg Med.* 2019;26:3-14.
- Chan M, Ostermann M. Outcomes of chronic hemodialysis patients in the intensive care unit. *Crit Care Res Pract.* 2013;2013:715807.
- D'Agata EM, Mount DB, Thayer V, Schaffner W. Hospital-acquired infections among chronic hemodialysis patients. *Am J Kidney Dis.* 2000;35:1083-8.
- Peres IT, Hamacher S, Oliveira FLC, Thomé AMT, Bozza FA. What factors predict length of stay in the intensive care unit? Systematic review and meta-analysis. *J Crit Care.* 2020;60:183-94.
- Almashrafi A, Elmontsri M, Aylin P. Systematic review of factors influencing length of stay in ICU after adult cardiac surgery. *BMC Health Serv Res.* 2016;16:1-12.
- Shaw C. How can hospital performance be measured and monitored? Copenhagen: World Health Organization, Regional Office for Europe; 2003.
- Abd-Elrazek MA, Eltahawi AA, Abd Elaziz MH, Abd-Elwhab MN. Predicting length of stay in hospitals intensive care unit using general admission features. *Ain Shams Eng J.* 2021;12:3691-702.
- Li C, Chen L, Feng J, Wu D, Wang Z, Liu J, et al. Prediction of length of stay on the intensive care unit based on least absolute shrinkage and selection operator. *IEEE Access.* 2019;7:110710-21.
- Daghistani TA, Elshawi R, Sakr S, Ahmed AM, Al-Thwayee A, Al-Mallah MH. Predictors of in-hospital length of stay among cardiac patients: a machine learning approach. *Int J Cardiol.* 2019;288:140-7.
- Orooji A, Shanbehzadeh M, Mirbagheri E, Kazemi-Arpanahi H. Comparing artificial neural network training algorithms to predict length of stay in hospitalized patients with COVID-19. *BMC Infect Dis.* 2022;22:1-13.
- Wang K, Yan LZ, Li WZ, Jiang C, Wang NN, Zheng Q, et al. Comparison of four machine learning techniques for prediction of intensive care unit length of stay in heart transplantation patients. *Front Cardiovasc Med.* 2022;9:863642.
- Afrash MR, Kazemi-Arpanahi H, Nopour R, Tabatabaei ES, Shanbehzadeh M. Proposing an intelligent monitoring system for early prediction of need for intubation among COVID-19 hospitalized patients. *J Environ Health Sustain Dev.* 2022.
- Hosseini SM, Rahimi M, Afrash MR, Ziaefar P, Yousefzadeh P, Pashapour S, et al. Prediction of acute organophosphate poisoning severity using machine learning techniques. *Toxicology.* 2023;486:153431.
- Wu J, Lin Y, Li P, Hu Y, Zhang L, Kong G. Predicting prolonged length of ICU stay through machine learning. *Diagnostics (Basel).* 2021;11:2242.
- Kaji DA, Zech JR, Kim JS, Cho SK, Dangayach NS, Costa AB, et al. An attention based deep learning model of clinical events in the intensive care unit. *PloS One.* 2019;14:e0211057.
- Syed M, Syed S, Sexton K, Syeda HB, Garza M, Zozus M, et al. Application of machine learning in intensive care unit (icu) settings using MIMIC dataset: systematic review. *Informatics (MDPI).* 2021;8:16.
- Ayouni S, Maddeh M, Al-Otaibi S, Alazzam MB, Alturki NM, Hajjej F. Development of a smart hospital bed based on deep learning to monitor patient conditions. *J Disabil Res.* 2023;2:25-36.
- Malakouti SM, Menhaj MB, Suratgar AA. The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. *Cleaner Eng Technol.* 2023;15:100664.
- Bardenet R, Brendel M, Kégl B, Sebag M. Collaborative hyperparameter tuning. In: *Proceedings of the International Conference on Machine Learning (ICML); 2013 May 13; Atlanta, USA.* PMLR; 2013:199-207.
- Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, Swami A. Practical black-box attacks against machine learning. In: *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security; 2017:506-19.*
- Sudmann A. On the media-political dimension of artificial intelligence: deep learning as a black box and OpenAI. *Digital Cult Soc.* 2018;4:181-200.
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30.
- Sun J, Sun CK, Tang YX, Liu TC, Lu CJ. Application of SHAP for explainable machine learning on age-based subgrouping mammography questionnaire data for positive mammography prediction and risk factor identification. *Healthcare (Basel).* 2023;11:2000.
- Wang F, Wang D, Xu Y, Jiang H, Liu Y, Zhang J. Potential of the non-contrast-enhanced chest CT radiomics to distinguish molecular

- subtypes of breast cancer: a retrospective study. *Front Oncol.* 2022;12:848726.
32. Huang JC, Tsai YC, Wu PY, Lien YH, Chien CY, Kuo CF, et al. Predictive modeling of blood pressure during hemodialysis: a comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method. *Comput Methods Programs Biomed.* 2020;195:105536.
33. Wang F, Wang Y, Tian Y, Zhang P, Chen J, Li J. Pattern recognition and prognostic analysis of longitudinal blood pressure records in hemodialysis treatment based on a convolutional neural network. *J Biomed Inform.* 2019;98:103271.
34. Ota K, Nishiura Y, Ishihara S, Adachi H, Yamamoto T, Hamano T. Evaluation of hemodialysis arteriovenous bruit by deep learning. *Sensors (Basel).* 2020;20:4852.
35. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science.* 2015;349:255-60.
36. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electron Mark.* 2021;31:685-95.
37. Rahimi M, Afrash MR, Shadnia S, Mostafazadeh B, Evini PET, Bardsiri MS, et al. Prediction the prognosis of the poisoned patients undergoing hemodialysis using machine learning algorithms. *BMC Med Inform Decis Mak.* 2024;24:38.
38. Brenner L, Singh AK, Campbell D, Frei F, Winkelmayer WC. Associations between demographic factors and provider structures on cost and length of stay for hemodialysis patients with vascular access failure. *Clin J Am Soc Nephrol.* 2006;1:455-61.
39. Yan G, Norris KC, Greene T, Alison JY, Ma JZ, Yu W, Cheung AK. Race/ethnicity, age, and risk of hospital admission and length of stay during the first year of maintenance hemodialysis. *Clin J Am Soc Nephrol.* 2014;9:1402-9.