Using Explainable Machine Learning for Diabetes Management in Emergency Departments

Silas Majyambere
Department of Computer and Systems Sciences
Stockholm University (SU)
Stockholm, Sweden
E-mail: majyambere@dsv.su.se

Celestin Twizere
Center for Biomedical Engineering and E-Health (CEBE)
University of Rwanda (UR)
Kigali, Rwanda
E-mail: celestintwizere@gmail.com

Tony Lindgren
Department of Computer and Systems Sciences
Stockholm University (SU)
Stockholm, Sweden
E-mail: tony@dsv.su.se

Gerard Nyiringango
Center for Biomedical Engineering and E-Health (CEBE)
University of Rwanda (UR)
Kigali, Rwanda
E-mail: nyiringangogerard@gmail.com

***Abstract*—Uncontrolled diabetes can lead to severe complications and Intensive Care Unit (ICU) admissions. This study presents an explainable machine learning model using electronic health records to predict ICU admissions and estimate hospital stay duration for diabetic patients. AdaBoost model outperformed other models on ICU admission prediction, while CatBoost exhibited superior performance in estimating ICU length of stays among diabetic patients admitted to the emergency departments. The results demonstrate the potential of explainable machine learning in ICU risk assessment and can aid healthcare providers in early intervention and resource utilization. The clinician and the proposed model agree on the top 25 features identified by Shapley Additive exPlanations (SHAP) methods for predicting ICU admission, but they differ in the ranking of the top five most significant predictors.**

Keywords-Explainable Machine Learning; Intensive Care Unit; Diabetes; Length of Stay; SHAP.

I. INTRODUCTION

Diabetes is a chronic disease affecting millions globally and is a major contributor to morbidity and mortality. There are three kinds of diabetes: Type 1 Diabetes Mellitus (T1DM), which results from the pancreas's inability to produce insulin; Type 2 Diabetes Mellitus (T2DM), characterized by the body's ineffective use of insulin [1][2]; and gestational diabetes, which occurs during pregnancy and may later be resolved or progress to T1DM or T2DM. Despite medical advancements, diabetes prevalence continues to rise, with cases estimated at 536 million in 2021 and projected to reach 783 million by 2045 [3]. Managing diabetes remains challenging due to complications, such as cardiovascular diseases, neuropathy, nephropathy, retinopathy, and glycemic complications, often leading to frequent Emergency Department (ED) visits and Intensive Care Unit (ICU) admissions [4]. This study focuses on T1DM and T2DM.

Emergency departments play a crucial role in managing critically ill diabetic patients and serve as valuable data sources for predicting ICU admissions and estimating hospital stays. Machine Learning (ML) models trained on ED data can assist clinicians in identifying high-risk patients in real-time, allowing for timely interventions. In the medical domain, Explainable Artificial Intelligence (XAI) techniques, such as SHAP, are essential to ensure the reliability and clinical applicability of AI-enabled tools. XAI fosters trust among ED healthcare providers by ensuring that ML models use relevant and medically validated features. Additionally, XAI provides valuable insights into critical risk factors, improving decision-making in emergency care settings.

Machine learning has been widely applied during the COVID-19 pandemic to predict ICU admissions [5], as well as to estimate ICU length of stay [6] and readmission risk among diabetic patients [7]. However, the application of explainable ML in emergency settings for diabetes management remains underexplored. The contribution of this study is in three folds: (1) Develop boosted tree-based ensemble machine learning models using ED data to predict ICU admission risk for T1DM and T2DM patients, (2) Build a machine learning model to estimate the length of hospital stays for diabetes patients upon ED admission, (3) Apply SHAP methods to provide interpretable explanations for the classification and regression models predicting ICU admissions and length of stay.

This study aims to develop explainable ML models using boosted tree ensemble algorithms and SHAP to predict ICU admission risk and estimate ICU length of stay for diabetic patients in the ED. The dataset for this study includes historical patient data, including demographics, vital signs, and lab results from electronic health records stored in the Medical Information Mart for Intensive Care (MIMIC)-IV database [19]. Performance evaluation metrics used for ICU

 ISBN: 978-1-68558-270-8

admission prediction (classification task) are accuracy, precision, recall, F1-score, and Area Under the Curve score (AUC), while hospital stay estimation (regression task) was evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics.

The remainder of this paper is structured as follows: In Section 2, we outline the methodology guiding our study, Section 3 presents the experimental results, Section 4 provides a discussion of the findings, and Section 5 concludes the study while outlining directions for future research.

II. METHODOLOGY

This study has developed ML models to predict diabetes ICU admissions and length of stay using MIMIC-IV data and five machine learning models, including Decision Tree, AdaBoost, CatBoost, XGBoost, and LightGBM. These boosted tree ensemble models were chosen for their superior performance on ICU and Length Of Stays (LOS) prediction, as reported in the literature [16][17][18]. Additionally, they effectively address class imbalance, making them well-suited for this study. Performance was evaluated using classification and regression metrics. SHAP was applied to ensure interpretability and clinical relevance. This study was conducted in six steps: data acquisition, dataset building, data preprocessing, model building, performance evaluation, and generating explanations of the best model, as shown in Figure 1.

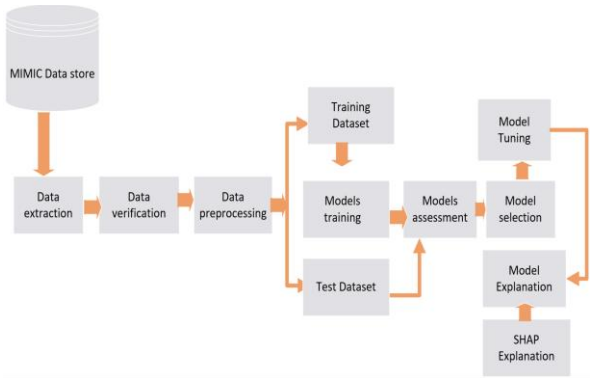

Figure 1. Design of the research processes.

A. *Data acquisition*

We used MIMIC-IV (version 2) [19], focusing on MIMICIV-ED, MIMICIV-ICU, and MIMICIV-Hospital datasets, extracted via BigQuery SQL. Diabetic patients were identified using the International Classification of Diseases (ICD)-10 codes: E10.XXX for T1DM and E11.XXX for T2DM [20]. MIMIC-IV extends MIMIC-III, providing electronic health records for Beth Israel Medical Center ICU patients from 2008 to 2019. The publicly available database contains deidentified patient data in compliance with the Health Insurance Portability and Accountability Act (HIPAA). Figure 2 highlights the tables and the type of data extracted from the MIMIC-IV database.

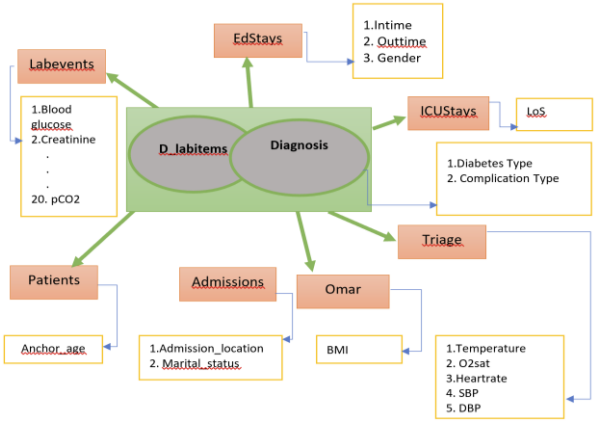

Figure 2. MIMIC-IV tables used dataset building.

Data was extracted using ICD-10 codes. Subject ID represents individual patients, while item ID corresponds to lab tests during hospitalization. More data was extracted from the Labevents table (20 features).

B. *Dataset building*

Between 2008 and 2019, 9339 diabetic patients were hospitalized, including 1090 with T1DM and 7097 with T2DM. Focusing on first-time emergency admissions, 4365 patients were selected after removing duplicates (863 T1DM, 3502 T2DM). The study includes demographic data, vital signs, diagnoses, and lab results for critically ill diabetic patients. We extracted 34 numerical features (detailed in Table I) and 6 categorical features, resulting in 20 features after label and one-hot encoding, as illustrated in Figure 3.

C. *Data preprocessing*

Data preprocessing is a crucial step in machine learning, involving missing data handling, feature correlation analysis, deduplication, outlier removal, and data scaling to enhance model performance. Categorical variables were encoded using label encoding for binary features and one-hot encoding for multi-value features. To maintain a focus on T1DM and T2DM, cases related to pregnancy, malnutrition, and unspecified causes were excluded, reducing the dataset to 4,317 patients. Outlier detection was performed by analyzing each feature individually for anomalies due to human error or lack of relevance to the study. Extreme deviations from expected distributions, such as a Body Mass Index (BMI) of 3,658.50, systolic blood pressure of 1.00, or a temperature of 9°F (as shown in Table I), were considered outliers and removed. Features with more than 50% missing values, including Protein C (PC), Protein Creatinine Ratio (PCR), and Hemato, were removed to prevent bias to the ML model. Finally, two datasets were created: one for ICU admission prediction (4,055 samples, 49 features) and another for ICU length of stay prediction, patients with a zero-day ICU stay were excluded from ICU LOS dataset, as they were not admitted leading to a dataset of 1,432 samples, and 49 features.

TABLE I. DESCRIPTIVE STATISTICS

Feature Name	Count	Min	Max	Mean	Missing Values (%)
Temp	3721.00	9.00	106.30	97.89	13.81
Hrate	3776.00	20.00	220.00	84.97	12.53
Resprate	3758.00	8.00	40.00	17.74	12.95
O2sat	3737.00	60.00	109.00	98.23	13.44
SBP	3687.00	1.00	249.00	143.64	14.59
DBP	3633.00	14.00	474.00	78.15	15.84
Age	4317.00	18.00	91.00	58.53	0.00
BMI	3964.00	0.67	3658.5	54.83	8.18
Alb	3765.00	1.20	5.85	3.81	12.79
Hemog	4148.00	3.20	19.60	11.64	3.91
Cr	4287.00	0.10	16.30	1.66	0.69
BG	4295.00	28.00	1083.00	200.01	0.51
Trig	3061.00	10.00	7140.00	189.43	29.09
HbA1C	3495.00	4.00	19.40	8.24	19.04
RBC	4285.00	1.77	6.65	4.08	0.74
WBC	4216.00	0.20	181.70	10.83	2.34
Sod	4275.00	82.00	177.00	136.72	0.97
Pot	4252.00	1.80	10.00	4.78	1.51
pH	2465.00	6.64	7.86	7.34	42.90
pO2	2395.00	12.00	536.00	120.34	44.52
pCO2	2387.00	8.00	199.00	40.25	44.71
LDL	2938.00	10.00	407.00	96.13	31.94
HDL	3017.00	2.10	214.00	48.35	30.11
EdLOS	4317.00	0.00	4.73	0.55	0.00
ICUStays	1553.00	0.00	50.33	2.76	64.03
BUN	4286.00	1.00	260.00	23.60	0.72
Bilir	3456.00	0.10	29.20	0.82	19.94
PC	1999.00	6.00	1697.00	250.28	53.69
CholR	3027.00	1.30	36.00	3.97	29.88
CK	3049.00	8.00	68510.00	332.95	29.37
VB12	2337.00	88.00	1976.00	671.05	45.87
Iron	2494.00	8.00	396.00	61.08	42.23
Hemato	1286.00	10.00	60.00	32.76	70.21
PCR	1065.00	0.10	355.50	1.86	75.33

Abbreviations: Temp: Temperature, Hrate: Heart rate, Resprate: Respiration rate, O2sat: Oxygen saturation, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, Alb: Albumin, Hemog: Hemoglobin, Cr: Creatinine, BG: Blood glucose, Trig: Triglycerides, HbA1C: Glycated Hemoglobin, RBC: Red Blood Cells, WBC: White Blood Cells, Sod: Sodium, Pot: Potassium, pO2: Partial pressure of oxygen, pCO2: Partial pressure of Carbon Dioxide, LDL: Low-Density Lipoprotein sometimes called bad cholesterol, HDL: High-Density Lipoprotein known as good cholesterol, BUN: Blood Urea Nitrogen, Bilir: Bilirubin (Total), CholR: Cholesterol ratio, CK: Creatinine kinase, VB12: Vitamin B12.

Our experiment shows the distribution of categorical features (Figure 3) in the dataset, which includes 2293 females and 2024 males, with 3464 type 2 diabetes and 853 type 1 diabetes patients. A total of 4317 T1DM and T2DM patients were admitted to the emergency department, with 1560 admitted to the ICU, indicating not a fully balanced dataset. The "Admit" feature has four categories for admission status, while the "CompType" feature includes nine diabetes-related complications. The "Marital Status" feature has four categories: married, widowed, single, and divorced. Widowed, single, and divorced are often grouped as single, but they are treated separately here due to psychological differences that may affect diabetes differently.

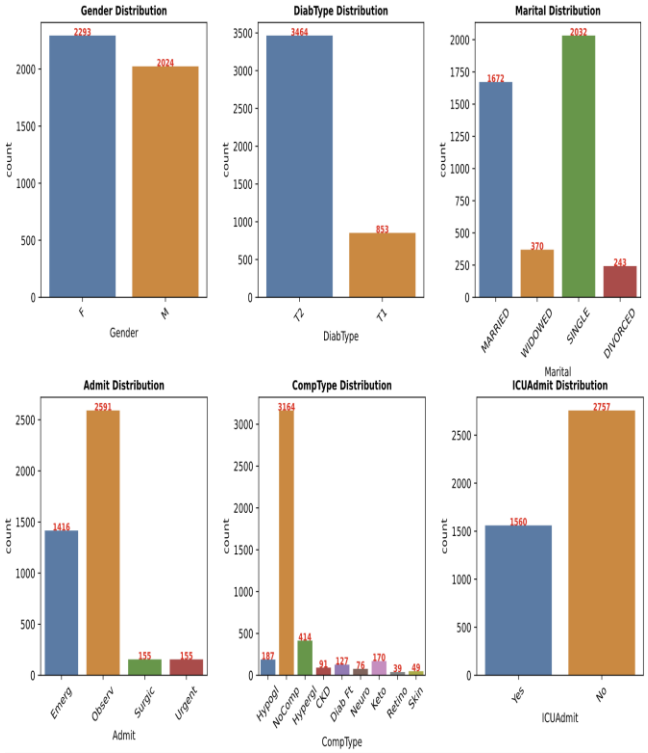

Figure 3. Visualizing the distribution of categorical data.

We used the Robust scaler method to minimize the effect of features with very low or very high values to put all numerical data on the same scale. The formula of a robust scaler is:

$$X = \frac{X - X_{median}}{IQR} \qquad (1)$$

A robust scaler is less affected by outliers. IQR means the InterQuartile Range between the first quartile (25%) and the third quartile (75%).

D. *Machine Learning Algorithms*

This study uses supervised machine learning with five models: Decision Tree [8], CatBoost [9], AdaBoost [10], XGBoost [12], and LightGBM [13] for ICU admission prediction and length of stay estimation. Initial experiments involve training these models with default parameters on cleaned datasets. The best model was selected and fine-tuned through hyperparameter optimization to improve predictive performance.

(1) *Decision Tree Algorithm*

A decision tree model is a simple predictive modeling tool that can be used for classification and regression tasks. It works by building a tree where the nodes are the decision rules, and the leaf nodes give the output of the prediction [8]. This type of model is known as the “white model” since it is easier to understand and interpret its prediction process. The decision tree model is affected by the curse of dimensionality, where a large number of features will increase the splitting process, which results in poor performance of the model.

Tree-based ensemble models were introduced to solve the limitations of decision tree models operating on large datasets.

(2) *CatBoost Algorithm*

CatBoost is an open-source gradient-boosted tree library that excels in handling categorical data without the need for transformation. Developed by Yandex, CatBoost is well-suited for machine learning tasks involving heterogeneous data, particularly when categorical features are present. It builds a model iteratively through gradient boosting, improving its performance step by step using weak learners, typically decision trees [9]. One of CatBoost's key advantages is its ability to handle categorical data efficiently, reducing the need for pre-processing transformations often required by other machine learning models.

(3) *AdaBoost Algorithm*

AdaBoost is a boosted tree ensemble algorithm that improves the weights of weak learners over multiple iterations to make predictions. It builds a strong classifier by training weak classifiers sequentially, with more emphasis on misclassified instances in each iteration. AdaBoost is particularly effective for binary classification and regression tasks [10]. Its drawback is that training weak learners sequentially leads to longer training time when the dataset is large.

(4) *XGBoost Algorithm*

XGBoost (Extreme Gradient Boosting) builds trees sequentially, with each tree correcting the errors of the previous ones through gradient optimization. It is an open-source library that uses distributed gradient-boosted trees for predictions. XGBoost combines regularization and optimization techniques to increase predictive performance and reduce the training time. Known for its high prediction accuracy and fast training speed, XGBoost is widely used in classification and regression tasks with structured data [11] [12]. The base model used in XGBoost is the Classification And Regression Tree (CART).

(5) *LightGBM Algorithm*

LightGBM is a gradient-boosting library developed by Microsoft for classification and regression tasks. It builds decision trees using a leaf-wise approach, enabling faster training. Its strong performance is driven by two key innovations: Gradient-Based One-Side Sampling (GOSS), which prioritizes informative data points by removing small-gradient samples, and Exclusive Feature Bundling (EFB), which reduces dimensionality by combining mutually exclusive features. Additionally, LightGBM addresses class imbalance by assigning higher weights to the minority class [13].

E. *Experiment Setup*

We used Python and its libraries, such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn, and the SHAP library for explainability and other libraries required for this study. The experiment began with exploratory data analysis, followed by feature engineering to prepare the final datasets. In collaboration with a clinician, we selected relevant features for use in the chosen machine learning models. Significant effort was dedicated to data preprocessing to ensure the creation of a high-quality dataset suitable for both ICU admission prediction and ICU length of stay estimation.

F. *Training and Evaluating Models*

We assessed the performance of the proposed model using standard evaluation metrics and analyzed the interpretability of its predictions through SHAP-based explanations

(1) Performance metrics

The selected machine learning models support both classification and regression tasks. In both cases, models were trained on 80% of the data and validated on the remaining 20%. For the classification task (ICU admission prediction), performance was evaluated using accuracy, precision, recall, F1-score, and AUC according to equations (2–5). The Receiver Operating Characteristics (ROC)-AUC curve was used for model comparison, and the best-performing model underwent hyperparameter tuning with GridSearchCV and internal validation via K-Fold cross-validation with K equals to 10. The optimized model was then used for final predictions, with its learning curve analyzed and feature importance compared to SHAP's global explanations. For the regression task (ICU length-of-stay estimation), models were evaluated using RMSE and MAE. The best model's hyperparameters were tuned before predicting ICU stays for diabetic patients. SHAP was applied to provide local explanations, illustrating why specific predictions were made.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = \frac{2*TP}{2*TP + FP + FN} \quad (5)$$

AUC and ROC-AUC curves are calculated using the probability distribution of True Positive Rate (TPR) and False Positive Rate (FPR), as indicated in Figure 5. The regression models used to estimate the LOS, were assessed using RMSE and MAE calculated using the below formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - y_p)^2} \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - y_p| \quad (7)$$

where n is the total number of samples (rows), y_i is the true value and y_p is the predicted value.

(2) Model Explanations using SHAP Methods

SHAP is a widely used post-model explanation method in XAI, using cooperative game theory to compute Shapley values that quantify each feature's contribution to model predictions. It is characterized by four key properties: Efficiency, Symmetry, Dummy, and Additivity, as outlined by Lundberg and Lee [14]. Shapley values are calculated using the following formula:

$$\phi_i(f) = \sum_{s \subseteq N \setminus i} \frac{|s|!(|N|-|S|-1)!}{N!} [f(S \cup \quad i) - f(s)] \quad (8)$$

Here, S is a subset of features excluding feature i, and N is the total number of features. $\phi_i(f)$ represents the contribution of feature i, calculated as the difference between the total prediction and the prediction without feature i.

Due to their computational complexity, Shapley values can be expensive to compute. To address this, we used TreeSHAP [15], an optimized version of SHAP for tree-based ensemble models. TreeSHAP employs path-dependent feature perturbation, reducing computational complexity to $O(TLD^2)$ in the worst case, compared to the original SHAP's complexity of $O(2^n)$, where n is the number of features, T is the number of decision trees, L is the number of leaves, and D is the depth.

III. EXPERIMENTAL RESULTS

This section presents findings on ICU admission prediction and length-of-stay estimation for diabetic patients visiting the emergency department.

A. ICU Admission Prediction

We illustrate feature correlations with ICU admission on Figure 4, highlighting the top ten predictors: Albumin (Alb), Hemoglobin (Hemog), Creatinine (Cr), Red Blood Cell count (RBC), White Blood Cell count (WBC), Blood Urea Nitrogen (BUN), Total Bilirubin (Bilir), pH, Emergency (Emerg), and Observation (Observ).

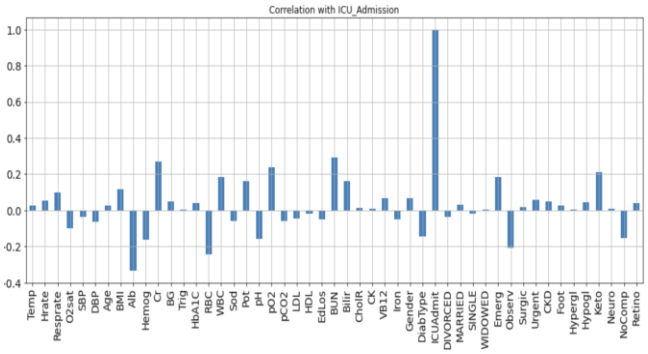

Figure 4. Features correlation to the ICU admission.

The five machine-learning models were trained on 80% of the data and tested on the remaining 20%. The experimental results demonstrate that the selected ensemble models effectively identify diabetic patients at risk of ICU admission, with all boosted tree models achieving an AUC above 0.85 (Figure. 5). CatBoost outperformed other models on the AUC metric. The dashed yellow line indicates the random guess, and its AUC is 0.5. All selected models performed well on ICU admission prediction as indicated in Table II. AdaBoost excels in most evaluation metrics.

TABLE II. MODEL PERFORMANCE ON ICU ADMISSION PREDICTION

Classifier	Accuracy	Precision	Recall	F1-score	AUC
Decision Tree	74.60	66.33	64.57	65.44	0.726
XGBoost	80.76	77.24	68.54	72.63	0.861
CatBoost	82.24	79.04	71.19	74.91	0.882
LightGBM	81.75	78.73	69.87	74.04	0.869
AdaBoost	82.74	80.92	70.20	75.18	0.876

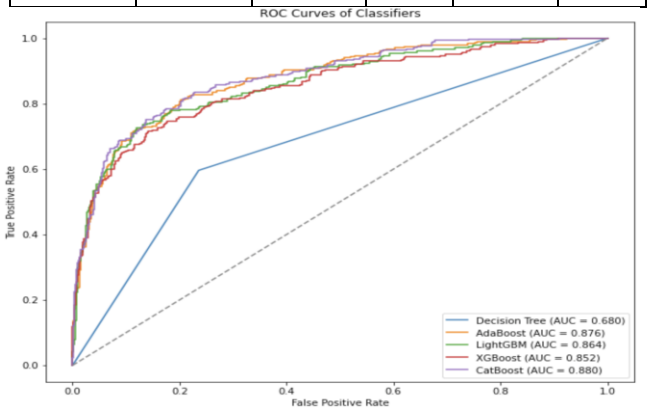

Figure 5. ROC Curve of ML models used to predict ICU admission with AUC between 0 and 1.

The learning curve shows that training and validation accuracy stabilize after 3,500 samples (Figure 6), indicating good generalization. This suggests the model reliably predicts ICU admissions.

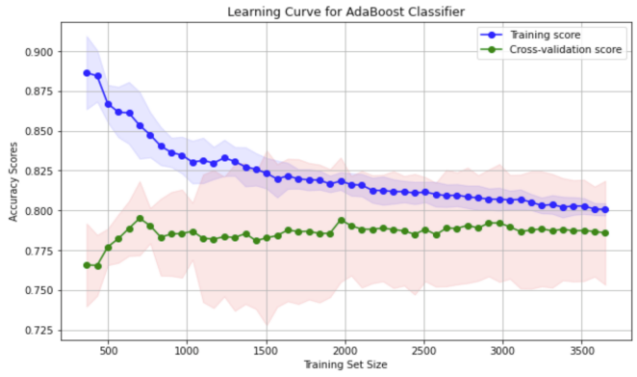

Figure 6. Accuracy of AdaBoost learning curve.

Using the internal architecture of AdaBoost and SHAP methods, we identified 25 most influential features for

patients at risk of ICU admission, as shown in Figure 7.

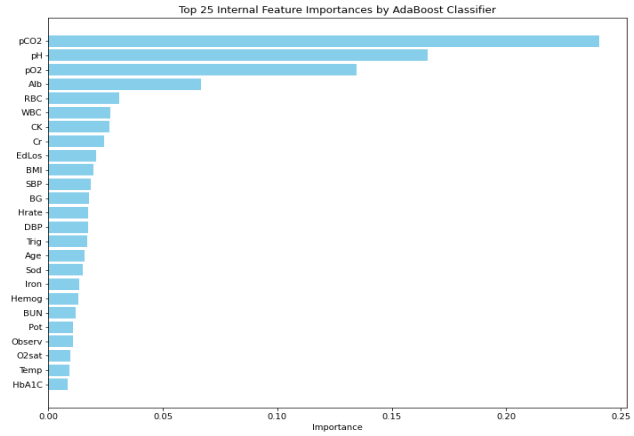

Figure 7. (a) Top 25 predictors by AdaBoost.

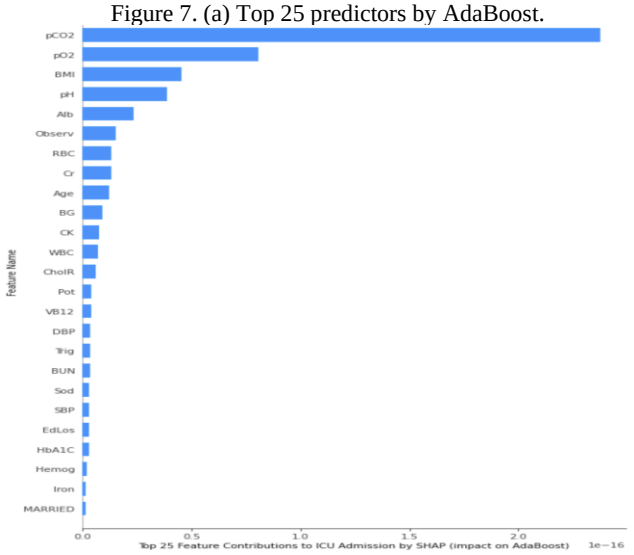

Figure 7. (b) Top 25 predictors by SHAP.

AdaBoost and SHAP identified the same four out of top five predictors of ICU admission in diabetic patients: pCO2, pH, pO_2, and Alb. The top 25 features in both AdaBoost and SHAP differed on five features: AdaBoost included O2sat and Temp, while SHAP added CholR, VB12, and MARRIED. pCO2 was further assessed as it is the top predictor of ICU admission (normal range is 35 – 45 mmHg). We found that most patients admitted to the ICU have pCO2 beyond the normal range.

We further trained the AdaBoost model using the top 25 predictors identified by SHAP. As shown in Table III, the model achieved performance comparable to that trained on the full feature set, demonstrating the effectiveness of SHAP in identifying key risk factors for ICU admission.

TABLE III. ADABOOST PREDICTIVE PERFORMANCE USING FEATURES IDENTIFIED BY THE SHAP METHOD.

Classifier	Accuracy	Precision	Recall	F1-score	AUC
AdaBoost	82.74	77.01	73.26	75.09	0.881

The clinician and SHAP agree on 25 key features associated with ICU admission, though they differ in their relevance rankings. A comparison of their rankings of top five features is summarized in Table V. Importantly, both agree on three of the top five features identified by SHAP.

TABLE IV. ANALYSIS OF SHAP EXPLANATIONS BY CLINICIAN

No.	Model's top five	Clinician's top 5	Agreement
1	pCO2	BG	No
2	pO2	pO2	Yes
3	BMI	BMI	Yes
4	pH	Cr	No
5	Alb	Alb	Yes

B. Length Of Stays (LOS) Prediction

We have used five regression ML models to estimate the length of stays among diabetic patients admitted to the hospital with the status of emergency. Table V shows the performance of the selected models on RMSE and MAE evaluation metrics. CatBoost model outperformed the other models with 2.454 and 1.695 days in ICU for RMSE and MAE on prediction of ICU LOS for diabetic patients admitted in the emergency department.

TABLE V. RSME AND MAE FOR FIVE REGRESSION MODELS WHILE ESTIMATING ICU LOS.

Regression Model	RMSE	MAE
CatBoost	2.454	1.695
LightGBM	2.739	1.937
XGBoost	2.863	1.951
Decision Tree	4.547	2.521
AdaBoost	4.864	4.537

Using SHAP, we visualized the explanations of predicted LOS of 2.62 days in ICU for a diabetic patient, as indicated by $f(x)$ on Figure 8. The ground truth is that this patient was previously hospitalized for 3.10 days in ICU.

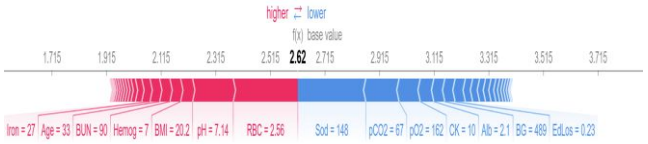

Figure 8. The SHAP force plot explaining ICU LOS for the patient on row number 51 in the test set.

SHAP was used to give details as to why this patient was predicted to stay in the ICU for 3.79 days in Figure 9. With the help of the highlighted features, we can see that features with Shapley values in red color will increase the number of days in the ICU, and features in blue will lower ICU stays.

This patient has a blood glucose of 264 and an HbA1C of 13.6, indicating poor control of blood sugar levels. In addition, the patient has pCO2 of 22 (very low) and Ketoacidosis complication, which is a major cause of prolonged stays in ICU among diabetic patients.

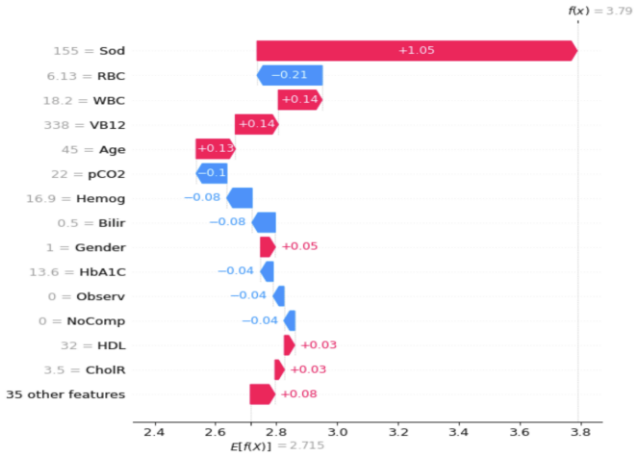

Figure 9. SHAP waterfall plot for estimating ICU stays for a patient with row 241 in the test set.

IV. Discussion

In this retrospective study, we employed supervised machine learning to predict ICU admission among diabetic patients in the emergency department and estimate their ICU Length Of Stays (LOS). We evaluated four boosted tree ensembles and one decision tree model for classification and regression tasks. Despite the class imbalance, the models performed well, with AdaBoost achieving superior performance for ICU admission prediction (Table II) and CatBoost outperforming other models in estimating ICU LOS (Table V).

AdaBoost was selected for further optimization in ICU admission prediction and fine-tuned using GridSearchCV with 10-fold cross-validation. The best hyperparameters for classification were base_estimator = DecisionTreeClassifier (max_depth=5), learning_rate = 0.005, and n_estimators = 300. For regression, CatBoost was optimized with learning_rate = 0.05, n_estimators = 100, and max_depth = 8. Model predictions were interpreted using SHAP to enhance explainability.

To validate the effectiveness of SHAP explanations, we retrained AdaBoost using only the top 25 SHAP-selected features and achieved similar performance (Table III). For LOS estimation, CatBoostRegressor outperformed other models. SHAP explanations demonstrated clinical relevance, helping interpret model decisions. Figure 9 illustrates a case where a female patient with WBC = 18.2 (above the normal range: 4.5-11), Sod = 155 (above the normal 135–145 range), Age = 45, and VB12 = 338 (150-399 pg/mL indicates low levels of VB12) had a predicted ICU stay of over 3.5 days. These predictors, alongside abnormal values for BG and HbA1C, are clinically significant, as they are associated with uncontrolled blood glucose, which is a primary cause of diabetic emergency admissions at the hospital. To ensure the relevancy of our study, a medical practitioner specializing in diabetes management confirmed that our model is fair and accurately identifies key risk factors associated with ICU admission and extended length of stay for diabetic patients and suggested its implementation in the clinical workflow.

V. Conclusion And Future Work

This study explored an explainable machine learning approach for diabetes management in emergency departments, focusing on early ICU admission prediction and length-of-stay estimation. Among five classifiers, the AdaBoost model demonstrated superior performance on three out of five metrics, as shown in Table II. CatBoost outperformed other tree-based models in regression tasks in predicting ICU stay duration as shown in Table V. SHAP analysis provided interpretability for both tasks, reinforcing the model's reliability. Our findings highlight the potential of ML integration in clinical workflows, most importantly in the emergency department since all critically ill patients start in this hospital unit, improving early ICU risk identification, optimizing hospital resource utilization, and enhancing emergency care for diabetic patients.

The experimental results successfully addressed the study's objectives: developing predictive models for ICU admission and length of stay among diabetic patients and generating interpretable explanations using SHAP. AdaBoost achieved the best performance for ICU admission prediction, while CatBoost excelled in estimating ICU length of stay. SHAP methods revealed the top 25 influential features for ICU admission prediction. A clinician with over 15 years of experience confirmed agreement with the features identified by SHAP, with minor differences in top feature rankings. Notably, the clinician appreciated that marital status was ranked lowest, aligning with its minimal clinical relevance in emergency diabetes care.

Considering the agreement and disagreement between the proposed model and clinical judgment in the Table IV, we recommend that emergency departments prioritize laboratory examinations of pCO_2, pO_2, BMI, Blood Glucose (BG), Creatinine (Cr), Albumin (Alb), and pH levels for diabetic patients presenting in emergency situations. This prioritization may enhance the accurate identification of patients at risk of ICU admission.

In future work, we aim to integrate the pretrained models into a Web application for real-world deployment in emergency departments. This will enable assessment of the model's effectiveness in identifying diabetic patients at risk of ICU admission. We also plan to evaluate the correlation between predicted and actual ICU length of stay. Finally, we will assess the impact on diabetes emergency care and the level of trust among diabetes care providers.

ETHICAL APPROVAL

MIMIC-IV is a publicly available dataset accessed through PhysioNet and does not require additional ethical approval. However, to ensure patient anonymity, identifiable features such as subject ID and item ID, which were used to extract diabetic patient data, were removed.

ACKNOWLEDGMENT

The authors express their gratitude to the University of Rwanda and Stockholm University for their support and facilitation during this study. This research was funded by the UR-Sweden Program for Research and Capacity Building Program through a scholarship to Silas Majyambere. All authors contributed equally to writing the manuscript. Silas was responsible for research design, data extraction, experiments, and drafting the write-up. Tony and Celestin supervised the study, enhanced the manuscript, and analyzed the experimental results. Gerard, leveraging his medical expertise in diabetes management for over 15 years, assisted in selecting relevant features for the dataset and assessing the clinical significance of the findings of the study.

REFERENCES

[1] A. Oyeranmi, F. Okikiola, N. Yekini, and B. Ronke, "Classification of Diabetes Types using Machine Learning", IJACSA, vol. 13, No. 9, pp. 152-161, 2022.

[2] L. Kopitar, P. Kocbek, L. Cilar, A. Sheikh, and G. Stiglic, "Early detection of type 2 diabetes mellitus using machine learning-based prediction models". Nature Scientific Reports, 2020. https://doi.org/10.1038/s41598-020-68771-z.

[3] IDF Diabetes Atlas, "International Diabetes Federation", 10th edn. Brussels, 2021, https://www.diabetesatlas.org.

[4] Y. Jian, M. Pasquier, A. Sagahyroon, and F. Aloul, "A Machine Learning Approach to Predicting Diabetes Complications", MDPI Healthcare, 2021, https://doi.org/10.3390/healthcare9121712

[5] S. Saadatmand, K. Salimifard K, R. Mohammadi, A. Kuiper, M. Marzban, and A. Farhadi, "Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients", Annals of Operation Research, 2022. doi: 10.1007/s10479-022-04984-x.

[6] Y. Hu, L. Zheng, and J. Wang, "Predicting ICU Length of Stay for Patients with Diabetes Using Machine Learning Techniques", ICCSI, pp. 417-422, 2022.

[7] T. L. Hu, C. M. Chao, C. C. Wu, T. N. Chien, and C. Li, "Machine Learning-Based Predictions of Mortality and Readmission in Type 2 Diabetes Patients in the ICU", MDPI Applied Sciences, vol. 14, issue 18, 2024, https://doi.org/10.3390/app14188443.

[8] M. D. Maniruzzaman, M. D. J. Rahman, A. Benojir, and M. D. A. Menhazul, "Classification and prediction of diabetes disease using machine learning paradigm", HISS, vol. 8, 2020, https://doi.org/10.1007/s13755-019-0095-z.

[9] H. E. Massari, N. Gherabi, F. Qanouni, and S. Mhammedi, "Diabetes Prediction Using Machine Learning with Feature Engineering and Hyperparameter Tuning", IJACSA, vol. 15, no.8, 2024.

[10] X. Dong, K. Yu, and Z. Cui, "Readmission prediction of diabetic patients based on AdaBoost-RandomForest mixed model", ICBAIE, pp. 130-134, 2022.

[11] M. R. Khurshid, S. Manzoor, T. Sadiq, L. Hussain, M. S. Khan, and A. K. Dutta, "Unveiling diabetes onset: Optimized XGBoost with Bayesian optimization for enhanced prediction". PLOS ONE, vol. 20, 2025. doi.org/10.1371/journal.pone.0310218

[12] M. Liu, C. Guo, and G. Sijia, "An explainable knowledge distillation method with XGBoost for ICU mortality prediction", Computers in Biology and Medicine, vol. 152, 2023. https://doi.org/10.1016/j.compbiomed.2022.106466.

[13] G. Ke et al., "Lightgbm: A highly efficient gradient boosting decision tree". NIPS, pp. 3149 - 3157, 2017.

[14] S. M. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions", NIPS, 2017.

[15] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees". Journal of NMI 2, pp. 56–67, 2020. https://doi.org/10.1038/s42256-019-0138-9.

[16] L. Yuxi and Q. Shaowen, "An Interpretable Machine Learning Approach for Predicting Hospital Length of Stay and Readmission", ADMA 17th International Conference, pp. 73–85, 2022. https://doi.org/10.1007/978-3-030-95405-5_6.

[17] A. Gandra, "Predicting Hospital Readmissions in Diabetes Patients: A Comparative Study of Machine Learning Models", 2024. IJHS, vol. 8, no.3, pp. 289-297, 2024. https://doi.org/10.53730/ijhs.v8n3.15189.

[18] D. M. Somya, L. Deborah, T. P. McCoy, M. Jenkins, and M. Prashanti, "Machine learning for predicting readmission risk among the frail: Explainable AI for healthcare", Patterns, vol. 3, 2021. doi:10.1016/j.patter.2021.100395.

[19] E. W. J. Alistair et al., "Author Correction: MIMIC-IV, a freely accessible electronic health record dataset." Scientific data, vol. 10, 2023. doi:10.1038/s41597-023-01945-2.

[20] J. Dugan and J. Shubrook, "International Classification of Diseases, 10th Revision, Coding for Diabetes", Clinical diabetes: a publication of the American Diabetes Association, vol. 35, no. 4, pp. 232-238, 2017. doi:10.2337/cd16-0052.