IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 18 November 2023, accepted 1 January 2024, date of publication 8 January 2024, date of current version 22 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3351188

RESEARCH ARTICLE

An Improved Ensemble Method for Predicting Hyperchloremia in Adults With Diabetic Ketoacidosis

GEORGE OBAIDO[1], (Member, IEEE), BLESSING OGBUOKIRI[2], CHIDOZIE WILLIAMS CHUKWU[3], FADEKEMI JANET OSAYE[4], OLUWASEUN FRANCIS EGBELOWO[5], MARK IZUCHUKWU UZOCHUKWU[4], IBOMOIYE DOMOR MIENYE[6], (Member, IEEE), KEHINDE ARULEBA[7], MPHO PRIMUS[8], (Member, IEEE), AND OKECHINYERE ACHILONU[9]

[1]Center for Human-Compatible Artificial Intelligence (CHAI), Berkeley Institute for Data Science (BIDS), University of California at Berkeley, Berkeley, CA 94720, USA
[2]Department of Mathematics and Statistics, York University, Toronto, ON M3J 1P3, Canada
[3]Department of Mathematics, Wake Forest University, Winston-Salem, NC 27109, USA
[4]Department of Mathematics and Computer Science, Alabama State University, Montgomery, AL 36104, USA
[5]Department of Integrative Biology, The University of Texas at Austin, Austin, TX 78712, USA
[6]Department of Information Technology, Walter Sisulu University, Buffalo City Campus, East London 5200, South Africa
[7]School of Computing and Mathematical Sciences, University of Leicester, LE1 7RH Leicester, U.K.
[8]Institute for Intelligent Systems, University of Johannesburg, Johannesburg 2006, South Africa
[9]School of Public Health, University of the Witwatersrand, Johannesburg 2193, South Africa

Corresponding author: George Obaido (gobaido@berkeley.edu)

ABSTRACT Diabetic ketoacidosis (DKA) is a serious complication that affects millions of individuals globally and presents significant health complications. *Hyperchloremia*, an electrolyte imbalance characterized by high levels of chloride in the blood, may result in gastrointestinal problems, kidney damage, and even death, especially in DKA patients. Early detection and treatment of hyperchloremia are of utmost importance in the management of DKA. This study explores the potential of the *bootstrap aggregating ensemble with random subspaces* machine learning approach to predict the occurrence of hyperchloremia, providing a basis for early intervention and improved patient outcomes. We tested our approach with the retrospective MIMIC-III database containing 1177 DKA patients and compared it with previous studies with an area under the curve (AUC) of 100%. Our approach showed significant performance outperforming other methods. The combination of this approach may enhance the early detection and timely intervention of hyperchloremia cases, ultimately leading to improved patient outcomes and a more effective management of DKA-associated complications. Our work aims to contribute to the development of decision support tools for healthcare professionals, assisting them in making informed decisions for DKA patients, with a focus on preventing and managing hyperchloremia.

INDEX TERMS Boosting aggregating or bagging classifier, diabetic ketoacidosis (DKA), hyperchloremia, machine learning, predictive modeling.

I. INTRODUCTION

Diabetes, a metabolic disorder, disrupts the regulation of blood glucose levels which can lead to both short-term and long-term health complications, and in severe cases, even death if not effectively managed [1], [2], [3], [4], [5]. This condition is broadly categorized into two types: type 1 diabetes (TY1D) and type 2 diabetes (TY2D). The pancreas plays a vital role in producing insulin and serves as a key regulator for blood sugar levels. Insulin acts as a major energy source for muscles and various tissues, facilitating the entry of blood sugar into the body's cells [6]. This process

The associate editor coordinating the review of this manuscript and approving it for publication was Ikramullah Lali.

is particularly significant in understanding both TY1D and TY2D as disruptions in insulin production or its effectiveness can lead to distinct metabolic imbalances and complications associated with each type. Both TY1D and TY2D diabetes are chronic diseases that affect millions of people worldwide and can lead to long-term health complications such as heart disease, nerve damage, and kidney failure if early detection and treatment are not initiated upon diagnosis. Diabetes is estimated to affect about 537 million adults worldwide, with a global prevalence of 10.5 percent among adults aged 20 to 79 years [7]. The number is predicted to rise to 643 million by 2030 and 783 million by 2045, while about 1.5 million deaths are directly attributed to diabetes annually [7], [8].

Diabetic ketoacidosis (DKA) is a severe complication of diabetes, a health condition characterized by inadequate insulin production or an excessive presence of ketones in the blood [9], [10], [11], [12], [13]. The diagnosis of Diabetic Ketoacidosis (DKA) involves considering values such as serum ketones at 3 mmol/L or higher, urine ketone strips at 2+ or above, serum bicarbonate less than 15 mmol/L, and blood glucose levels exceeding 11 mmol/L. Insufficient insulin prompts the body to break down fat for energy, affecting organ functions [6], [9], [14]. This process produces ketones, acidic compounds that accumulate in the bloodstream and lead to diabetic ketoacidosis (DKA) if untreated [15], [16], [17]. Symptoms of DKA include excessive thirst, frequent urination, vomiting, abdominal pain, weakness, shortness of breath, nausea, and fruity-scented breath, and can appear within 24 hours [6], [9]. DKA commonly affects individuals with TY1D or TY2D diabetes [18], [19], [20]. Treatment for DKA involves replenishing lost fluids to reduce blood sugar levels, restoring essential electrolytes for proper nerve, muscle, heart, and brain function, administering insulin injections to halt ketone production, and addressing underlying causes, such as infections, by eliminating their sources using antibiotics, antifungal, or antiviral medications [21]. The use of saline for resuscitation in DKA patients can worsen electrolyte imbalances [22], specifically leading to the development of hyperchloremia. Recent studies have shown that hyperchloremia can affect various organ systems and may contribute to adverse outcomes [23], [24], [25]. It can lead to renal vasoconstriction, decreased glomerular filtration rate, and impaired renal function. In a similar vein, hyperchloremia can also impact cardiac contractility, promote systemic inflammation, and potentially prolong the duration of DKA [23], [26], [27], [28].

Several authors have explored the prevalence of hyperchloremia in DKA management and have explored different modeling techniques to predict its likelihood. Some of these methods used a variety of models ranging from statistical models [28], [29], [30] and artificial intelligence and machine learning techniques [31], [32], [33], [34]. Statistical models use observed data and assumptions for estimating and analyzing relationships, making predictions, and drawing inferences about a phenomenon of interest. A statistical approach for determining the prevalence of hyperchloremia in DKA was proposed by Toledo et al. [35]. In particular, the study employed the Wilcoxon test for comparing non-parametric quantitative outcome measures, and the Fisher's exact test for comparing categorical outcome measures. Another study by Goad et al. [29] employed Cox regression analysis to examine the relationship between peak serum chloride levels and the time to resolution of diabetic ketoacidosis (DKA) while controlling for other potential risk factors for a prolonged duration of DKA and insulin infusion. The Cox statistical analysis revealed that peak serum chloride was independently associated with a prolonged time to DKA resolution.

Artificial intelligence and machine learning (ML) tools have also been increasingly used to analyze healthcare data, including electronic health records (EHR) [36], [37], [38], [39], [40]. Machine learning has emerged as a powerful tool with significant implications for healthcare [41], [42], [43], [44]. By analyzing large-scale datasets, machine learning algorithms can uncover hidden relationships and patterns that may not be readily apparent through traditional statistical approaches. This enables the identification of novel risk factors that contribute to the development, progression, or severity of diseases. By understanding these risk factors, healthcare professionals can intervene earlier, implement preventive measures, and provide targeted interventions to improve patient outcomes. Furthermore, there has been a shift in the perception of machine learning models from being considered "black-box" approaches to ones that can provide interpretable results [45], [46], [47]. In the past, the inner workings of complex machine learning models were often challenging to interpret, hindering their adoption in critical healthcare decision-making processes [48], [49], [50]. However, recent advances in model explainability have allowed researchers to shed light on how machine learning models arrive at their predictions [46], [51], [52].

In the context of predicting hyperchloremia in DKA (diabetic ketoacidosis) patients, the use of machine learning is limited, and predictive performance remains poor due to several factors, including the complexity of the condition, the multifactorial nature of electrolyte imbalances, the variability in patient responses to treatment, the heterogeneity of patient populations, and the challenges in obtaining large, high-quality datasets that are representative of the general DKA patient population. Similarly, the interpretation of these predictive models and the factors influencing hyperchloremia in DKA patients remain largely unknown. To address this challenge, we introduce a *bootstrap aggregating ensemble with random subspaces* to diversify the models' exposure to the training data, both in terms of the samples and the features used. This method operates by generating numerous subsets from the original dataset through bootstrap sampling, allowing each model in the ensemble to train on slightly different data points. This approach aims to improve model

accuracy by reducing the likelihood of overfitting to noise in the data and by ensuring that the ensemble does not become overly dependent on any particular feature, which is critical in medical datasets where non-linear interactions between features can be important.

In this paper, we propose the use of ensemble machine learning techniques to develop a predictive model for determining the likelihood of hyperchloremia occurring during the management of DKA. By leveraging the power of the ensemble machine learning algorithms and analyzing relevant clinical data, we seek to identify key factors and patterns that can predict the occurrence of hyperchloremia in DKA patients. The primary contributions of this study are as follows:

1) Development of an ensemble predictive model: We present an ensemble-based predictive model using bagging and random subspaces, specifically tailored to address the challenge of predicting hyperchloremia in patients with DKA.
2) Identification of risk factors: We identify and analyze various potential risk factors associated with the development of hyperchloremia in DKA patients.
3) Advance machine learning applications in DKA management: This study contributes towards the advancement of machine learning applications in the field of DKA management.

The remainder of the manuscript is organized as follows: Section II describes the proposed methods for data collection, pre-processing, and the models used. Section III presents the experimental results. In Section IV, the discussion of this paper is presented. Finally, Section V concludes the study and provides future directions.

II. RELATED WORK

Many studies have applied machine learning techniques in the detection and management of hyperchloremia and DKA complications. Many of these studies utilized publicly available datasets to measure the DKA effect. For example, Liu et al. [53] utilized the logistic regression (LR), multilayer perceptron (MLP), Random Forest (RF), and eXtreme gradient boosting (XGBoost) for the predicting hyperchloremia and validating their performance with distinct demographic information, especially with race, gender and insurance subgroups in the MIMIC-III dataset. The XGBoost achieved the highest performance in terms of AUC at 0.797%. Thongprayoon et al. [31] applied unsupervised machine learning using the consensus cluster analysis (CCA) to evaluate mortality risk factors among hyperchloremic patients and categorized them into clusters. The study presented the efficacy of an unsupervised learning task, which aided in categorizing patients with varying risks of mortality. Furthermore, the study motivated towards a more focused follow-up approach for cases where mortality risk may be high. In another study, Yeh et al. [54] explored four classifiers, including MLP, RF, ridge regression (RR), and XGBoost, to predict hyperchloremia in critically ill patients. The results showed that the machine learning techniques helped clinicians effectively manage patients at high risk of hyperchloremia complications, potentially improving patient outcomes.

Peng et al. [55] developed a predictive model using neural networks (NN), LR, RF, Naive Bayes (NB), and MLP for mortality prediction in ICU-admitted patients within a 28-day period. The study's findings were consistent with prior research, emphasizing chloride as a prominent contributor to unfavorable outcomes. Fan et al. [56] applied the univariate analysis and machine learning methods using artificial neural networks (ANN), Bayesian networks (BN), chi-square automatic interaction detection (CHAID), classification and regression trees (CRT), quick unbiased efficient statistical tree (QUEST), discriminate, and ensemble (XF) to predict complication and poor glycemic control risks in TY2D. Predicting hyperchloremia is not limited to adult patient diagnosis. Eid et al. [57] examined various machine learning classifiers, including Decision Trees (DT), RF, KNN, XGBoost, Rule Induction Algorithm (CN2), and AdaBoost, to predict DKA cases in pediatric patients, achieving an AUC of 97%. Williams et al. [58] extended the Long-Short Term Memory (LSTM) to predict a 180-day risk of DKA-related hospitalization for youth with TY1D.

III. METHODOLOGY

This section aims to offer a detailed account of the dataset including the approach taken for data collection and different methods used for data preprocessing and machine learning models.

A. DATA COLLECTION

This study utilized the MIMIC-III dataset [59], a large, freely available health dataset containing information about patients admitted to the Beth Israel Medical Center between 2001 and 2012. The dataset includes information about patients' demographics, laboratory results, procedures, caregiver notes, imaging results, and mortality information, including discharge details. Our study involved a retrospective cohort of 1177 patients admitted during this period. Out of these cases, 978 (83%) were classified as `normochloremia`, while 199 (17%) were classified as `hyperchloremia`. The peak chloride levels in the blood are measured in milliequivalents per liter (mEq/L). The normal range for blood chloride is typically between 96 and 106 mEq/L; however, these values can vary slightly depending on the laboratory reference standards. A chloride level of 107 mEq/L is slightly above the normal range.

Demographic information, such as age, gender, BMI, diabetes types (TY1D and TY2D), and Apache scores were collected for each patient. Laboratory values, including chloride levels and admission creatinine levels, along with other features, were used to assess and monitor various aspects of patient health. Fluid administration and clinical outcomes were also considered to evaluate the impact of intravenous fluid therapy on patient well-being.

B. DATA PREPROCESSING

In this study, we implemented two preprocessing pipelines to ensure the data was suitable for the application of machine learning algorithms. The first preprocessing pipeline focused on handling missing data and initial feature extraction. Missing data are a common issue in healthcare datasets, which can interfere with the reliability and performance of machine learning models [60], [61]. The technique of using imputation techniques helped ensure that the imputed values were reflective of the underlying data distribution and did not introduce bias into the analysis. Given the high-dimensional nature of the clinical data, feature extraction was a crucial step in reducing dimensionality and improving the efficiency of our machine learning models. This process involved the extraction of relevant features based on their correlation with the outcome variable (Chloride).

The second preprocessing pipeline involved feature scaling and further feature extraction. Feature scaling was conducted to standardize the range of the continuous features. This process is critical when applying machine learning algorithms that are sensitive to the scale of the features, such as support vector machines (SVM) and k-nearest neighbors (KNN) [62]. The methods employed for feature scaling included standardization and normalization. After the feature scaling process, we carried out additional feature extraction to enrich the data representation and glean further pertinent information for further analysis. Owing to the data imbalance, we performed the Synthetic Minority Over-sampling Technique (SMOTE) [63], a technique used in the field of machine learning and data mining to address the class imbalance problem. We utilize the boosting ensemble method with KNN as the base learner for the ensemble task. The random subspace method was applied to randomly select subsets of features for each iteration, enhancing diversity among the ensemble members and potentially improving generalization capabilities. Further, to ensure the robustness and generalizability of the models, we employed K-fold cross-validation with 10 splits. This validation technique provides an unbiased evaluation of their performance on unseen data, which is crucial for assessing the true predictive power of the ensemble method.

The final step in our data preprocessing involved the choice of hyperparameters for training our machine learning models. The selection of optimal hyperparameters is crucial in ensuring the performance of machine learning models. In this study, we systematically explored a range of potential hyperparameters and identified the combination that yielded the best model performance. These preprocessing steps laid the groundwork for the successful application of machine learning techniques to predict the likelihood of hyperchloremia in DKA patients.

C. MACHINE LEARNING ALGORITHMS

In this section, we discussed a short review of the seven machine learning classification algorithms used for this project. These algorithms include Support Vector Machines, K-nearest Neighbour, Logistic Regression, Random Forest, Extreme Gradient Boost, and the Bootstrap Aggregating Ensemble method with Random Subspaces.

1) SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are traditional machine learning algorithms commonly employed for classification and regression tasks on numeric or text data [64]. SVMs aim to find an optimal hyperplane that separates different classes of data points or predicts a continuous output value, all while maximizing the margin between them. When provided with a training dataset containing numeric features and corresponding class labels, SVMs attempt to learn a decision boundary that best fits the data. This decision boundary, or function, is represented by a hyperplane in a high-dimensional space [65], [66]. To train an SVM model on numeric data, we followed these steps: data preprocessing, model formulation, kernel trick application, training, and prediction. The decision function for SVM classification can be represented as:

$$f(x) = \text{sign}\left(\sum_i \alpha_i y_i K(x_i, x) + b\right) \tag{1}$$

where $f(x)$ represents the predicted class label for a new data point x, α_i are the learned coefficients (support vector weights), y_i are the corresponding class labels of the training data, $K(x_i, x)$ is the kernel function, and b is the bias term. Meanwhile, the *sign* function helps determine the direction and polarity of the prediction, enabling SVMs to classify new data points based on the learned hyperplane or decision boundary.

2) K-NEAREST NEIGHBOUR

Given a dataset with numeric features and corresponding class labels, the K-nearest Neighbour (KNN) algorithm classifies a new data point by considering its K nearest neighbors based on a distance metric (e.g., Euclidean distance) [64]. The majority class among the K nearest neighbors determines the class label assigned to the new data point [64], [65], [66]. The formula for KNN classification can be represented as follows:

$$\hat{y} = \arg\max_{y_i} \left(\sum_{i=1}^{K} \delta(y_i, y)\right) \tag{2}$$

where $\hat{y}$ represents the predicted class label for the new data point, y_i denotes the class label of the i-th nearest neighbor, y is a specific class label, and $\delta(y_i, y)$ is a function that returns 1 if y_i is equal to y and 0 otherwise. In simpler terms, the formula finds the class label that appears most frequently among the K nearest neighbors. The new data point is assigned that majority class label as its predicted class.

3) LOGISTIC REGRESSION

Logistic Regression (LR) is a popular machine learning algorithm used for binary classification on numeric datasets [64]. It models the relationship between the independent variables (features) and the probability of the binary outcome using the logistic function. Given a numeric dataset with features X and corresponding binary labels y, the logistic regression algorithm estimates the parameters (θ) of a logistic function to predict the probability of the positive class ($y = 1$). The logistic function (also known as the sigmoid function) transforms the linear combination of the features and parameters into a value between 0 and 1. It is represented as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{3}$$

where $z = \theta^T X$ is the linear combination of the features and parameters. The logistic regression formula calculates the probability of the positive class as:

$$P(y = 1 \mid X; \theta) = \frac{1}{1 + e^{-\theta^T X}} \tag{4}$$

To train our logistic regression model, the parameters θ are estimated by maximizing the likelihood of the observed data. This is usually achieved through optimization algorithms such as gradient descent or Newton's method. Once the model is trained, it can predict the probability of the positive class for new data points. Subsequently, a threshold is set to convert these probabilities into binary class labels.

4) RANDOM FOREST

Random Forest (RF) is an ensemble learning algorithm that combines multiple decision trees to make predictions [40], [67]. Although RF is a combination of multiple decision trees, we explain its prediction process using an aggregated formula. The formula for the prediction can be represented as follows:

$$\hat{y} = \text{MajorityVote}(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n) \tag{5}$$

where $\hat{y}$ represents the predicted class label or numerical value for a new data point. $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are the predicted class labels or numerical values from the individual decision trees within the RF ensemble. *MajorityVote* is a function that selects the most frequent class label or computes the average among the predictions of the individual decision trees.

In essence, RF combines the predictions of multiple decision trees to make a final prediction. For classification tasks, the class label that occurs most frequently across the ensemble is selected as the predicted class label. For regression tasks, the predicted numerical values from the individual trees are averaged to obtain the final prediction.

5) EXTREME GRADIENT BOOST

Extreme Gradient Boosting (XGBoost) is a gradient-boosting algorithm commonly used for numeric datasets [37], [68], [69]. XGBoost is an ensemble learning algorithm that combines weak prediction models, typically decision trees, into a better predictive model. It iteratively trains models, learning from the mistakes of previous models to make more accurate predictions.

The formula for XGBoost can be represented as follows:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) = \sum_{k=1}^{K} w_k h_k(x_i) \tag{6}$$

where $\hat{y}_i$ represents the predicted value for a specific data point x_i. K is the number of weak prediction models (decision trees) used in the ensemble. $f_k(x_i)$ represents the prediction of the $k - th$ weak model for the data point x_i. w_k represents the weight associated with the $k - th$ weak model. $h_k(x_i)$ represents the prediction made by the $k - th$ weak model for the data point x_i. XGBoost trains weak models sequentially, with each subsequent model attempting to correct the mistakes made by the previous models. The final prediction for a data point is the sum of the predictions of all weak models weighted by their respective weights. During the training process, XGBoost minimizes a loss function that quantifies the errors made by the ensemble. Common loss functions include squared error loss for regression and logistic loss for binary classification.

6) BOOTSTRAP AGGREGATING ENSEMBLE WITH RANDOM SUBSPACES

The Bootstrap aggregating method, also known as bagging, was introduced by Leo Breiman [70] in his influential work titled "Bagging Predictors". The study described that the bagging ensemble was designed to improve the stability and accuracy of machine learning models. The method involves training multiple instances of the same learning algorithm on different subsets of the training data, then aggregating their predictions to produce a more robust and accurate model. It can be applied to both classification and regression problems [71], [72], [73]. For classification problems, it functions as a majority voting scheme, while for regression tasks, it involves averaging predictions.

To create better ensembles, Panov and Džeroski [74] proposed a combination of bagging and random subspaces to resample the training set and generate random independent bootstrap replicates, which can be applied to any base-level algorithm. Their work argued that this approach performs better than the random forest algorithm in terms of both accuracy and diversity of the constituent learners. Through empirical experiments, they demonstrated that their combined method outperforms traditional random forests on a variety of datasets, especially when handling high-dimensional data. The proposed approach is outlined in Algorithm 1.

The algorithm constructs an ensemble of models by integrating bootstrap aggregating (bagging) with the random subspace method. It takes a training dataset D, generates M different bootstrap samples each with N instances, and selects k features for each model. Each of the M base

Algorithm 1 Bootstrap Aggregating Ensemble With Random Subspaces

Data: Training dataset D, Number of base models M, Number of features to select k, Number of samples N
Result: Ensemble Model consisting of M base models
for $m \leftarrow 1$ **to** M **do**
  $D_m \leftarrow$ Randomly select N samples from D with replacement (Bootstrap sampling)
  $F_m \leftarrow$ Randomly select k features from the feature set
  Train base model B_m on D_m using only features F_m
end
Function `EnsemblePrediction(x)`**:**
  Data: Test instance x
  Result: Predicted outcome for x
  Initialize an empty list *predictions*
  for $m \leftarrow 1$ **to** M **do**
    $y_m \leftarrow$ Predict outcome for x using model B_m with features F_m
    Append y_m to *predictions*
  end
  return *Combine predictions using majority voting*

models is trained on its unique combination of instances and features. This process captures various aspects of the data and promotes model diversity. When making predictions, the algorithm aggregates individual predictions from all models using majority voting.

D. LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS

Local Interpretable Model-Agnostic Explanations (LIME) is a technique that provides local explanations for the predictions of any black-box machine learning model [75], [76]. LIME aims to explain the predictions of a model by generating interpretable explanations in the local neighborhood of a specific instance. The formula for LIME can be represented as follows:

$$\phi(x) = \arg\min_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{7}$$

where $\phi(x)$ represents the interpretable explanation for the instance x. g is an interpretable model that approximates the behavior of the black-box model within a local neighborhood. f is the black-box model to be explained. $L(f, g, \pi_x)$ is the loss function that measures the discrepancy between the predictions of the black-box model f and the interpretable model g on the instance x, weighted by the proximity measure π_x. $\Omega(g)$ is a complexity penalty term that encourages the interpretable model g to be simpler and more interpretable. LIME generates explanations by optimizing the above formula to find an interpretable model that explains the behavior of the black-box model within a local neighborhood of the instance x.

E. PERFORMANCE EVALUATION

To assess the performance of our model in terms of accuracy, speed, resource utilization, and other relevant metrics, we elaborate on common techniques and metrics employed for evaluating the performance of the machine learning models. These include hyperparameter tuning, AUC-ROC, accuracy, specificity, and sensitivity. The formulas in this part were adapted from [32], [36], and [39].

1) HYPERPARAMETER TUNING

Hyperparameters are parameters that are not learned by the model but are set before training. We assessed various hyperparameter configurations using the grid search technique to identify the optimal set for enhancing our model's performance.

2) AUC-ROC

The Area under the Receiver Operating Characteristic (ROC) Curve is a graphical representation of a binary classifier's performance at various classification thresholds. ROC plots the true positive rate (TPR) against the false positive rate (FPR). The area under the ROC curve (AUC-ROC) is commonly used as a metric to compare and evaluate different models. The True Positive Rate (TPR) and False Positive Rate (FPR) are calculated using the following formulas:

$$TPR = \frac{TP}{TP + FN} \tag{8}$$

$$FPR = \frac{FP}{FP + TN} \tag{9}$$

TP (True Positives) represents the number of correctly predicted positive instances. FN (False Negatives) represents the number of incorrectly predicted negative instances. FP (False Positives) represents the number of incorrectly predicted positive instances. TN (True Negatives) represents the number of correctly predicted negative instances.

3) ACCURACY

Accuracy measures the correctness of model predictions, representing the ratio of correct predictions to the total number of predictions. However, accuracy alone may be insufficient for evaluating model performance, particularly with imbalanced datasets. Model accuracy is calculated using the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

4) SPECIFICITY

Specificity, also known as True Negative Rate (TNR) or Selectivity, is a performance metric commonly used in binary classification problems to evaluate the model's performance [36]. We employed specificity to measure the proportion of actual negative instances correctly identified as

TABLE 1. Experimental evaluation of the models on the DKA dataset without oversampling.

Classifier	Accuracy	AUC	Sensitivity	Specificity
LR	0.83	0.84	0.80	0.83
RF	0.94	0.93	0.93	0.93
SVM	0.91	0.90	0.70	0.93
KNN	0.76	0.82	0.65	0.79
XGBoost	0.90	0.92	0.90	0.91
Bagging Ensemble with Random Subspaces	0.94	0.93	0.90	0.92

TABLE 2. Experimental evaluation of the models on the DKA dataset with oversampling.

Classifier	Accuracy	AUC	Sensitivity	Specificity
LR	0.93	0.83	0.80	0.82
RF	0.96	0.96	0.95	0.96
SVM	0.94	0.91	0.79	0.85
KNN	0.86	0.82	0.70	0.87
XGBoost	0.91	0.92	0.90	0.92
Bagging Ensemble with Random Subspaces	1.00	1.00	0.99	0.99

negative by the model. It also quantifies the model's ability to correctly identify the negative class, with high specificity indicating a low rate of false positives. Equation 11 outlines the formula for calculating the model's specificity.

$$Specificity = \frac{TN}{TN + FP} \tag{11}$$

5) SENSITIVITY

In binary classification, sensitivity, also known as True Positive Rate (TPR) or Recall, measures the proportion of actual positive instances correctly identified as positive by the model. Sensitivity quantifies the model's ability to correctly identify the positive class, with high sensitivity indicating a low rate of false negatives. The model's sensitivity is calculated using Equation 12

$$Sensitivity = \frac{TP}{TP + FN} \tag{12}$$

IV. RESULTS

The experimental results obtained from various classifiers are presented in Tables 1 and 2, showcasing the performance measures on the DKA dataset. These results were obtained using the k-fold cross-validation procedure to ensure a reliable evaluation of the classifiers. The hyperparameters of the classifiers were carefully tuned to achieve optimal performance.

Upon examining the results in Table 1, where the models were trained without the SMOTE oversampling, both the Random Forest model and the Bagging method achieved similar accuracies of 94%. These were closely followed by the XGBoost classifier with an accuracy of 90%, Logistic Regression at 83%, and KNN at 76%. We believe this performance disparity could be largely attributed to the imbalance in the dataset.

Table 2 presents the experimental results after implementing the SMOTE oversampling technique. We observed an overall improvement in performance: the Bagging classifier outperformed others with an accuracy of 100%, followed closely by the random forest at 96%, the support vector machine at 94%, logistic regression at 93%, and KNN at 86%. It is evident from these results that the application of oversampling techniques like SMOTE can be crucial in scenarios where dataset imbalance is a significant challenge.

In evaluating the sensitivity scores, we analyzed the performance of each classifier both before and after implementing the resampling method, as depicted in Figure 1.

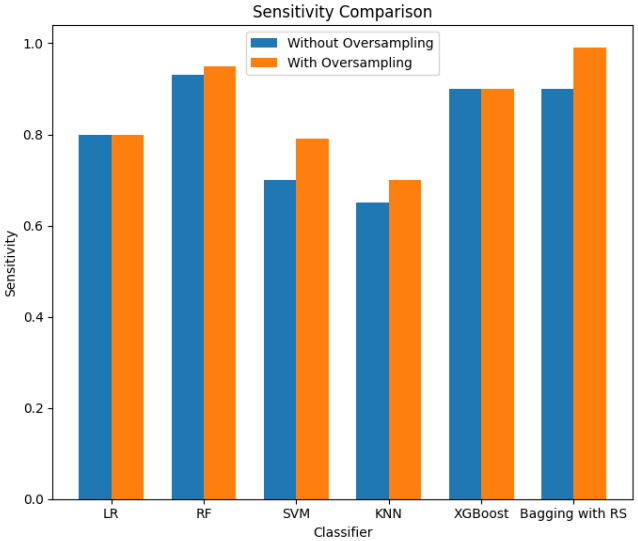


FIGURE 1. Sensitivity before and after oversampling.

It was observed that our proposed method exhibited a notable improvement, increasing from 90% to 99%. This was followed by the random forest classifier, which showed a modest rise from 93% to 95%. The XGBoost classifier remained stable at 90% across both scenarios. Similarly, specificity scores also demonstrated improvements across various classifiers, as illustrated in Figure 2. Our proposed method showed a significant increase from 92% to 99%, closely followed by the random forest and XGBoost classifiers. However, the support vector machine exhibited a considerable decrease, dropping from 93% to 85%.

We assessed the overall efficacy of the classifiers using ROC-AUC scores, which illustrate the balance between the true positive rate (sensitivity) and the false positive rate (1 - specificity). Figures 3 and 4 demonstrate this trade-off. In the scenario without oversampling, as shown in Figure 3, both the random forest and our proposed method achieved commendable performance, each scoring 93%. They were closely followed by XGBoost with 92% and the support vector machine with 90%. Upon implementing oversampling, as presented in Figure 4, there was a notable enhancement in performance. Our proposed methodology excelled with a perfect score of 100%, with the random forest classifier reaching 96%. The SVM also showed improvement at 91%. In contrast, the logistic regression experienced a slight decline, dropping from 84% to 83%, while the performance of the KNN remained consistent at 82%.

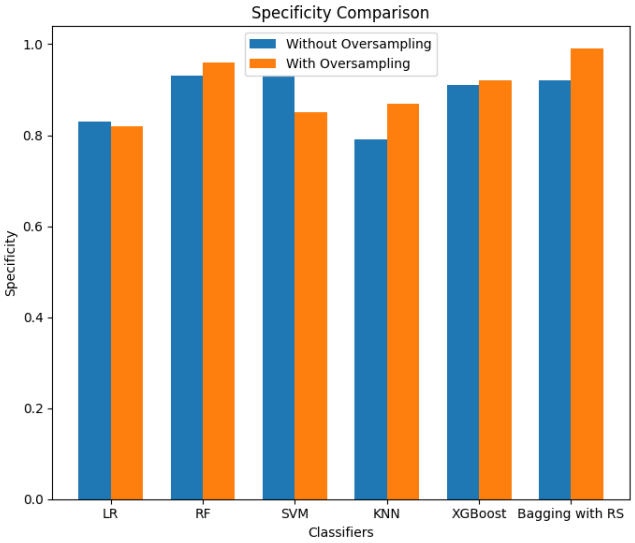


FIGURE 2. Specificity before and after oversampling.

These results underscore the effectiveness of the proposed method and other classifiers in identifying instances of hyperchloremia and normochloremia in the dataset, demonstrating their potential for clinical applications in DKA management.

A. MODEL AGNOSTIC METHOD

In this section, we use the LIME model-agnostic method to interpret hyperchloremia predictions made by the machine learning model applied to the DKA dataset. With LIME, our goal is to leverage the predictive capabilities of the model and gain valuable insights into its inner workings.

We initiate the process by training the bagging ensemble classifier on the DKA dataset, known for its proficiency in handling complex datasets and providing accurate predictions. Subsequently, we utilize the LIME explainer, specifically designed for tabular data. This explainer employs a perturbation-based approach to generate local, interpretable models that approximate the behavior of the Random Forest classifier near each instance. Following the fitting of the classifier to the dataset, we interpret its predictions using LIME. For each instance, we generate an explanation highlighting the most influential features in the decision-making process. These explanations are then visualized in Figures 5 and 6, demonstrating the instance-level interpretability achieved through LIME.

When the bar is indicated to the left (blue) of the prediction, it signifies that the corresponding feature negatively impacts the prediction, decreasing the probability or likelihood of the positive class. In other words, the feature represented by the bar contributes to the model predicting the negative class or a lower probability of the positive class. Conversely, when the bar is on the right (orange) of the prediction, it implies that the feature positively impacts the prediction, increasing the probability or likelihood of the positive class.

In Figure 5, the analysis reveals a 95% probability of normochloremia when the Chloride level is between 0.27 and 0.05, contributing 16% to the prediction. Additionally, blood sodium greater than -0.54 contributes 4% to the prediction, Bicarbonate exceeding 0.56 contributes 2%, coronary heart disease (CHD) with no myocardial infarction (MI) greater than -0.31 contributes up to 2%, and Mean Corpuscular Hemoglobin Concentration (MCHC) greater than 70% contributes 1%. These factors collectively contribute to classifying the chloride level in the blood within the normal range. In contrast, Figure 6 presents a different scenario where the instance is classified as hyperchloremia with a predictive probability of 75%. This classification is influenced by specific conditions: Chloride value exceeding 0.59 contributes 49% to the prediction, Bicarbonate less than or equal to -0.64 contributes 8%, Blood calcium less than or equal to -0.62, partial pressure of carbon dioxide in the blood less than -0.59 contributes 2%, and temperature less than or equal to -0.64. When these conditions hold, the model predicts hyperchloremia as the likely classification.

These findings illustrate how various combinations of features influence the classification outcome, underscoring the significance of individual features in determining whether an instance is classified as normochloremia or hyperchloremia. The observations depicted in Figures 5 and 6 offer valuable insights into the model's decision-making process, contributing to a more comprehensive understanding of the factors driving the classification results.

V. DISCUSSION

Hyperchloremia is a common complication in the management of DKA, carrying significant implications for patient outcomes [77]. The ability to predict the likelihood of its occurrence, as demonstrated in this study, can lead to more informed and proactive management strategies, potentially reducing the incidence and improving patient outcomes. Our research has illuminated the complex interplay between various factors in the onset and management of DKA, with a particular focus on hyperchloremia. Through the application of machine learning techniques, we have gained a better understanding of these relationships and developed a promising predictive and explainable method for hyperchloremia.

One significant outcome of our correlation analysis was the negative association between age and DKA. This association is well-documented in the literature, where DKA can affect both younger and older individuals [56], [57]. This finding has profound implications for the management of DKA, suggesting that age is not a determinant of DKA manifestation, despite the higher mortality rate observed in older patients [78], [79].

Our study also revealed that chloride is a crucial marker for DKA. This finding holds significant implications for predicting and managing hyperchloremia in DKA patients, indicating that individuals with elevated chloride levels upon admission are more likely to experience higher peak chloride levels during treatment. This aligns with a recent study that emphasized the importance of early chloride monitoring in DKA management [29].

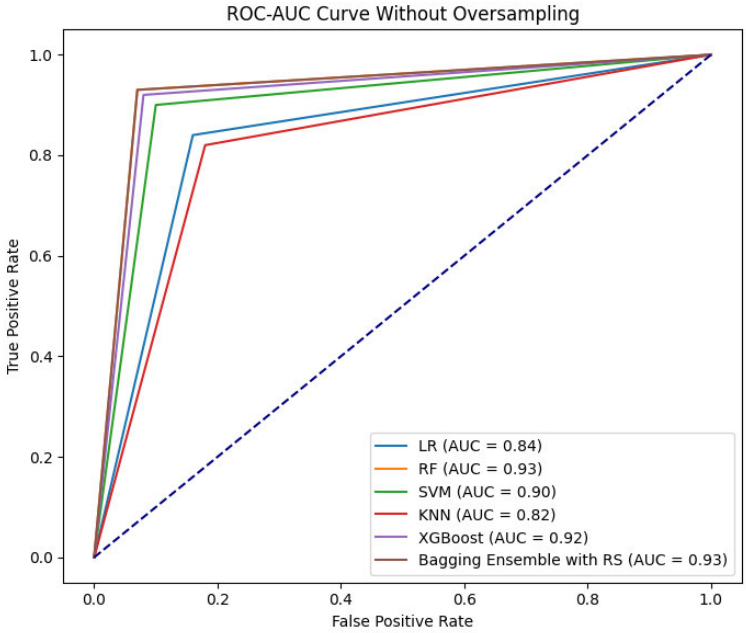


FIGURE 3. ROC-AUC of the classifiers without oversampling.

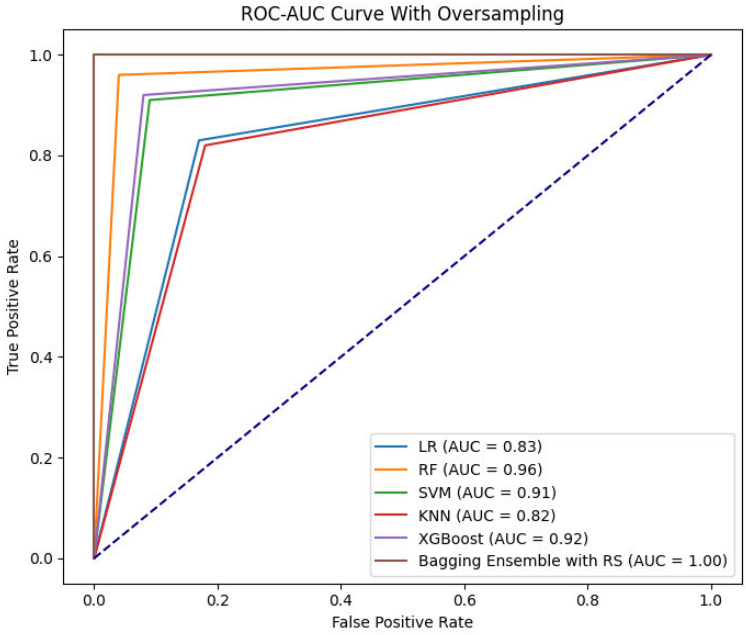


FIGURE 4. ROC-AUC of the classifiers with oversampling.

Regarding predictive models, the utilization of various machine learning classifiers has underscored their potential in predicting hyperchloremia with varying levels of accuracy. Notably, our proposed method, along with random forest and extreme gradient boost classifiers, exhibited superior performance. These results align with a growing body of evidence suggesting that machine learning holds immense potential in improving clinical decision-making and disease prediction [37], [39], [80]. For the interpretability of our model, we employed LIME, an approach that provides interpretable and faithful explanations of machine learning classifiers. LIME revealed the complex interactions of features within our model, offering valuable insights into the factors contributing to the prediction of hyperchloremia.

A. COMPARISON WITH OTHER STUDIES

In Table 3, we compare our proposed study work with a few state-of-the-art machine learning models and dataset utilized in hyperchloremia and DKA studies and their use of

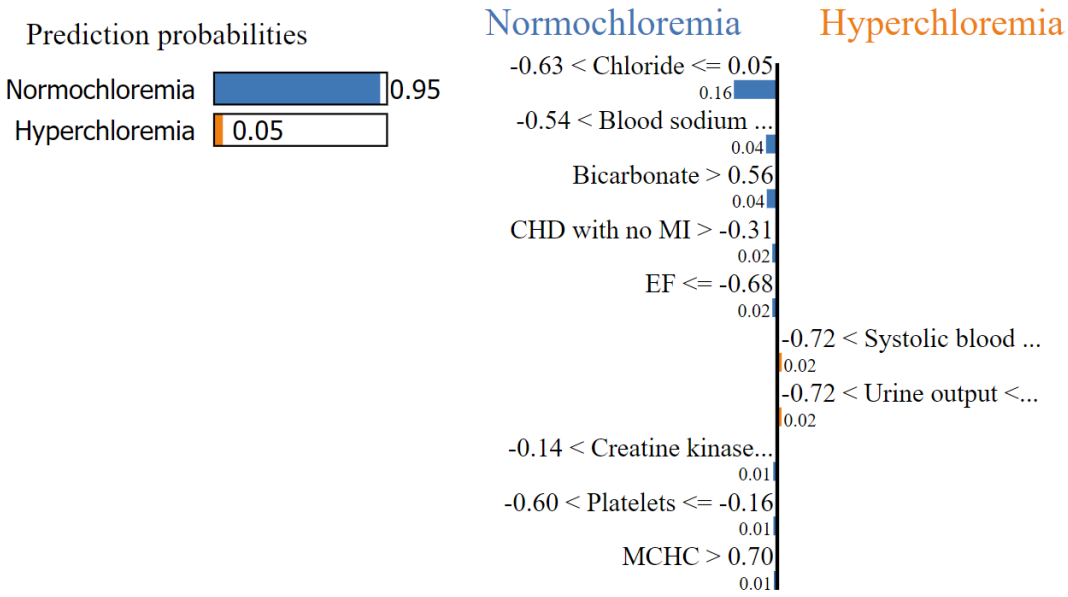


FIGURE 5. An instance showing Normachloremia.

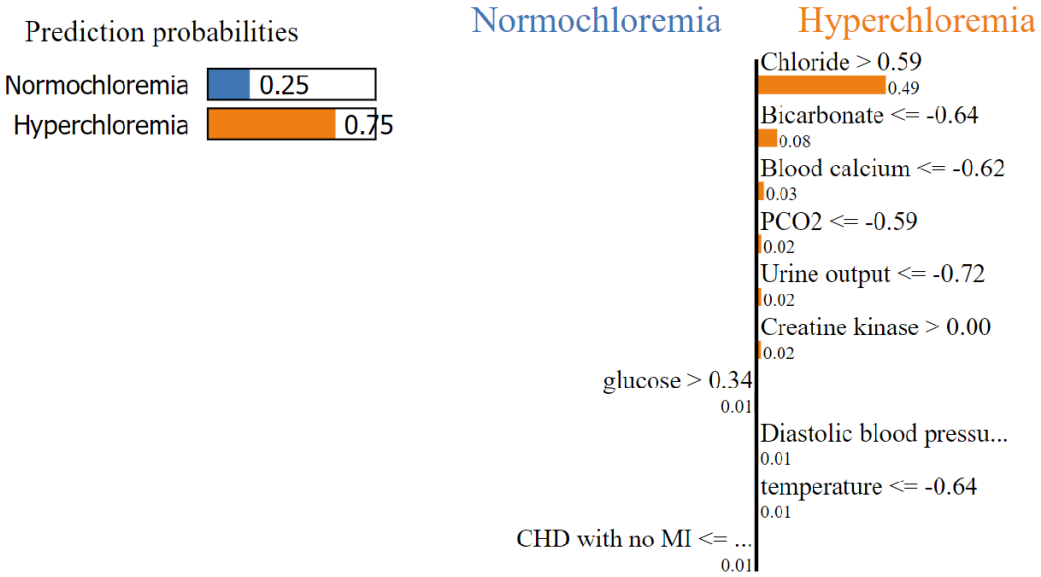


FIGURE 6. An instance showing Hyperchloremia.

explainability feature. Liu et al. [53] utilized LR, multi-Layer perceptron (MLP), RF, and XGBoost and achieved an AUC of 0.90 with explainability features. Thongprayoon et al. [31] used consensus cluster analysis (CCA) without providing sensitivity, specificity, or AUC values and did not incorporate an explainer. In contrast, Yeh et al. [54] implemented MLP, RF, ridge regression (RR), and XGBoost with a sensitivity of 0.67 and an AUC of 0.19 but without model explainability. Peng et al. [55] used neural networks (NN), LR, Naive Bayes (NB), RF, and MLP, and achieved a sensitivity of 0.18, a perfect specificity of 1.00, and an AUC of 0.76. This model also included explainability features. Fan et al. [56] used artificial neural networks (ANN), Bayesian networks (BN), chi-square automatic interaction detection (CHAID), classification and regression trees (CRT), quick unbiased efficient statistical tree (QUEST), discriminate (D), and ensemble (XF), achieving a sensitivity of 0.87 and an AUC of 0.90, but did not include an explainer. Eid et al. [57] used Decision Trees (DT), RF, KNN, XGBoost, Rule Induction Algorithm (CN2), and AdaBoost, obtaining an AUC of 0.97, but did not provide sensitivity or specificity values and lacked explainability features.

The bagging ensemble with random subspaces technique outperformed all the aforementioned techniques. While it would be beneficial to observe how the proposed approach performs on multiple datasets, our method demonstrated significant improvement after employing the oversampling technique.

TABLE 3. Comparison with some existing methods.

Reference	Dataset	Method	Sensitivity	Specificity	AUC	Explainer
Liu et al. [53]	MIMIC III	LR, MLP, RF, XGBoost	-	-	0.90	✓
Thongprayoon et al. [31]	Private	CCA	-	-	-	
Yeh et al. [54]	MIMIC-III	MLP, RF, RR, XGBoost	0.67	-	0.19	
Peng et al. [55]	MIMIC-IV	NN, LR, NB, RF, MLP	0.18	1.00	0.76	✓
Fan et al. [56]	Private	ANN, BN, CHAID, CRT, QUEST, D, XF	0.87	-	0.90	
Eid et al. [57]	Private	DT, RF, kNN, GB, CN2, AdaBoost	-	-	0.97	
Fan et al. [81]	MIMIC-IV	XGBoost, LR, LGM, GBM, CNB, MLP, SVM	0.77	0.82	0.83	
Li et al. [32]	Optum EHR	LR, LASSO, XGBoost, DRF, FFN	0.83	0.41	0.95	
Our paper	MIMIC III	Bagging ensemble with random subspaces	0.99	0.99	1.00	✓

VI. CONCLUSION

The application of machine learning methodologies in managing diabetic ketoacidosis (DKA) presents a promising approach to predicting hyperchloremia, a condition that can exacerbate DKA and lead to undesirable outcomes. By examining an extensive array of clinical data, our research has highlighted crucial factors and patterns that can aid in predicting the occurrence of hyperchloremia in DKA patients. These insights have the potential to guide clinicians in promptly detecting and managing hyperchloremia, thereby enhancing patient outcomes. Furthermore, LIME, the interpretability technique, has proven valuable in understanding the predictive power of the machine learning model we developed.

Despite the promising results, several opportunities for future research exist. Firstly, enhancing the performance of our machine learning model could be achieved through the integration of more patient data and the exploration of additional predictive algorithms. Secondly, the implementation of our predictive model within a clinical environment requires further scrutiny, including the seamless integration of the model into electronic health record systems. Thirdly, we propose that future research endeavors expand the model to forecast additional complications associated with DKA. Lastly, we advocate for more extensive research into the practical application of machine learning interpretability techniques in the healthcare sector. Such advancements could strengthen confidence in and the adoption of machine learning models in clinical decision-making processes.

ACKNOWLEDGMENT

George Obaido wishes to express gratitude to the reviewers for their insightful comments and to the Center for Human-Compatible Artificial Intelligence for covering the publication fee.

REFERENCES

[1] A. Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, and G. Hartvigsen, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes," *Artif. Intell. Med.*, vol. 98, pp. 109–134, Jul. 2019.

[2] D. Dabelea, E. J. Mayer-Davis, S. Saydah, G. Imperatore, B. Linder, J. Divers, R. Bell, A. Badaru, J. W. Talton, T. Crume, A. D. Liese, A. T. Merchant, J. M. Lawrence, K. Reynolds, L. Dolan, L. L. Liu, and R. F. Hamman, "Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009," *JAMA*, vol. 311, no. 17, p. 1778, May 2014.

[3] M. Janghorbani, R. M. Van Dam, W. C. Willett, and F. B. Hu, "Systematic review of type 1 and type 2 diabetes mellitus and risk of fracture," *Amer. J. Epidemiol.*, vol. 166, no. 5, pp. 495–505, Jun. 2007.

[4] M. Cnop, N. Welsh, J.-C. Jonas, A. Jorns, S. Lenzen, and D. L. Eizirik, "Mechanisms of pancreatic β-cell death in type 1 and type 2 diabetes: Many differences, few similarities," *Diabetes*, vol. 54, no. 2, pp. S97–S107, 2005.

[5] C. A. Newton and P. Raskin, "Diabetic ketoacidosis in type 1 and type 2 diabetes mellitus: Clinical and biochemical differences," *Arch. Internal Med.*, vol. 164, no. 17, pp. 1925–1931, 2004.

[6] Mayo Clinic. (2022). *Diabetic Ketoacidosis*. Accessed: May 22, 2023. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/diabetic-ketoacidosis/symptoms-causes/syc-20371551

[7] IDF Diabetes Atlas. (2023). *Diabetes Around the World in 2021*. Accessed: May 22, 2023. [Online]. Available: https://diabetesatlas.org/

[8] World Health Organization. (May 24, 2023). *Diabetes*. [Online]. Available: https://www.who.int/health-topics/diabetes#tab=tab_1

[9] K. K. Dhatariya, N. S. Glaser, E. Codner, and G. E. Umpierrez, "Diabetic ketoacidosis," *Nature Rev. Disease Primers*, vol. 6, no. 1, p. 40, 2020.

[10] S. F. Farsani, K. Brodovicz, N. Soleymanlou, J. Marquard, E. Wissinger, and B. A. Maiese, "Incidence and prevalence of diabetic ketoacidosis (DKA) among adults with type 1 diabetes mellitus (T1D): A systematic literature review," *BMJ Open*, vol. 7, no. 7, Jul. 2017, Art. no. e016587.

[11] T. T. Alamuri, S. Mahesh, K. Dell'Aquila, T. J. Leong, R. Jennings, and T. Q. Duong, "COVID-19 associated ketosis and diabetic ketoacidosis: A rapid review," *Diabetes, Obesity Metabolism*, vol. 25, pp. 1785–1793, 2023.

[12] W. H. Albuali, A. A. Yousef, M. H. Al-Qahtani, F. O. AlQurashi, H. W. Albuali, H. A. Yousef, A. A. Aldajani, M. A. A. Ghamdi, and B. N. AlBassam, "A clinical and biochemical comparative study of diabetic ketoacidosis (DKA) in newly diagnosed vs known cases of type 1 diabetic children," *Rev. Diabetic Stud.*, vol. 19, no. 1, pp. 28–33, Mar. 2023.

[13] K. Evans, "Diabetic ketoacidosis: Update on management," *Clin. Med.*, vol. 19, no. 5, pp. 396–398, Sep. 2019.

[14] I. P. Kaur, S. Inkollu, A. Prakash, H. Gandhi, M. S. Mughal, and D. Du, "Pantoea agglomerans bacteremia: Is it dangerous?" *Case Rep. Infectious Diseases*, vol. 2020, pp. 1–4, Apr. 2020.

[15] R. L. Layell and K. A. Lane, "Systematically treating the extra sweet pediatric patient, without inadvertently giving them a big head," *Air Med. J.*, vol. 42, no. 4, pp. 268–270, Jul. 2023.

[16] A. Ahsan, A. Khan, M. A. Farooq, M. Naveed, M. M. F. A. Baig, and W.-X. Tian, "Physiology of endocrine system and related metabolic disorders," in *Endocrine Disrupting Chemicals-Induced Metabolic Disorders and Treatment Strategies*. Springer, 2021, pp. 3–41.

[17] A. Bedaso, Z. Oltaye, E. Geja, and M. Ayalew, "Diabetic ketoacidosis among adult patients with diabetes mellitus admitted to emergency unit of Hawassa University comprehensive specialized hospital," *BMC Res. Notes*, vol. 12, no. 1, pp. 1–5, Dec. 2019.

[18] E. Ooi, K. Nash, L. Rengarajan, E. Melson, L. Thomas, A. Johnson, D. Zhou, L. Wallett, S. Ghosh, P. Narendran, and P. Kempegowda, "Clinical and biochemical profile of 786 sequential episodes of diabetic ketoacidosis in adults with type 1 and type 2 diabetes mellitus," *BMJ Open Diabetes Res. Care*, vol. 9, no. 2, Dec. 2021, Art. no. e002451.

[19] V. W. Zhong, J. Juhaeri, and E. J. Mayer-Davis, "Trends in hospital admission for diabetic ketoacidosis in adults with type 1 and type 2 diabetes in England, 1998–2013: A retrospective cohort study," *Diabetes Care*, vol. 41, no. 9, pp. 1870–1877, Sep. 2018.

[20] L. Barski, R. Nevzorov, A. Jotkowitz, E. Rabaev, M. Zektser, L. Zeller, E. Shleyfer, I. Harman-Boehm, and Y. Almog, "Comparison of diabetic ketoacidosis in patients with type-1 and type-2 diabetes mellitus," *Amer. J. Med. Sci.*, vol. 345, no. 4, pp. 326–330, 2013.

[21] D. P. Westerberg, "Diabetic ketoacidosis: Evaluation and treatment," *Amer. Family Physician*, vol. 87, no. 5, pp. 337–346, 2013.

[22] M. Jayashree, V. Williams, and R. Iyer, "Fluid therapy for pediatric patients with diabetic ketoacidosis: Current Perspectives," *Diabetes, Metabolic Syndrome Obesity, Targets Therapy*, vol. 12, pp. 2355–2361, Nov. 2019.

[23] J. A. Kraut and N. E. Madias, "Metabolic acidosis: Pathophysiology, diagnosis and management," *Nature Rev. Nephrol.*, vol. 6, no. 5, pp. 274–285, May 2010.

[24] G. Bandak and K. B. Kashani, "Chloride in intensive care units: A key electrolyte," *FResearch*, vol. 6, p. 1930, Nov. 2017.

[25] S. A. McCluskey, K. Karkouti, D. Wijeysundera, L. Minkovich, G. Tait, and W. S. Beattie, "Hyperchloremia after noncardiac surgery is independently associated with increased morbidity and mortality: A propensity-matched cohort study," *Anesthesia Analgesia*, vol. 117, no. 2, pp. 412–421, 2013.

[26] J. M. Handy and N. Soni, "Physiological effects of hyperchloraemia and acidosis," *Brit. J. Anaesthesia*, vol. 101, no. 2, pp. 141–150, Aug. 2008.

[27] L. Castellanos, M. Tuffaha, D. Koren, and L. L. Levitsky, "Management of diabetic ketoacidosis in children and adolescents with type 1 diabetes mellitus," *Pediatric Drugs*, vol. 22, no. 4, pp. 357–367, Aug. 2020.

[28] H. M. Ahmed, H. R. H. Elnaby, R. M. A. El Kareem, and M. Hodeib, "The relationship between hyperchloremia and acute kidney injury in pediatric diabetic ketoacidosis and its impact on clinical outcomes," *Pediatric Nephrol.*, vol. 37, no. 6, pp. 1407–1413, Jun. 2022.

[29] N. T. Goad, R. N. Bakhru, J. L. Pirkle, and M. T. Kenes, "Association of hyperchloremia with unfavorable clinical outcomes in adults with diabetic ketoacidosis," *J. Intensive Care Med.*, vol. 35, no. 11, pp. 1307–1313, Nov. 2020.

[30] J. P. Ferreira, M. Hamui, M. Torrents, R. Carrano, M. Ferraro, and I. Toledo, "The influence of chloride for the interpretation of plasma bicarbonate during the treatment of diabetic ketoacidosis," *Pediatric Emergency Care*, vol. 36, no. 3, pp. e143–e145, 2020.

[31] C. Thongprayoon, V. Nissaisorakarn, P. Pattharanitima, M. A. Mao, A. G. Kattah, M. T. Keddis, C. Y. Dumancas, S. Vallabhajosyula, T. Petnak, S. B. Erickson, J. J. Dillon, V. D. Garovic, K. B. Kashani, and W. Cheungpasitporn, "Subtyping hyperchloremia among hospitalized patients by machine learning consensus clustering," *Medicina*, vol. 57, no. 9, p. 903, Aug. 2021.

[32] L. Li, C. Lee, F. L. Zhou, C. Molony, Z. Doder, E. Zalmover, K. Sharma, J. Juhaeri, and C. Wu, "Performance assessment of different machine learning approaches in predicting diabetic ketoacidosis in adults with type 1 diabetes using electronic health records data," *Pharmacoepidemiol. Drug Saf.*, vol. 30, no. 5, pp. 610–618, May 2021.

[33] R. Cheheltani, N. King, S. Lee, B. North, D. Kovarik, C. Evans-Molina, N. Leavitt, and S. Dutta, "Predicting misdiagnosed adult-onset type 1 diabetes using machine learning," *Diabetes Res. Clin. Pract.*, vol. 191, Sep. 2022, Art. no. 110029.

[34] P. Yeh, Y. Pan, L. N. Sanchez-Pinto, and Y. Luo, "Using machine learning to predict hyperchloremia in critically ill patients," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 1703–1707.

[35] I. Toledo, R. Wainsztein, C. Mannucci, M. Ferraro, J. Ferreira, and A. Balestracci, "Impact of the hyperchloremic component of metabolic acidosis on the patient's hydration status and the treatment of diabetic ketoacidosis," *Arch Argent Pediatr*, vol. 116, no. 3, pp. e365–e370, 2018.

[36] G. Obaido, B. Ogbuokiri, T. G. Swart, N. Ayawei, S. M. Kasongo, K. Aruleba, I. D. Mienye, I. Aruleba, W. Chukwu, F. Osaye, O. F. Egbelowo, S. Simphiwe, and E. Esenogho, "An interpretable machine learning approach for hepatitis B diagnosis," *Appl. Sci.*, vol. 12, no. 21, p. 11127, Nov. 2022.

[37] G. Obaido, B. Ogbuokiri, I. D. Mienye, and S. M. Kasongo, "A voting classifier for mortality prediction post-thoracic surgery," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, 2022, pp. 263–272.

[38] R. T. Aruleba, T. A. Adekiya, N. Ayawei, G. Obaido, K. Aruleba, I. D. Mienye, I. Aruleba, and B. Ogbuokiri, "COVID-19 diagnosis: A review of rapid antigen, RT-PCR and artificial intelligence methods," *Bioengineering*, vol. 9, no. 4, p. 153, Apr. 2022.

[39] I. D. Mienye, G. Obaido, K. Aruleba, and O. A. Dada, "Enhanced prediction of chronic kidney disease using feature selection and boosted classifiers," in *Proc. Int. Conf. Intell. Syst. Design Appl.* Cham, Switzerland: Springer, 2021, pp. 527–537.

[40] I. D. Mienye and Y. Sun, "Heart disease prediction using enhanced machine learning techniques," in *Intelligent Systems and Machine Learning for Industry*. Boca Raton, FL, USA: CRC Press, 2022, pp. 93–114.

[41] S. Lee, N. Mohr, N. Street, and P. Nadkarni, "Machine learning in relation to emergency medicine clinical and operational scenarios: An overview," *Western J. Emergency Med.*, vol. 20, no. 2, pp. 219–227, Feb. 2019.

[42] A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," *JAMA*, vol. 319, no. 13, p. 1317, Apr. 2018.

[43] Y. K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, and A. Eirug, "Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manage.*, vol. 57, Jan. 2021, Art. no. 101994.

[44] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 156–180, 2021.

[45] G. S. Bullock, T. Hughes, A. H. Arundale, P. Ward, G. S. Collins, and S. Kluzek, "Black box prediction methods in sports medicine deserve a red card for reckless practice: A change of tactics is needed to advance athlete care," *Sports Med.*, vol. 52, no. 8, pp. 1729–1735, Aug. 2022.

[46] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[47] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020.

[48] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: A multidisciplinary perspective," *BMC Med. Informat. Decis. Making*, vol. 20, no. 1, pp. 1–9, Dec. 2020.

[49] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[50] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman, "Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities," *Inf. Fusion*, vol. 50, pp. 71–91, Oct. 2019.

[51] M. Honegger, "Shedding light on black box machine learning algorithms: Development of an axiomatic framework to assess the quality of methods that explain individual predictions," 2018, *arXiv:1808.05054*.

[52] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, "Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review," *Appl. Sci.*, vol. 11, no. 11, p. 5088, May 2021.

[53] S. Liu and Y. Luo, "Assessing social determinants-related performance bias of machine learning models: A case of hyperchloremia prediction in ICU population," 2021, *arXiv:2111.09507*.

[54] P. Yeh, Y. Pan, L. N. Sanchez-Pinto, and Y. Luo, "Hyperchloremia in critically ill patients: Association with outcomes and prediction using electronic health record data," *BMC Med. Informat. Decis. Making*, vol. 20, no. S14, pp. 1–10, Dec. 2020.

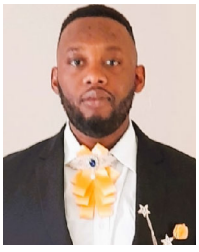
[55] S. Peng, J. Huang, X. Liu, J. Deng, C. Sun, J. Tang, H. Chen, W. Cao, W. Wang, X. Duan, X. Luo, and S. Peng, "Interpretable machine learning for 28-day all-cause in-hospital mortality prediction in critically ill patients with heart failure combined with hypertension: A retrospective cohort study based on medical information mart for intensive care database-IV and eICU databases," *Frontiers Cardiovascular Med.*, vol. 9, Oct. 2022, Art. no. 994359.

[56] Y. Fan, E. Long, L. Cai, Q. Cao, X. Wu, and R. Tong, "Machine learning approaches to predict risks of diabetic complications and poor glycemic control in nonadherent type 2 diabetes," *Frontiers Pharmacol.*, vol. 12, Jun. 2021, Art. no. 665951.

[57] W. M. Eid, H. Alharthi, N. Aslam, I. U. Abdur Rab, and A. Madani, "Predicting diabetic ketoacidosis in pediatric patients using machine learning," *FResearch*, vol. 12, p. 611, Jun. 2023.
[58] D. D. Williams, D. Ferro, C. Mullaney, L. Skrabonja, M. S. Barnes, S. R. Patton, B. Lockee, E. M. Tallon, C. A. Vandervelden, and C. Schweisberger, "An 'all-data-on-hand' deep learning model to predict hospitalization for diabetic ketoacidosis in youth with type 1 diabetes: Development and validation study," *JMIR Diabetes*, vol. 8, Mar. 2023, Art. no. e47592.
[59] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W.-H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, May 2016.
[60] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
[61] M. Ghassemi, T. Naumann, P. Schulam, A. L. Beam, I. Y. Chen, and R. Ranganath, "A review of challenges and opportunities in machine learning for health," *AMIA Summits Transl. Sci.*, vol. 2020, p. 191, Jan. 2020.
[62] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Syst. Appl.*, vol. 80, pp. 340–355, Sep. 2017.
[63] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.
[64] B. Ogbuokiri, A. Ahmadi, Z. M. Nia, B. Mellado, J. Wu, J. Orbinski, A. Asgary, and J. Kong, "Vaccine hesitancy hotspots in Africa: An insight from geotagged Twitter posts," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 1, pp. 1–14, Jan. 2023.
[65] B. Ogbuokiri, A. Ahmadi, N. L. Bragazzi, Z. Movahedi Nia, B. Mellado, J. Wu, J. Orbinski, A. Asgary, and J. Kong, "Public sentiments toward COVID-19 vaccines in South African cities: An analysis of Twitter posts," *Frontiers Public Health*, vol. 10, Aug. 2022, Art. no. 987376.
[66] B. Ogbuokiri, A. Ahmadi, B. Mellado, J. Wu, J. Orbinski, A. Asgary, and J. Kong, "Can post-vaccination sentiment affect the acceptance of booster jab?" in *Intelligent Systems Design and Applications*, A. Abraham, S. Pllana, G. Casalino, K. Ma, and A. Bajaj, Eds. Cham, Switzerland: Springer, 2023, pp. 200–211.
[67] I. D. Mienye and Y. Sun, "A deep learning ensemble with data resampling for credit card fraud detection," *IEEE Access*, vol. 11, pp. 30628–30638, 2023.
[68] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, and H. Cho, "XGBoost: Extreme gradient boosting," *R Package Version*, vol. 1, no. 4, pp. 1–4, Aug. 2015.
[69] A. Ogunleye and Q.-G. Wang, "XGBoost model for chronic kidney disease diagnosis," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 17, no. 6, pp. 2131–2140, Nov. 2020.
[70] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
[71] C. D. Sutton, "Classification and regression trees, bagging, and boosting," *Handbook Statist.*, vol. 24, pp. 303–329, Dec. 2005.
[72] Y. Ren, L. Zhang, and P. N. Suganthan, "Ensemble classification and regression-recent developments, applications and future directions [review article]," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 41–53, Feb. 2016.
[73] N. Altman and M. Krzywinski, "Ensemble methods: Bagging and random forests," *Nature Methods*, vol. 14, no. 10, pp. 933–934, Oct. 2017.
[74] P. Panov and S. Džeroski, "Combining bagging and random subspaces to create better ensembles," in *Proc. Int. Symp. Intell. Data Anal.* Cham, Switzerland: Springer, 2007, pp. 118–129.
[75] S. Mishra, B. L. Sturm, and S. Dixon, "Local interpretable model-agnostic explanations for music content analysis," in *Proc. ISMIR*, vol. 53, 2017, pp. 537–543.
[76] I. P. de Sousa, M. M. B. R. Vellasco, and E. C. da Silva, "Local interpretable model-agnostic explanations for classification of lymph node metastases," *Sensors*, vol. 19, no. 13, p. 2969, Jul. 2019.
[77] S. A. Mahler, S. A. Conrad, H. Wang, and T. C. Arnold, "Resuscitation with balanced electrolyte solution prevents hyperchloremic metabolic acidosis in patients with diabetic ketoacidosis," *Amer. J. Emergency Med.*, vol. 29, no. 6, pp. 670–674, Jul. 2011.
[78] M. L. Malone, V. Gennis, and J. S. Goodwin, "Characteristics of diabetic ketoacidosis in older versus younger adults," *J. Amer. Geriatrics Soc.*, vol. 40, no. 11, pp. 1100–1104, Nov. 1992.
[79] V. Sehgal and B. Ulmer, "Clinical conundrums in the management of diabetic ketoacidosis in the elderly," *J. Transl. Internal Med.*, vol. 7, no. 1, pp. 10–14, Mar. 2019.
[80] Y. Zhang, D. Hong, D. McClement, O. Oladosu, G. Pridham, and G. Slaney, "Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging," *J. Neurosci. Methods*, vol. 353, Apr. 2021, Art. no. 109098.
[81] T. Fan, J. Wang, L. Li, J. Kang, W. Wang, and C. Zhang, "Predicting the risk factors of diabetic ketoacidosis-associated acute kidney injury: A machine learning approach using XGBoost," *Frontiers Public Health*, vol. 11, Apr. 2023, Art. no. 1087297.

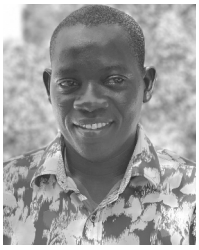
GEORGE OBAIDO (Member, IEEE) received the master's and Ph.D. degrees in computer science from the University of the Witwatersrand, Johannesburg, South Africa. He is a Berkeley Institute for Data Science (BIDS)–Center for Human-Compatible Artificial Intelligence (CHAI) Postdoctoral Fellow with the University of California at Berkeley, USA. His research interest includes using machine learning to find solutions to problems of societal importance.

BLESSING OGBUOKIRI received the Ph.D. degree in computer science from the University of the Witwatersrand, Johannesburg, South Africa. He is a Postdoctoral Fellow and an Instructor with the York University's African–Canada Artificial and Data Innovation Consortium Laboratory, Toronto, Canada, within the Department of Mathematics and Statistics. His work focuses on the intersection of AI and health, helping communities, and governments tackle infectious diseases. He collaborates with researchers from various fields and is currently leading NeurIPS 2023 Affinity workshops and organizing the black in AI workshop. This shows his commitment to making AI inclusive and sharing knowledge in the community. Through his efforts, he is using AI to address real health issues effectively.



CHIDOZIE WILLIAMS CHUKWU received the master's and Ph.D. degrees in applied mathematics from the University of Johannesburg, South Africa. He is a Visiting Assistant Professor with Wake Forest University, USA. His research interest spans the field of mathematical biology and the application of machine learning to predict future trends in the emergence and resurgence of epidemics.

FADEKEMI JANET OSAYE received the M.Sc. degree in applied mathematics from the University of KwaZulu-Natal, South Africa, and the Ph.D. degree in mathematics from the University of Johannesburg, South Africa. She is an Assistant Professor with the Department of Mathematics and Computer Science, Alabama State University, USA. Her research interests include graph theory, network science, machine learning, and their application to disease modeling.



OLUWASEUN FRANCIS EGBELOWO is a Postdoctoral Researcher with the Department of Integrative Biology, The University of Texas at Austin. He uses the application of mathematical and statistical techniques to aid in decision-making for the control of infectious diseases.

MARK IZUCHUKWU UZOCHUKWU received the first M.Sc. degree in data science from Auburn University, USA, the second M.Sc. degree in pure and applied mathematics from the African University of Science and Technology, Nigeria, and the Ph.D. degree in statistics from Auburn University. He is an Assistant Professor of statistics with Alabama State University. His research spans nonparametric statistics, high dimensional data analysis, big data, business analytics, and machine learning, including supervised and unsupervised learning, deep learning, predictive modeling, and operations research. He is adept in nonlinear operator theory, time series forecasting, and causal inference, focusing on iterative methods for statistical estimation and interdisciplinary collaboration.

IBOMOIYE DOMOR MIENYE (Member, IEEE) received the B.Eng. degree in electrical and electronic engineering and the M.Sc. degree (cum laude) in computer systems engineering from the University of East London, in 2012 and 2014, respectively, and the Ph.D. degree in electrical and electronic engineering from the University of Johannesburg, South Africa. He is currently a Postdoctoral Research Fellow with the Department of Information Technology, Walter Sisulu University, South Africa. His research interests include machine learning and deep learning for finance and healthcare applications.

KEHINDE ARULEBA received the Ph.D. degree in computer science from the University of the Witwatersrand, Johannesburg, South Africa. He is currently a Lecturer with the University of Leicester, U.K. Before that, he was a Postdoctoral Fellow with Walter Sisulu University, South Africa. His expertise extends beyond computer science. He actively collaborates with researchers across disciplines. His research interests include machine learning, data ethics, and ICT4D.

MPHO PRIMUS (Member, IEEE) is an NRF-rated researcher specializing in computational linguistics. She is the co-Director with the Institute for Intelligent Systems, University of Johannesburg. Amongst numerous awards, she has received the L'Oréal-UNESCO Women in Science Sub-Saharan Regional Fellow and the Department of Science and Technology (South Africa) Women in Science Alumni, and most recently the National Research Foundation (South Africa) Research Excellence Award for Early Career/Emerging Researcher. She has worked in numerous projects (multidisciplinary) with project partners across three continents. She serves on the IDEMIA Black Women's Ownership Trust Board of Trustees and the Advisory Board for the Pan African Information Communication Technology Association.

OKECHINYERE ACHILONU received the M.Sc. degree in statistics and actuarial science and the Ph.D. degree in biostatistics from the University of the Witwatersrand, Johannesburg, South Africa. She is a Biostatistics Lecturer with the University of the Witwatersrand. She has several years of experience in clinical analysis, statistical support, and teaching. She is passionate about developing statistical and machine learning models from structured and unstructured clinical and epidemiological research data. Her research focuses on developing and assessing analytical frameworks for knowledge discovery toward improving predictive and personalized medicine, focusing on non-communicable diseases, such as cancer, chronic kidney failure, and diabetes.