Predictive Modeling of Kuwaiti Chronic Kidney Diseases (KCKD): Leveraging Electronic Health Records for Clinical Decision-Making

Talal M. Alenezi[1*], Taiseer H. Sulaiman[2], Mohamed Abdelrazek[3], Amr M. AbdelAziz[4]
Faculty of Computers and Information, Assiut University, Assiut, Egypt[1]
Information Science Dept-Faculty of Computers and Information, Assiut University, Assiut, Egypt[2]
Information Systems Dept-Faculty of Computer and Artificial Intelligence, Benha University, Egypt[3]
Faculty of Computers and AI, Beni-Suef University, Beni-Suef, Egypt[4]

***Abstract*—Chronic kidney disease (CDK) represents a significant public health concern globally, and its prevalence is on the rise. In the context of Kuwait, this study addresses the imperative of predicting CKD by leveraging the wealth of information embedded in electronic health records (EHRs). The primary objective is to develop a predictive model capable of early identification of individuals at risk for CKD, thereby enabling timely interventions and personalized healthcare strategies and equip clinicians with information that enhances their ability to make well-informed decisions regarding prognoses or therapeutic interventions. In this study, a dataset has been created from Kuwaiti healthcare institutions, emphasizing the richness and diversity of patient information encapsulated in EHRs and a feature engineering step has been applied for labeling it. Various ensemble learning algorithms, Ada Boost, Extreme Gradient Boosting, Extra Trees, Gradient Boosting, Random Forest, and various single learning algorithms, Decision Tree, K-Nearest Neighbors, Logistic Regression, Multilayer Perceptron, Stochastic Gradient Descent, Support Vector Machines, have been implemented. By examining the empirical findings of our tests, our results showcase the models' capability to identify individuals at risk for CKD at an early stage, facilitating targeted healthcare interventions. Decision Tree was the best classifier achieving 99.5% accuracy and 99.3% macro averaged f1-score.**

***Keywords*—Chronic kidney diseases; Electronic Health Records (EHR); classification; machine learning**

I. Introduction

The digitalization of patient health records has brought about a new era in healthcare, one that offers previously unheard-of possibilities for data-driven research and medical improvements [1]. With the right use, Electronic Health Records (EHR) can become a veritable gold mine of detailed, longitudinal patient data. With the right application, this data can revolutionize the way to anticipate and prevent disease. A major obstacle confronting the healthcare sector is the increasing prevalence of chronic diseases, which contribute significantly to worldwide morbidity and mortality [2]. A growing number of people are interested in using the potential of EHR to create strong predictive models that target early identification, risk assessment, and tailored intervention for chronic diseases. Millions of people worldwide suffer from CKD, a widespread, frequently silent illness that places a heavy burden on healthcare systems around the globe. Innovative methods for identifying those at risk are desperately needed as the frequency of CKD rises, as this will allow for early intervention and individualized care. In this quest, EHRs, which comprise an extensive patient data repository, proves to be a vital asset. They provide a dynamic platform for the creation of predictive models that have the potential to revolutionize the management CKD [3].

When it comes to identifying subtle signs and patterns that precede overt clinical symptoms, traditional diagnostic techniques frequently fall short. Within this framework, EHRs function as a repository for longitudinal patient data, encompassing test findings, medication records, and demographic details, offering a comprehensive perspective of a person's medical journey. By using EHR data, it is possible to identify complex patterns and risk factors related to CKD, which can lead to tailored interventions that can be implemented in a timely manner [3].

In earlier research [4], the same authors suggested using EHRs instead of paper ones to record patients' health information. Also, they highlighted the application of predictive analytics models, which use electronic health data to predict the emergence of chronic diseases early on. According to this study, CKD affects around 700 million people globally each year, and it causes nearly 1.2 million fatalities [4]. The current research contributes to improve the quality of life for those with or at risk of CKD in Kuwait by highlighting the revolutionary potential of predictive ensemble learning and single learning algorithms models using actual Kuwaiti EHRs which are collected from hospitals and health institutes in Kuwait and altering clinical workflows and resource allocation.

Through patient follow-up, changes in several clinical markers could be seen over time and their relationship to the course of the disease. By using this method, we may record the time dynamics and find any patterns or trends that might point to deteriorating CKD. While our study primarily focused on the development and validation of the predictive model, we acknowledge the importance of discussing the practical aspects of its clinical implementation. Addressing issues related to data integration, workflow adaptation, and acceptance by healthcare

professionals is critical for the effective deployment of predictive models in routine clinical care.

The following are the primary contributions of the article:

- Using electronic health records instead of paper records.
- Using accessible datasets from patients' medical records, machine learning techniques are used to predict the existence of chronic illnesses.
- Examining medical records of all patients to ensure proper diagnosis of chronic disorders.
- Identifying new patients with comparable symptoms and illness development phases based on physician supervision and medical record analysis for a specific type of chronic disease.

The latter part of the manuscript will delve into related research in Section II, followed by an examination of the datasets employed in this study in Section III. Section IV will provide a comprehensive description of the proposed technique. Subsequently, Section V will present the test results and evaluate the effectiveness of the proposed strategy. Lastly, Section VI and Section VII will present the discussion and conclusions respectively.

II. Related Work

Considerable work has been done to anticipate CKD. This section will include descriptions of a few of these works.

To predict CKD using clinical data, Ekanayake and Herath [5] investigated the use of machine learning techniques. They noted the need of feature engineering, handling missing values, and integrating domain knowledge in the study. They presented a procedure that includes attribute selection, handling of missing values, and data preprocessing. The application of a KNN-based technique to handle missing values in datasets pertaining to several diseases was also taken into consideration in this work. According to the study, the random forest and extra trees classifiers produced the best results for predicting CKD, obtaining 100% accuracy for both training and testing. Furthermore, the study made no mention of any potential privacy or ethical issues with using patient data for predictive modeling.

Q. Bai et al. [6] developed a predictive model for end-stage kidney disease (ESKD) using a dataset of 748 people with chronic kidney disease (CKD). To manage missing data, the authors used a five-set multiple imputation method. They then examined each model's performance on each imputed set, combining the findings to get the result. At 81%, the random forest algorithm produced the best overall performance as determined by the AUC score. On the other hand, the Kidney Failure Risk Equation (KFRE) model, which is based on three straightforward variables, showed the highest accuracy, specificity, and precision along with equivalent AUC scores. The research found a void in the literature about the applicability of predictive models for ESKD in other ethnic groups, including the Chinese population. It also brought attention to the possibility of predicting ESKD without the need for urine testing, which could result in a more straightforward model with comparable reliability. The KFRE model's default threshold sensitivity and the lack of previous attempts to use machine learning techniques to predict the occurrence of ESKD in CKD patients are among the study's shortcomings.

Y. Zhu et al. [7] presented a unique method utilizing longitudinal patient Electronic Health Records (EHRs) to predict the course of CKD. They forecasted the course of CKD with impressive accuracy by combining an AI prediction model with an EHR preprocessing pipeline. Preprocessing the EHR incorporates multiple clinical factors and transforms them into time series data that may be used in Recurrent Neural Network (RNN) modeling. Their main goal was to forecast how quickly CKD will advance from early to late stages. Feature vectors that represent patient data prior to a given period are analyzed for each case patient. Based on patient race, sex, age, and duration of time series, control patients are matched when utilizing the time series of a single variable, eGFR, the RNN model predicts disease development within a year with an average AUROC of 0.957. Due to patient privacy issues and the proprietary nature of the data, there is a research gap in the lack of publicly available datasets.

H. Nayeem et al. [8] applied machine learning approaches to predict chronic kidney disease (CKD). The 400 examples in the sample comprise 25 attributes total one dependent attribute and 24 independent attributes. To predict CKD, the study used methods from Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN). An analysis of the classifiers' comparative performance revealed that ANN performed better than the other techniques, achieving 98.6% and 98% accuracy and f1-score, respectively. The study failed to examine the possibility of biases in the dataset or the applicability of the results to different groups with chronic kidney disease.

D. Chicco et al. [9] examined CKD and found independent risk factors linked to stages 3-5 of the disease by examining a dataset of 491 individuals from the United Arab Emirates. The authors used two different methodologies, one based on machine learning techniques and the other on conventional univariate biostatistics testing. The results of the biostatistical tests showed that while 68.42% of the clinical parameters were significant, they were not precise. As a result, the writers ranked features using Random Forests. The study showed that, independent of temporal information, computational intelligence could predict the development of severe CKD, suggesting that the significance of clinical factors changes when the temporal component is considered. The study is significant because it uses machine learning to uncover critical clinical traits while thoroughly examining risk factors linked to CKD at stages 3-5. Remarkably, the study concentrated on developing and enhancing computational intelligence technologies rather than discussing the therapeutic implications of the results.

In another study on the prediction of chronic renal illness, S. Pal [10] used three machine learning classifiers: Support Vector Machine (SVM), Decision Tree, and Logistic Regression. After the classifiers were assessed, it was discovered that the Decision Tree classifier performed the best, obtaining 95.92% accuracy, 0.99 precision, 0.98 recall, and

0.98 F1-score. To enhance the performance of the base classifiers, bagging ensemble methods were also used in this study. The Decision Tree classifier achieved the greatest accuracy of 97.23%. There may be a research gap because the study did not specify the precise dataset that was utilized to train and test the classifiers.

M. Klamrowski et al. [11] used machine learning to create a prediction model for individuals with advanced chronic renal disease who are at risk of short-term kidney failure. They were intended to be integrated into electronic medical records for clinical use, and they made use of dynamic laboratory data to increase prediction accuracy. They showed that the prediction of short-term kidney failure requires the inclusion of more current follow-up data, especially dynamic lab data. The study also demonstrated how the prediction model may be used to lower the rates of unscheduled dialysis and the negative consequences that are linked with it. The study found that using machine learning methods, such as Cox regression, to take into consideration the complex interrelationships and complexity of the data was the best approach. This study found a knowledge gap on the requirement to increase the generalizability of prediction models to various populations. Furthermore, there is a lack of validating the prediction model's efficacy within the typical renal disease clinic workflow and integration of the model into clinical practice.

A. Islam et al. [12] used machine learning techniques to forecast chronic kidney disease (CKD) in its early stages. To minimize the number of features and get rid of unnecessary data, relationships between various aspects were investigated and the models were trained and validated using input parameters. After preprocessing the dataset, principal component analysis (PCA) was used to determine which features were most important in CKD detection. This study discovered that, after using a filter feature selection approach, hemoglobin, albumin, and specific gravity had the greatest influence on CKD prediction. The best approach involved using machine learning algorithms to predict CKD at an early stage, with a focus on identifying the most dominant features for detecting the disease. The performance of the proposed model was evaluated, and it was suggested that CKD risk factor predicting could be used to identify individuals at risk within the community without the need for hospital admission. The challenge of locating a different dataset with comparable properties for a useful comparison was noted by the authors.

III. DATA

In this study, an EHR dataset comprising information about patients in Kuwaiti hospitals was constructed. Each row in the dataset represents a single patient, and the columns indicate all the patient's attributes (laboratory analysis), as detailed in Table I. This dataset was created for the purpose of predicting Kuwaiti Chronic Kidney Diseases (KCKD) by combining all characteristics from original datasets into a single csv file for use in training and testing prediction models, in addition to the personal information of all patients during all hospital visits. This EHR dataset has been acquired from the Department of Health Studies and Research at the Kuwaiti Ministry of Health to get clearance to access medical data and publish under the supervision of National Center for Health Information, and the Department of Prevention and Control of Non-Communicable Diseases in Kuwait. Data authorization has been obtained from the director of Al-Adan Hospital and forward the Request to the appropriate departments. Obtain clearance from the heads of the Medical Board, the Department of Clinical Radiology, the Department of Medical Laboratories, and the Department of Information Systems at Al-Adan Hospital for various departments. The dataset is available on Kaggle, KCKD, in the final version after feature engineering process. Another online labeled CKD dataset with the same features has been used for labeling the first one using feature engineering process that will be discussed in the next section. For the findings of this study to be applicable and generalizable to a wide range of populations, we must take ethnic and cultural heterogeneity into account in our research. We can evaluate potential differences in illness susceptibility, progression trends, and response to therapies by stratifying our dataset according to ethnicity or cultural background. Furthermore, our predictive models may be more accurate and relevant for demographic groups if pertinent cultural determinants of health are included.

TABLE I. NAME AND TYPE OF EACH FEATURE OF THE EHR DATASET

Feature Name	Type
Potassium	Numeric
Sodium	Numeric
CL	Numeric
Ceriatinin	Numeric
Blood Urea	Numeric

IV. METHODOLOGY

In this section, the methodology of the proposed work will be represented and provide an explanation for every step. Fig. 1 represents the block diagram of the proposed work and steps including two phases.

A. Feature Engineering

This paper introduces a feature engineering approach [13] for kidney disease classification, focusing on leveraging known features from a labeled dataset to enhance predictive modeling on an unlabeled dataset. The methodology involves meticulous extraction and refinement of relevant features, employing preprocessing techniques to ensure data quality, and training multiple machine learning [14] classification algorithms. The best-performing model is then selected based on cross-validation results on the labeled dataset. Subsequently, this chosen model is applied to predict classes for the unlabeled dataset, providing a seamless transfer of knowledge between labeled and unlabeled data. The paper concludes with an analysis of the model's performance, highlighting the efficacy of the proposed feature engineering process in improving the accuracy and generalization of kidney disease classification.

B. Data Preprocessing

Preprocessing steps are crucial for kidney datasets as they play a vital role in enhancing the quality of the data and ensuring that machine learning models can effectively learn patterns and make accurate predictions. Here are some key preprocessing steps and their importance for kidney datasets.

1) Handling missing value: Kidney datasets may often have missing values due to various reasons such as incomplete sample collection or laboratory errors. Imputing or handling missing values is critical to maintain the integrity of the dataset and ensure that the analysis is based on as much relevant information as possible. By checking for null values in the aggregated dataset we noticed that there are some missing values in the input features as shown in Table II.

TABLE II. NUMBER OF NULL VALUES IN THE AGGREGATED DATASET

Feature Name	# Null values	# All records
Blood Urea (mgs/dL)	14	680
Serum Creatinine (mgs/dL)	12	
Sodium (mEq/L)	67	
Potassium (mEq/L)	68	

2) Normalization / Scaling: Different features in the dataset may have different scales. Normalizing or scaling features, especially numeric ones like blood pressure or serum creatinine, helps in bringing them to a similar scale, preventing certain features from dominating others during model training. We utilized the Standard Scaler [2] during preprocessing for the kidney dataset. This technique normalizes features to have a mean of 0 and a standard deviation of 1, ensuring uniform scales and enhancing the effectiveness of machine learning models, especially those reliant on distance measures.

3) Data splitting: In the experimentation of this study, for the aggregated dataset, we divided the aggregated kidney dataset into training (85%) and testing (15%) sets. This resulted in 571 instances for training and 101 instances for testing, out of the total 672 instances in the dataset. After labeling the second dataset, we divided it into two portions 80% (1600 samples) for training and 20% (401 samples) for testing as shown in Table III. This approach ensures a comprehensive evaluation of model performance, balancing training, and testing for reliable insights into the effectiveness of the proposed models for kidney disease classification.

TABLE III. NUMBER OF INSTANCES IN AGGREGATED CKD AND KCKD DATASETS

Dataset	Split ratio		Training instances	Testing instances	Total instances
Aggregated Dataset	85% Train	15% Test	571	101	672
Dataset after labeling	80% Train	20% Test	1600	401	2001

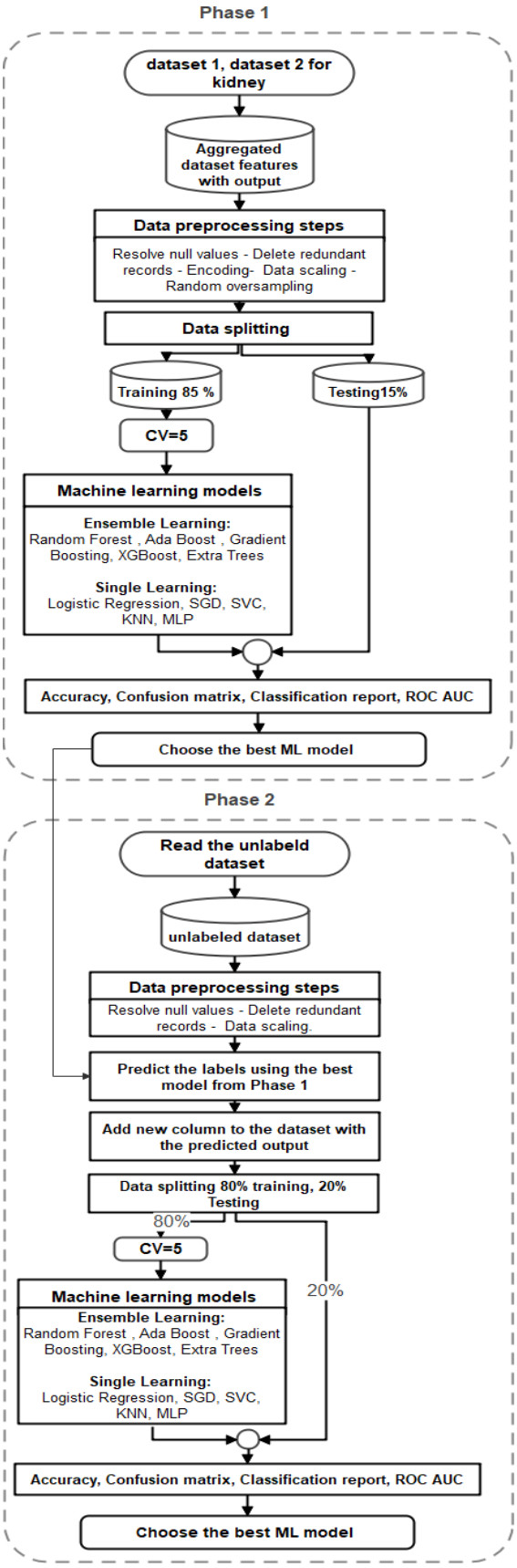

Fig. 1. Proposed method architecture.

TABLE IV. NUMBER OF INSTANCS BEFORE AND AFTER OVER SAMPLING IN THE AGGREGATED DATASET AND NEW DATASET

Dataset	Label	# samples before balancing	# samples after balancing
Aggregated Dataset	non-CKD	216	355
	CKD	355	355
Dataset after labeling	non-CKD	345	1255
	CKD	1255	1255
Total		2171	3220

Table IV displays the impact of oversampling on instance counts in both the Aggregated Dataset and new labeled dataset after feature engineering. In the Aggregated Dataset section, the initial counts show 216 instances for the "non-CKD" class and 355 instances for the "CKD" class. After balancing, both classes have 355 instances, resulting in a total of 710 instances. Moving to the dataset after labeling using feature engineering, the “non-CKD” class initially has 345 instances, while the "CKD" class has 1255 instances. Following oversampling, both classes achieve balance with 1255 instances each, contributing to a total of 2510 instances. This oversampling strategy aims to ensure a more equitable representation of classes for enhanced model training and evaluation.

4) Cross Validation (CV): To rigorously assess proposed machine learning models, we adopted a five-fold cross-validation approach [15]. This method divides the dataset into five subsets, iteratively training the model on four and testing on the remaining one. By calculating and averaging performance metrics, such as accuracy and precision, across all iterations, we obtain a robust evaluation of proposed model's generalizability. This strategy ensures reliability by preventing over-sensitivity to a particular training set composition and guides hyperparameter tuning efforts for optimized model performance.

C. Machine Learning Methods

In this study, we employed two categories of machine learning algorithms, namely ensemble learning algorithms and single learning algorithms, as outlined below.

1) Single learning: In this research, we employed a diverse set of single learning algorithms, each contributing distinct strengths to analysis. The single learning algorithms used are illustrated below.

a) Decision Tree (DT) [16]: A clever and straightforward machine learning predictive model technique called a decision tree classifier uses a tree representation to go from an item's observation to a judgment about the item's target value. The decision tree is a tool for classification, description, and generalization of a given collection of data that combines mathematics and computational techniques.

b) Logistic Regression (LR) [17]: Predicting Binary Probabilities: Logistic Regression serves as a linear classification method that predicts the likelihood of a binary outcome. It accomplishes this by fitting a logistic curve to the data, making it particularly suitable for applications like binary classification tasks such as spam detection or medical diagnoses.

c) Stochastic Gradient Descent (SGD) [18]: SGD stands out as an optimization technique widely employed in machine learning model training. It refines model parameters through iterative and stochastic updates, proving notably efficient for handling extensive datasets. SGD finds frequent use in tasks involving neural network training and other iterative optimization challenges.

d) Support Vector Classifier (SVC) [19]: Using training data at class boundaries, the SVM is a linear classifier. Radial Basis Function (RBF) kernels, which were employed in this work, sigmoid, linear, and other kernel functions are used by the SVM model to classify non-linear data. Assuming that the new sample and the existing samples are similar, the KNN algorithm assigns the new sample to the category that most closely matches the existing categories [19].

e) K-Nearest Neighbours (KNN) [20]: KNN, a straightforward yet powerful algorithm, is adept at classification and regression duties. Its principle lies in classifying data points by considering the majority class among their k-nearest neighbours. KNN's simplicity and ease of implementation make it suitable for diverse applications, ranging from recommendation systems to pattern recognition.

f) Multi-Layer Perceptron (MLP) [21]: MLP, categorized as a neural network with multiple layers, exhibits proficiency in discerning intricate patterns and relationships within datasets. This algorithm's versatility is evident across applications like image recognition, natural language processing, and speech recognition. The depth of the network facilitates the capture of intricate hierarchical features in the data.

These models operate independently, with each algorithm focusing on learning patterns and relationships within the data individually. The application of these single learning techniques allows us to harness the specific capabilities of each algorithm to enhance the understanding of the intricate dynamics within the kidney dataset.

2) Ensemble learning: To further fortify the predictive capabilities of proposed models, we incorporated ensemble learning algorithms, a category renowned for amalgamating multiple models to achieve superior performance. The ensemble learning algorithms utilized in this research encompassed the following types:

a) Random Forest [22]: Random Forest, an ensemble learning algorithm, builds numerous decision trees during training, consolidating predictions to improve reliability and mitigate overfitting. Its versatility has proven effective across diverse domains, including the focus of this research.

b) Ada Boost [23]: Ada Boost, a boosting algorithm, combines weak learners sequentially to form a robust model. Its iterative approach corrects errors from previous models, with a focus on challenging instances. In this research, Ada Boost plays a pivotal role in elevating accuracy in ensemble predictions.

c) Gradient Boosting (GBoost) [24]: Gradient Boosting, an iterative ensemble algorithm, constructs decision trees sequentially to rectify the errors of preceding trees. Known for achieving high precision, GBoost is particularly valuable in scenarios requiring accurate predictions, as exemplified in this research.

d) XGBoost [25]: XGBoost, or Extreme Gradient Boosting, represents an optimized form of gradient boosting with a focus on speed and efficiency. Its parallelized training and regularization techniques make it scalable and efficient for handling extensive datasets. In this research, XGBoost enhances the effectiveness of ensemble learning.

e) Extra Trees [26]: Extra Trees, or Extremely Randomized Trees, is an ensemble algorithm introducing additional randomization during tree construction. This intentional randomness enhances model robustness and generalization. In this research, Extra Trees contributes to the ensemble's diversity, fostering a more resilient predictive model.

Leveraging the collective wisdom of diverse models, these ensemble learning techniques aimed to amplify the robustness and accuracy of the predictions, particularly in the context of kidney disease classification.

D. Performance Metrics

In this section, we detail the evaluation metrics used in the thesis to assess the performance of statistical and machine learning algorithms, including accuracy, the confusion matrix, and the classification report.

1) Confusion Matrix: This matrix [27], a vital tool for classification model evaluation, provides a comprehensive summary of predictions versus actual labels. Comprising elements like True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN), it illuminates the model's performance for positive and negative classes.

2) Classification Report: Offering a detailed assessment across different classes, the classification report presents various metrics per class. It includes equations for accuracy in Eq. (1), precision in Eq. (2), recall in Eq. (3), and F1 score in Eq. (4) derived from the report [28]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (2)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (3)$$

$$\text{F-measure} = \frac{2 * TP}{(2 * TP + FP + FN)} \quad (4)$$

3) ROC AUC Curve: Illustrating a binary classification model's performance across decision thresholds, the ROC curve [29] and its AUC metric (ranging from 0 to 1) indicate discrimination ability, with higher values denoting superior performance.

V. Experimental Results

Several experiments have been conducted to assess the proposed model. The Python programming language and various machine learning toolboxes, such as scikit-learn, imblearn, NumPy, and matplotlib, were used for all experiments, which were conducted using the Jupiter notebook editor.

As shown in Table V, a feature engineering process has been applied for an aggregated CKD dataset with the same attributes to obtain the optimal labels for new EHRs dataset. The aggregated CKD dataset has been trained using nine classifiers, five ensemble learning algorithms (Ada Boost, XGBoost, Extra Trees, GBoost, and RF) and six single learning algorithms (DT, KNN, LR, MLP, SGD, and SVM), and the highest performance was obtained by GBoost classifier, which achieved 97.7%, 97.8, 98%, and 97.8% for precision, recall, accuracy, and macro-averaged f1-score respectively. The RF classifier was in second place achieving 95.6%, 95.6%, 96%, and 95.5% for precision, recall, accuracy, and macro-averaged f1-score respectively.

After obtaining the best labels for each patient in KCKD dataset, it was ready for building the classification model with the same classifiers mentioned above, which can predict the case of patients in Kuwait hospitals, who have the same symptoms with different values.

TABLE V. Training Performance of All Classifiers for the Aggregated CKD Dataset

	Classifier	Precision	Recall	Accuracy	F1-score
Single Learning	DT	94.3%	94.9%	95.0%	94.6%
	KNN	92.2%	92.7%	93.1%	92.4%
	LR	86.5%	85.1%	87.1%	85.6%
	MLP	94.1%	90.9%	93.1%	92.2%
	SGD	87.0%	87.0%	88.1%	87.0%
	SVM	86.8%	87.6%	88.2%	87.2%
Ensemble Learning	Ada Boost	86.8%	82.6%	77.6%	74.7%
	XGBoost	**94.9%**	**94.5%**	**90.2%**	**92.6%**
	Extra Trees	81.8%	85.7%	53.6%	54.8%
	GBoost	89.9%	88.5%	68.4%	82.3%
	RF	86.8%	84.6%	71.0%	75.3%

Table VI shows the performance of all classifiers for KCKD dataset. The highest performance was obtained by DT classifier, which achieved 98.9%, 99.7%, 99.5%, and 99.3% for precision, recall, accuracy, and macro-averaged f1-score respectively. The second place for Ada Boost classifier, which we achieved 98.4%, 99.5%, 99.2%, and 98.9% for precision, recall, accuracy, and macro-averaged f1-score respectively.

A recognized confusion matrix is obtained in Table VII and Table VIII for the purpose of estimating four different measures: recall, accuracy, precision, and f-score. The confusion matrix displays the classification results as a matrix. Information for both existing and anticipated classes created with the classification framework is included. The cell shows

the sample size that was mistakenly identified as false while quiet (i.e., TN) and as true when it was truly true (i.e., TP). The number of pieces that were erroneously classified is indicated by the two remaining cells.

TABLE VI. TRAINING PERFORMANCE OF ALL CLASSIFIERS FOR KCKD DATASET AFTER LABELING

	Classifier	Precision	Recall	Accuracy	F1-score
Single Learning	**DT**	**98.9%**	**99.7%**	**99.5%**	**99.3%**
	KNN	91.3%	95.1%	94.7%	92.9%
	LR	95.2%	96.5%	97.0%	95.8%
	MLP	97.4%	97.7%	98.2%	97.5%
	SGD	94.1%	96.6%	96.5%	95.3%
	SVM	91.8%	97.1%	95.5%	94.1%
Ensemble Learning	Ada Boost	98.4%	99.5%	99.2%	98.9%
	XGBoost	97.9%	96.4%	98.0%	97.1%
	Extra Trees	38.4%	50.0%	76.8%	43.4%
	GBoost	98.2%	98.9%	99.0%	98.6%
	RF	98.3%	99.2%	99.2%	98.9%

TABLE VII. CLASSIFICATION REPORTS OF SVM, DT, AND MLP CLASSIFIERS FOR AGGREGATED CKD DATASET

Dataset	Classifier	Precision	Recall	F-measure	Class
Single Learning	DT	91.89	94.44	93.15	0
		96.88	95.38	96.12	1
	KNN	89.19	91.67	90.41	0
		95.31	93.85	94.57	1
	LR	84.85	77.78	81.16	0
		88.24	92.31	90.23	1
	MLP	96.77	83.33	89.55	0
		91.43	98.46	94.81	1
	SGD	83.33	83.33	83.33	0
		90.77	90.77	90.77	1
	SVM	81.58	86.11	83.78	0
		92.06	89.23	90.62	1
Ensemble Learning	Ada Boost	96.67	80.56	87.88	0
		90.14	98.46	94.12	1
	XGBoost	89.74	97.22	93.33	0
		98.39	93.85	96.06	1
	Extra Trees	89.47	47.22	61.82	0
		76.83	96.92	85.71	1
	GBoost	97.22	97.22	97.22	0
		98.46	98.46	98.46	1
	RF	94.44	94.44	94.44	0
		96.92	96.92	96.92	1

TABLE VIII. CLASSIFICATION REPORTS OF SVM, DT, AND MLP CLASSIFIERS FOR KCKD DATASET

Dataset	Classifier	Precision	Recall	F-measure	Class
Single Learning	DT	97.89	100.0	98.94	0
		100.0	99.35	96.67	1
	KNN	93.96	95.70	89.45	0
		98.64	94.48	96.52	1
	LR	91.75	95.70	93.68	0
		98.86	97.40	98.04	1
	MLP	95.74	96.77	96.26	0
		99.02	98.70	99.86	1
	SGD	89.11	96.77	92.78	0
		99.00	96.43	97.70	1
	SVM	83.87	100.0	91.18	0
		100.0	94.16	96.99	1
Ensemble Learning	Ada Boost	96.88	100.0	98.41	0
		100.0	99.03	99.51	1
	XGBoost	97.75	93.55	95.60	0
		98.08	99.35	98.71	1
	Extra Trees	00.00	00.00	00.00	0
		76.81	100.0	86.88	1
	GBoost	96.84	98.92	97.87	0
		99.67	99.03	99.35	1
	RF	96.88	100.0	98.41	0
		100.0	99.03	99.51	1

The cells indicating the sample size labeled true when it was incorrect (i.e., FP) and false when it was true (i.e., FN). All measures were calculated using the formulas listed in the previous subsection, 2.4.2. In the class column, "1" means CKD patient and "0" means non-CKD patient.

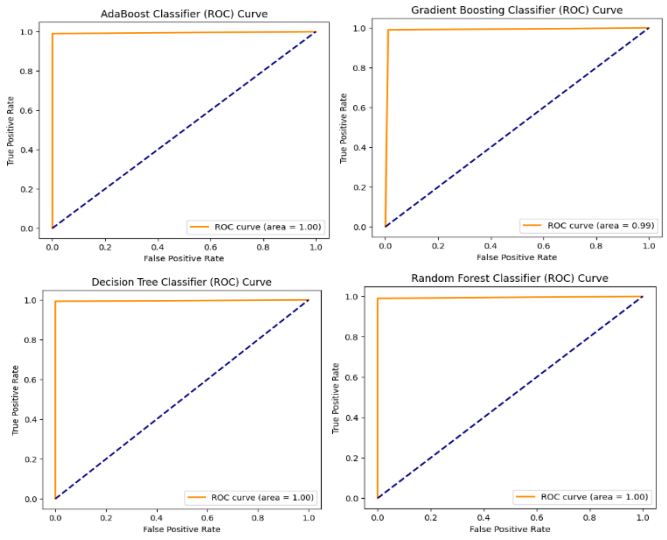

Fig. 2. ROC curve of DT, RF, AdaBoost, GBoost curve for Kuwaiti CKD, from left to right, respectively.

Fig. 2 shows the ROC curve of the best ensemble and single learning classifiers applied for Kuwaiti CKD, DT, RF, AdaBoost, GBoost.

VI. Discussion

The labeled dataset of CKD, as demonstrated in Table VI, presents a comprehensive evaluation of various classifiers, each assessed on key performance metrics including precision, recall, accuracy, and F1-score. Among these classifiers, the results distinctly highlight the exceptional performance of the Decision Trees (DT) algorithm. With a precision of 98.9%, recall of 99.7%, accuracy of 99.5%, and an F1-score of 99.3%, DT emerges as the standout performer across all metrics. The superiority of DT can be attributed to several inherent advantages it offers. Notably, its innate interpretability lends itself well to domains such as CKD, where comprehensible decision-making is crucial for clinical applications. Moreover, DT's ability to effectively capture non-linear relationships within the data proves invaluable in handling the complex patterns often present in CKD datasets. Furthermore, its robustness to irrelevant features ensures efficient feature selection, enhancing model performance and generalization. Given the scalability and efficiency of DT, particularly in managing large datasets, it emerges as not only the best-performing classifier in this evaluation but also a pragmatic choice for real-world CKD classification tasks. This robust performance underscores the utility of Decision Trees as a reliable and effective tool for medical diagnosis and decision support in the context of chronic kidney disease. We carefully selected algorithms based on their proven effectiveness for the task. Through rigorous testing, we found that our chosen ensemble and single learning algorithms consistently delivered high performance. While we acknowledge the potential for different results with alternative algorithms, our focus was on leveraging well-established methods known for their reliability. Our thorough validation process supports the confidence in the efficacy of our selected algorithms for this study. We recognize the importance of clinical interpretability in healthcare settings. While our study primarily focused on performance metrics like accuracy, precision, recall, and F1-score, we understand the need to understand model predictions. By addressing interpretability, we aim to bridge the gap between model performance and real-world healthcare applications, enhancing trust among healthcare professionals.

VII. Conclusion and Future Work

To predict KCKD, we created EHRs dataset using patients' symptoms collected from Kuwait hospitals and health institutions, a feature engineering process has been utilized for this dataset to obtain the optimal labels for each patient by training another CKD dataset with the same attributes using several ensemble learning classifiers, Ada Boost, XGBoost, Extra Trees, GBoost, RF, and several single learning classifiers, DT, KNN, LR, MLP, SGD, and SVM. The study's findings suggest that chronic disease identification and prediction can be accomplished with the help of data mining tools. According to the findings, the DT algorithm was the best option with the highest performance for predicting Kuwait CKD, achieving 99.5% accuracy and 99.3% f1-score, while the GBoost algorithm was the most effective for training the aggregated CKD dataset and obtaining the optimal labels of Kuwait CKD dataset, achieving 98% accuracy and 97.8% f1-score. Strong performance was also shown by the RF and Ada Boost algorithms on both datasets. In further work, we intend to include a portion addressing the practical issues and difficulties related to applying our predictive model in clinical settings. The predictive model's seamless integration into clinical decision-making processes, workflow adaptation to ensure healthcare professionals' acceptance and adoption of the model, and data compatibility and integration with current electronic health record systems will all be covered in this.

References

[1] M. E. Hossain, A. Khan, M. A. Moni, and S. Uddin, "Use of Electronic Health Data for Disease Prediction: A Comprehensive Literature Review," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 2. Institute of Electrical and Electronics Engineers Inc., pp. 745–758, Mar. 01, 2021. doi: 10.1109/TCBB.2019.2937862.

[2] A. Perotte, R. Ranganath, J. S. Hirsch, D. Blei, and N. Elhadad, "Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis," Journal of the American Medical Informatics Association, vol. 22, no. 4, pp. 872–880, Jul. 2015, doi: 10.1093/jamia/ocv024.

[3] D. Chicco, C. A. Lovejoy, and L. Oneto, "A Machine Learning Analysis of Health Records of Patients with Chronic Kidney Disease at Risk of Cardiovascular Disease," IEEE Access, vol. 9, pp. 165132–165144, 2021, doi: 10.1109/ACCESS.2021.3133700.

[4] T. M. Alenezi, T. H. Sulaiman, and A. M. Abdelaziz, "Applying Machine Learning Models to Electronic Health Records for Chronic Disease Diagnosis in Kuwait." [Online]. Available: www.ijacsa.thesai.org

[5] I. U. Ekanayake and D. Herath, "Chronic Kidney Disease Prediction Using Machine Learning Methods," in MERCon 2020 - 6th International Multidisciplinary Moratuwa Engineering Research Conference, Proceedings, Institute of Electrical and Electronics Engineers Inc., Jul. 2020, pp. 260–265. doi: 10.1109/MERCon50084.2020.9185249.

[6] Q. Bai, C. Su, W. Tang, and Y. Li, "Machine learning to predict end stage kidney disease in chronic kidney disease," Sci Rep, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-12316-z.

[7] Y. Zhu, D. Bi, M. Saunders, and Y. Ji, "Prediction of chronic kidney disease progression using recurrent neural network and electronic health records," Sci Rep, vol. 13, no. 1, p. 22091, Dec. 2023, doi: 10.1038/s41598-023-49271-2.

[8] M. Ariful, I. Mozumder, H.-C. Kim, N. Hosen, A. I. Mozumder, and R. I. Sumon, "Prediction of Chronic Kidney Disease Using Machine Learning," 2023. [Online]. Available: https://www.researchgate.net/publication/373642582.

[9] D. Chicco, C. A. Lovejoy, and L. Oneto, "A Machine Learning Analysis of Health Records of Patients with Chronic Kidney Disease at Risk of Cardiovascular Disease," IEEE Access, vol. 9, pp. 165132–165144, 2021, doi: 10.1109/ACCESS.2021.3133700.

[10] M. B. Nirmala, D. K. Priyamvada, P. R. Shetty, and S. Pallavi Singh, "Chronic kidney disease prediction using machine learning techniques," in 12th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2021, Grenze Scientific Society, 2021, pp. 185–190. doi: 10.1007/s44174-022-00027-y.

[11] M. M. Klamrowski et al., "Short Timeframe Prediction of Kidney Failure among Patients with Advanced Chronic Kidney Disease," Clin Chem, vol. 69, no. 10, pp. 1163–1173, Oct. 2023, doi: 10.1093/clinchem/hvad112.

[12] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," J Pathol Inform, vol. 14, Jan. 2023, doi: 10.1016/j.jpi.2023.100189.

[13] H. Liu, "Feature Engineering for Machine Learning and Data Analytics," Feature Engineering for Machine Learning and Data Analytics, 2018, doi: 10.1201/9781315181080.

[14] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," 2020 11th International Conference on Information and Communication Systems, ICICS 2020, pp. 243–248, 2020, doi: 10.1109/ICICS49469.2020.239556.

[15] N. Ganapathy, R. Swaminathan, and T. M. Deserno, "Adaptive learning and cross training improves R-wave detection in ECG," Comput Methods Programs Biomed, vol. 200, 2021, doi: 10.1016/j.cmpb.2021.105931.

[16] M. A. Abdelaal, M. A. Fattah, and M. M. Arafa, "Predicting Sarcasm and Polarity in Arabic Text Automatically: Supervised Machine Learning Approach," Article in Journal of Theoretical and Applied Information Technology, vol. 100, no. 8, 2022, [Online]. Available: www.jatit.org

[17] T. G. Nick and K. M. Campbell, "Logistic regression.," Methods Mol Biol, vol. 404, pp. 273–301, 2007, doi: 10.1007/978-1-59745-530-5_14.

[18] R. Roy, "ML | Stochastic Gradient Descent (SGD)," Geeks for geeks, 2020.

[19] A. H. Abuelatta, M. Sobhy, A. A. El-Sawy, and H. Nayel, "Arabic Regional Dialect Identification (ARDI) using Pair of Continuous Bag-of-Words and Data Augmentation." [Online]. Available: www.ijacsa.thesai.org

[20] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 2888, pp. 986–996, 2003, doi: 10.1007/978-3-540-39964-3_62.

[21] M. Verleysen, "Multi - Layer Perceptron (MLP)," no. September, pp. 1–21, 2005.

[22] Gerard Biau, "Analysis of a Random Forests Model," Journal of Machine Learning Research, vol. 13, pp. 1063–1095, 2012.

[23] V. Kurama, "Guide to AdaBoost: Boosting To Save The Day," 2019.

[24] S. Touzani, J. Granderson, and S. Fernandes, "Gradient boosting machine for modeling the energy consumption of commercial buildings," Energy Build, vol. 158, no. 510, pp. 1533–1543, 2018, doi: 10.1016/j.enbuild.2017.11.039.

[25] T. Chen, T. He, and M. Benesty, "XGBoost : eXtreme Gradient Boosting," R package version 0.71-2, pp. 1–4, 2018.

[26] S. M. Mastelini, F. K. Nakano, C. Vens, and A. C. P. de L. F. de Carvalho, "Online Extra Trees Regressor," IEEE Trans Neural Netw Learn Syst, 2022, doi: 10.1109/TNNLS.2022.3212859.

[27] Ž. Vujović, "Classification Model Evaluation Metrics," International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.

[28] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," Electronics (Switzerland), vol. 8, no. 8, pp. 1–34, 2019, doi: 10.3390/electronics8080832.

[29] Sarang Narkhede, "Understanding AUC - ROC Curve," Towards Data Science, pp. 6–11, 2019.