A Dual-stage Approach for Chronic Kidney Disease Detection Using Shapley Additive Explanations and an Evolutionary Light Gradient Boosting Machine

Moumita Pramanik[1], Ranjit Panigrahi[1*], Bidita Khandelwal[2], Joseph Bamidele Awotunde[3], Akash Kumar Bhoi[4,*]

[1]*Department of Computer Applications*
Sikkim Manipal Institute of Technology, Sikkim Manipal University
Majitar, Sikkim, 737136, India
E-mails: moumita.pramanik@gmail.com,
ranjit.panigrahi@gmail.com

[2]*Department of Medicine*
Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University
Tadong, Sikkim, 737102, India
E-mail: bidita.k@smims.smu.edu.in

[3]*Department of Computer Science*
Faculty of Information and Communication Sciences, University of Ilorin
Ilorin, 240003, Kwara State, Nigeria
E-mail: awotunde.jb@unilorin.edu.ng

[4]*Directorate of Research*
Sikkim Manipal University
Tadong, Sikkim, 737102, India
E-mail: akashkrbhoi@gmail.com

[*]*Corresponding authors*

Received: July 23, 2024 **Accepted: February 06, 2025**

Published: December 31, 2025

Abstract: *This article discusses significant advancements in diagnosing chronic kidney disease (CKD) using state-of-the-art machine learning methods. It employs two CKD detection models: the gradient boosting decision tree (GBDT) model and dropout additive regression tree (DART) model. The models are used for CKD detection within the light gradient boosting machine (LightGBM) framework. This article also describes a new dual-stage feature selection strategy to improve model inputs by selecting relevant features while maintaining transparency. The LightGBM feature importance score was used to identify critical features of CKD patients during the feature selection phase. Additionally, the SHapley Additive exPlanations (SHAP) method is utilised to assess the significance of individual attributes, making the model predictions easier to understand. The experimental evaluation utilised a dataset containing 24 features related to CKD. The developed GBDT and DART models demonstrated high accuracy, sensitivity, and specificity levels. The GBDT model exhibited a sensitivity of 99.20%, a specificity of 100%, and an accuracy of 99.50%. Similarly, the GBDT-based model achieved a precision of 98.80%, a sensitivity of 99.20%, and a specificity of 100%.*

Keywords: *Chronic kidney disease, Machine learning, Gradient boosting decision trees, Light gradient boosting machine framework, SHapley Additive exPlanations.*

Introduction

A gradual decline in renal function leads to chronic kidney disease (CKD), resulting in impaired filtration of waste products and surplus fluids from the bloodstream [31]. This condition is progressive and tends to deteriorate over time, potentially resulting in kidney failure if not treated appropriately [21]. CKD can be influenced by various factors, such as diabetes, high blood pressure, autoimmune disorders, genetic anomalies, infections, and other medical conditions. Symptoms of CKD include fatigue, weakness, diminished appetite, nausea, vomiting, hives, muscular cramps, and changes in urinary output [9]. These symptoms may not become apparent until the disease advances. CKD comprises various stages, ranging from moderate to severe. The standard approach to managing CKD generally involves addressing the root cause, regulating blood pressure and blood sugar levels, limiting protein consumption, and administering drugs to mitigate potential problems, such as anaemia, bone disease, and cardiovascular ailments. In severe cases, kidney transplantation or dialysis may be required. Many people with CKD are unaware of their condition until it progresses to an advanced stage, primarily because of the absence of noticeable symptoms during the early phases of the disease [17]. The initial phase of CKD may be asymptomatic. Generally, the human body can experience a considerable reduction in kidney function [10]. Hence, unless routine screening for an unrelated ailment arises, such as blood or urine tests, kidney disease typically occurs in advanced stages.

It is important to promptly acknowledge a medical condition, as it can prevent CKD from progressing to a more serious stage. Timely detection and intervention in patients with CKD have advantages such as alleviating or halting the development of the condition and potentially avoiding the need for costly dialysis or kidney transplantation, both of which carry significant risks of morbidity and mortality. CKD may lead to various unfavourable consequences such as anaemia, skeletal abnormalities, and cardiovascular complications [5, 29]. Prompt identification and intervention can help individuals effectively manage these outcomes and enhance their overall well-being. Screening techniques allow for the timely identification of individuals with an increased risk profile, facilitating swift intervention measures. Prompt detection of CKD is crucial for preventing its progression, addressing related issues, determining and addressing root causes, and carrying out screenings for at-risk individuals. Machine learning (ML) can improve the prediction and diagnosis of CKD and its consequences while also enhancing treatment effectiveness in the initial stages of the disease [24, 33, 35]. This involves using ML algorithms trained on comprehensive patient datasets to detect risk factors and predict the likelihood of CKD development. In doing so, healthcare providers can implement timely interventions and hinder the progression of this ailment. Moreover, ML has the potential to identify specific indications in patient data that may suggest either CKD or its associated complications [12, 40].

Biomarkers play an important role in identifying and monitoring early disease progression. ML methods can be used to analyse patient information to customise treatment strategies. This process helps to identify patterns that guide personalised treatment plans, leading to benefits such as optimised dosage, reduced side effects, and improved patient outcomes. ML algorithms are not limited to customising pharmaceuticals, but also extend to the evaluation of medical imaging data, including ultrasound, computed tomography (CT) scans, and magnetic resonance imaging (MRI). These algorithms help to identify signs of kidney damage or illness, contributing to the timely identification and monitoring of CKD progression [8, 23]. ML has the potential to enhance our understanding of CKD and its effects, as well as to assess the effectiveness of treatments for patients with this condition. However, it is important to note that ML should be viewed as a supplementary tool alongside other diagnostic and therapeutic

approaches rather than as a replacement for clinical expertise.

Artificial intelligence (AI) models have great potential to detect and predict CKD and identify individuals at risk for unfavourable outcomes. However, the lack of interpretability and comprehensibility of these models presents challenges that limit their acceptance and effectiveness in the clinical environment [37, 44]. Explainable artificial intelligence (XAI) approaches can be used to address these challenges and enhance the transparency and interpretability of AI models [25]. This enables physicians to better understand how these models generate predictions and recommendations. For example, XAI can help identifying the main factors driving a model's predictions, analyse the relationship between different components, and clarify the decision-making process leading to the model's final output [27]. Implementing XAI techniques can improve the transparency and interpretability of AI models, assisting clinicians in making well-informed decisions about patient treatment and ultimately leading to improved outcomes for individuals with CKD.

This study employs two models designed to identify CKD, utilising the light gradient boosting machine (LightGBM) [19] and the gradient boosting decision tree (GBDT) [39, 41], as well as the dropout additive regression tree (DART) [38]. These models employ a dual-stage feature selection approach. Initially, important features associated with CKD were identified by assessing the relevance score of each feature. The utilisation of LightGBM is crucial for identifying the essential elements for classification procedures. In the secondary feature selection stage, a limited number of interpreted features are selected using the XAI technique [28]. The SHapley Additive exPlanations (SHAP) technique, which is a fundamental component of XAI, plays a central role in the feature selection procedure. Incorporating XAI elements into the detection process is believed to enhance the trust and confidence of medical professionals in the proposed model.

The contributions of this article are as follows. This study proposes the use of GBDT and DART as evolutionary ML models for identifying CKD. The identification procedure for CKD has been enhanced using the GBDT and DART models within the architecture of LightGBM. A dual-stage feature selection technique was proposed using both the LightGBM feature importance and the feature score provided by the SHAP explainable approach. In this analysis, feature importance and SHAP serve as dual-stage feature selection techniques. In this study, we empirically assessed a dataset containing 24 CKD-related features. The models built using the GBDT and DART showed impressive accuracy, sensitivity, and specificity.

Literature review

Several approaches have been proposed by various authors to identify CKD. Recently, different ML techniques have been explored for predicting CKD [17]. A study [17] gathered data from 660 individuals diagnosed with CKD and 440 healthy individuals. To analyse data and predict the occurrence of CKD, researchers have utilised multiple ML methods, such as logistic regression, decision trees (DTs), random forests (RFs), artificial neural networks (ANNs), and support vector machine (SVM) methods. According to these findings, the accuracy of the RF algorithm was 95.80% and the highest. The primary risk variables for CKD, included age, blood urea nitrogen (BUN), serum creatinine (SCr), uric acid (UA), and phosphorus (P). We developed a hybrid methodology for CKD identification by fusing the RF algorithm with the multi-objective firefly optimisation algorithm (MOFFA) in a different investigation [18]. The hybrid approach achieved 90.00% accuracy in predicting CKD, surpassing other algorithms. This powerful predictive tool has the potential to create individualised treatment regimens for patients with chronic diseases, thereby supporting

medical practitioners in enhancing patient care strategies. Similarly, [11] revealed an innovative strategy for applying ML algorithms to predict CKD. Patients with CKD who participated in the study provided information about their population demographics, laboratory test results, and medical history. Using feature selection techniques, ML algorithms have created predictive models for CKD forecasting, outperforming traditional techniques and achieving remarkable accuracy, suggesting that ML based models are valuable tools for timely identification and effective management.

Using different classifiers, such as DT, regression tree (RT), k-nearest neighbours (k-NN), and SVM, based on patient records ML can improve CKD diagnosis [30]. The authors pre-processed the data using feature-selection techniques and trained several classification models to produce prediction models for CKD diagnosis. This study suggests that merging ML approaches can significantly increase the efficacy and accuracy of CKD diagnosis. A corresponding investigation [20] assess the efficacy of ML methodologies in forecasting CKD by employing patient related data. Researchers have used many ML techniques to design CKD detection models. The results revealed that, compared to logistic regression and DTs, the SVM method exhibited the highest degree of accuracy in predicting chronic renal ailments. A study [7] significantly advanced the state of healthcare by creating a smart system that employs ML to predict and categorize CKD. Using a patient dataset, the program accurately forecasts CKD. DTs are the most effective techniques, demonstrating the potential of ML for accurate diagnostic tools and early detection and treatment.

In addition to individual ML algorithms, ensemble methods have been proposed to identify CKD. Recently, authors [15] suggested an ensemble approach for CKD detection. Their research utilised a dataset containing the demographic and clinical information of individuals with and without CKD. The ensemble design employed six classification algorithms: DT, RF, naive Bayes, k-NN, SVM, and ANN. The findings indicate that the performance of the ensemble classifiers, specifically the RF and ANN classifiers, surpassed that of the basic classifiers (DT, naive Bayes, k-NN, and SVM classifiers) in terms of the area under the curve receiver operating characteristic (AUC-ROC) score, sensitivity level, and F1-score calculation. Based on the findings of this study, we suggest that ensemble classifiers are more suitable for predicting CKD.

However, this ensemble method has several limitations. There is a limitation in terms of the size of the dataset used in this study [15], which may undermine its ability to accurately represent a wider population. Additionally, important variables, such as genetic susceptibility, lifestyle choices, and environmental factors, were not considered when predicting CKD. Future studies should address these issues in order to develop more accurate CKD prediction algorithms. Researchers conducted a study [32] using AI-based algorithms for CKD prediction and monitoring. This study used data pre-processing, ML-based algorithms, and interpretability analysis to improve the prediction of CKD progression. The AI algorithms achieved 87.10% and 81.90% accuracy, highlighting the importance of interpretation in clinical decision-making and patient outcomes.

CKD has been predicted using deep artificial neural networks (DANNs). A recent study compared the ANN and SVM models for CKD prediction [1]. According to the study findings, the ANN and SVM models had substantial predictive power for CKD detection. Regarding the accuracy and AUC, the SVM model performed better than the ANN model. DANN refers to a unique approach for predicting CKD [22]. Authors determined that the DANN technique had an amazing prediction accuracy rate of 92.50% and an exceptional sensitivity rate of 95.50%,

which examined a sample of 400 people who had been evaluated for CKD. These results demonstrate the intriguing potential of DANN for predicting CKD. DANN was used in a novel manner to effectively detect CKD [34]. This DANN model, which had 4 hidden layers was built using patient data from electronic health records, and was trained using a sizable dataset of more than 160 000 individuals. When performance was measured against several parameters, it became clear that the model could forecast CKD progression up to 1 year in advance. The authors recommend incorporating this sophisticated DANN model into the clinical decision-making processes to improve early CKD diagnosis and therapy. The authors suggested a method for detecting CKD using a multilayer perceptron algorithm. For datasets with an uneven distribution of classes, this method is intended to forecast CKD [42]. Only a small percentage of more than 3 000 participants with medical records included in the dataset for the study had symptoms of CKD. The authors used oversampling techniques on the minority class and trained a hybrid multilayer perceptron model to handle the issue of imbalanced data. Multiple metrics were used to assess the performance of the model and showed outstanding accuracy, sensitivity, and specificity in predicting CKD. According to the authors, the use of this multilayer perceptron model as a screening tool for early CKD detection in high-risk patients is promising.

Materials and methods

The operating procedure of the concepts of LightGBM [19] is briefly discussed in this section concerning GBDTs [41] and dropout meets DART [38]. LightGBM is a gradient boosting architecture that combines GBDT and DART techniques. This framework is superior to other implementations of gradient boosting techniques in that it is explicitly designed to increase the training speed and efficiency. GBDT and DART algorithms, chosen for their ability to successfully handle CKD, serve as the foundation for the proposed model. These features were identified using a feature importance analysis and explainable AI techniques. Following the utilisation of the dataset, a comprehensive elucidation was conducted, which was then succeeded by implementing dual-stage feature selection. The CKD detection module suggested in this study incorporates feature selection in two distinct steps. Consequently, each stage was explained individually, accompanied by its corresponding explanations.

Dataset

The dataset used in this study was derived from the Unified Configuration Interface machine learning repository (UCI ML repository) [22], which contains 400 samples, corresponding to a patient with 24 attributes. Attribute descriptions are presented in Table 1.

All the clinical features outlined in Table 1 are causally linked, either directly or indirectly, to the development of CKD. However, not all features are essential for CKD classification. The significance of feature selection lies in its ability to enhance the accuracy and efficiency of the classification model by reducing feature space dimensionality and eliminating irrelevant or redundant information.

Table 1. Chronic kidney disease dataset

Features		Mean	Std	Min	25%	50%	75%	Max
Age of subjects	(age)	51.54	16.98	2	42	55	64	90
Blood pressure	(bp)	76.58	13.51	50	70	80	80	180
Specific gravity	(sg)	1.02	0.01	1.005	1.01	1.015	1.02	1.025
Albumin	(al)	1.11	1.31	0	0	0	2	5
Sugar	(su)	0.51	1.05	0	0	0	1	5
Red blood cells	(rbc)	0.12	0.32	0	0	0	0	1
Pus cell	(pc)	0.19	0.39	0	0	0	0	1
Pus cell clumps	(pcc)	0.11	0.31	0	0	0	0	1
Bacteria	(ba)	0.06	0.23	0	0	0	0	1
Blood glucose random	(bgr)	131.75	88.00	0	93	114.5	150	490
Blood urea	(bu)	57.53	49.45	1.5	27	43	70.25	391
Serum creatinine	(sc)	3.08	5.63	0.4	0.9	1.3	3.05	76
Sodium	(sod)	136.86	9.33	4.5	134	136	141	163
Potassium	(pot)	4.70	2.82	2.5	4	4.7	5	47
Hemoglobin	(hemo)	12.37	2.79	3.1	10.8375	12.1	14.8	17.8
Packed cell volume	(pcv)	38.14	8.39	9	34	37	44	54
White blood cell count	(wbcc)	8353.28	2526.03	2200	6975	8248	9400	26400
Red blood cell count	(rbcc)	4.50	0.91	2.1	4	4.1	5.2	8
Hypertension	(htn)	0.37	0.48	0	0	0	1	1
Diabetes mellitus	(dm)	0.34	0.48	0	0	0	1	1
Coronary artery disease	(cad)	0.09	0.28	0	0	0	0	1
Appetite	(appet)	0.21	0.40	0	0	0	0	1
Pedal edema	(pe)	0.19	0.39	0	0	0	0	1
Anemia	(ane)	0.15	0.36	0	0	0	0	1

Stage I feature selection using feature importance

Within the framework of a classifier such as LightGBM, feature importance scores quantitatively assess the relative significance of each feature in facilitating precise predictions. These scores offer insights into the influence of each feature on the model's decision-making process, and are calculated using two primary approaches: split importance and gain importance. In the split importance method, the significance of feature i is quantified based on its frequency of use for splitting the dataset across all DTs in the model. Let T denote the total number of DTs and n_t denote the number of splits in tree t. The split importance S_i for feature i is mathematically expressed as follows:

$$S_i = \sum_{t=1}^{T} \sum_{j=1}^{n_t} \mathbb{I}(split\ at\ j\ in\ tree\ t\ uses\ feature\ i), \quad (1)$$

where $\mathbb{I}(split\ at\ j\ in\ tree\ t\ uses\ feature\ i)$ is an indicator function that equals *1* if the split uses feature i, and 0 otherwise. Features with a higher split importance are considered more relevant because they frequently partition the data, facilitating accurate predictions.

In contrast, the gain importance method evaluates the improvement in the model's performance, as measured by the loss function $\mathcal{L}$, owing to the splits involving a given feature. For a split j in tree t using feature i, the gain $G_{i,j}$ is calculated as follows:

$$G_{i,j} = \mathcal{L}_{before\ split} - \mathcal{L}_{after\ split}, \quad (2)$$

where $\mathcal{L}_{before\ split}$ is the loss before the split and $\mathcal{L}_{after\ split}$ is the weighted sum of the losses in the child nodes after the split. The total gain importance G_i for feature i is given by the following Eq. (3):

$$G_i = \sum_{t=1}^{T} \sum_{j=1}^{n_t} \mathbb{I}(split\ at\ j\ in\ tree\ t\ uses\ feature\ i)\ G_{i,j}. \quad (3)$$

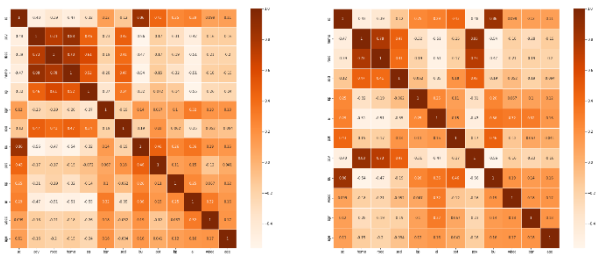
Features with higher gain importance are prioritised because they contribute significantly to reducing the loss function, thereby improving the model's predictive performance. These complementary approaches enable LightGBM to assign higher significance to features that play a critical role in segmenting data and enhancing prediction accuracy.

Two independent techniques, GBDT and DART, were used to identify the essential features associated with CKD. These approaches were applied using a LightGBM classifier. The feature significance scores are listed in Table 2. It is noteworthy that the DART algorithm chooses fewer features than the GBDT method. Given the fewer features, the analysis becomes more intriguing when considering whether DART's selective feature choice leads to maximum detection accuracy in the later detection phases. Examining DART's effectiveness of DART in feature selection and its potential influence on the final detection accuracy is an exciting area of inquiry.

Table 2. CKD features with feature importance scores generated by GBDT and DART

Features	**GBDT**	**DART**	**Features**	**GBDT**	**DART**
Age of subjects	1	1	Sodium	20	43
Blood pressure	4	26	Potassium	8	17
Specific gravity	38	0	Hemoglobin	55	80
Albumin	4	19	Packed cell volume	61	12
Sugar	0	0	White blood cell count	2	10
Red blood cells	0	0	Red blood cell count	55	77
Pus cell	0	0	Hypertension	0	0
Pus cell clumps	0	0	Diabetes mellitus	0	0
Bacteria	0	0	Coronary artery disease	0	0
Blood glucose random	23	9	Appetite	0	0
Blood urea	16	10	Pedal edema	0	0
Serum creatinine	106	89	Anemia	0	0

To assess the effectiveness of feature selection using feature importance ratings obtained from GBDT and DART, we generated a visual representation of the feature correlation matrix (Fig. 1). The sparsity pattern of the matrix is apparent, indicating a lack of interrelations among all features. In addition, the Stage I feature selection process, which is led by feature importance scores, successfully eliminates a significant portion of correlated features, strengthening their contribution to improving the capacity of the model to distinguish between different classes.



a) GBDT b) DART

Fig. 1 Feature correlation of selected features after Stage I feature selection

Stage II feature selection using SHAP as an explainable AI approach

At this stage, SHAP were employed to identify the most clinically relevant features in predicting CKD. The SHAP values quantify the contribution of each feature to model's prediction, enabling the selection of features based on their importance in a clinically interpretable manner.

The SHAP values are derived using concepts from cooperative game theory, in which ML model is considered as additive function. The model output $f(x)$ is decomposed into contributions from individual features as follows:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i, \quad (4)$$

where ϕ_0 is the base value, the average model prediction across the dataset; ϕ_i is the SHAP value for feature i, representing its contribution to the prediction; M is the total number of features.

For a specific feature i and data point x, the SHAP value is computed using the following:

$$SHAP(i,x) = \sum_{z \subseteq Z \setminus \{i\}} \frac{|z|!(|Z|-|z|-1)!}{|Z|!} [f(z_i, x) - f(z, x)], \quad (5)$$

where Z is the set of all features; z is the subset of features excluding feature i; $|z|$ is the size of the subset z; $f(z_i, x)$ is the model prediction where feature i is included in subset z; $f(z, x)$ is the model prediction when feature i is excluded.

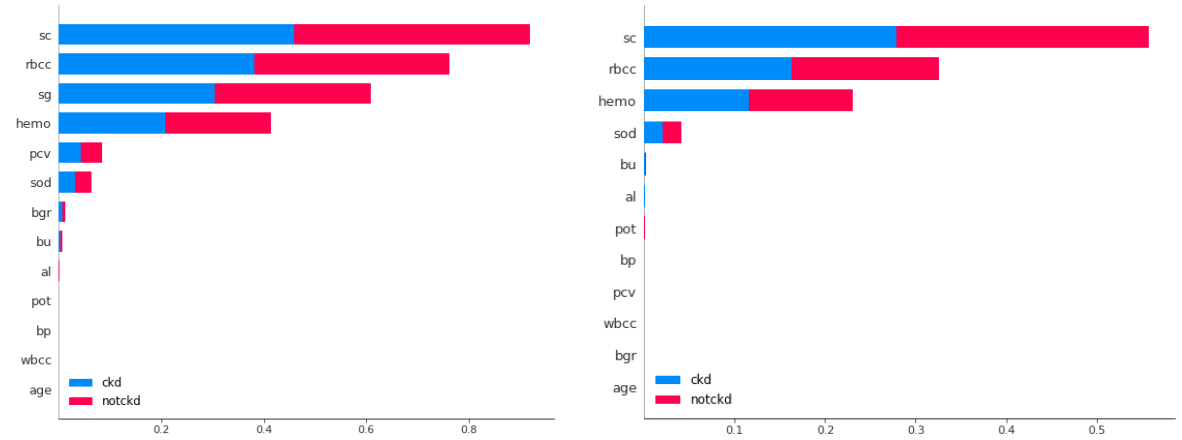
This summation averages the marginal contributions of feature i over all possible subsets z, weighted by subset sizes. To make the computation of SHAP values efficient, instead of evaluating all possible subsets of features (which are computationally expensive), a background dataset was used. The background dataset represents a sample of the feature space, against which the feature contributions are compared. Using this dataset, the SHAP value for a specific feature i and data point x is approximated as follows:

$$SHAP(i,x) = \frac{1}{S} \sum_{s=1}^{S} \left[f\left(z_i^{(s)}, x\right) - f\left(z^{(s)}, x\right) \right], \quad (6)$$

where S is the number of samples in the background dataset; $z_i^{(s)}$ is a binary vector indicating that feature i is present for sample s, whereas other features may be included or excluded; $z^{(s)}$ is a binary vector indicating that feature i is absent for sample s; $f\left(z_i^{(s)}, x\right)$ is the model's prediction when feature i is included for the data point x; $f\left(z^{(s)}, x\right)$ is the model's prediction when feature i is excluded from the same data point.
The approximation reduces the computational complexity by leveraging the background dataset S to evaluate the impact of each feature instead of iterating through all subsets of features in the dataset. Background datasets often consist of representative or sampled training data points. This makes SHAP feasible for real-world applications such as CKD prediction.

SHAP values serve as powerful tools for both local and global interpretabilities in feature selection [26, 40]. Locally, SHAP values explain the impact of each feature i on the model prediction $f(x)$ for a specific data point x. Globally, aggregating SHAP values across all data points provides a ranking of feature importance, enabling the identification of the most influential features [40]. These features were visualised in descending order of their SHAP scores, highlighting those with the greatest clinical relevance. For CKD prediction, applying SHAP enhances model interpretability by linking predictions to clinically significant

factors and focusing on key features, thereby reducing dimensionality and improving model efficiency without sacrificing performance. This systematic approach ensures that the model remains both predictive and clinically aligned, thereby offering a robust explainable AI framework for CKD detection. The SHAP values also provide a quantitative measure of the influence of each feature on the model forecast for a certain data point. These values can be used to elucidate the rationale behind a given prediction by emphasising the key characteristics and their corresponding effects on the prediction. Fig. 2 illustrates the features arranged in descending order based on their SHAP scores. Fig. 3 shows the force plot of each feature.



a) GBDT b) DART

Fig. 2 Features of GBDT and DART models are arranged according to their sharp values

a) GBDT

b) DART

Fig. 3 The force plot of various features of the models

The SHAP force diagram fully explains the ML model prediction. The prediction is broken down into distinct contributions from each feature, making it easier to explain the specific conclusions of the model. The force plots show each feature as a vertical bar. The bar size and direction indicate how much an attribute's value deviates from its baseline. When the bar expanded to the right, the prediction increased, and when it expanded to the left, it decreased. The features sc, bcc, sg, hemo, pcv, and sod contribute most to GBDT in our situation. The features sc, rbcc, hemo, and sod contributed the most to dropouts in DART. To make a choice, the decision-making modules received each feature separately.

The adoption of a two-stage feature selection process, utilising feature importance scores in Stage I and SHAP values in Stage II, markedly enhanced the model's performance compared to single-stage selection. Stage I, driven by GBDT and DART feature importance metrics, efficiently filtered the features based on their split and gain contributions, thereby eliminating redundant and less significant variables. This step ensures the initial reduction of noise and improves data segmentation for accurate predictions. However, single-stage selection methods often fail to account for the interpretability and contextual relevance of features in complex clinical datasets, such as CKD. By incorporating SHAP values in Stage II, the model refined its focus to clinically explainable and predictive features, ensuring that the selected features not only had statistical significance but also aligned with domain knowledge. This additional refinement eliminated residual correlated or redundant features, further reducing overfitting risks and enhancing the generalizability of the model. Together, the complementary strengths of the two stages synergistically improve the classifier's predictive accuracy and interpretability, demonstrating the robustness of this approach in identifying meaningful and impactful features.

Decision making using a LightGBM

LightGBM is a gradient boosting framework developed by Microsoft Research with the primary objective of achieving high speed, accuracy, and scalability. The use of many optimisation strategies enhances the efficiency of the training process and diminishes memory consumption, rendering it well-suited for datasets that are extensive in scale and high in dimensionality. LightGBM employs a gradient-based methodology to construct a collection of DTs to enhance the residual error of the preceding tree. Regarding working with sizable datasets with high-dimensional features, LightGBM has several benefits. This method involves converting continuous features into discrete features using histograms, which reduces the number of unique feature values and conserves memory. In addition, this algorithm builds DTs using a leafwise approach, which requires fewer nodes and splits while maintaining an accuracy comparable to that of the conventional level-wise method.

GBDT and dropouts that meet DART are methods employed within the LightGBM framework. The GBDT is a boosting technique that uses an iterative approach to construct DTs. The algorithm aims to minimise the gradient of the loss function throughout each iteration. The procedure was initiated by fitting an initial model to the dataset, followed by the progressive incorporation of further models. Each subsequent model aims to rectify errors made by its predecessor. The ultimate prediction is derived from the aggregate of all models with each model's contribution being weighted. GBDTs are widely recognised for their exceptional precision and have been extensively utilised in various practical domains, including but not limited to web search ranking and targeted advertising. To predict a given subject S_i whether the subject suffers from CKD can be determined by LightGBM (GBDT) as follows:

$$D_i = sigmoid\left(\sum\left[w_j * f_j(S_i)\right]\right), \quad (7)$$

where D_i is the predicted output for subject S_i. The sigmoid function that converts the sum of the predictions to a probability value between 0 and 1 is represented by $sigmoid\left(\sum\left[w_j * f_j(S_i)\right]\right)$. The weight assigned to the j^{th} tree in the ensemble is denoted as w_j. The prediction of the j^{th} tree for subject S_i is represented as $f_j(S_i)$. The weight assigned to each tree was dependent on its performance throughout the training process. The weight of a tree increases proportionally to its performance. The LightGBM employs a loss function, such as binary cross-entropy, to quantify the discrepancy between the predicted and actual values of the model during the training process. The weights of the trees in the ensemble are updated using the gradient of the loss function.

By contrast, DART can be regarded as a variation of GBDT that incorporates a novel regularisation approach known as dropout. The dropout technique was employed to eliminate a portion of the trees randomly during the training phase. This strategy effectively mitigated overfitting and enhanced the overall generalisation capability. In addition to its primary enhancements, DART incorporates supplementary advancements such as feature subsampling and a more adaptable tree construction approach, which can further enhance the accuracy of the model. DART algorithm demonstrates notable efficacy when confronted with datasets that exhibit high levels of noise or that possess numerous dimensions. Consequently, it has become a popular choice in the context of ML contests. DART presents the output of a given subject S_i as follows:

$$D_i = sigmoid\left(\frac{\sum[w_j * d_j * f_j(S_i)]}{\sum[w_j * d_j]}\right), \tag{8}$$

where d_j represents the dropout rate for the j^{th} tree, which is a value between 0 and 1 that determines the probability of dropping a node in the tree during the training process.

XGBoost [4] and CatBoost [14], two commonly used gradient boosting frameworks, have not been able to match the performance of LightGBM in terms of training speed and precision across a variety of test datasets. The successful application of this technique has also been demonstrated in various fields, such as image classification, object recognition, natural language processing, and biological data analysis. In this study, GBDT and DART from LightGBM were chosen over other ensemble techniques, such as XGBoost and CatBoost, because of their computational efficiency, adaptability, and interpretability. LightGBM's histogram-based learning technique significantly reduces the training and prediction times, enabling rapid experimentation and model optimisation. While the CKD dataset is not highly imbalanced, LightGBM's flexible boosting mechanisms and customizable loss functions provide robust handling of its characteristics. Additionally, DART mitigates overfitting by introducing tree dropouts during training, thereby enhancing generalisation without requiring extensive hyperparameter tuning. Importantly, LightGBM integrates seamlessly with SHAP, facilitating feature importance ranking and clinically meaningful insights. Comparative experiments demonstrated that LightGBM achieved comparable or slightly better predictive performance than XGBoost and CatBoost while maintaining lower computational costs. This combination of efficiency, accuracy, and interpretability made the GBDT and DART suitable choices for CKD detection in this study.

LightGBM technique iteratively builds CKD detection DTs. This promotes DT creation from samples with higher classification difficulty, improving the model's ability to detect complex patterns and correlations in the CKD dataset. LightGBM iteratively combines the predictions of many DTs to create a robust predictive model that can accurately identify CKD patients. Fig. 4 shows the LightGBM framework for CKD and non-CKD classification.

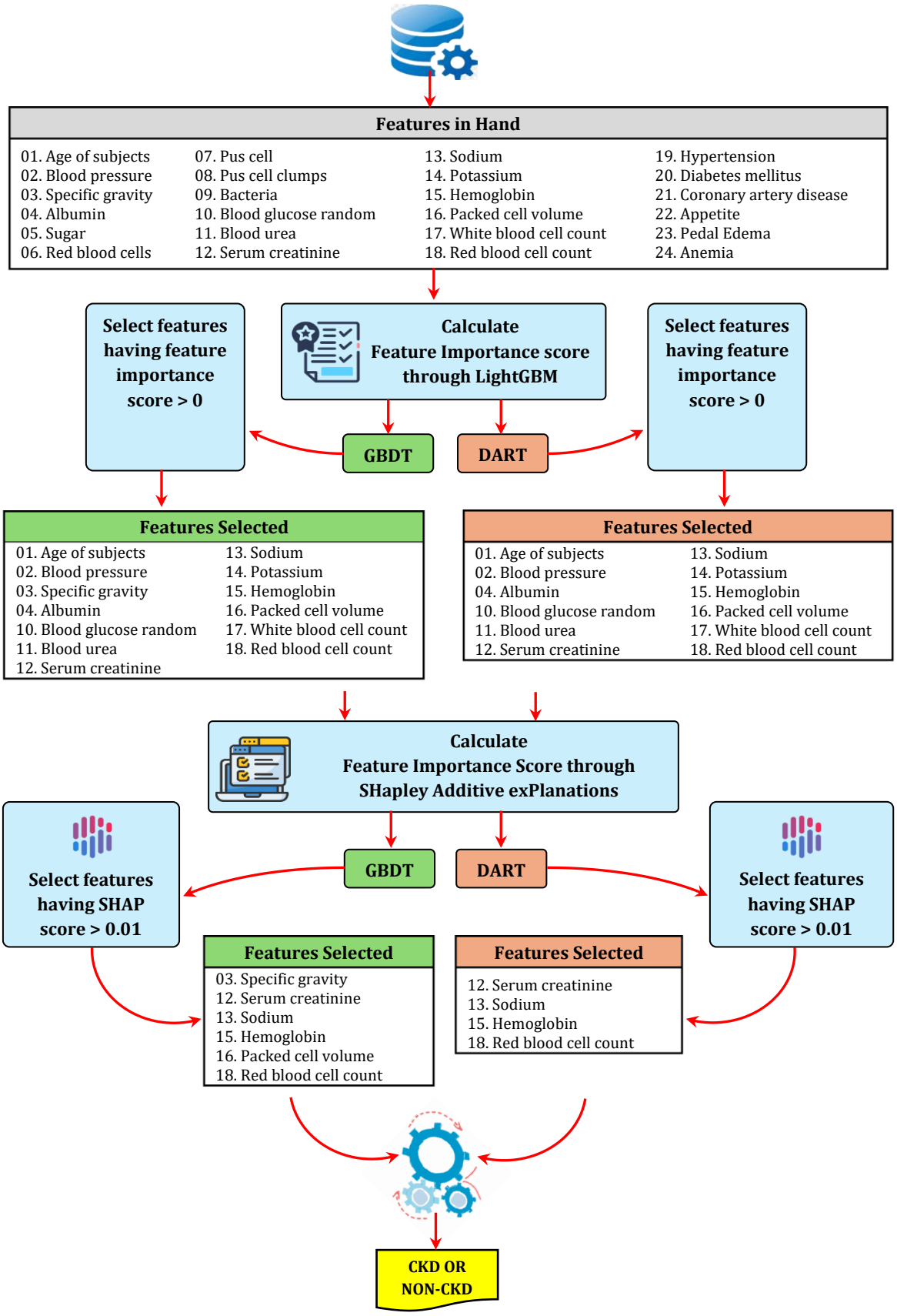


Fig. 4 The proposed LightGBM method for CKD detection

The information presented in Fig. 4 is readily understandable without further explanation. The CKD dataset used in this study consists of 24 unique features. The aforementioned features are designated as unique serial numbers for ease of understanding. To improve the detection process, a dual-stage feature selection strategy was introduced. During the preliminary phase of feature selection, the value of each feature is quantified using a feature importance score. Features with a score greater than zero were advanced to the next step of the selection process. The feature importance scores were determined using the LightGBM framework which incorporates both the GBDT and DART approaches. Notably, when utilising LightGBM with GBDT, 13 features were selected, whereas with DART, only 12 features were chosen. The selected attributes were subsequently arranged in descending order depending on their relevance scores. The planned ordering of the prioritising features with the highest relevance is processed first during the detection phase.

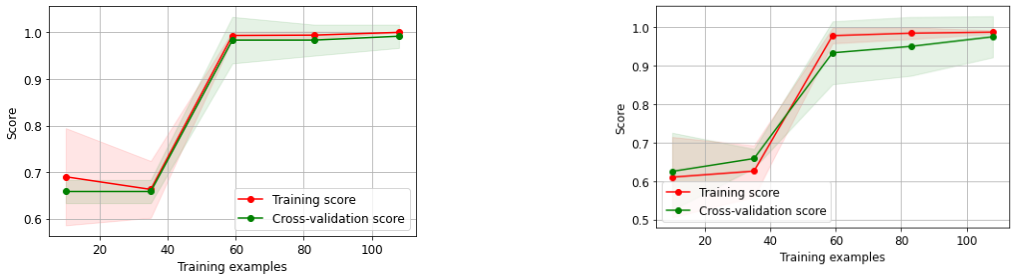
After the initial feature selection phase, the dataset was optimised using 13 features identified by the GBDT algorithm and 12 features selected by the DART algorithm. The last phase of feature selection is the assessment of SHAP scores for each feature. The SHAP scores provide valuable insights into the individual contributions of the features in model predictions. The LightGBM model, which incorporates the GBDT and DART, was utilised to compute the SHAP scores for each feature. Features with SHAP scores exceeding 0.01 were considered when developing the predictive model. In the context of GBDT, the SHAP algorithm identifies 6 features. Conversely, in the case of DART, the selection of features was based on declining SHAP scores, resulting in the identification of four features. The set of features derived from the aforementioned selection procedure was fed into the LightGBM detection model to obtain a precise diagnosis of CKD. This model incorporates the GBDT and DART techniques and has undergone thorough evaluation.

Results and discussion

The results of the proposed LightGBM-based CKD model were derived and analysed before and after feature selection.

Performance of the detection module before feature selection

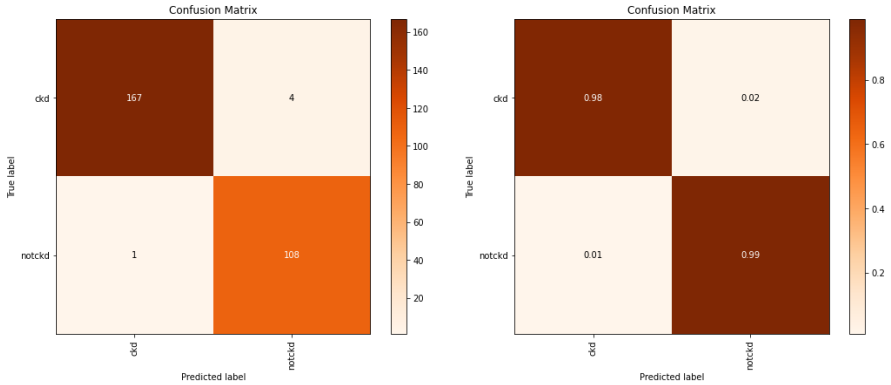
A series of experiments was conducted on the proposed framework to evaluate its effectiveness of the recommended model. These experiments were conducted both before and after the feature selection. Fig. 5 shows that the GBDT and DART models were utilized in the experiment before feature selection. Fig. 5 shows that for the GBDT model, the training score was 99.10% and the cross-validation score was 98.60%. Similarly, for the DART model, the training score was 98.30%, and cross-validation score was 95.70%, respectively. These findings indicate that the GBDT model outperformed the DART model in terms of performance.



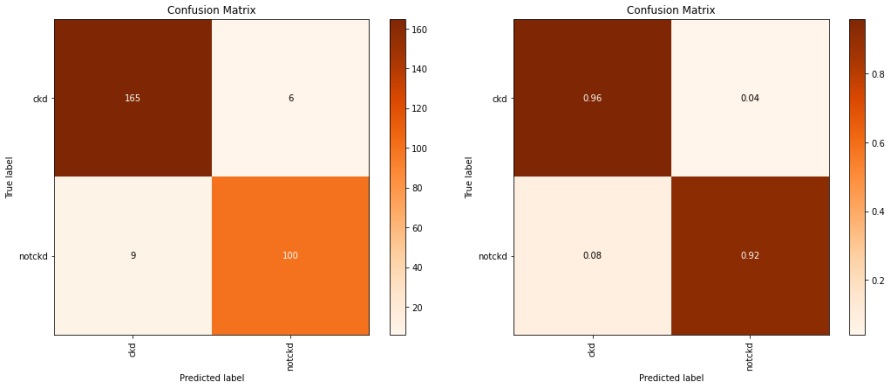
a) GBDT b) DART

Fig. 5 Learning curves of GBDT and DART on the original features before feature selection

Before selecting the features, the confusion matrices for both GBDT and DART were determined and are shown in Fig. 6. The confusion matrix utilized in this study offers valuable insights into the performance of GBDT model for identifying CKD. The model demonstrated exceptional accuracy by correctly identifying 167 patients as positive for the condition. However, it committed oversight by incorrectly designating only one case as positive when it was truly negative. Moreover, the model failed to include four instances in which the disease was present, while it accurately recognized 108 cases as negative for the condition. An evaluation of these data revealed that the GBDT model exhibited exceptional performance. The classification model demonstrated a high level of accuracy (approximately 98.21%), indicating its consistent and reliable performance in accurately categorizing the majority of the cases. Precision, which measures the degree of accuracy in positive predictions achieves remarkable accuracy of 99.40%. Recall, which quantifies the accuracy of properly identifying genuine positives, is noteworthy, with a value of 97.66%. Moreover, the F1-score, a metric that combines precision and recall, was approximately 98.52%. Similarly, when the original feature data were input into the DART model within the LightGBM ensemble architecture, the model successfully classified 165 occurrences as positive for the disease, demonstrating its ability to precisely identify individuals with CKD precisely. Nevertheless, the model also exhibited 9 instances of false positives, erroneously identifying healthy persons as having a condition. Moreover, the model failed to detect 6 instances in which individuals were afflicted with the disease, thereby emphasizing the potential hazards associated with undetected medical disorders. One notable advantage is that the model accurately classified 100 instances as negative for the condition, demonstrating its proficiency in discerning those who are free from ailments.



a) GBDT



b) DART

Fig. 6 Confusion matrix of GBDT and DART under the LightGBM framework before feature selection

The model had a commendable level of accuracy, with an overall classification rate of approximately 94.64%. This suggests that this approach can effectively and precisely categorize a substantial proportion of cases. A precision of approximately 94.83% indicates that the model's predictions of positive cases are typically accurate. The model exhibited a recall or sensitivity rate of approximately 96.51%, indicating its ability to accurately identify a significant proportion of true-positive cases. Finally, the F1-score, which balances precision and recall, was approximately 95.67%, suggesting a favourable compromise between these two parameters. The combined findings demonstrate the efficacy of the model in detecting CKD before the feature selection process. The GBDT model demonstrated notable proficiency in identifying CKD, surpassing the DART model in various crucial performance metrics before implementing any feature selection technique.

Table 3 presents the performance of the LightGBM (GBDT) and LightGBM (DART) models before feature selection. Examination of specificity and sensitivity, revealed that GBDT model outperformed DART model in identifying CKD. In terms of sensitivity, GBDT model had greater sensitivity (99.08%) than DART model (91.74%). This disparity indicates that GBDT is more effective at detecting individuals with this ailment. Additionally, GBDT model exhibited a greater degree of specificity (97.66%) compared with DART (96.49%), highlighting its superior ability to accurately categorise people as healthy. The results of this study indicate that GBDT strikes a more balanced compromise between decreasing false negatives and false positives, resulting in an enhanced F1-score and overall accuracy. Across key indices, the GBDT showed improved performance metrics, confirming its position as the preferred choice for identifying CKD within this specific context.

Table 3. Sensitivity and specificity analysis of GBDT and DART models for CKD detection before any feature selection

Performance measures	**GBDT**	**DART**
Specificity	97.66%	96.49%
Sensitivity	99.08%	91.74%

Performance of the post-stage I feature selection detection module

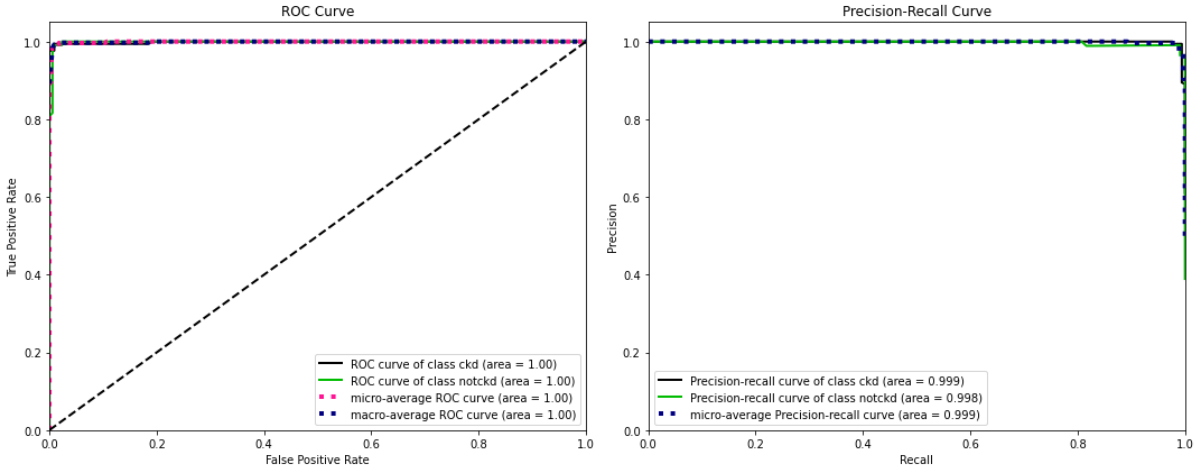
This study presents a complete examination of CKD identification of CKD utilising LightGBM frameworks, specifically employing GBDT and DART models. Following the initial feature-selection step, a meticulous review was performed. This study aimed to assess the significance of various features and consequently enhance the efficacy of the model in differentiating between patients with and without CKD. The overall outcomes are presented in Table 4.

Table 4. The detection ability of GBDT and DART models after stage I feature selection

Results	**GBDT**		**DART**	
	CKD	**Non-CKD**	**CKD**	**Non-CKD**
Precision	97.66%	99.08%	96.49%	91.74%
Recall	99.41%	96.43%	94.83%	94.34%
F1-score	98.53%	97.74%	95.65%	93.02%
Support	168	112	174	106
Accuracy	98.21%		94.64%	
Specificity	99.08%		96.49%	
Sensitivity	97.66%		91.74%	

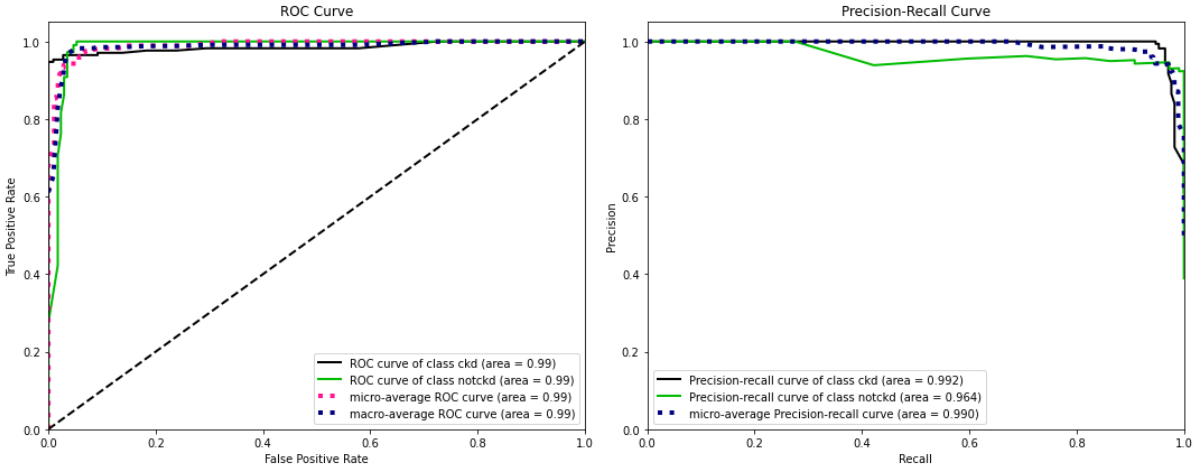
The GBDT model was initially employed to identify 13 potentially significant features. Subsequently, the model was used to distinguish patients with CKD. Remarkably, the model demonstrated a precision of 97.66%, signifying that its predictions of positive cases were accurate for nearly 98.00% of occurrences. Recall, also known as sensitivity, exhibited a notably high value of 99.41%, hence emphasising its efficacy in identifying a significant proportion of true-positive instances. This observation was further demonstrated by the F1-score, a statistical measure that combines precision and recall through a harmonic mean, yielding a substantial value of 98.53%. The predictions were substantiated by the model, which yielded a support value of 168. The performance of the model was equally promising for separating the non-CKD participants. It achieved a precision of 99.08%, which indicates its accuracy in making positive predictions. Additionally, it achieved a recall of 96.43%, which signifies its ability to properly identify the majority of genuine negative cases The F1-score for this class was 97.74%, substantiated by a support value of 112. The overall accuracy exhibited a significant level of performance, reaching 98.21%, thereby demonstrating its overall effectiveness. The specificity of the model, which indicates its accuracy in properly identifying negative situations, was 99.08%. Similarly, the model's sensitivity which represents its effectiveness in capturing positive cases, was 97.66%. Similarly, after undergoing stage I feature selection, which yielded 12 characteristics, DART model was proficient in detecting CKD. The precision, recall, and F-scores for patients with CKD were 96.49%, 94.83%, and 95.65%, respectively. The aforementioned measures, in conjunction with a support value of 174, highlight the model's proficiency in accurately discerning CKD cases. For individuals without CKD, the precision metric yielded a value of 91.74%, indicating the accuracy of correctly identifying non-CKD patients. Similarly, the recall metric achieved a value of 94.34%, indicating the ability to correctly capture a high proportion of non-CKD patients. The F1-score, which combines precision and recall, reached 93.02%, further highlighting the ability of the classification model to accurately categorise non-CKD patients. The model achieved an overall accuracy of 94.64%, a specificity of 96.49%, and a sensitivity of 91.74%.

A comparison of the two models revealed that the GBDT model demonstrated marginally superior performance in terms of recall, F1-score, and specificity in both CKD and non-CKD patients. Nevertheless, the DART model exhibited notable efficacy, particularly in individuals without CKD. Both models demonstrated exceptional accuracy, sensitivity, and specificity, emphasizing their potential for CKD detection. The analysis was augmented by the inclusion of ROC and precision-recall curves (PRCs). Fig. 7a and Fig. 8a show ROCs of GBDT and DART algorithms within the LightGBM framework. Fig. 7b and Fig. 8b illustrate PRCs for GBDT and DART, respectively. The graphical representations illustrate that the LightGBM framework has notable separability, as seen by the positioning of ROC and PRCs near the upper left corner. The observed performance was quantified by a probability value ranging from 0 to 1, with a tendency toward values closer to 1. The results depicted in Fig. 7a demonstrate the outstanding performance of the model, as evidenced by ROC value of 1 and PRC values approaching 1. The observed performance exceeded the threshold point of an AUC of 0.9 with a significance level of 0.5. The results of this study support the effectiveness of GBDT model in generating precise individual predictions for CKD datasets. Fig. 8 presents the analysis of ROCs and PRCs of DART model. The results demonstrated a noteworthy 99.00% performance for both CKD and non-CKD categories for ROCs and PRCs. In addition, the dataset included in DART model produced 99.20% ROC and 96.4% non-CKD PRC. Remarkably, the obtained outcomes exceeded the critical AUC value of 0.9, thus achieving statistical significance at the level of 0.5. Therefore, DART model has emerged as a strong competitor, providing superior individual ML predictions for CKD dataset. This highlights the capacity of the system to effectively classify and distinguish between cases in the field of CKD identification.



a) ROC b) PRC

Fig. 7 ROC and PRC of GBDT model after stage I feature selection using the feature importance score

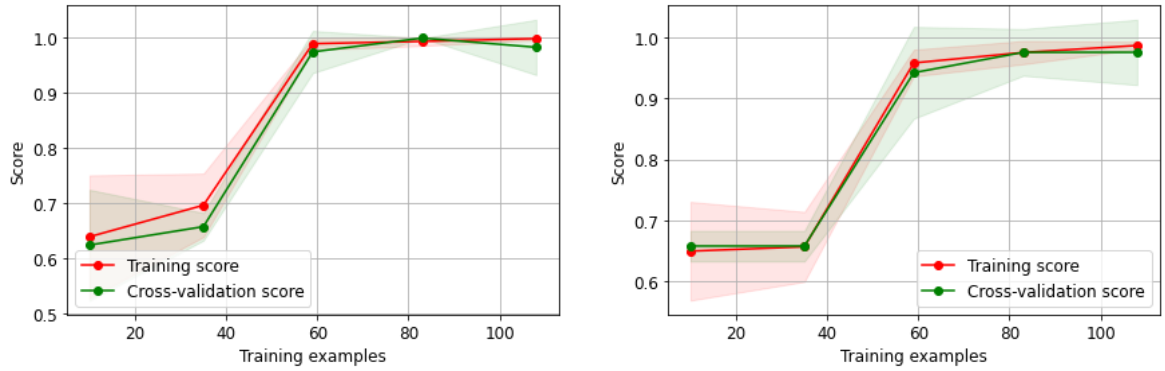


a) ROC b) PRC

Fig. 8 ROC and PRC of DART model after stage I feature selection using the feature importance score

Performance of the detection module after stage II feature selection

This subsection examines the results of the second feature-selection phase, which was conducted using SHAP model. The objective of SHAP model in this investigation was to provide an in-depth understanding of the key features that significantly affect CKD prediction. This was accomplished by quantifying the individual contribution of each variable to the prediction process. Subsequently, GBDT and DART models were used to predict CKD based on the selected features. Fig. 9 shows the learning curves obtained from GBDT and DART models implemented within the LightGBM ensemble framework. The findings demonstrate that GBDT model achieved a training score of 100% and a cross-validation score of 99.60%. On the other hand, DART model displayed training and cross-validation scores of 97.20% and 98.90%, respectively. Collectively, the results show how much better GBDT model performs than DART model in the particular situations under study.



a) GBDT b) DART

Fig. 9 Learning curves of GBDT and DART for the original features following stage II feature selection

Six significant and comprehensible features were found after employing SHAP to implement the stage II feature selection approach. These characteristics were then included in the LightGBM (GBDT) model to enhance the accuracy of distinguishing patients with CKD from healthy individuals. GBDT model exhibited notable performance in the provided scenario. The model successfully identified 98.83% of the patients with CKD, demonstrating high accuracy in its optimistic predictions. The model's accuracy in identifying CKD positive patients was excellent, reaching a perfect recall rate of 100%, which was also significant. The F1-score, which skilfully blends recall and precision into a single metric, attained a remarkable value of 99.41%.

The GBDT model has been proven to be extremely successful in identifying patients without CKD. Accurate predictions for non-CKD patients achieved an accuracy of 100%, indicating a remarkable degree of precision. Additionally, it had a recall rate of 98.20%, indicating that it can correctly identify the most real negative situations. The F1-score which considers both precision and recall, obtained an impressive value of 99.09%. This outcome underscores the ability of the model to precisely predict non-CKD patients. In addition, with an accuracy rate of 99.29%, the GBDT model attained a high level of overall accuracy in consistently classifying both CKD patients with and non-CKD patients. Its specificity score was 98.83%, demonstrating its capacity to successfully differentiate genuine negatives among non-CKD participants. Furthermore, with a sensitivity score of 100%, this model is incredibly effective for correctly categorising true positive patients with CKD.

Upon completing stage II feature selection, the SHAP analysis revealed that the LightGBM (DART) explainer identified a total of four interpretable features. These selected features were subsequently incorporated into DART model for accurate detection of CKD. Remarkably, DART model exhibited high performance in this regard, with an impressive precision of 95.32% and a recall of 99.39%. These outcomes provide strong evidence that the model is efficient in precisely detecting a noteworthy percentage of true positive patients with CKD. Moreover, when considering the F1-score as an indicator of the balance between precision and recall rates, it showed exceptional accuracy of 97.31%, confirming its ability to correctly identify instances of CKD. In patients without CKD, the results demonstrated a commendable precision of 99.08%, along with a recall rate of 93.10%. This shows how effectively DART model can predict cases where no CKD exists while also successfully identifying significant proportions where there is indeed no condition present. The performance of the model in this classification was demonstrated by the resulting F1-score of 96.00%.

According to Table 3, when evaluating the specificity and sensitivity, it is evident that GBDT model outperforms the DART model in identifying CKD. The GBDT model demonstrated a sensitivity of 99.08%, which was higher than that of DART model (91.74%). This difference implies that GBDT possesses a superior ability to effectively detect individuals suffering from this disease. Additionally, GBDT model showed greater specificity, with a value of 97.66%, than did DART model, with a value of 96.49%, indicating its enhanced accuracy in categorising individuals as healthy. These findings suggest that GBDT achieves an improved balance between reducing false negatives and false positives, thus leading to enhanced F1-scores and overall accuracy. Based on a comprehensive examination of the performance metrics and outcomes, GBDT model proved to be the optimal selection for the identification of CKD when employing SHAP explainable features. GBDT model demonstrated high precision and recall in identifying CKD patients, highlighting its accurate positive predictions and capacity to identify all true positive patients. A strong F1-score underscores its equitable effectiveness in mitigating both false positives and false negatives. Furthermore, GBDT model exhibits a remarkable level of accuracy, sensitivity, and precision in detecting non-CKD, thereby showing its constant and dependable performance in both CKD and non-CKD detection scenarios. The comprehensive performance observations are shown in Table 5.

Table 5. The detection ability of GBDT and DART models after stage II feature selection

Results	GBDT		DART	
	CKD	Non-CKD	CKD	Non-CKD
Precision	98.83%	100.00%	95.32%	99.08%
Recall	100.00%	98.20%	99.39%	93.10%
F1-score	99.41%	99.09%	97.31%	96.00%
Support	169	111	164	116
Accuracy	99.29%		96.79%	
Specificity	98.83%		95.32%	
Sensitivity	100%		99.08%	

Final results of the detection module (10-fold cross-validation)

We thoroughly assessed methodologies utilising SHAP enabled explainable features, integrated classification, and model-based techniques. The cross-validation results were evaluated using GBDT and DART algorithms. This evaluation included all input features and considered the features selected through the two steps of feature elimination, as illustrated in Table 6. A comparison analysis comparing the GBDT and DART models demonstrated the superiority of GBDT in effectively handling the problem, even in scenarios including several classes. Although the performance of the best-performing DART model was close to that of the GBDT model, this might be attributed to the inherent challenge of accurately predicting the precise number of relevant features. The DART model exhibited a cross-validation accuracy of 98.00%. Hence, both models demonstrate promise for utilisation in the classification of CKD datasets.

The findings in Table 6 highlight the greater accuracy, precision, specificity, and sensitivity of the GBDT model than those of the DART model. These results suggest that the GBDT model exhibits superior overall performance. Although the DART model demonstrates commendable performance, the consistently superior performance of the GBDT model indicates its ability to identify CKD accurately.

Table 6. Performance results obtained by GBDT and DART under SHAP explainable features through 10-fold cross-validation

Results	GBDT	DART
Accuracy	99.50%	98.00%
Precision	98.80%	96.20%
Specificity	100.00%	98.70%
Sensitivity	99.20%	97.60%

Comparison with existing models

Before conducting a comprehensive comparison, it is imperative to acknowledge the wide array of methodologies employed in predicting CKD, as outlined in the dataset. Different methods use diverse algorithms, methodologies, and approaches to address the intricate challenges of CKD identification. Measurements of accuracy, sensitivity, and specificity offer significant insights into the performance of these approaches. In the present context, LightGBM (GBDT) with dual-stage feature selection and LightGBM (DART) with dual-stage feature selection have been proposed for CKD detection. These unique methodologies incorporate a dual-stage feature selection technique that combines classifier feature importance and SHAP feature importance scores. The objective of this study was to improve the accuracy and interpretability of CKD prediction. Based on this understanding, we can now evaluate the efficacy and distinctiveness of the suggested methodologies for prevailing procedures for predicting CKD. Table 7 presents the methods used for the comparison and validation.

The comparative analysis presented in Table 7 highlights the superior performance of the proposed LightGBM models (GBDT and DART) with dual-stage feature selection compared to the baseline methods in terms of accuracy, sensitivity, and specificity. The proposed models outperform traditional ML approaches, such as regression-based models, k-NN, and DTs, largely because of the integration of a dual-stage feature selection process that identifies a minimal yet clinically relevant subset of features. For instance, the proposed GBDT model achieved an accuracy of 99.50%, sensitivity of 99.20%, and specificity of 100%, outperforming state-of-the-art methods, such as principal component analysis (PCA) combined with XGBoost or CatBoost, while using fewer features (5 vs. 10). Compared to deep learning approaches, such as convolutional neural network (CNN) with long short-term memory (LSTM), which also achieved high accuracy (99.17%), the proposed GBDT model demonstrated comparable or better results with significantly fewer features, making it computationally efficient. DART variant achieved a slightly lower performance (98.00% accuracy) but offered robustness and efficiency, addressing overfitting through tree dropouts.

Comparing the proposed methods with existing approaches reveals several surprising discoveries. Despite the simplicity of the latter, the proposed LightGBM (GBDT) model with dual-stage feature selection consistently outperformed the regression-based model across all evaluation criteria. The C4.5 DT demonstrated a slightly higher accuracy but lacked sensitivity and specificity. Furthermore, the LightGBM (GBDT) and LightGBM (DART) algorithms trained on dual-stage feature selection have demonstrated superior performance compared to the k-NN algorithm, thereby emphasising their capacity to capture complex patterns in CKD prediction effectively. Combining PCA with different classifiers exhibits diverse performance outcomes. However, the proposed LightGBM (GBDT) and LightGBM (DART) trained on dual-stage feature selection models regularly outperform these classifiers in terms of accuracy, sensitivity, and specificity. When comparing the proposed LightGBM models to ensemble techniques such as bagging, RF, and XGBoost, it was observed that the LightGBM models with

dual-stage feature selection demonstrated similar accuracy levels while outperforming in sensitivity and specificity. Within the domain of SVM, the methods that have been proposed provide superior levels of accuracy, sensitivity, and specificity compared with both Laplacian and radial basis function SVMs.

Table 7. Comparative analysis of the proposed method with existing state-of-the-art methods

Ref.	Methods	#f	Accuracy [%]	Sensitivity [%]	Specificity [%]
[30]	Regression-based model	24	98.95	98.44	99.80
[3]	C4.5 DT	24	99.00	99.60	98.00
[3]	k-NN	24	95.75	93.20	100.00
[17]	PCA and AdaBoost	10	98.33	97.30	100.00
[17]	PCA and DT	10	97.50	96.00	100.00
[17]	PCA and XGBoost	10	99.17	98.63	100.00
[17]	PCA and CatBoost	10	97.50	96.00	100.00
[17]	PCA and k-NN	10	59.17	64.94	48.84
[17]	PCA and RF	10	97.50	96.00	100.00
[17]	PCA and Naive bayes	10	88.33	93.94	81.48
[17]	PCA and LightGBM	10	98.33	97.30	100.00
[36]	RF	14	80.20	82.00	81.88
[36]	XGBoost	14	82.27	83.29	82.91
[16]	SVM (Laplacian)	24	91.71	94.84	86.71
[16]	SVM (Radial basis function)	24	89.95	87.94	91.27
[43]	CNN with LSTM	13	99.17	100.00	98.70
[6]	SVM with recursive feature elimination	9	95.50	92.20	98.8
[6]	DT with recursive feature elimination	16	98.60	97.70	99.40
[2]	Deep belief network with extreme ML	24	96.91	96.80	97.02
Proposed GBDT with dual-stage feature selection		5	99.50	99.20	100.00
Proposed DART with dual-stage feature selection		4	98.00	97.60	97.60

It has been shown that a CNN integrated with LSTM has a notable level of accuracy. However, GBDT and DART, which are trained on dual-stage feature-selection models, have been suggested to exhibit superior sensitivity and specificity. The accuracy and sensitivity of the proposed GBDT and DART with dual-stage feature selection approaches surpass feature elimination strategies, such as SVM with recursive feature elimination and DT with recursive feature elimination.

In summary, combining a deep belief network and extreme ML exhibits similar accuracy to the proposed LightGBM approaches. However, the LightGBM (GBDT) and LightGBM (DART) models trained in dual-stage feature selection methods outperformed the former in terms of sensitivity and specificity. These comparisons highlight the effectiveness of the LightGBM models described in this study when combined with dual-stage feature selection. These models demonstrated impressive performance in predicting CKD by utilizing classifier feature importance and SHAP feature importance scores. They exhibit robustness across multiple evaluation criteria, indicating their potential for accurate and reliable disease detection.

Although the CKD dataset used in this study was not highly imbalanced, imbalanced datasets are common in medical applications, including chronic disease prediction. In cases where class

imbalance is present, models such as LightGBM may exhibit bias toward the majority class, potentially affecting the performance in the minority class. To address this, techniques such as class weighting, oversampling, and undersampling can be employed to improve the handling of imbalanced data by the model. Additionally, using evaluation metrics, such as F1-score, precision, and recall, along with accuracy, can offer a more comprehensive assessment of model performance. Future studies will explore these methods to enhance the robustness and generalizability of the proposed models for imbalanced datasets.

The proposed LightGBM-based CKD detection models, with their high accuracy and interpretability, have significant potential for integration into healthcare settings, especially for the early detection and diagnosis of CKD. By identifying the key clinical and laboratory features that contribute to CKD risk, the model can assist healthcare professionals in making data-driven decisions, potentially reducing the need for invasive diagnostic procedures. Furthermore, the model's interpretability, facilitated by SHAP values, enables clinicians to understand the rationale behind predictions, fostering trust and adoption in clinical practice. This interpretability is crucial to ensure that the predictions are clinically accurate. The model's ability to process large patient datasets efficiently allows for timely identification of at-risk individuals, enabling early intervention and improving patient outcomes. Additionally, the model can be integrated into decision support systems, providing real-time predictions and personalized treatment recommendations, ultimately contributing to more accurate, efficient, and patient-centred care. These aspects make the proposed approach relevant in clinical practice as it could support healthcare providers in both preventive and diagnostic capacities.

Furthermore, LightGBM-based CKD detection models offer significant clinical value by accurately predicting CKD based on a minimal set of clinically relevant features. These models can be integrated into electronic health record (EHR) systems to provide real-time data-driven decision support for healthcare providers. For example, the system could flag high-risk patients during routine screenings, enabling early intervention and tailored treatment plans. The interpretable nature of the model, facilitated by SHAP-based explanations, ensures that healthcare providers can understand the rationale behind predictions. This transparency enhances trust in the model's recommendations and allows practitioners to cross-verify predictions with clinical judgment. Furthermore, the computational efficiency of the proposed approach enables rapid predictions even in resource-constrained settings, which is particularly beneficial in rural and underserved areas. By focusing on clinically significant features, the model minimizes data collection burden, ensuring that its integration does not disrupt existing workflows. Ultimately, the adoption of such models can improve patient outcomes by facilitating earlier diagnoses, optimising resource allocation, and supporting informed decision-making processes.

Conclusion

This study presents a detailed examination of the diagnosis of CKD using LightGBM models, leveraging evolutionary approaches such as GBDT and DART frameworks. The primary goal of our investigation was to determine the most effective model and combination of features that could reliably distinguish between individuals with CKD and those without CKD. Several key findings have emerged from meticulous experimentation and evaluation. Before applying the feature selection, the GBDT model consistently demonstrated superior performance compared to the DART model. This superiority was validated by detailed analyses of the learning curves and confusion matrices, emphasising the capacity of the GBDT model to make accurate predictions and robustly differentiate between CKD and non-CKD cases. After feature selection, the GBDT model maintained its dominance by achieving high precision, recall,

and F1-scores for both classes. Moreover, employing the SHAP model to evaluate feature importance not only enhanced the interpretability of the model but also improved its overall performance and reliability by ensuring that critical features contributed effectively to the prediction process. Beyond these observations, this study underscores the necessity of adopting robust ML models, such as GBDT, in CKD detection when coupled with advanced feature selection techniques. Although traditional metrics such as accuracy, precision, specificity, and sensitivity highlight the exceptional performance of the GBDT model, additional evaluations using other well-known objective criteria, such as the Matthews correlation coefficient (MCC) and AUC-ROC, further validated its reliability. The results from these metrics confirmed the robustness of the model, even under varying experimental conditions, thus enhancing its credibility as a practical diagnostic tool.

The consistent performance of our proposed CKD detection model across multiple experimental setups and rounds of feature selection reaffirms its potential as a reliable and accurate CKD detector. This study not only contributes to advancing ML applications in medical diagnostics but also highlights the importance of combining effective feature selection and robust modelling techniques to enhance predictive accuracy. By focusing on real-world applicability and reliability, our findings pave the way for future research and potential clinical integration of ML frameworks in CKD detection.

Data Availability: The chronic kidney disease dataset used in this study is derived from previously published work [22].

References

1. Almansour N. A., H. F. Syed, N. R. Khayat, R. K. Altheeb, et al. (2019). Neural Network and Support Vector Machine for the Prediction of Chronic Kidney Disease: A Comparative Study, Computers in Biology and Medicine, 109,101-111.
2. Alsuhibany S. A., S. Abdel-Khalek, A. Algarni, A. Fayomi, et al. (2021). Ensemble of Deep Learning Based Clinical Decision Support System for Chronic Kidney Disease Diagnosis in Medical Internet of Things Environment, Computational Intelligence and Neuroscience, 2021(1), 4931450.
3. Boukenze B., A. Haqiq, H. Mousannif (2017). Predicting Chronic Kidney Failure Disease Using Data Mining Techniques, International Symposium on Ubiquitous Networking, 701-712.
4. Chen T., C. Guestrin (2016). XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, 785-794.
5. Cunillera-Puértolas O., D. Vizcaya, M. J. Cerain-Herrero, N. Gil-Terrón, et al. (2022). Cardiovascular Events and Mortality in Chronic Kidney Disease in Primary Care Patients with Previous Type 2 Diabetes and/or Hypertension. A Population-based Epidemiological Study (KIDNEES), BMC Nephrology, 23(1), 376.
6. Debal D. A., T. M. Sitote (2022). Chronic Kidney Disease Prediction Using Machine Learning Techniques, Journal of Big Data, 9(1), 109.
7. Elhoseny M., K. Shankar, J. Uthayakumar (2019). Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease, Scientific Reports, 9(1), 9583.
8. Erickson B. J., P. Korfiatis, Z. Akkus, T. L. Kline (2017). Machine Learning for Medical Imaging, Radiographics, 37(2), 505-515.
9. Evans M., R. D. Lewis, A. R. Morgan, M. B. Whyte, et al. (2022). A Narrative Review of Chronic Kidney Disease in Clinical Practice: Current Challenges and Future Perspectives, Advances in Therapy, 39(1), 33-43.

10. Falke L. L., S. Gholizadeh, R. Goldschmeding, R. J. Kok, et al. (2015). Diverse Origins of the Myofibroblast-implications for Kidney Fibrosis, Nature Reviews Nephrology, 11(4), 233-244.
11. Gudeti B., S. Mishra, S. Malik, T. F. Fernandez, et al. (2020). A Novel Approach to Predict Chronic Kidney Disease Using Machine Learning Algorithms, 4th International Conference on Electronics, Communication and Aerospace Technology, 1630-1635.
12. Gulamali F. F., A. S. Sawant, G. N. Nadkarni (2022). Machine Learning for Risk Stratification in Kidney Disease, Current Opinion in Nephrology and Hypertension, 31(6), 548-552.
13. Guleria P., P. N. Srinivasu, M. Hassaballah (2024). Diabetes Prediction Using Shapley Additive Explanations and DSaaS Over Machine Learning Classifiers: A Novel Healthcare Paradigm, Multimedia Tools and Applications, 83 (14), 40677-40712.
14. Hancock J. T., T. M. Khoshgoftaar (2020). CatBoost for Big Data: An Interdisciplinary Review, Journal of Big Data, 7 (1), 1-45.
15. Hassan M. M., M. M. Hassan, S. Mollick, M. A. R. Khan, et al. (2023). A Comparative Study, Prediction and Development of Chronic Kidney Disease Using Machine Learning on Patients Clinical Records, Human-Centric Intelligent Systems, 3(2), 92-104.
16. Iftikhar H., M. Khan, Z. Khan, F. Khan, et al. (2023). A Comparative Analysis of Machine Learning Models: A Case Study in Predicting Chronic Kidney Disease, Sustainability, 15(3), 2754.
17. Islam M. A., M. Z. H. Majumder, M. A. Hussein (2023). Chronic Kidney Disease Prediction Based on Machine Learning Algorithms, Journal of Pathology Informatics, 14, 100189.
18. Kavi Priya S., N. Saranya (2024). An Effective Chronic Disease Prediction Using Multi-objective Firefly Optimisation Random Forest Algorithm, IETE Journal of Research, 70(1), 307-321.
19. Ke G., Q. Meng, T. Finley, T. Wang, et al. (2017). LightGBM: A highly Efficient Gradient Boosting Decision Tree, Advances in Neural Information Processing Systems, 30.
20. Khan B., R. Naseem, F. Muhammad, G. Abbas, S. Kim (2020). An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy, IEEE Access, 8, 55012-55022.
21. Kovesdy C. P. (2022). Epidemiology of Chronic Kidney Disease: An Update 2022, Kidney International Supplements, 12(1), 7-11.
22. Kriplani H., B. Patel, S. Roy (2019). Prediction of Chronic Kidney Diseases Using Deep Artificial Neural Network Technique, Computer Aided Intervention and Diagnostics in Clinical and Medical Images, 179-187.
23. Li Q., G. Luo, J. Li (2022). Evaluation of Therapeutic Effects of Computed Tomography Imaging Classification Algorithm-based Transcatheter Arterial Chemoembolization on Primary Hepatocellular Carcinoma, Computational Intelligence and Neuroscience, 2022(1), 5639820.
24. Liang P., J. Yang, W. Wang, G. Yuan, et al. (2023). Deep Learning Identifies Intelligible Predictors of Poor Prognosis in Chronic Kidney Disease, IEEE Journal of Biomedical and Health Informatics, 27(7), 3677-3685.
25. Linardatos P., V. Papastefanopoulos, S. Kotsiantis (2020). Explainable AI: A Review of Machine Learning Interpretability Methods, Entropy, 23(1), 18.
26. Lundberg S. M., S. I. Lee (2017). A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems, 30.
27. Miller T. (2019). Explanation in Artificial Intelligence: Insights from the Social Sciences, Artificial Intelligence, 267, 1-38.

28. Nohara Y., K. Matsumoto, H. Soejima, N. Nakashima (2019). Explanation of Machine Learning Models Using Improved Shapley Additive Explanation, Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, 546-546.
29. Palaka E., S. Grandy, H. van Haalen, P. McEwan, et al. (2020). The Impact of CKD Anaemia on Patients: Incidence, Risk Factors, and Clinical Outcomes – A Systematic Literature Review, International Journal of Nephrology, 2020(1), 7692376.
30. Qin J., L. Chen, Y. Liu, C. Liu, et al. (2020). A Machine Learning Methodology for Diagnosing Chronic Kidney Disease, IEEE Access, 8, 20991-21002.
31. Sanmarchi F., C. Fanconi, D. Golinelli, D. Gori, et al. (2023). Predict, Diagnose, and Treat Chronic Kidney Disease with Machine Learning: A Systematic Literature Review, Journal of Nephrology, 36(4), 1101-1117.
32. Schena F. P., V. W. Anelli, D. I. Abbrescia, T. Di Noia (2022). Prediction of Chronic Kidney Disease and Its Progression by Artificial Intelligence Algorithms, Journal of Nephrology, 35(8), 1953-1971.
33. Senan E. M., M. H. Al-Adhaileh, F. W. Alsaade, T. H. Aldhyani, et al. (2021). Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques, Journal of Healthcare Engineering, 2021(1), 1004767.
34. Singh V., V. K. Asari, R. Rajasekaran (2022). A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease, Diagnostics, 12 (1), 116.
35. Sobrinho A., A. C. D. S. Queiroz, L. D. Da Silva, E. D. B. Costa, et al. (2020). Computer-aided Diagnosis of Chronic Kidney Disease in Developing Countries: A Comparative Analysis of Machine Learning Techniques, IEEE Access, 8, 25407-25419.
36. Song W., Y. Liu, L. Qiu, J. Qing, et al. (2023). Machine Learning-based Warning Model for Chronic Kidney Disease in Individuals Over 40 Years Old in Underprivileged Areas, Shanxi Province, Frontiers in Medicine, 9, 930541.
37. Tang R., S. Zhang, C. Ding, M. Zhu, et al. (2022). Artificial Intelligence in Intensive Care Medicine: Bibliometric Analysis, Journal of Medical Internet Research, 24(11), e42185.
38. Vinayak R. K., R. Gilad-Bachrach (2015). DART: Dropouts Meet Multiple Additive Regression Trees, Artificial Intelligence and Statistics, 489-497.
39. Wang J., P. Li, R. Ran, Y. Che, et al. (2018). A Short-term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree, Applied Sciences, 8(5), 689.
40. Wu I. W., T. H. Tsai, C. J. Lo, Y. J. Chou, et al. (2022). Discovering a Trans-omics Biomarker Signature that Predisposes High Risk Diabetic Patients to Diabetic Kidney Disease, NPJ Digital Medicine, 5(1), 166.
41. Xuan P., C. Sun, T. Zhang, Y. Ye, et al. (2019). Gradient Boosting Decision Tree-based Method for Predicting Interactions Between Target Genes and Drugs, Frontiers in Genetics, 10, 459.
42. Yildirim P. (2017). Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction, 41st Annual Computer Software and Applications Conference, 2, 193-198.
43. Yildiz E. N., E. Cengil, M. Yildirim, H. Bingol (2023). Diagnosis of Chronic Kidney Disease Based on CNN and LSTM, Acadlore Transactions on AI and Machine Learning, 2(2), 66-74.
44. Zhang K., X. Liu, J. Xu, J. Yuan, et al. (2021). Deep-learning Models for the Detection and Incidence Prediction of Chronic Kidney Disease and Type 2 Diabetes from Retinal Fundus Images, Nature Biomedical Engineering, 5(6), 533-545.

Assist. Prof. Moumita Pramanik, Ph.D.
E-mail: moumita.pramanik@gmail.com

Dr. Moumita Pramanik currently serves as an Assistant Professor in the Department of Computer Applications at Sikkim Manipal Institute of Technology, India. She holds a Ph.D. Degree in Computer Applications from Sikkim Manipal University, India. Her research focused on machine learning and data science in the field of biomedical engineering. Her work is particularly noted for its innovative approaches to addressing class imbalance and improving classification accuracy in neurodegenerative disease detection. Dr. Pramanik's contributions to the field have been recognised through numerous publications and interdisciplinary collaborations, where she continues to push the boundaries of computational intelligence and its real-world applications.

Assist. Prof. Ranjit Panigrahi, Ph.D.
E-mail: ranjit.panigrahi@gmail.com

Dr. Ranjit Panigrahi is an accomplished academician and a researcher in the field of computer applications, specialising in biomedical data analysis and cyber security. Dr. Panigrahi's has a M.Sc. Degree of Technology in Computer Sciences and Engineering from Sikkim Manipal Institute of Technology in 2013. He has a Ph.D. Degree in Computer Applications from Sikkim Manipal University in 2020. Dr. Panigrahi is an Assistant Professor in the Department of Computer Applications at Sikkim Manipal Institute of Technology, Majitar, Sikkim, India since 2010. Currently he serves as the Head of the IT Council at Sikkim Manipal Institute of Technology, India. He has numerous publications in reputed journals and conferences, and published many articles in SCI and Scopus-indexed journals. He is a sought-after reviewer for various technical publications and has chaired sessions at prominent international conferences. Beyond his academic achievements, Dr. Panigrahi has also actively engaged in intellectual property research and secured five patents.

Prof. Bidita Khandelwal, M.D.
E-mail: bidita.k@smims.smu.edu.in

Dr. Bidita Khandelwal is a distinguished Professor and Head of the Department of Medicine at Sikkim Manipal Institute of Medical Sciences (SMIMS), Sikkim Manipal University, Sikkim, India, where she plays a pivotal role in shaping the academic landscape. She imparts knowledge to both undergraduate and postgraduate students and coordinates postgraduate studies. Dr. Khandelwal is a Ph.D. supervisor and an influential member of the Governing Council and Academic Senate at Sikkim Manipal University. Her involvement extends to the Pre-publication Review Committee and the Board of Study of the Department of Psychiatry at SMIMS. Additionally, she is an active member of the Departmental Research Committees for Medicine and Psychiatry at SMIMS. Dr. Khandelwal currently teaches MBBS students subjects such as Genetics and Chromosomal Diseases, Fluids and Electrolytes, the Renal System, and Physical Medicine. Dr. Khandelwal has a MBBS Degree and a M.D. Degree in Medicine from Silchar Medical College, Silchar, India. She has held positions ranging from Senior Resident in Neurology at Institute of Human Behaviour and Allied Sciences, New Delhi, India to Assistant Professor, Associate Professor, and now Professor in Medicine at Central Referral Hospital and SMIMS, Sikkim, India. Her contributions highlight her dedication and expertise in the field of medicine. Dr. Khandelwal has an extensive publication record, contributing significantly to medical literature. Her research spans various areas of medicine, focusing on genetics, renal systems, and physical medicine. She has published numerous papers in reputable peer-reviewed journals, providing valuable insights and advancing knowledge in her field. Dr. Khandelwal's work is frequently cited by her peers, highlighting the impact and importance of her research contributions.

Joseph Bamidele Awotunde, Ph.D.
E-mail: awotunde.jb@unilorin.edu.ng

Joseph Bamidele Awotunde is a Lecturer in the Department of Computer Science, Faculty of Communication and Information Sciences, University of Ilorin, Ilorin, Nigeria. His area of research interest cuts across Artificial Intelligence, Internet of Things, Cybersecurity, Information Security, Social Computing, Bioinformatics, and Biometrics. He has over 150 publications in reputable publishers like Elsevier, Springer, and Hindawi, among others, covering journals, edited conference proceedings, and chapters in books. He was part of the team that won the Artificial Intelligence for Females in Science, Technology, Engineering and Mathematics (AI4FS) Grant sponsored by the Royal Academy of Engineering (Higher Education Partnerships in sub-Saharan Africa (HEP SSA) 22/24). Dr. Awotunde is a member of the International Association of Engineers and Computer Scientists (MIAENG), the Computer Professional Registration Council of Nigeria (MCPN) the Nigerian Computer Society (MNCS), Internet Society, etc.

Akash Kumar Bhoi, Ph.D.
E-mail: akashkrbhoi@gmail.com

Akash Kumar Bhoi (B.Tech, M.Tech, Ph.D.) is listed in the World's top 2% scientists for single-year impact for the year 2022 (compiled by John P. A. Ioannidis, Stanford University and published by Elsevier BV) and currently associated with Directorate of Research, Sikkim Manipal University, India as Adjunct Research Faculty and also working as a Research Associate at Wireless Networks (WN) Research Laboratory, Institute of Information Science and Technologies, National Research Council, Pisa, Italy. He was appointed as the honorary title "Adjunct Fellow" of Institute for Sustainable Industries and Liveable Cities, Victoria University, Melbourne, Australia for the period from 1 August 2021 to 31 July 2022. He was the University Ph.D. Course Coordinator for "Research and Publication Ethics" at SMU, India. He is a former Assistant Professor (SG) of Sikkim Manipal Institute of Technology and served for about 10 years. He is a member of IEEE, ISEIS, and IAENG, an Associate Member of IEI, UACEE, and an editorial board reviewer of Indian and International journals. He is also a regular reviewer of reputed journals, namely IEEE, Springer, Elsevier, Taylor and Francis, Inderscience, etc. His research areas are biomedical technologies, internet of things, computational intelligence, antenna, and renewable energy. He has published several papers in national and international journals and conferences, and over 150 publications have been registered in the Scopus database. He has also served on many organising panels for international conferences and workshops. He is currently editing several books with Springer Nature, Elsevier, Routledge and CRC Press. He is also serving as a Guest Editor for special issues of journals like Springer Nature, Wiley Hindawi, and Inderscience.