Journal of Current Science and Technology, January – March 2025 Vol. 15 No. 1, Article 76

ISSN 2630-0656 (Online)

Cite this article: Pechprasarn, S., Wetchasit, P., & Pongsuwan, S. (2025). Optimizing chronic kidney disease prediction: a machine learning approach with minimal diagnostic predictors. *Journal of Current Science and Technology*, *15*(1), Article 76. https://doi.org/10.59796/jcst.V15N1.2025.76

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Optimizing Chronic Kidney Disease Prediction: A Machine Learning Approach with Minimal Diagnostic Predictors

Suejit Pechprasarn[1,*], Peeraya Wetchasit[2], and Suphornthip Pongsuwan[2]

[1]College of Biomedical Engineering, Rangsit University, Pathum Thani 12000, Thailand
[2]Satriwithaya School, Wat Bowon Niwet, Phra Nakhon, Bangkok 10200, Thailand

[*]Corresponding author; E-mail: suejit.p@rsu.ac.th

Received 4 March 2024; Revised 15 April 2024; Accepted 16 May 2024; Published online 24 December 2024

Abstract

Chronic kidney disease (CKD) is a major public health issue that necessitates accurate diagnostic methods for effective management. This study involved training an open-source clinical dataset of 200 patients from Enam Medical College, comprising 28 clinical features, obtained from the UCI machine learning repository. After preprocessing to ensure a balanced dataset for objectivity, the data was split into training and testing sets in an 80:20 ratio. The research trained 22 machine learning models, including Naïve Bayes, decision trees, support vector machines (SVM), logistic regression, ensemble methods, kernel models, and neural networks. These models were evaluated using several metrics-accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic (ROC) curve-computed through 5-fold cross-validation to assess their performance and ensure they were not overfitting or underfitting. The best-performing model was the Kernel Naïve Bayes, achieving a 96.55% accuracy, 95% precision, 98.28% recall, and 96.61% F1-score on the training dataset. For the test dataset, it showed a slight performance drop but remained robust with 92.86% accuracy, 87.50% precision, 100% recall, and 93.33% F1-score. Furthermore, feature selection techniques such as minimum-redundancy-maximum-relevance, Chi2, ANOVA, and Kruskal-Wallis tests were used to determine the most significant predictors. It was found that only four features-packed cell value, stages of glomerular filtration rate, specific gravity of urine, and albumin content in urine-were necessary for maintaining similar model performance. This systematic approach not only highlighted critical clinical features but also helped in simplifying the model complexity, which could benefit broader medical applications like lung cancer screening by reducing screen time, resources, and medical costs.

Keywords**:** *chronic kidney disease classification; chronic kidney disease; machine learning; feature selection methods; artificial intelligence*

1. Introduction

Chronic kidney disease (CKD), often known as chronic kidney failure, is a medical disorder with substantial morbidity and no cure wherein the kidneys lose their ability to function optimally (Kalantar-Zadeh et al., 2021). Consequently, waste products and excess fluids accumulate in the bloodstream, leading to electrolyte imbalances. If these imbalances are left unaddressed, they can contribute to various health complications such as heart disease, anemia, and nausea, and may ultimately result in the complete cessation of kidney function (Drawz, & Rahman, 2015). It poses significant global health challenges, affecting countless individuals worldwide, with the global estimated prevalence of CKD being 13.4% and between 4.902 and 7.083 million individuals requiring renal replacement therapy (Lv, & Zhang, 2019).

Additionally, there are various forms of kidney diseases, including Polycystic kidney disease, Lupus nephritis, Interstitial nephritis, Glomerulonephritis, APOL1-mediated kidney disease, conditions associated with long-lasting viral illnesses, and Pyelonephritis (Murphy et al., 2016). Each of these conditions presents distinct challenges and implications for kidney health. Signs and symptoms of CKD can develop over time depending on how severe it is and can cause vomiting, urinating more or less, chest pain, and swelling of feet and ankles (Rainey, 2019).

However, patients with CKD are usually asymptomatic until the final stage, so early diagnosis becomes crucial (George et al., 2022). According to The National Kidney Foundation, numerous clinical practice guidelines advocate for a risk-based screening approach, especially for people over 60 years old or with a history of diabetes (Levey et al., 2003). The utilization of artificial intelligence (AI) in diagnosing CKD marks a paradigm shift in healthcare. AI algorithms, particularly machine learning models, exhibit significant potential in analyzing datasets, including patient records and lab results, to identify early signs of CKD. These advanced technologies offer a more efficient and accurate detection method than traditional methods (Sawhney et al., 2023).

CKD is often hard to detect early due to reliance on traditional diagnostic methods like serum creatinine and proteinuria, which have limitations. CKD's asymptomatic early stages delay diagnosis until it's advanced (Khwanchum et al., 2024). Artificial intelligence, especially machine learning, offers promise by analyzing complex data to detect subtle CKD patterns earlier (Durga &, Karthikeyan, 2023). However, challenges like selecting relevant features and translating predictions into clinical action remain. This study addresses these challenges by evaluating machine learning models and feature selection methods for efficient and early CKD detection

According to Islam et al., (2020), their study examines twelve machine learning-based classifiers, achieving an accuracy of 0.98, as well as precision, recall, and an F1 score of 0.98 for the XgBoost classifier. Haratian et al., (2022) developed several models, with their optimal models being random forest and LightGBM, which can achieve an AUC of 0.90 and an accuracy of 0.74. Ashafuddula et al., (2023) established the dominance of adaptive boosting, logistic regression, and passive-aggressive ML classifiers with 96.48% accuracy. Dritsas, & Trigka (2022) aimed to build efficient tools for predicting CKD occurrence, following an approach that exploits ML techniques and highlights the Rotation Forests (RotF), which prevailed in comparison to other models with an AUC of 100%, a precision, a recall, an F1-score, and an accuracy equal to 99.2%. However, the research on CKD detection is based on a single dataset published in the UCI Machine learning repository. In addition to enhancing classification performance using this dataset, this study attempted to reduce the number of input characteristics using feature selection methods and to construct a machine learning-based model that achieved the best accuracy. A total of 22 different machine learning-based classifiers were evaluated.

The growth of artificial intelligence (AI) and machine learning (ML) in the field of chronic kidney disease diagnosis has resulted in the development of various models (Pradeepa, & Jeyakumar, 2022; Dubey et al., 2023), each claiming to be superior in terms of accuracy, precision, and efficiency (Iftikhar et al., 2023). However, a large gap exists in systematically evaluating these models' relative efficacy in CKD prediction (Nishat et al., 2018). Prior research has frequently concentrated primarily on individual or a small number of models, preventing a comprehensive understanding of their performance under varied settings and datasets. A comparative study is necessary for a variety of reasons. First, it facilitates it easier to identify the most successful models for CKD prediction based on various parameters. Second, it assists researchers in picking the best model depending on study objectives or data features. Finally, it promotes innovation by identifying opportunities for model improvement and stimulating the development of novel approaches (Kumari & Singh, 2022). To close this gap, we thoroughly tested 22 machine learning models, ranging from classical methods to more complex ensemble and kernel models. This comprehensive comparison identifies the most effective CKD prediction models and evaluates their scalability and real-world applications. Through this approach, we hope to advance the field of CKD diagnostics towards early, accurate, and efficient detection techniques, ultimately improving patient outcomes and management approaches.

This study employs feature selection methods and ML models to identify the essential predictors that should be considered when diagnosing CKD. Strategic feature selection is critical for improving ML model efficacy in CKD diagnosis, aiming to reduce complexity and increase clinical applicability. The existing literature lacks integration of various feature selection strategies. Our unique approach uses

numerous techniques to prioritize diagnostic indicators, improving prediction accuracy and interpretability. This discovery closes a major gap in CKD diagnostics, promising to accelerate early detection and management. The prediction of CKD using ML models is particularly efficient, especially when boosting methods are used. A 5-fold cross-validation procedure was used to quantify precision, recall, F1 score, and accuracy. Supervised machine learning models were trained to assess their prediction accuracy through an unseen test dataset to ensure that the trained models were generalized. In this paper, we utilize the dataset from a publicly available database from the University of California Irvine's website (UCI Machine Learning Repository website) to analyze a dataset of 200 individuals with CKD (Islam et al., 2020). We utilize various ML models to predict chronic kidney disease and highlight the most relevant links to the disease.

2. Objectives

To enhance the diagnosis of chronic kidney disease (CKD) by developing and evaluating machine learning models that accurately classify CKD and non-CKD cases. This involves curating and cleaning the dataset to ensure unbiased training, identifying crucial predictors to reduce model complexity, and demonstrating the potential impact of these simplified models on clinical practice.

3. Materials and Methods

This study started from data curation and data splitting into training and test datasets, followed by data preparation for supervised classification tasks and the application of feature selection methods. There is no widely recognized standard criterion in machine learning to determine whether a model outperforms another based on dataset analysis, as there is no one-size-fits-all solution. All 22 models available in MATLAB R2022b have been included here. The flow of this research is depicted in Figure 1.

3.1 Dataset Details

The CKD dataset examined in this paper was obtained from the UCI database on January 1st, 2024. It includes data from 200 observations, including 28 predictors and 1 label, to identify risk factors of CKD. Although we acknowledge that the obtained dataset for this work was available under a Creative Commons (CC) license, we know that ethical issues go beyond the license. Although a study of publicly available data does not require ethical approval, we took preemptive steps to guarantee the highest ethical standards surrounding patient data privacy. We have carefully followed the ethical standards for research involving human subjects and deleting any potentially identifiable information from the dataset.

The study encounters potential limitations, including biases in the dataset, which will be addressed by preparing a balanced training dataset, constraints of machine learning models, and uncertainties about the broader applicability of the results. These could impact the model's accuracy. Furthermore, the complexity of certain machine learning approaches presents challenges in interpretation, particularly in clinical settings. Concerns arise regarding how well the findings can be applied across different contexts. Future research should focus on gathering more diverse datasets, improving model interpretability, and validating findings with external data to enhance their generalizability.

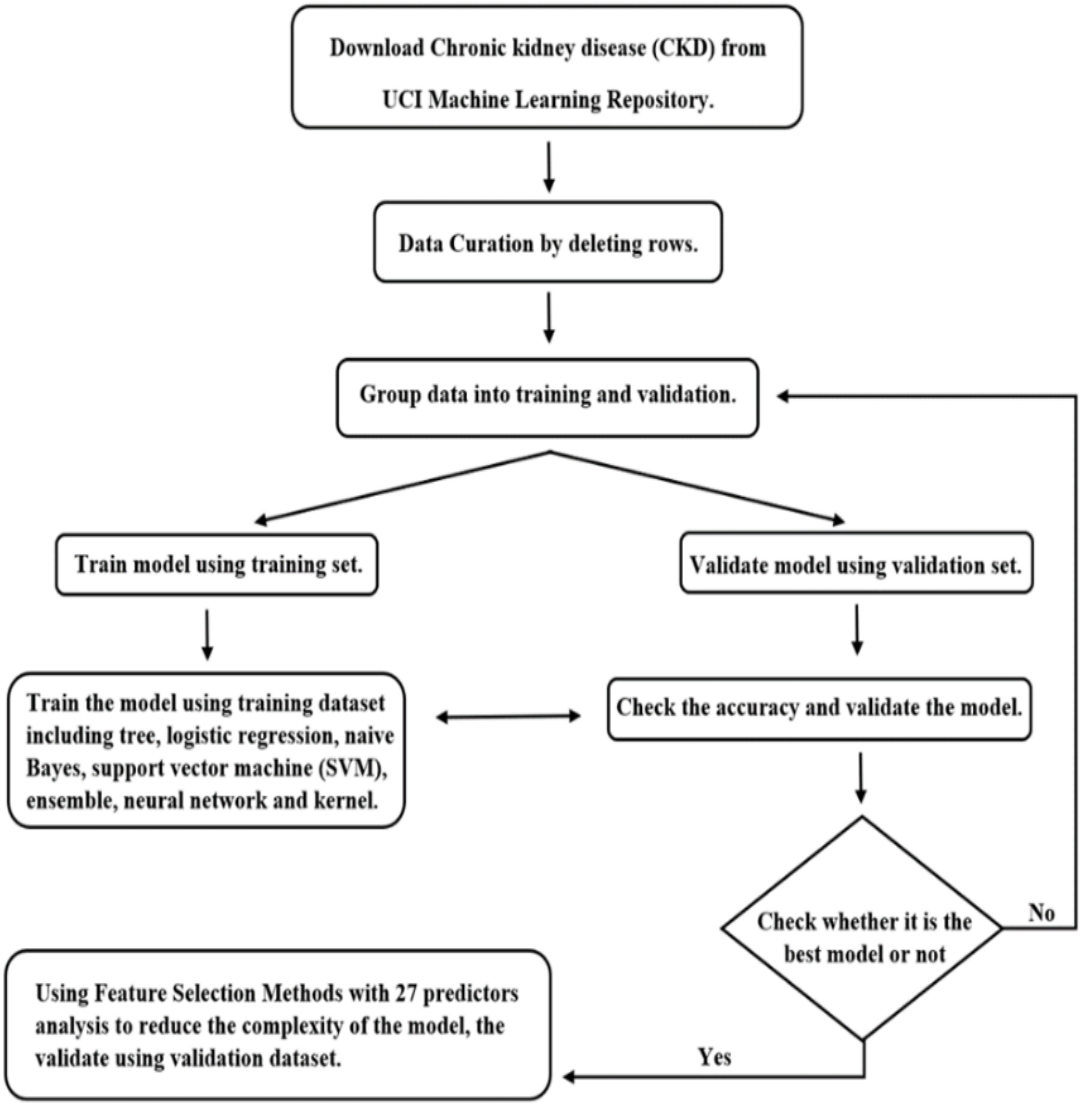

Figure 1 Process flow of the data. Data curation, ML training, ML testing, and performance metrics analysis in this study. Adapted from "Predicting Parkinson's Disease Severity using Telemonitoring Data and Machine Learning Models: A Principal Component Analysis-based Approach for Remote Healthcare Services during COVID-19 Pandemic". Published in *J. Curr. Sci. Technol*, 2023, 13(2), 465-485.

Table 1 Predictors, their definitions, data types, and values

Variables & Definition	Predictor or Label in ML training	Category	Values/Value Range
class	Label	Nominal	ckd = Chronic Kidney Disease not-ckd = Non-Chronic Kidney Disease
bp (diastolic): diastolic blood pressure	Predictor	Numerical	In mm/Hg unit
bp limit: blood pressure	Predictor	Numerical	In mm/Hg unit
sg: specific gravity	Predictor	Nominal	The range is from 1.005 to 1.025
al: albumin	Predictor	Nominal	The range is from 0 to 5 noted that the higher, the better
rbc: red blood cell	Predictor	Nominal	0 = normal, 1 = abnormal
su: sugar	Predictor	Nominal	The range is from 0 to 5
pc: pus cell	Predictor	Nominal	0 = normal, 1= abnormal
pcc: pus cell clumps	Predictor	Nominal	0 = not present, 1 = present
ba: bacteria	Predictor	Nominal	0 = not present, 1 = present
bgr: blood glucose random	Predictor	Numerical	In mg/dl unit
bu: blood urea	Predictor	Numerical	In mg/dl unit
sod: sodium	Predictor	Numerical	In mEq/L unit
sc: serum creatinine	Predictor	Numerical	In mg/dl unit
pot: potassium	Predictor	Numerical	In mEq/L unit
hemo: hemoglobin	Predictor	Numerical	In gms, *less than 15 indicates kidney failure
pcv: packed cell volume	Predictor	Numerical	Percentage of pcv
rbcc: red blood cell count	Predictor	Numerical	In millions/cumm unit
wbcc: white blood cell count	Predictor	Numerical	In cells/cumm unit
htn: hypertension	Predictor	Nominal	0 = no, 1 = yes
dm: diabetes mellitus	Predictor	Nominal	0 = no, 1 = yes
cad: coronary artery disease	Predictor	Nominal	0 = no, 1 = yes
appet: appetite	Predictor	Nominal	0 = good, 1= poor
pe : peda edema	Predictor	Nominal	0 = no, 1 = yes
ane : anemia	Predictor	Nominal	0 = no, 1 = yes
gfr : glomerular filtration rate	Predictor	Numerical	In mL/min unit
stage: 5 stages (G1-G5)	Predictor	Nominal	G1 = normal or high G2 = mildly decreased G3a = mildly to moderately decreased G3b = moderately to severely decreased G4 = severely decreased G5 = kidney failure
age	Predictor	Numerical	Age in years unit
affected* (Removed attribute)	Predictor	Nominal	0 = not-CKD, 1 = CKD

3.2 Data Curation

The initial dataset included 200 patient records, with a significant imbalance of 128 'CKD' instances and 72 'not-CKD' cases. This imbalance may bias machine learning models towards the more prevalent class (Masko, & Hensman, 2015). To create a balanced dataset with 72 examples for each class, we used under-sampling to remedy this problem by eliminating 56 'CKD' cases. While this method resulted in a more objective learning process, it lowered the dataset size. Furthermore, the dataset originally comprised 29 clinical variables as possible predictors.

Nevertheless, one variable, designated 'affected', introduced redundancy, which was discovered to duplicate the class names precisely. This variable was removed, leaving a final set of 28 predictors to improve efficiency and minimize overfitting. These predictors cover different clinical parameters and conditions crucial to the diagnosis of CKD, as shown in Table 1.

3.3 Dataset for Training and Testing

Following dataset curation, the dataset was split into training and test sets in an 80:20 ratio, with 116 rows for training and 28 for testing. This split is critical for effective machine learning because it allows models to learn from a large chunk of the data while evaluating performance on a distinct subset. Random record selection was critical to avoiding bias and maintaining an even distribution of 'CKD' and 'not-CKD' cases in both sets to ensure unbiased evaluation. This thorough dataset preparation offers the groundwork for robust model training and testing, addressing issues such as skewed data and predictor redundancy while enabling precise prediction of CKD presence.

3.4 Machine Learning Training and Testing

As mentioned in the introduction section, one of the gaps in the CKD diagnosis using ML is that there has still been no direct performance comparison between different machine learning models using the same dataset and quantitative settings. In addition, one of our goals is to train, validate, and compare different machine-learning models available in the built-in Classification Learner program in MATLAB R2022b, as listed in Table 2.

The curated training dataset was then employed to train the specified ML models. Several models were trained and tested against the preprocessed dataset. Afterward, we employed a separate test dataset to verify whether the learned models were generalized and could furnish an accurate prediction for the unseen dataset. Including a wide range of models in this study serves multiple purposes: comprehensively evaluating different ML approaches for CKD prediction, identifying models with optimal accuracy and clinical feasibility, and contributing valuable insights to refine ML strategies for enhancing CKD diagnostics and patient outcomes.

Table 2 ML models available in Matlab2022b and trained in this study

Models	Details	Models	Details
Logistic Regression	Logistic Regression	Tree	Last Change: Fine Tree
Ensemble	Boosted Trees		Last Change: Medium Tree
	Bagged Trees		Last Change: Coarse Tree
	RUSBoosted Trees	Naïve Bayes	Kernel Naïve Bayes
Neural Network	Trilayered Neural Network		Gaussian Naïve Bayes
	Bilayered Neural Network	Support vector machine (SVM)	Coarse Gaussian SVM
	Wide Neural Network		Medium Gaussian SVM
	Medium Neural Network		Fine Gaussian SVM
	Narrow Neural Network		Cubic SVM
Kernel	SVM Kernel		Quadratic SVM
	Logistic Regression Kernel		Linear SVM

For the test dataset, we employed the same performance metrics as the training dataset for a direct comparison, including a K-fold cross-validation with a K-fold of 5 to calculate performance metrics including, precision, recall, accuracy, and F1-score, using the training dataset as demonstrated in Equations (1)-(4):

$$\text{Precision}=\frac{T_p}{T_p+F_p} \quad (1)$$

$$\text{Recall}=\frac{T_p}{T_p+F_n} \quad (2)$$

$$F_1=2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (3)$$

$$\text{Accuracy}=\frac{T_p+T_n}{T_p+T_n+F_p+F_n} \quad (4)$$

where T_p and T_n are true positive cases and true negative cases, respectively, and F_p and F_n are false positive cases and false negative cases.

The trained networks and models were then evaluated using the separated test dataset to determine their generalization and capability of accurately classifying the unseen test set. The 5-fold cross-validation approach was used to validate the model performance metrics by comparing the training dataset to the test dataset, obtaining similar performance metrics in precision, recall, F1-score, and accuracy for the test dataset.

The implementation of 5-fold cross-validation in this study serves numerous important functions, including ensuring model stability and generalizability, preventing overfitting, and optimizing model parameters for enhanced performance. This rigorous validation method maintains scientific integrity and robustness, aiming to improve CKD diagnosis predictions and provide significant insights to the medical community.

3.5 Feature Selection Methods

The built-in feature selection tool in MATLABR2022b was then used to identify key predictors contributing to model classification accuracy. Following identification of critical predictors, a less sophisticated ML model was trained to prove that with fewer predictors, the model could still perform comparably to models built with all predictors. The feature selection methods are based on statistical approaches, including minimum Redundancy-Maximum Relevance (MRMR), Chi^2, ANOVA, and Kruskal-Wallis. The statistical analysis in this study was thorough, employing both descriptive statistics for the training dataset and inferential statistics for the test dataset to explore the data and evaluate machine learning model performance, ensuring result validity and methodological rigor. By combining descriptive and inferential statistics, utilizing advanced software tools, and adhering to fundamental statistical principles, this study provides a strong foundation for future CKD prediction research. Detailed descriptions of the statistical methodology, tools, and assumptions enable replication and expansion of the research, fostering innovative machine learning approaches in CKD diagnosis advancement.

Combining diverse feature selection techniques provides a comprehensive approach, accommodating numerical and categorical data while addressing redundancies and inter-variable relationships. This ensures that the final predictors for the CKD model are both statistically significant and varied, improving predictive accuracy and generalizability. By utilizing these methods, the study effectively handles the complexities of medical datasets, resulting in a robust, interpretable model at the forefront of CKD diagnostic methods.

3.6 Statistical Analysis Using Bootstrap Analysis

Here, we employed bootstrap analysis in Python using Jupyter Notebook to scrutinize the confusion matrices further. This statistical technique randomly resamples our dataset with replacement, using a sample size of 1,000. This process is repeated numerous times to ensure statistical reliability. The bootstrap method helps us estimate the accuracy and variability of our network's performance in classifying CKD cases. We aim to establish a 95% confidence interval for these estimations, providing a robust measure of the error range associated with CKD classification.

This analysis is particularly critical as it also assesses the performance of a complexity-reduced model. This simplified model uses only the most critical factors previously identified through a rigorous feature selection process (detailed in section 3.5). By applying bootstrap analysis, we can verify the stability and justify the performance of this streamlined model under varying data conditions. The outcomes highlight the reliability of error estimates and underscore the effectiveness of feature selection in enhancing model efficiency.

4. Results

4.1 Classification Models Trained using All 27 Predictors

All the labels and 27 clinical features from Table 1 were initially used to train the ML models listed in Table 3. Kernel Naïve Bayes was the best-performing ML model, with classification accuracy, precision, recall, and F1 scores of 96.55%, 95.00%, 98.28%, and 96.61%, respectively, as shown in Table 3. These results were determined using a 5-fold cross-validation approach from the training dataset.

Table 3 Classification performance metrics of the trained models using all 27 predictors and the training dataset for validation using a 5-fold cross-validation method

Models	Details	Precision	Recall	F1-score	Accuracy
Tree	Last Change: Fine Tree	96.43%	93.10%	94.74%	94.83%
	Last Change: Medium Tree	96.43%	93.10%	94.74%	94.83%
	Last Change: Coarse Tree	96.43%	93.10%	94.74%	94.83%
Naïve Bayes	Gaussian Naïve Bayes	FAILED	FAILED	FAILED	FAILED
	Kernel Naïve Bayes*	95.00%	98.28%	96.61%	96.55%
SVM	Linear SVM	100%	81.03%	89.52%	90.52%
	Quadratic SVM	100%	82.76%	90.57%	91.38%
	Cubic SVM	100%	82.76%	90.57%	91.38%
	Fine Gaussian SVM	68.42%	89.65%	77.61%	74.14%
	Medium Gaussian SVM	100%	81.03%	89.52%	90.52%
	Coarse Gaussian SVM	100%	67.24%	80.41%	83.62%
Ensemble	Boosted Trees	48.57%	58.62%	53.13%	48.28%
	Bagged Trees	96.43%	93.11%	94.74%	94.83%
	RUSBoosted Trees	57.63%	58.62%	58.12%	57.76%
Neural Network	Narrow Neural Network	98.00%	84.48%	90.74%	91.38%
	Medium Neural Network	98.00%	84.48%	90.74%	91.38%
	Wide Neural Network	98.00%	84.48%	90.74%	91.38%
	Bilayered Neural Network	96.15%	86.21%	90.91%	91.38%
	Trilayered Neural Network	98.00%	84.48%	90.74%	91.38%
Kernel	SVM Kernel	100%	89.66%	94.54%	94.83%
	Logistic Regression Kernel	98.04%	86.21%	91.74%	92.24%

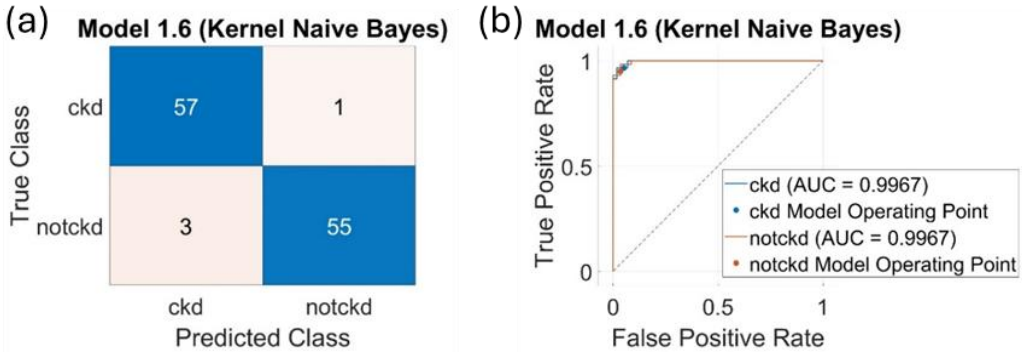

Figure 2 (a) Confusion matrix of the trained Kernel Naïve Bayes model
(b) ROC plots of the trained Kernel Naïve Bayes model

The models listed above have the best average performance metric of 96.55%, calculated by averaging the values of the four metrics. Figures 2a and 2b illustrate the model's confusion matrix and receiver operating characteristic (ROC) plots.

The analysis comparing machine learning models for predicting CKD highlights subtle differences in performance and their practical implications, with Kernel Naïve Bayes stands out for its ability to handle complex interactions effectively. However, models like Bagged Trees and SVMs also hold promise despite potential difficulties in capturing all CKD cases. Gaussian Naïve Bayes and Boosted Trees encounter obstacles related to data distribution and overfitting assumptions. Choosing the right model for clinical use involves weighing precision and recall, considering factors such as the cost of false positives and how easily the model can be integrated into healthcare systems. Meanwhile, Kernel Naïve Bayes appears to be the strongest candidate in this study, which will be employed and tested in further sections.

The confusion matrix presented in Figure 2a was subjected to a rigorous bootstrap analysis to validate the robustness of the Kernel Naïve Bayes model with a 95% confidence interval. The performance metrics-accuracy, precision, recall, and F1-score-demonstrated excellent stability, falling within the ranges of 94.87% to 98.26%, 93.33% to 96.67%, 96.55% to 100.00%, and 94.92% to 98.31%, respectively. Additionally, the standard deviations for these metrics were notably low, all below 0.86%, underscoring the model's exceptional classification performance. This analysis confirms the Kernel Naïve Bayes model's high reliability in classifying data and highlights its precision and efficiency in handling diverse datasets.

4.2 Validation of the Trained Model using the Test Dataset

The separated test dataset was used to predict classification results compared to known labels. For the 22 validation cases, the trained Kernel Naïve Bayes model can predict 70 T_p cases, 12 T_n cases, 2 F_p cases, and 0 F_n cases, as shown in Figure3a. These four variables offer the following performance metrics, including a precision of 87.50%, recall of 100%, F1 score of 93.33%, and accuracy of and 92.86%, respectively. Note that the test dataset was imbalanced. The confusion matrix was therefore normalized by the total number for each class, as shown in Figure3b, to calculate the performance metrics. The models listed above have the highest percentage of ML classification accuracy performance based on all 27 features. The test dataset's performance parameters and the validation cases' performance parameters in the preceding section differed by only 4%; in other words, the trained model produced a slightly overfitting model. The AUC values generated from the ROC curves of the test dataset demonstrate comparable performance to the training dataset, as illustrated in Figure 3c in contrast to Figure 2b.

Integrating statistical tests for significance into machine learning model evaluation ensures more reliable conclusions about their effectiveness by determining if performance differences are meaningful or random. This involves selecting tests, setting hypotheses, determining significance levels, computing statistics, and interpreting results. Applied to CKD prediction models like Kernel Naïve Bayes, these tests provide insights into model superiority and guide further research and clinical use.

The confusion matrix depicted in Figure 3a was analyzed using the bootstrap method described in section 3.6. The analysis yielded the following 95% confidence intervals for the model's performance metrics: accuracy ranged from 91.06% to 93.43%, precision from 85.59% to 89.32%, recall remained at 100% due to the absence of false negatives, and the F1-score varied between 92.24% and 94.36%. The narrow range of these confidence intervals indicates a high level of stability in the model's performance when validated against an unseen test dataset.

4.3 Number of Predictors Reduction using Feature Selection Methods

We then employed several feature selection methods to identify the relevance of each predictor. To begin, the Feature Selection Tools comprise all 27 predictors, and Table 4 indicates which variable has the highest priority among all the variables. The feature selection tool, categorized into four types of statistical feature selection methods, including MRMR, Chi^2, ANOVA, and Kruskal-Wallis, was employed to analyze and present the outcomes of different predictors, as illustrated in Table 4.

Analyzing the feature selection outcomes gives a detailed understanding of how each method contributed to pinpointing essential predictors for CKD prognosis and their clinical significance. Through techniques like MRMR, Chi^2, ANOVA, and Kruskal-Wallis, critical predictors such as packed cell volume (pcv), hemoglobin (hemo), specific gravity (sg), albumin (al), and serum creatinine (sc) were identified, highlighting their diagnostic importance and physiological implications in CKD. These predictors reflect kidney function, erythropoiesis, urine concentration ability, and markers of kidney damage and clearance efficiency, aiding in early detection and intervention strategies. This thorough feature selection process underscores the potential of machine learning to enhance disease diagnosis by leveraging complex clinical data for actionable insights.

Based on the table above, each selector has common top-ranking criteria: hemo and pcv. On the other hand, there is a unique overall value for the selection of MRMR compared to the others. Thus, to ascertain the overall significance of the predictors ranked from highest to lowest, the choice was made to utilize Chi^2, ANOVA, and Kruskal-Wallis for averaging their statistical values, as shown in Table 5.

Using the first predictor (pcv) shown in Table 5, we trained the ML models in the initial training round until the highest classification accuracy was reached. Note that by training the model using the first four variables listed in Table 6, the Kernel Naïve Bayes model reached up to 98.28%.

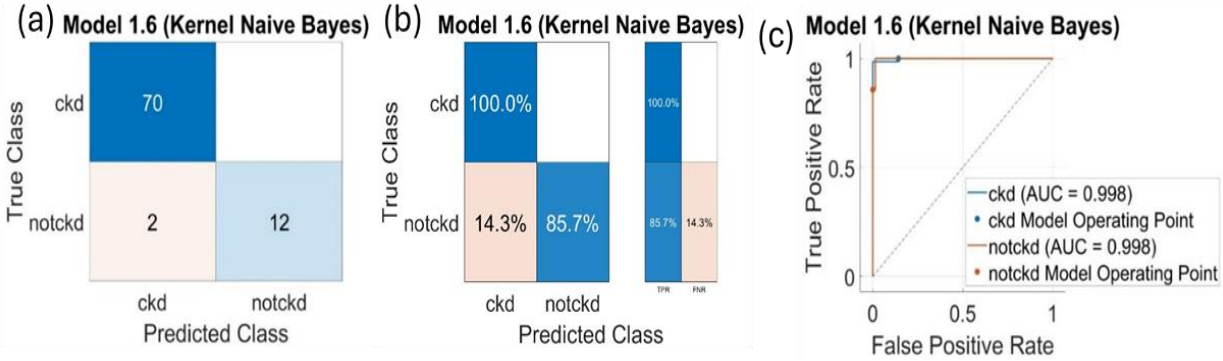

Figure 3 (a) Confusion Matrix of the trained Kernel Naïve Bayes model for the test dataset
(b) Normalized Confusion Matrix of the trained Kernel Naïve Bayes model for the test dataset
(c) ROC plots of the trained Kernel Naïve Bayes using the test dataset

Table 4 The significance of each predictor within the feature selection tool

No.	MRMR		Chi^2		ANOVA		Kruskal-Wallis	
1	hemo	0.5183	hemo	35.0091	pcv	35.8042	pcv	31.5909
2	bpDiastolic	0.0093	pcv	27.4987	stage	35.303	al	27.9565
3	appet	0.0082	sg	26.4943	sg	31.7091	sg	27.4497
4	dm	0.0071	stage	25.5119	al	29.8613	stage	26.5012
5	sg	0.0052	al	23.8265	htn	21.4283	rbcc	22.8058
6	rbc	0.005	grf	21.4112	rbcc	20.935	htn	18.6166
7	pcc	0.0049	rbcc	19.2477	dm	19.1773	bu	17.4308
8	al	0.0043	htn	18.7652	pc	18.1162	dm	16.9573
9	pe	0.0039	bpLimit	18.1971	grf	16.6418	grf	16.5318
10	htn	0.0034	dm	17.0918	bu	16.3071	pc	16.1526
11	grf	0.0033	pc	16.2804	ane	15.1583	sc	15.1959
12	ane	0.0031	bu	14.3381	sc	14.7984	sod	15.0455
13	rbcc	0.0031	ane	13.9394	appet	14.2398	ane	13.8312
14	pcv	0.0028	appet	13.1883	sod	13.21	appet	13.0863
15	bpLimit	0.0027	sod	13.0446	bpLimit	12.2145	pe	10.9304
16	cad	0.0026	pe	11.0143	pe	11.659	bpLimit	9.6004
17	ba	0.0025	sc	9.8385	rbc	10.0668	rbc	9.553
18	sod	0.0025	rbc	9.6255	pcc	6.4448	pcc	6.2818
19	bgr	0.0023	age	7.1316	su	4.5913	su	5.0196
20	pc	0.0021	pcc	6.3276	ba	3.8085	hemo	4.0517
21	stage	0.0021	bgr	4.8228	wbcc	3.7318	ba	3.779
22	age	0.0019	wbcc	3.9043	bpDiastolic	3.2204	bpDiastolic	3.2069
23	sc	0.0019	ba	3.8049	cad	2.5191	cad	2.5185
24	bu	0.0019	bpDiastolic	3.2284	hemo	2.4358	wbcc	1.5869
25	su	0.0016	cad	2.5348	pot	1.1412	age	1.2532
26	wbcc	0.0012	su	2.172	age	0.7259	pot	1.1479
27	pot	0.0009	pot	1.1545	bgr	0.6187	bgr	0.3196

Table 5 The top rank for each predictor across three distinct Feature Selection Methods

Priority	Predictors	Rate	Priority	Predictors	Rate
1	pcv	31.63	15	bpLimit	13.34
2	stage	29.11	16	sc	13.28
3	sg	28.55	17	pe	11.20
4	al	27.24	18	rbc	9.75
5	rbcc	20.99	19	pcc	6.35
6	htn	19.63	20	su	3.93
7	grf	18.94	21	ba	3.79
8	dm	17.74	22	bpDiastolic	3.22
9	pc	16.85	23	wbcc	3.07

Table 5 Cont.

Priority	Predictors	Rate	Priority	Predictors	Rate
10	bu	16.03	24	age	3.04
11	ane	14.31	25	cad	2.52
12	hemo	13.83	26	bgr	1.92
13	sod	13.77	27	pot	1.15
14	appet	13.50			

Table 6 The Twenty-two models were trained using the top four priority predictors

Model Number	Model Type	Accuracy % (Validation)
1	Last Change: Fine Tree	96.55%
2	Last Change: Medium Tree	96.55%
3	Last Change: Coarse Tree	95.69%
4	Logistic Regression	93.97%
5	Gaussian Naïve Bayes	98.28%
6	Kernel Naïve Bayes*	98.28%
7	Linear SVM	97.41%
8	Quadratic SVM	96.55%
9	Cubic SVM	97.41%
10	Fine Gaussian SVM	95.69%
11	Medium Gaussian SVM	97.41%
12	Coarse Gaussian SVM	87.93%
13	Boosted Trees	48.28%
14	Bagged Trees	95.69%
15	RUSBoosted Trees	48.28%
16	Narrow Neural Network	98.28%
17	Medium Neural Network	99.14%
18	Wide Neural Network	98.28%
19	Bilayered Neural Network	97.41%
20	Trilayered Neural Network	97.41%
21	SVM Kernel	98.28%
22	Logistic Regression Kernel	98.28%

The inclusion of the 5th, 6th, and 7th predictors emphasized the relevance of the first four predictors. However, the classification accuracy of Kernel Naïve Bayes was significantly lowered. Thus, it can be summarized that four predictors are the optimal number for our model. Additionally, it is widely recognized that a larger number of parameters does not always provide better performance since some parameters might not be consistent with the label or may contain more noise, especially if the optional data gives more noise rather than useful information for ML training.

Adding the 7th predictor reinforced the significance of the first 4 predictors. The classification accuracy of Kernel Naïve Bayes was a consistent percentage, which gradually diminishes by adding the 5th and continues, as shown in Table 7. As a result, we can conclude that 4 predictors are the optimal quantity for our model.

Table 7 The percentage of Kernel Naïve Bayes model by using the first 8 predictors

Number of Predictors	Kernel Naïve Bayes
1	87.93%
2	88.79 %
3	96.55 %
4	98.28% *
5	98.28%
6	98.28%
7	95.69%

Analyzing the study's outcomes within its objectives enriches its depth and practical relevance, showcasing the findings' efficiency and potential clinical usefulness. The revelation that the Kernel Naïve Bayes model achieves an accuracy of up to 98.28% in predicting CKD using just four predictors perfectly aligns with the intention to optimize CKD prediction.

An important question to consider is the performance equivalence between the complexity-reduced model, which utilizes only four clinical features, and the full model, which includes all predictors. The confusion matrix of the Kernel Naïve Bayes model, trained on these four crucial factors, underwent bootstrap analysis, revealing performance metrics within the following 95% confidence intervals: accuracy ranged from 96.58% to 99.15%, precision from 98.21% to 100%, recall from 94.83% to 98.28%, and F1-score from 96.49% to 99.13%. These metrics closely mirror those reported for the full model in section 4.1, suggesting that the simplified model maintains similar classification efficacy as the more comprehensive model.

This efficiency underscores the model's strength and simplifies the diagnostic process, making it more accessible for healthcare providers. Identifying the four key predictors—packed cell volume, hemoglobin, specific gravity, and albumin—holds significant implications, streamlining diagnostic testing and offering insights into CKD's underlying processes. Furthermore, these findings contribute to the broader aim of personalized medicine by laying the groundwork for tailored risk assessments and treatment plans. Ultimately, this study's results have profound implications for CKD diagnosis, suggesting the potential for machine learning models to effectively transform healthcare practices in detecting and managing chronic diseases like CKD.

5. Discussion

CKD is a degenerative disorder in which kidney function gradually diminishes, frequently without any symptoms in the early stages, resulting in a delayed diagnosis. Traditional diagnostic procedures, such as creatinine level monitoring, have limits, highlighting the need for novel approaches like artificial intelligence (AI). Using patient data from a dataset of 28 clinical variables, we created an AI model with an impressive 96.55% classification accuracy, demonstrating machine learning's (ML) potential for improving diagnosis.

Comparing our findings to past research reveals both similarities and differences. Islam et al., (2020) and Haratian et al., (2022) found that ML models like XGBoost and Random Forest effectively predicted CKD with reasonable accuracy. Similarly, Ashafuddula et al., (2023) and Dritsas, & Trigka (2022) emphasized the need for advanced ML classifiers for early CKD detection. Our study's use of the Kernel Naïve Bayes model, which achieved 98.28% accuracy with only four predictors, is consistent with these findings, highlighting ML's diagnostic capability and the efficiency of limited predictors.

Regarding clinical consequences, the global lack of nephrologists highlights the importance of our findings in improving early CKD diagnosis. Integrating AI and ML models into clinical practice could help solve this gap by automating initial screenings and prioritizing patients for further assessment, thus lowering the demand for already overburdened healthcare facilities. Furthermore, seamless integration of these technologies with existing Electronic Health Record (EHR) systems can streamline diagnostic workflows, resulting in more efficient patient care and resource allocation. Furthermore, our model's capacity to rely on only four important variables indicates a realistic approach to diagnosis that might be easily deployed in various clinical situations. This streamlined diagnostic tool can improve testing processes and allow for early intervention by identifying at-risk individuals earlier in their illness. By providing doctors with actionable information, our model may enable them to make better decisions about patient care, ultimately improving outcomes and lowering the burden of CKD on patients and healthcare systems.

Recognizing constraints such as dataset biases and the danger of overfitting is critical for understanding the scope and trustworthiness of our findings. Dataset biases in all datasets can skew

results and limit our model's applicability to larger patient populations. Furthermore, the risk of overfitting, in which a model performs well on training data but fails to generalize to new data, emphasizes the importance of rigorous validation across varied patient cohorts. Moreover, although our model shows remarkable accuracy with a small set of predictors, it might not account for all pertinent variables influencing the diagnosis of chronic kidney disease. Variations in patient demographics, comorbidities, and environmental factors may affect model performance and necessitate further inquiry. In addition, addressing these limitations is critical to ensuring our model's dependability and usefulness in real-world clinical scenarios. Future research should focus on minimizing dataset biases, maximizing model generalizability, and refining feature selection approaches to improve our approach's robustness.

The unique method we used in our work to apply the Kernel Naïve Bayes model with minimum predictors goes beyond the diagnosis of CKD and provides a model for effective diagnostic instruments for various chronic illnesses. This breakthrough could revolutionize illness identification worldwide and have wide-ranging effects. By incorporating AI and ML into clinical practice, we can progress toward proactive, tailored healthcare, lowering inequities and increasing outcomes. Through telehealth technologies and remote monitoring, this change can potentially democratize access to healthcare, especially for underprivileged people. In short, our research ushers in a new era of precision medicine and better patient care while advancing machine learning and artificial intelligence in the healthcare industry.

6. Conclusion

This study highlights the critical importance of precise CKD diagnosis, a significant public health concern, through machine learning techniques on a comprehensive clinical dataset from Enam Medical College, comprising 28 features and records of 200 Bangladeshi patients. Firstly, an unbiased training dataset was provided by curating the data and training multiple ML models, including kernel models, neural network tree-based models, Naïve Bayes, logistic regression, support vector machine, and ensemble using MATLAB R2022b. Model performance was assessed using various performance metrics such as confusion matrix, accuracy, precision, sensitivity, recall, F1-score, and receiver operating curve (ROC) analysis with the area under the curve (AUC). This evaluation also employed a 5-fold cross-validation technique to ensure the reliability and generalizability of the results. Furthermore, feature selection methods, including minimum-redundancy-maximum-relevance, Chi^2, ANOVA, and Kruskal-Wallis, were employed to identify the most significant predictors for CKD classification. Intriguingly, the analysis reveals that only four predictors, including (1) packed cell value, (2) stages of glomerular filtration rate, (3) specific gravity of urine sample, and (4) albumin content of urine sample suffices to achieve a comparable performance level in training the model. It was found that Kernal Naïve Bayes emerged as the optimal model, achieving an impressive accuracy of 96.55%. The application of AI in CKD diagnosis and management, as demonstrated in this study, holds immense promise. By leveraging machine learning techniques and feature selection methods on clinical datasets, AI facilitates the development of accurate CKD classification models. Kernel Naïve Bayes highlights the potential of AI in enhancing CKD diagnosis, with high accuracy, precision, recall, and F1-score achieved across both training and test datasets. Moreover, identifying key predictors through feature selection techniques highlights AI's ability to streamline diagnostic processes, potentially leading to more effective CKD management strategies. Overall, this study exemplifies the impactful role of AI in advancing healthcare analytics and improving CKD diagnosis and treatment outcomes.

7. Acknowledgements

We sincerely thank Enam Medical College in Bangladesh for providing the dataset utilized in this study, which was obtained via the UCI Machine Learning Repository, an open-source database. We also express our deepest gratitude to Satriwithaya School and Institute of Research, Rangsit University, for the research and publication grant.

8. References

Ashafuddula, N. I., Islam, B., & Islam, R. (2023). An Intelligent Diagnostic System to Analyze Early-Stage Chronic Kidney Disease for Clinical Application. *Applied Computational Intelligence and Soft Computing, 2023*, Article 3140270. https://doi.org/10.1155/2023/3140270

Dubey, Y., Mange, P., Barapatre, Y., Sable, B., Palsodkar, P., & Umate, R. (2023). Unlocking Precision Medicine for Prognosis of Chronic Kidney Disease Using Machine Learning.

Diagnostics, *13*(19), Article 3151, 1-22. https://doi.org/10.3390/diagnostics13193151

Durga, P., Karthikeyan, S. (2023). Comparative analysis for augmented decision-making applications using deep learning models. *Journal of Current Science and Technology*, *13*(3), 791-803 https://doi.org/10.59796/jcst.V13N3.2023.2273

Drawz, P., & Rahman, M. (2015). Chronic kidney disease. *Annals of Internal Medicine*, *162*(11), ITC1-ITC16. https://doi.org/10.7326/AITC201506020

Dritsas, E., & Trigka, M. (2022). Machine learning techniques for chronic kidney disease risk prediction. *Big Data and Cognitive Computing*, *6*(3), Article 98. https://doi.org/10.3390/bdcc6030098

George, C., Echouffo-Tcheugui, J. B., Jaar, B. G., Okpechi, I. G., & Kengne, A. P. (2022). The need for screening, early diagnosis, and prediction of chronic kidney disease in people with diabetes in low-and middle-income countries-a review of the current literature. *BMC Medicine*, *20*(1), 1-12. https://doi.org/10.1186/s12916-022-02438-6

Haratian, A., Maleki, Z., Shayegh, F., & Safaeian, A. (2022). Detection of factors affecting kidney function using machine learning methods. *Scientific Reports*, *12*(1), Article 21740. https://doi.org/10.1038/s41598-022-26160-8

Iftikhar, H., Khan, M., Khan, Z., Khan, F., Alshanbari, H., & Ahmad, Z. (2023). A Comparative Analysis of Machine Learning Models: A Case Study in Predicting Chronic Kidney Disease. *Sustainability*. 15(3), Article 2754. https://doi.org/10.3390/su15032754.

Islam, M. A., Akter, S., Hossen, M. S., Keya, S. A., Tisha, S. A., & Hossain, S. (2020, December 3-4). *Risk factor prediction of chronic kidney disease based on machine learning algorithms* [Conference presentation]. Paper presented at the 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India. https://doi.org/10.1109/ICISS49785.2020.9315878

Kalantar-Zadeh, K., Jafar, T. H., Nitsch, D., Neuen, B. L., & Perkovic, V. (2021). Chronic kidney disease. *Lancet*, *398*(10302), 786-802. https://doi.org/10.1016/s0140-6736(21)00519-5

Khwanchum, R., Pothiban, L., Wonghongkul, T., Lirtmulikaporn, S. (2024). Effectiveness of the Nurse-led Self and Family Management Support Program among Adults with Early-stage Chronic Kidney Disease: A Randomized Controlled Trial. *Pacific Rim International Journal of Nursing Research*, *28*(1), 219-33. Available from: https://he02.tci-thaijo.org/index.php/PRIJNR/article/view/264735

Kumari, S., & Singh, S. (2022, November 4-5). *An ensemble learning-based model for effective chronic kidney disease prediction* [Conference presentation]. 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India. https://doi.org/10.1109/ICCCIS56430.2022.10037698

Levey, A. S., Coresh, J., Balk, E., Kausz, A. T., Levin, A., Steffes, M. W., ... & Eknoyan, G. (2003). National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Annals of Internal Medicine*, *139*(2), 137-147. https://doi.org/10.7326/0003-4819-139-2-200307150-00013

Lv, J.-C., & Zhang, L.-X. (2019). Prevalence and disease burden of chronic kidney disease. *Renal Fibrosis: Mechanisms and Therapies*, *1165*, 3-15. https://doi.org/10.1007/978-981-13-8871-2_1

Masko, D., & Hensman, P. (2015). The impact of imbalanced training data for convolutional neural networks. *In KTH Royal Institute of Technology CSC School*. https://www.kth.se/social/files/588617ebf27654 01cfcc478c/PHensmanDMasko_dkand15.pdf

Murphy, D., McCulloch, C. E., Lin, F., Banerjee, T., Bragg-Gresham, J. L., Eberhardt, M. S., ... & Centers for Disease Control and Prevention Chronic Kidney Disease Surveillance Team. (2016). Trends in prevalence of chronic kidney disease in the United States. *Annals of Internal Medicine*, *165*(7), 473-481. https://doi.org/10.7326/m16-0273

Nishat, M., Faisal, F., Dip, R., Nasrullah, S., Ahsan, R., Shikder, F., Asif, M., & Hoque, M. (2018). A Comprehensive Analysis on Detecting Chronic Kidney Disease by Employing Machine Learning Algorithms. *EAI Endorsed Trans. Pervasive Health Technol*, *21*(29), Article e1. https://doi.org/10.4108/eai.13-8-2021.170671.

Pechprasarn, S., Manavibool, L., Supmool, N., Vechpanich, N., & Meepadung, P. (2023). Predicting Parkinson's Disease Severity using Telemonitoring Data and Machine Learning Models: A Principal Component Analysis-based Approach for Remote Healthcare

Services during COVID-19 Pandemic. *Journal of Current Science and Technology*, *13*(2), 465-485. https://ph04.tci-thaijo.org/index.php/JCST/article/view/694

Pradeepa, P., & Jeyakumar, M. K. (2022). Data redundancy removal using K-MAD based self-tuning spectral clustering and CKD prediction using ML techniques. *Journal of Current Science and Technology*, *12*(3), 517-537. https://ph04.tci-thaijo.org/index.php/JCST/article/view/291

Rainey, H. (2019). Preventing complications and managing symptoms of CKD. *Practice Nursing*, *30*(6), 276-281. https://doi.org/10.12968/pnur.2019.30.6.276

Sawhney, R., Malik, A., Sharma, S., & Narayan, V. (2023). A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease. *Decision Analytics Journal*, *6*, Article 100169. https://doi.org/10.1016/j.dajour.2023.100169