Vol. 7, No. 2 | July – December 2024

Predictive Modeling of Chronic Kidney Disease Progression with Ensemble Learning Techniques

Engr. Priha Bhatti[1], Dr. Muhammad Affan Alim[2]

Abstract:

The present study aims to tackle the significant issue of prompt identification of chronic kidney disease (CKD), a highly prevalent and potentially fatal medical illness. Given the crucial function of the kidneys in maintaining homeostasis, we put forth a novel ensemble learning model to forecast the onset of chronic kidney disease (CKD). Utilizing an extensive dataset, the study employs ten carefully designed stages, covering data analysis, missing data management, normalization, and training of machine learning models. The model that we have proposed exhibits superior performance compared to the existing approaches, attaining a noteworthy accuracy rate of 98.74%. Additionally, it demonstrates a sensitivity rate of 100%, a specificity rate of 96.54%, and an F1 score of 99.02%. The visual representation of the confusion matrix effectively showcases the strong performance of the model. The results of this study indicate that our ensemble technique holds promise as a valuable tool for the prompt detection of chronic kidney disease (CKD). It has the potential to improve diagnostic accuracy in clinical settings and alleviate the financial burden associated with advanced CKD treatments.

Keywords: *Ensemble model, Random Forest, K nearest neighbor, Decision tree, Support vector machine, Adaboost ensemble method.*

1. Introduction

The kidney is a crucial organ because it acts as the body's primary filtration system, helping to get rid of waste and excess water by making urine. Chronic Kidney Disease (CKD), alternatively called renal impairment, is a life-threatening illness characterized by the kidney's diminished capacity to conduct its vital physiological processes [1]. The significance of this condition is emphasized by acknowledging the pivotal role of the kidney in maintaining homeostasis and enabling the process of excretion. In human organs, the kidneys may interact with the oxidizing system of the human body which can effectively and efficiently process the elimination of harmful and excess substances from the human body.

[1] Department of Computer Science, Muhammad Ali Jinnah University, Karachi, Pakistan
[2] Department of Computer Science, IQRA University, Karachi, Pakistan
Corresponding Author: FA23PHCS0001@maju.edu.pk

Chronic kidney disease (CKD) is a progressive pathological medical condition that can cause dysfunction in the renal function system below the threshold of 60 mL/min/1.7 m² [2].

CKD is a serious and irreversible disorder defined by a continuous deterioration in kidney function, resulting in the body's inability to maintain a healthy fluid and electrolyte balance throughout metabolism. The renal function of the human body is responsible for the regulation of blood composition, as well as the elimination of metabolic and cancerous or infectious wastes through the urination process of the human body through the urinary tract system. This process plays a crucial action using in the maintenance of the body's delicate acid-base balance [3]. In this case, the presence of an electrolyte imbalance may necessitate the use of dialysis. According to the Global Burden of Disease Report [4], CKD was classified as one of the world's leading causes of death in 2015 in several nations.

The reported genetic-based cases found in Pakistan range from 12.5% to 31.2% [5]. Given the conservative growth in both the incidence and prevalence of end-stage renal disease (ESKD) for the last thirty-five years, chronic kidney disease has emerged as a major life-threatening issue [6]. Chronic kidney disease (CKD) is very common in Pakistan as well as globally, with an estimated more than one million cases diagnosed every year, which can lead to the serious diagnosis of renal failure and dysfunction. This medical emergent condition may further proceed towards a substantial and forthcoming risk, which can be the main cause of a a slow decrease in kidney function and having long-term consequences for people facing kidney problems. As the glomerular filtration rate (GFR) decreases, which acts as an indicator of worsening kidney function, the disease advances through successive stages, Finally, Stage 5 is known as End-Stage Renal Disease (ESRD). Individuals in this stage require continual kidney transplantation or dialysis to survive. In Table 1 we depict the risk factors of chronic kidney disease (CKD) prediction based on the risk variables thorough dataset analysis. This table contains several risk factors, including hypertension, diabetes, Genetics, and family history. These findings give the climax for the complexity of chronic kidney disease progression, and identification of the importance of early detection and diagnosis of the Chronic disease, followed by the risk factors associated with the several stages of chronic kidney disease (CKD), which are classified based on the Glomerular Filtration Rate.

Table 1: Chronic Kidney disease risk factor

Phase	Description	Range
----	Increased Risk	>=90 with risk factors
One	Normal/Increased kidney damage	GFR >=90
Two	GFR (Mild decrease)	60-89
Three	GFR (Moderate decrease)	30-59
Four	GFR (Severely decreased)	15-29
Five	Kidney Failure	Dialysis

Chronic kidney disease manifests gradually, with a progressive decline in renal function over an extended time, spanning months or even years. There are a few complex types of nephrons present in the kidneys that can play a crucial role in the filtration of bad and unusual waste substances from the bloodstream, which can lead to impairment when affected by chronic kidney disease (CKD). The occurrence of this dysfunction rapidly results in a diminished filtration capacity, therefore impairing the functionality of the immune system and continuously increasing the symptoms associated with chronic kidney disease (CKD). Significantly, a huge number indicates that almost 10% of the population may be affected by chronic kidney disease (CKD), often without their knowledge, emphasizing the significance of increased consciousness and timely identification [7]. The transition of chronic kidney disease (CKD) from its early phases to End-Stage Renal Disease (ESRD) depicts a pivotal stage characterized by the heightened demonstration of symptoms and

consequences [8]. Frequently the most important observed symptoms of encircle are the blood pressure elevation, anemia, compromised skeletal integrity, suboptimal nutritional status, peripheral neuropathy, diabetes, edema in the lower extremities, and fatigue. Early detection of these symptoms provides a chance for interference, facilitating the implementation of treatment following by the proper intake of drugs, dietary habits, and behavioral adjustments [9]. The timely detection of a medical condition serves as a crucial stepping in, providing a viable substitute for economically numerous renal replacement therapies (RRT), such are kidney dialysis or kidney transplantation. The imperative for a dependable prognostic algorithm to detect the latent presence of chronic kidney disease in its Emerging stages is accentuated by its secretive and gradual quality [10]. This study investigates the application of ensemble learning techniques as a novel approach for building the model to predict development of CKD. The goal is to participate in the Machine Learning application challenge by developing an accurate predictive model that can provide a clear interpretability and robustness, By using advanced Machine Learning Algorithms which can be implemented properly can led to the improvement of the diagnostic capability of medical practitioner in devising treatment strategies for patients who are on the high-risk and may be improve the outcomes on diagnosis as well as cost associated with this common and chronic condition of life threatening disease.

2. Literature Review

To maximize the performance of their models, some researchers have focused on the projection of CKD using various categorization tactics. Researchers in the work by [11] employed machine learning methods to forecast the occurrence of CKD using clinical data. The study employed a wide variety of logistic regression, SVMs, decision trees, and K-nearest neighbors (KNN) classifiers. The researchers performed a thorough evaluation of various predictive models by comparing information about chronic renal sickness. This allowed them to identify the most effective classifier. According to their findings, the support vector machine (SVM) classifier was more accurate in diagnosing chronic kidney disease (CKD). The primary objective of the study [12] was to contrast the performance of K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) classifiers within the framework of a model for predicting chronic kidney disease (CKD), taking accuracy, precision, and processing time into account. The results demonstrated that the KNN classification model outperformed the SVM classifier. Important new information regarding the relative efficacy of these two approaches is provided by this finding. To use clinical data in future studies to forecast the status of CKD, another study [13] laid out a thorough procedure. Finding relevant attributes, resolving missing values through a collaborative filtering system, and properly processing the data were all steps in the process. The study's findings demonstrated that the additional tree classifier and random forest classifier outperformed the other eleven machine learning approaches examined in terms of bias and accuracy. An earlier study [14] developed a unique method to forecast renal function by combining transfer learning with kidney medical imaging. The research used a ResNet model inside a neural network architecture; this model had previously been trained on the Imagenet database. Researchers were able to successfully establish a connection between 4,505 medical pictures of the kidneys that were tagged with estimated glomerular filtration rates (eGFRs), allowing them to make predictions about renal function. In their study, the researchers presented a new way to detect chronic kidney disease (CKD) [15]. This approach integrated a feature selection strategy based on information gain with a cost-sensitive adaptive boosting classifier. The novel strategy's stated goal was to reduce effort, expense, and time spent screening for CKD. This was accomplished by using a small set of clinical test findings to aid in diagnosis. In an experiment utilizing deep learning methods, [16] looked at the prediction and categorization of CKD according to certain clinical signs. Seven sophisticated deep learning approaches were employed in this study: artificial neural networks (ANN), multilayer perceptron (MLP), gated recurrent units (GRU), long short-term

memory (LSTM), bidirectional LSTM, and bidirectional GRU. The algorithms under consideration were successfully executed utilizing artificial intelligence techniques, hence showcasing their effectiveness in accurately predicting and categorizing chronic kidney disease (CKD) based on thorough assessment criteria. Similarly, to this, researchers used deep learning, feedback on neural networks, logistic regression, and other conventional machine learning approaches [17] to identify CKD. The study aimed to ascertain the most effective strategy by conducting a thorough analysis of their categorization performance. The authors introduced a unique methodology [18] that presents a system utilizing artificial intelligence to predict renal disease inside a cloud-IOT environment. The architectural design used two expert systems, namely neural network (NN) and linear regression (LR). The LR model was employed to identify the significant parameters that put up the development of CKD. The studies collectively contribute to the evolving field of CKD prediction, showcasing diverse methodologies and insights derived from various machine learning and artificial intelligence techniques.

3. Methodology

In Our study, the proposed model is structured in an organized manner, consisting of ten separate phases that have been meticulously crafted to optimize the efficacy of the overarching framework.

3.1 Dataset

The basis of our study is the dataset known as the CKD dataset, which consists of 24 attributes and 400 readings. This dataset was obtained from the machine learning repository - UCI. Eleven numerical features and fourteen notional attributes make up this dataset, which serves as the foundation for further analysis [19].

3.2 Data Analysis

Part two of our study involved analyzing the data extensively to learn more about the distribution of the independent and target variables [20].

Figure 1 depicts the data distribution for our study population. The histogram is a visual depiction of the frequency distribution, displaying the concentration of data points over various ranges. This distribution pattern demonstrates [insights or patterns], emphasizing [important observations] from our collection. The shaded area in the graphic represents [any specific regions of interest, such as outliers or clusters], which provide useful information about the overall properties of the data.

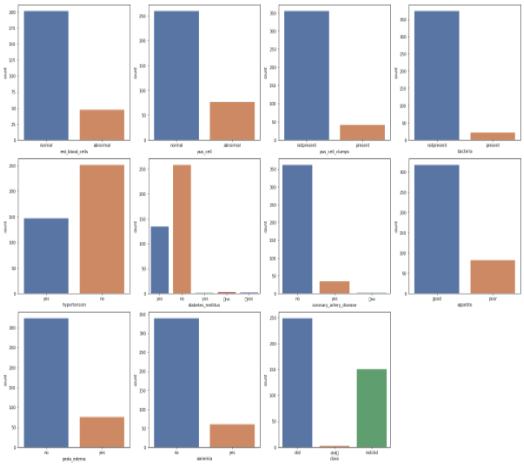

Figure 1: Data distribution

3.3 Handle Missing Data

The frequent issue of absent data was the primary focus of the third portion of our investigation. It was found that the original dataset included missing values after a more comprehensive analysis. To successfully handle the missing value problem, a thorough examination was carried out in this step. Afterward, Understanding the distribution of missing data, as demonstrated in Figure 2, allows us to make informed decisions about how to close these gaps in our research and ensure the credibility of our conclusions. In Figure 3, we show the results of a mean analysis of major variables from our dataset. The results were presented employing mean value data analysis [21,22].

M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe	ane	classif
1.2			15.4	44	7800	5.2	yes	yes	no	good	no	no	ckd
0.8			11.3	38	6000		no	no	no	good	no	no	ckd
1.8			9.6	31	7500		no	yes	no	poor	no	yes	ckd
3.8	111	2.5	11.2	32	6700	3.9	yes	no	no	poor	yes	yes	ckd
1.4			11.6	35	7300	4.6	no	no	no	good	no	no	ckd
1.1	142	3.2	12.2	39	7800	4.4	yes	yes	no	good	yes	no	ckd
24	104	4	12.4	36			no	no	no	good	no	no	ckd
1.1			12.4	44	6900	5	no	yes	no	good	yes	no	ckd
1.9			10.8	33	9600	4	yes	yes	no	good	no	yes	ckd
7.2	114	3.7	9.5	29	12100	3.7	yes	yes	no	poor	no	yes	ckd
4			9.4	28			yes	yes	no	good	no	yes	ckd
2.7	131	4.2	10.8	32	4500	3.8	yes	yes	no	poor	yes	no	ckd
2.1	138	5.8	9.7	28	12200	3.4	yes	yes	yes	poor	yes	no	ckd
4.6	135	3.4	9.8				yes	yes	yes	poor	yes	no	ckd
4.1	130	6.4	5.6	16	11000	2.6	yes	yes	yes	poor	yes	no	ckd
9.6	141	4.9	7.6	24	3800	2.8	yes	no	no	good	no	yes	ckd
2.2	138	4.1	12.6				no	no	no	good	no	no	ckd
5.2	139	3.7	12.1				yes	no	no	poor	no	no	ckd
1.3	135	4.3	12.7	37	11400	4.3	yes	yes	yes	good	no	no	ckd
1.6			10.3	30	5300	3.7	yes	no	yes	good	no	no	ckd
3.9	135	5.2	7.7	24	9200	3.2	yes	yes	yes	poor	yes	yes	ckd

Figure 2: Missing data

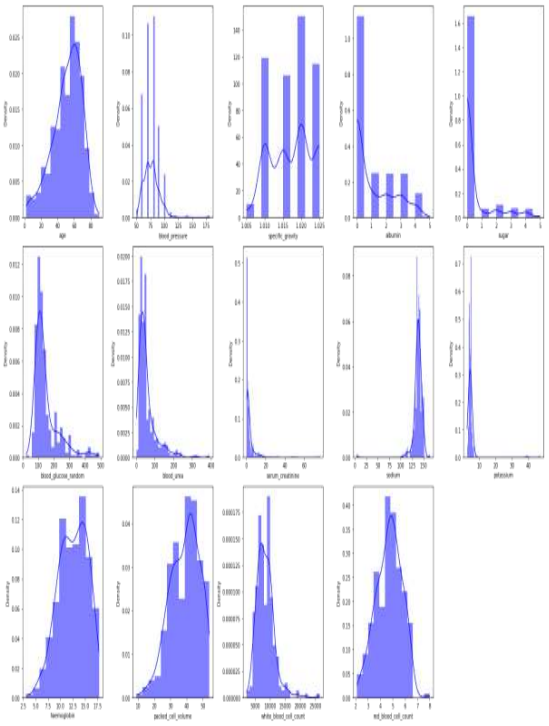

Fig 3: Mean Analysis

3.4 Normalization

The fourth phase involved the implementation of normalization, a vital strategy for scalability in machine learning. The aforementioned methodology involves the normalization of numerical field values to a standardized scale as part of the data pre-processing stage. In the previous phase, missing values were addressed, and subsequently, the min-max scaler was employed to normalize the variable ranges, so ensuring consistency [23, 24].

3.5 Train Test Split

In the fifth phase, the dataset was partitioned into two distinct subsets, namely the training subset and the testing subset. We decided to implement a split ratio of 70-30, where 70% of data was appointed for training intention, while the left 30% was set aside for testing [25, 26, 27].

3.6 Machine Learning Models

During the sixth phase, the emphasis was redirected towards the training of diverse machine learning models. The models included in our study included SVM, KNN, Random Forest, Decision Tree, and AdaBoost. The training procedure utilized the normalized dataset obtained from the preceding step, and thereafter, the models were rated based on their training accuracy [28].

3.7 Ranking

The introduction of the seventh phase involved the implementation of an ensemble technique that was based on the process of rating the individual models. The procedure entailed the allocation of rankings based on the in-sample statistical error of the models, whereby models with lower ranks were considered to possess superior overall accuracy. The ranking process identified the most suitable subset of forecasts for consolidation [29].

3.8 Weighted Average Ensemble

An ensemble method based on weighted averages was revealed in the eighth phase. The employed methodology was the integration of projections generated by multiple distinct models, wherein weights were assigned to each model depending on its respective rank. The development of the ensemble involved the process of optimizing hyperparameters and determining suitable weights to achieve optimal performance. In the given context, the variable

"w" represents the weight assigned to a model's rank depending on its accuracy percentage, whereas the variable "x" denotes the count of votes [30,31,32].

$$(W1X1 + W2X2 + W3X3 + \cdots \ldots . + WnXn) \quad \ldots (1)$$

3.9 Final Prediction

In the ninth phase, the implementation of the ensemble strategy was observed, which involved utilizing the weighted average ensemble model to generate final predictions. This phase encompassed utilizing the model to generate predictions for novel instances of input data [29].

3.10 Model Evaluation

The final step, referred to as the tenth phase, was devoted to the assessment of the performance of the proposed model using data that had not been previously encountered. The testing data was subjected to the same rigorous pre-processing procedures, and the evaluation was performed using parameters derived from the confusion matrix [28, 29]. The measures, namely accuracy, sensitivity, and specificity, collectively offered a full evaluation of the predictive capabilities of the ensemble learning framework that was developed. The comprehensive structure of our proposed model guarantees a thorough and enlightening examination of CKD prediction using ensemble learning techniques.

4. Results

In the evaluation of our proposed method, performance measurement is crucial for assessing the effectiveness of the classifier. The primary metrics used in this paper include sensitivity, specificity, accuracy, and the F1 score.

4.1 Accuracy (ACC)

Accuracy, denoted as ACC, represents the overall success rate of the classifier. It is determined as the ratio of correctly detected to the total number of instances. The formula for accuracy:

$$ACC = TP + TN / (P + N) \ldots (2)$$

In this case, P denotes the positive class, also known as the yes class, and N denotes the negative class, often known as the no class [33]. TP stands for a truly positive rate.

4.2 Sensitivity (TPR)

The fraction of positive cases that the model correctly predicts is measured by sensitivity, commonly known as the true positive rate (TPR). It gauges how well the classifier can recognize positive cases. Equation 3 is the formula used to calculate sensitivity: [34]

$$\text{Sensitivity} = TP/ (TP + FN) \ldots (3)$$

True positives are indicated here by TP and false negatives by FN.

4.3 Specificity (TNR)

The fraction of negative occurrences that the model reliably anticipates is known as specificity, or true negative rate, or TNR. It indicates how well the classifier can recognize negative situations. Equation 4 (the formula for specificity) expressed where TN represents true negatives, and FP stands for false positives [35]:

$$\text{Specificity} = TN/ (FP + TN) \ldots (4)$$

4.4 F1 Score (F1):

The F1 score provides a comprehensive perspective of a classifier's performance. It is a metric that balances both precision and recall. The F1 score (equation 5) is calculated as follows:

$$F1 = \text{Precision} + \text{Recal2} \cdot (\text{Precision} \cdot \text{Recall}) \ldots (5)$$

A classifier with a higher F1 score is more balanced, efficient, and has better recall and

precision. On the other hand, a lower F1 value indicates that recall or precision may be problematic [36].

$$F1=\frac{2TP}{2TP+FP+FN} \ldots . (6)$$

These all-encompassing performance indicators give a full and nuanced assessment of the suggested strategy, shedding light on its precision/recall balance, sensitivity/specificity, and accuracy.

4.5 Confusion Matrix Results

Various models' predictive capacities in the identification of chronic kidney disease are illustrated visually in Figures 4–9, which display the results of the confusion matrix. The given matrices provide a comprehensive analysis through the visuals of the models' classification accuracy by depicting the interplay between real and predicted labels. Figure 4 depicts the Confusion Matrix of the Support Vector Machine (SVM), Figure 5 depicts the Confusion Matrix of the K-Nearest Neighbor (KNN), Figure 6 depicts the Confusion Matrix of the decision Tree, Figure 7 depicts the Confusion Matrix of the random Forest, and Figure 8 depict the Confusion Matrix of (AdaBoost) demonstrate the distribution of all four parameters as shown by the confusion matrices. The utilization of matrices in this context is of utmost importance as they function as essential diagnostic instruments, facilitating a comprehensive evaluation of the efficacy of each model in accurately distinguishing between instances of CKD and non-CKD cases. The efficiency of the suggested ensemble model, which incorporates both ranking and weighted average ensemble strategies, is demonstrated in Figure 9 which depicts the Confusion Matrix of Proposed Model. It shows the cumulative impact of these strategies in achieving correct predictions. The utilization of visual representations in this context serves as a detailed improvement in the comprehensibility of the models' performance, hence facilitating the process of refining and optimizing predictive algorithms for renal dysfunction or disease.

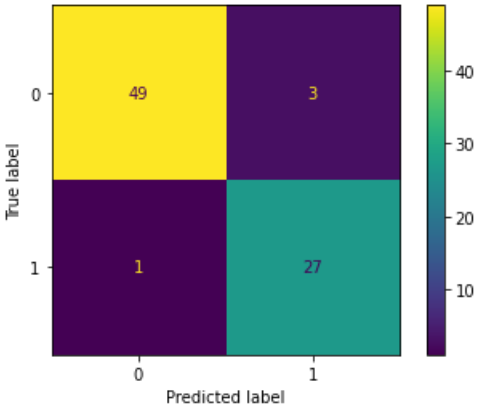

Fig 4: Confusion Matrix of SVM (Support Vector Machine)

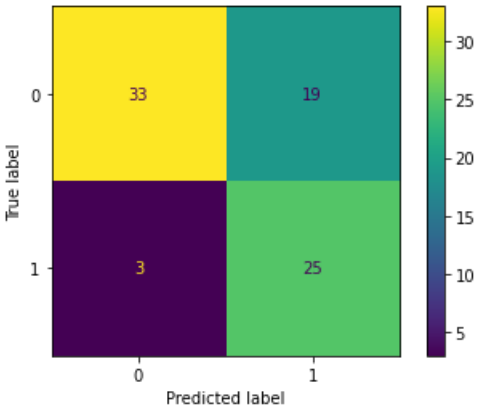

Fig 5: Confusion Matrix of KNN (K-Nearest Neighbor)

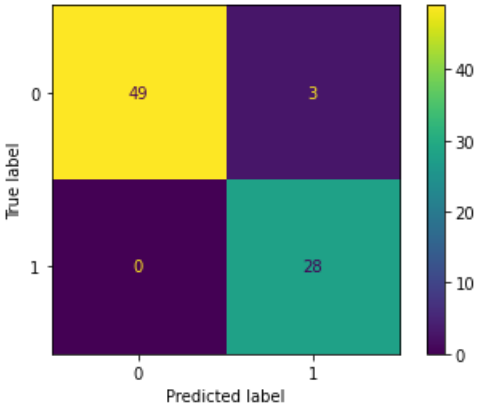

Fig 6: Confusion Matrix of Decision Tree

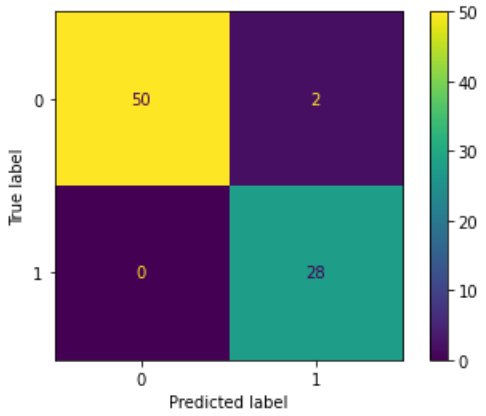

Fig 7: Confusion Matrix of Random Forest

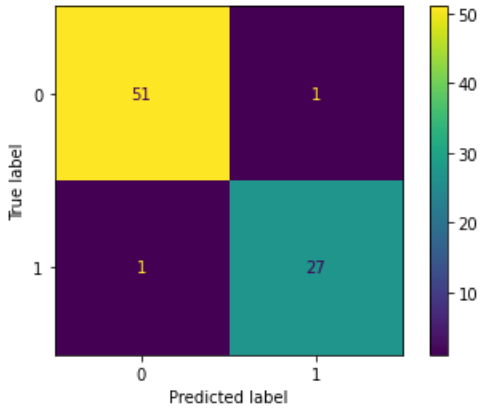

Fig 8: Confusion Matrix of AdABoost

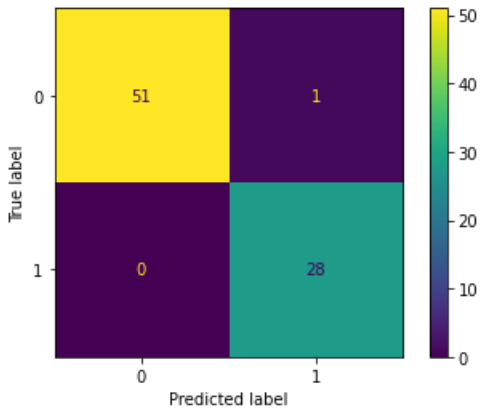

Fig 9: Confusion Matrix of Proposed Model

4.6 Performance Evaluation

The performance evaluation table identifies a concise concept of the primary parameters associated with each machine learning model's ability to predict chronic kidney disease effectively. The metrics encompassed in this study consist of sensitivity, specificity, overall accuracy, and the F1 score. Sensitivity, also known as the true positive rate, measures the proportion of actual positive cases correctly identified. Specificity, or the true negative rate, quantifies the proportion of actual negative cases correctly identified. Overall accuracy represents the overall correctness and robustness of the classification model. In the end, the F1 score is a statistic that strikes a balance between precision and recall, two important performance measures in classification tasks. Significantly, the ensemble model that has been proposed exhibits amazing performance across all criteria, attaining flawless sensitivity, high specificity, spectacular accuracy, and an exceptional F1 score of 99.02%. Our findings in this study focus on the potency of the ensemble methods technique which is an outstanding performer among the individual models, thereby demonstrating its potential results as a dependable tool for the accurate prediction of chronic kidney disease in clinical settings for the more interpretable form.

Table 2: Performance Evaluation

Model	Sensitivity	Specificity	Accuracy	F1 Score
SVM	0.98	0.9	0.94	0.9607
KNN	0.9167	0.5683	0.726	0.74
Decision Tree	1	0.9031	0.9624	0.9702
Random Forest	1	0.9332	0.973	0.9805
AdaBoost	0.9808	0.9642	0.974	0.9806
Proposed Model	1	0.9654	0.9874	0.9902

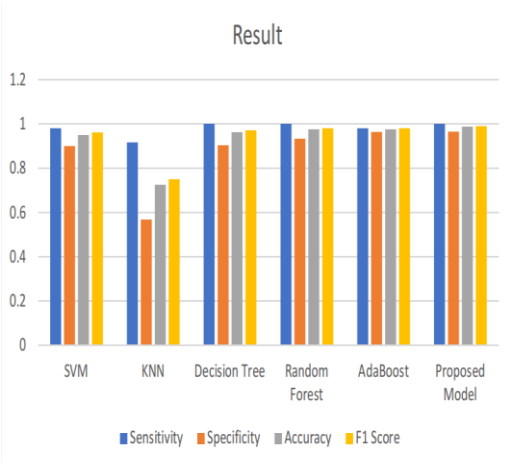

Figure 10: Performance Evaluation visualization

4.7 Analysis and Discussion

To build a strong and robust predictive model for chronic kidney disease, this study compiled a thorough ten-phase technique. The Chronic Kidney Disease dataset is the primary subject that tends to the objective of our work. Several areas are explored in this research, including data analysis, missing data remediation, normalization techniques, and training various machine learning models. Random Forest, AdaBoost, Decision Trees, and SVM are all part of the set of models. A ranking and weighted average sum strategy is introduced in the following phases of the ensemble process. The last step of this approach is to evaluate and forecast the model. In Table 2 and Figure 10 we depict the results, which show the performance measures for each model. The Developed model with a remarkable F1 score of 99.02%, is the ensemble model which performs outstanding and more accurate among the individual models significantly in terms of sensitivity, specificity, accuracy, and overall performance. The ensemble method technique shows a reliable tool in clinical settings for early decision making, as this research highlights its efficient working for predicting early identification of chronic kidney disease (CKD).

Figures 4–9 depict the findings of the confusion matrices, which is graphically represent the classification accuracy of all the models. This helps in evaluating their efficiency in differentiating between non-CKD and chronic kidney disease (CKD) cases. As a diagnostic tool, matrices are vital because they improve the interpretability of model results and help refine algorithms that predict renal illness. A robust and efficient method for chronic kidney disease prediction is an ensemble model that combines ranking and weighted average sum ensemble methods techniques. An outstanding performance, as presented by metrics like sensitivity, specificity, accuracy, and F1 score, shown the potential importance of this diagnostic tool in enhancing the precision of chronic renal disease detection accompanied with kidney disease. However, our research work identifies the necessary need to adapt an ensemble modelling for better improvement of models' performance.

5. Conclusion

Our research demonstrates a new weighed average ensemble model that gives the best and robust performing results on crucial metrics like sensitivity, specificity, accuracy and F1 score, for effective diagnosis and detection of chronic kidney disease (CKD) at early stage. This outstanding approach led to the development of Effective Patient-Specific Decision-Making process for CKD detection, which performed best among the individual machine learning models by making earlier and more accurate diagnoses of Chronic kidney disease. However, despite its promising results, the application of this model in real-world clinical applications faces several technical and operational challenges, particularly when scaled to larger populations globally. The model’s accuracy and reliability may depend strongly on the quality, quantity and consistency and clarity of dataset included features classified with CKD and Non-CKD patients, which can vary widely across different healthcare domains. An Imbalance dataset during data collection approach, there is a evidence of presence of missing and incomplete data, and anomalies in patient demographics may limit the model's effectiveness in clinical applications. Whereas the computational complexity of the ensemble method requires significant resources for both training and real-

time prediction for chronic kidney disease, which may not be easily available in all healthcare facilities, especially to the clinical practitioners. Moreover, the critical limitation is the interpretability of the ensemble model. While it gains high accuracy, the complexity of its decision-making process can make it difficult for clinicians to fully understand and trust its predictions. To overcome these challenges and stimulate further research in this field, several future directions are proposed. At the initial stage of the work, there is a need to improve data quality and standardization across various healthcare environments to enhance the model's capabilities. Developing scalable and computationally efficient algorithms will also be crucial in making these models more accessible in diverse clinical settings. Additionally, efforts should be made to increase the interpretability of ensemble models, possibly through the development of simplified models or visualization tools that clarify how predictions are made. Future research should also focus on validating the model's performance through large-scale clinical trials in real-world settings to ensure its robustness across different populations. Moreover, integrating these models into existing clinical workflows, while considering the ethical implications of their use in critical health decisions, is essential for their successful adoption. However, by highlighting the few limitations and pursuing these research areas, the full potential of ensemble learning models for chronic kidney disease detection can lead to more effective, scalable, and trustworthy diagnostic tools that could be widely implemented in clinical practice for the healthcare providers.

DECLARATIONS

FUNDING – Not sponsored.

CONFLICTS OF INTEREST – The authors included in this research mentioned no conflict of interest

AVAILABILITY OF DATA AND MATERIAL – All the Dataset is available at UCI machine learning repository and cited in methodology section.

AUTHORS' CONTRIBUTIONS – Concept & Idea: Priha Bhatti, Methodology: deign by Priha Bhatti, Experiment: performed by Priha Bhatti; Writing original Draft: Priha Bhatti, Supervision & Guidance: Dr. Affan Alim.

REFERENCES

[1] D. A. Rubenstein, W. Yin, and M. D. Frame, "Flow through the kidney," in *Elsevier eBooks*, 2015, pp. 403–441. doi: 10.1016/b978-0-12-800944-4.00012-3.

[2] S. R. Vaidya, "Chronic kidney disease," *StatPearls - NCBI Bookshelf*, Oct. 24, 2022. Available: https://www.ncbi.nlm.nih.gov/books/NBK535404/

[3] I. Ogobuiro, "Physiology, renal," *StatPearls - NCBI Bookshelf*, Jul. 24, 2023. Available: https://www.ncbi.nlm.nih.gov/books/NBK538339/

[4] B. Bikbov *et al.*, "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," *The Lancet*, vol. 395, no. 10225, pp. 709–733, Feb. 2020, doi: 10.1016/s0140-6736(20)30045-3.

[5] S. Imtiaz, B. Salman, R. Qureshi, M. F. Drohlia, and A. Ahmad, "A review of the epidemiology of chronic kidney disease in Pakistan: A global and regional perspective," *Saudi Journal of Kidney Diseases and Transplantation*, vol. 29, no. 6, p. 1441, Jan. 2018, doi: 10.4103/1319-2442.248307.

[6] Kovesdy, C.P., 2022. Epidemiology of chronic kidney disease: an update 2022. *Kidney International Supplements*, *12*(1), pp.7-11.

[7] V. A. Luyckx, M. Tonelli, and J. W. Stanifer, "The global burden of kidney disease and the sustainable development goals," *Bulletin of the World Health Organization*, vol. 96, no. 6, pp. 414-422D, Apr. 2019, doi: 10.2471/blt.17.206441.

[8] I. R. Rao *et al.*, "Chronic kidney disease of unknown aetiology: A comprehensive review of a global public health problem," *Tropical Medicine & International Health*, vol. 28, no. 8, pp. 588–600, Jul. 2023, doi: 10.1111/tmi.13913.

[9] M. F. Hashmi, "End-Stage renal disease," *StatPearls - NCBI Bookshelf*, Aug. 28, 2023. https://www.ncbi.nlm.nih.gov/books/NBK499861/

[10] S. R. Vaidya, "Chronic kidney disease," *StatPearls - NCBI Bookshelf*, Oct. 24, 2022. https://www.ncbi.nlm.nih.gov/books/NBK535404/

[11] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," *Journal of Big Data*, vol. 9, no. 1, Nov. 2022, doi: 10.1186/s40537-022-00657-5. Available: https://doi.org/10.1186/s40537-022-00657-5

[12] D. Saif, A. Sarhan, and N. M. Elshennawy, "Deep-kidney: an effective deep learning framework for chronic kidney disease prediction," *Health Information Science and Systems*, vol. 12, no. 1, Dec. 2023, doi: 10.1007/s13755-023-00261-8. Available: https://doi.org/10.1007/s13755-023-00261-8

[13] G. M. Ifraz, M. H. Rashid, T. Tazin, S. Bourouis, and M. M. Khan, "Comparative analysis for prediction of kidney disease using intelligent Machine learning methods," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–10, Dec. 2021, doi: 10.1155/2021/6141470.

[14] C. Kuo *et al.*, "Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning," *Npj Digital Medicine*, vol. 2, no. 1, Apr. 2019, doi: 10.1038/s41746-019-0104-2.

[15] S. A. Ebiaredoh-Mienye, T. G. Swart, E. Esenogho, and I. D. Mienye, "A Machine Learning Method with Filter-Based Feature Selection for Improved Prediction of Chronic Kidney Disease," *Bioengineering*, vol. 9, no. 8, p. 350, Jul. 2022, doi: 10.3390/bioengineering9080350.

[16] S. Akter, "Comprehensive performance assessment of deep learning models in early prediction and risk identification of chronic kidney disease," *IEEE Access*, vol. 9, pp. 165184–165206, Jan. 2021, doi: 10.1109/access.2021.3129491.

[17] A. A. Imran, N. Amin, and F. T. Johora, "Classification of Chronic Kidney Disease using Logistic Regression, Feedforward Neural Network and Wide & Deep Learning," *International Conference on Innovation in Engineering and Technology (ICIET)*, Dec. 2018, doi: 10.1109/ciet.2018.8660844.

[18] S. S. R. K. Y. Mall Vipul Narayan ,. Pawan Kumar, "An Ensemble Learning Approach For Chronic Kidney Disease Classification," *pnrjournal.com*, Dec. 2022, doi: 10.47750/pnr.2022.13.S10.279.

[19] Md. A. Islam, Md. Z. H. Majumder, and Md. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," *Journal of Pathology Informatics*, vol. 14, p. 100189, Jan. 2023, doi: 10.1016/j.jpi.2023.100189.

[20] J. Rahnenführer *et al.*, "Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges," *BMC Medicine*, vol. 21, no. 1, May 2023, doi: 10.1186/s12916-023-02858-y.

[21] K. J. Lee *et al.*, "Framework for the treatment and reporting of missing data in observational studies: The Treatment And Reporting of Missing data in Observational Studies framework," *Journal of Clinical Epidemiology*, vol. 134, pp. 79–88, Jun. 2021, doi:10.1016/j.jclinepi.2021.01.008.Available: https://doi.org/10.1016/j.jclinepi.2021.01.008

[22] https://ieeexplore.ieee.org/iel7/6287639/9668973/09730885.pdf

[23] Alghamdi, T.A. and Javaid, N., 2022. A survey of preprocessing methods used for analysis of big data originated from smart grids. *IEEE Access*, *10*, pp.29149-29171.

[24] L. K. T. A. K. Narayan C. K. Jha ,. Ashok Kumar Rai ,. Vipul, "Energy Efficient Routing Protocol Using Next Cluster Head Selection Process In Two-Level Hierarchy For Wireless Sensor Network," *pnrjournal.com*, Dec. 2022, doi: 10.47750/pnr.2022.13.S10.578.

[25] A. C. A. B. Panwar Nishant Ahlawat, Avi, "Research Paper On Image Caption Generator Using Deep Learning," *pnrjournal.com*, Dec. 2022, doi: 10.47750/pnr.2022.13.S10.702.

[26] J. Brownlee, "Train-Test split for evaluating machine learning algorithms," *MachineLearningMastery.com*, Aug. 26, 2020. https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/

[27] "Fig. 2. Example of a process to divide the dataset into two subsets:...," *ResearchGate (2022)*. https://www.researchgate.net/figure/Example-of-a-process-to-divide-the-dataset-into-two-subsets-train-and-test-This-process_fig2_327630039

[28] A. Korchi, M. Youness, A. Abatal, and E. E. A. Ahmed, "Hybrid Machine learning algorithms at the service of student performance," *Research Square (Research Square)*, Jul. 2023, doi: 10.21203/rs.3.rs-3192790/v1.

[29] S. M. F. D. S. Mustapha, "Predictive analysis of students' learning performance using data mining techniques: A comparative study of feature selection methods," *Applied System Innovation*, vol. 6, no. 5, p. 86, Sep. 2023, doi: 10.3390/asi6050086. Available: https://doi.org/10.3390/asi6050086

[30] J. Brownlee, "How to develop a weighted average ensemble with Python," *MachineLearningMastery.com*, May 07, 2021. https://machinelearningmastery.com/weighted-average-ensemble-with-python/

[31] M. Yamasaki, R. Z. Freire, L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. D. S. Coelho, "Optimized hybrid ensemble learning approaches applied to very short-term load forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 155, p. 109579, Jan. 2024, doi: 10.1016/j.ijepes.2023.109579.

[32] M. Yamasaki, R. Z. Freire, L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. D. S. Coelho, "Optimized hybrid ensemble learning approaches applied to very short-term load forecasting," *International Journal of Electrical*

Power & Energy Systems, vol. 155, p. 109579, Jan. 2024, doi: 10.1016/j.ijepes.2023.109579.

[33] J. Brownlee, "How to calculate precision, recall, and F-Measure for imbalanced classification," *MachineLearningMastery.com*, Aug. 01, 2020. https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/

[34] P. Mall, "GLCM Based Feature Extraction and Medical X-RAY Image Classification using Machine Learning Techniques," 2019. https://www.semanticscholar.org/paper/GLCM-Based-Feature-Extraction-and-Medical-X-RAY-Mall-Singh/cb2dd98a1ed95b48ef89d905af4bbf70b29a51d0

[35] A. Baratloo, "Part 1: Simple definition and calculation of accuracy, sensitivity and specificity," *PubMed Central (PMC)*, Jan. 01, 2015. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4614595/

[36] J. K, "The F1 score | Towards Data Science," *Medium*, Oct. 11, 2022. [Online]. Available: https://towardsdatascience.com/the-f1-score-bec2bbc38aa6