International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Classifier Ensembling-Based Machine Learning System for Analysis of Enduring Kidney Disease

Mr. Smeet D. Thakur[1], Dr. Archana T. Bhise[2], Dr. Amit K. Gaikwad[3]

[1]*Research Scholar, Shri JJTU Rajasthan, India,333001*

[2] *Assistant Professor, H.V. Desai College, Pune and Research Scholar, JJTU, Rajasthan,333001*

[3]*Associate Professor, G.H. Raisoni College of Engineering, Amravati, India, 444701*

Email id: [1]smeet87@gmail.com, [2]archanab34@rediffmail.com, [3]amitgaikwad1730@gmail.com

ABSTRACT

People die every year from the failure to diagnose Enduring Kidney Disease (EKD) quickly and accurately, particularly in developing countries where there aren't enough medical personnel and resources to assist in the analysis and treatment of EKD. In this research paper, ensembled classifiers are suggested for the analysis of EKD. Six Machine Learning (ML) algorithms were used to fill in an incomplete data set Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes (NB) Classifier. The SVM, LR classifier, RF, and KNN classifiers have 96.25, 99, 99.87, and 97.5 percent accuracy, respectively. The Naive Bayes classifier has 95% accuracy. These models' accuracy varies when linked in various ways. The mix accuracy of SVM and LR is 98.75 percent, RF and SVM is 99.85 percent, LR and Nave Bayes is 98.75 percent, and KNN and LR is 99.7 percent. The full model is analyzed and get the accuracy of the whole is 99.5%. Comparison findings reveal suggested delivers more accurate numbers than the base paper.

Keywords: *LR, KNN, RF, Machine Learning, and EKD.*

1. Introduction

Machine learning is a cutting-edge technology that employs a variety of factors to analyze large amounts of data. Information processing and computer learning in intelligent devices have been the primary sources of inspiration for machine learning [1]. Computational methodologies, algorithms, and analytical techniques are required. Computer-aided analysis and treatment, as well as the identification of individuals at high risk, are some of the goals of machine learning in the medical field. Ultimately, the goal is to improve patients' physical health at the lowest possible cost [2].

A wide variety of applications, from voice recognition to computer vision to medical diagnostics and engineering, have shown impressive results using ML. Since it has a high mortality rate, EKD has garnered a great deal of attention. World Health Organization (WHO) says that underdeveloped nations are more at risk [3]. EKD could proceed to the point of kidney failure if left untreated and it is curable if detected and addressed promptly. Seven hundred and fifty-three million people worldwide died from enduring renal disease, including 336 million men and 417 million women. Early detection and treatment of enduring renal illness halt its progression to kidney failure [4]. End-stage renal disease treatment is becoming more widely available as part of universal health coverage routes, as shown in Figure 1.

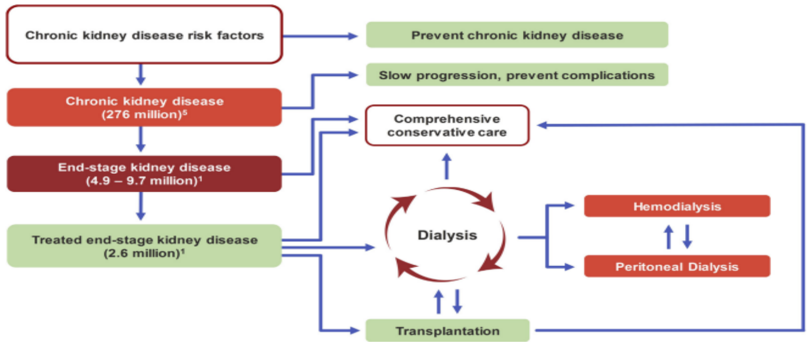

Figure 1: End-Stage Kidney Disease Pathways [5].

EKD must be detected and treated as soon as possible to prevent renal failure, which requires dialysis or kidney transplantation. As a result, approaches to the early analysis and treatment of kidney disease are needed worldwide [6]. People who reside in rural and difficult-to-reach areas have an even greater challenge when it comes to early detection of EKD because of a lack of or insecure primary care. The decline of kidney function in the human body is a progressive process that occurs over some time. For therapy to begin, there must be a three-month history of symptoms. In the first stages of EKD, the phenomenon of continuous degeneration is mostly unnamed and undiscovered [7]. Even if EKD doesn't show any symptoms at first, the patient might be identified by a gradual alteration. Delays in identifying and resolving the problem lead not only to kidney damage but also to malfunctions in the brain and immunological systems of humans [8].

ML has been used in both disease prediction and analysis. Increased public health expenses are a result of the widespread use of dialysis and kidney transplantation, which are now commonplace because of this, EKD has a significant influence on health costs. Costs and the increasing number of people in need of kidney transplants imply that the government should spend more money on treating renal disease [9]. EKD is a condition in which the kidneys are unable to carry out their normal functions of filtering blood and other waste products. Kidney cell degeneration is referred to be enduring when it occurs over a long period [10]. Increased levels of potassium and calcareous ions in the body are a result of this High concentrations of these salts cause a wide range of health problems [11]. Dialysis or kidney transplantation are often required in patients with end-stage EKD. Having a family history of renal disease increases the likelihood of developing EKD. EKD is seen in almost one-third of diabetics, according to research [12]. Patients live with early detection and treatment for enduring kidney disease [13].

EKD is an indication of kidney abnormality and the disease's progression are fewer symptoms and indicators, but this is also the time when the cure is most likely [14]. A human doctor makes an incorrect analysis because of the symptoms, which are often imprecise or ill-defined. Negligence, weariness, a lack of expertise, and bewilderment are some potential causes of human diagnostic mistakes [15]. The models are created using ML and the models were trained using data that has been verified many times. Instead, with computer-based analysis applying machine learning, just a few mouse clicks are required to mine through a massive collection of medical data to get the information required [16].

2. Review of Literature

The following study expands on classifier ensembling-based machine learning system for diagnosis of chronic kidney disease. Several researchers explained their findings as seen below.

Silveira et al., (2022) [17] explained one of the most common and debilitating diseases of the kidneys, CKD is often only discovered in the last stages of the illness. Such a problem necessitates an investment in early prediction. inequitable data sets of a certain size. Holdout validation, multiple stratified Cross-Validations (CV), and nested CV were used to evaluate the models' performance. With manual augmentation and the Delirium tremens (DT) model, the best accuracy is 98.99 percent. Designing early CKD prediction systems with unbalanced and small datasets benefits from the method.

Gokiladevi et al., (2022) [18] described that CKD is becoming a major public health issue that is spreading around the globe. It's a long-term sickness that's associated with higher rates of mortality and morbidity, as well as cardiovascular disease and a correspondingly higher cost of treatment. SVM, RF, LR, and KNN are just a few of the five ML-based categorization models that are used for Diphtheria-Tetanus (DT). With an F-score of 0.99 and an error rate of less than 0.012, the RF model has outperformed the other classifiers. Its accuracy, recall, and F-score are all above 0.99.

Suri et al., (2022) [19] described that Cardiovascular Disease (CVD) is the leading cause of death in the world. In light of rising healthcare expenses, non-invasive CVD risk assessment is essential. In comparison to more modern and rapidly expanding Artificial Intelligence (AI) technologies, conventional methods have exhibited low performance. Office-based, laboratory-based, and image-based phenotypes were the most often employed biomarkers. Carotid artery risk prediction has demonstrated good results with the use of surrogate carotid scans.

Yashwanth et al., (2022) [20] explained that biosciences have progressed to a greater level, and electronic health records have created vast volumes of data. This has resulted in an urgent requirement for the development of information from this vast volume of data. Data mining and machine learning play a significant part in this area of biosciences. Mean, mode, and median-based pre-processing approaches using neural networks were shown to be much more accurate than K-NN, SVM, Regression Tree, and Classification Tree in classification and detection accuracy.

Singh et al., (2022) [21] suggested that there is a strong correlation between diabetes and high blood pressure, and CKD. The disease is known as CKD diagnosed by measuring the Glomerular Filtration Rate (GFR) and kidney damage indicators. Death occurs more often in the early stages of CKD. Doctors must work hard to catch the many disorders associated with CKD early on if want to avoid it. There was no difference in accuracy between the deep neural model and the other four classifiers.

Moreno-Sanchez et al., (2021) [22] stated that CKD, which leads to early death if diagnosed late, is now a worldwide concern that costs healthcare systems a lot of money. A major contribution of the work was the development of a method for choosing the optimal prediction model while maintaining a balance between accuracy and explainability. Using cross-validation and fresh unknown data, the best-balanced explainable prediction model uses an extreme gradient boosting classifier across three features, with an accuracy rate of 99.2 percent. As a consequence of this investigation, the packed cell value, followed by specific gravity and hypertension, are the most critical factors influencing the model's prediction findings.

Jongbo, O. A. et al., (2020) [23] studied that millions of people die of CKD, which is mostly caused by a combination of bad lifestyle choices and genetic predispositions. The necessity for speedy and exact diagnosis prompted the introduction of data mining tools. These strategies are evaluated using an available dataset from the Unique Client Identifier (UCI) ML library. Recently, data drilling algorithms have been extensively researched in the identification of chronic renal disease, with an emphasis on accuracy, either Preprocessing, Feature Extraction (FE), and classification make up a conceptual D-ACO structure. Therefore, the model has the potential to be utilized to properly detect chronic renal illness.

Ghosh et al., (2020) [24] stated that CKD, a slowly progressive and sometimes misdiagnosed illness is one of contemporary medicine's major causes of mortality. This critical issue has not been addressed, a great number of men and women are now suffering each year. Early detection and treatment of sickness save patients' lives. Furthermore, with a large dataset, the evaluation technique of a machine learning system can determine the stage of this dangerous illness much sooner. Gradient Boosting (GB) Classifiers have the highest predictability, with an accuracy of 99.81 percent. A range of evaluation metrics has been provided to provide a more thorough picture of the outcomes. These metrics are used to choose the most efficient and effective algorithms for the job at hand.

Elhoseny et al., (2019) [25] suggested a useful healthcare forecasting and categorization method based on the Diplomate, Academy of Chiropractic Orthopedists (D-ACO) algorithm for general Use of CKD for density-based feature selection. Before the introduction of the ACO-based classifier, the suggested canny system removes duplicate and unneeded functionality from DFS. There are three processes in the proposed D-ACO structure: preprocessing, Feature Extraction (FE), and classification.

Koyner, Jay L et al., (2018) [26] established that a model for estimating the risk of serious kidney damage based on observational longitudinal data from hospital patients' electronic health records. Through serum creatinine verification and numerous clinical areas, publicly available Electronic Health Record (EHR) information is utilized to forecast imminent severe renal damage underlying serum creatinine changes with remarkable cross-sectional accuracy. This model's continuous usage would allow for early renal intervention for individuals at high risk of serious damage.

This section contains the comparative study of the literature review shown in table 1.

Table 1. Comparative analysis of literature review.

Author [Ref.]	**Technique**	**Outcome**
Silveira et al., (2022) [17]	DT model	Holdout validation, multiple stratified Cross-Validations (CV), nested CV, manual augmentation, and the Delirium Tremens (DT) model had the best accuracy is 98.99 percent.
Gokiladevi et al., (2022) [18]	RF model	F-score of 0.99 and an error rate of less than 0.012, the RF model has outperformed the other classifiers. Its accuracy, recall, and F-score are all above 0.99.
Suri et al., (2022) [19]	Carotid Scan	Carotid artery risk prediction has demonstrated good results with the use of surrogate carotid scans.

Yashwanth et al., (2022) [20]	Electronic Health Records	Pre-processing approaches using neural networks were shown to be much more accurate than K-NN, SVM, Regression Tree, and Classification Tree in classification and detection accuracy.
Singh et al., (2022) [21]	GFR	Detection accuracy extended, and there were no differences among the classifiers.
Moreno-Sanchez et al., (2021) [22]	KNN Nearest Neighbor	With cross-validation and fresh unknown data, the extreme gradient boosting classifiers are used across three features in the best-balanced explainable prediction model, with an accuracy rate of 99.2 percent.
Jongbo, O. A., et al. (2020) [23]	UCI	The model has the potential to be utilized to properly detect chronic renal illness.
Ghosh, P., et al. (2020) [24]	GB	Gradient Boosting (GB) Classifiers have the highest predictability, with an accuracy of 99.81 percent.
Elhoseny et al. (2019) [25]	DFS	the model of multivariate analysis, which is useful to predict the existence or the absence of a function or consequence. Based on the values of several different predictor variables.
Koyner, Jay L et al. (2018) [26]	EHR	This model's continuous usage would allow for early renal intervention for individuals at high risk of serious damage.

3. Different Types of Classifiers

Following are the classifiers that are used in the intended methodology.

- **Feed-Forward Back Propagation Neural Networks (FFBPNN)**

It comprises more than 1 level of neurons, much as neural networks of cascade-forward backpropagation. To construct the model and train the model, a technique such as a cascade forward neural network is taken out. One layer of sigmoid neurons has single-layered feed-forward neural networks, accompanied by linear neurons' an output layer [27].

$$X_k = \{x_1, \dots, X_k\} \qquad (1)$$

Where $g(X_k) \approx Y_k$ for training pairs (X_n, Y_n) for all $1 \leq k \leq n$, where (x,y) are training pairs and g is computing parameters.

- **Logistic Regression**

The LR method is widely used for linear classification. It allows a relationship between an independent variable and dependent variables to form a multivariate regression. LR is the model of multivariate analysis, which is useful to predict the existence, or the absence of a function or consequence based on the values of a series of various predictor variables.

$$\text{Log} = \left[\frac{p}{1-p}\right] = \beta_0 + \beta(\text{Age}) \qquad (2)$$

Where p is the probability and β_0 is intercept value. The LR simply divides the data into two parts with the help of a line of regression as shown in Figure 2 [28].

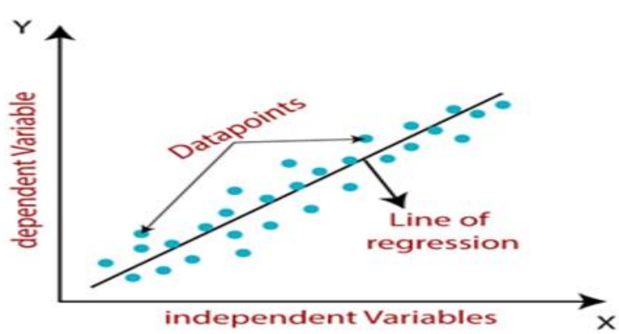

Figure 2. Logistics Regression

- **Random Forest**

Random forest is widely used in machine learning algorithms to solve problems. It is established on the principle of collective learning, which is a technique for merging multiple classifiers to resolve various complicated problems and enhance the model's accuracy [29]. The basic diagram of the RF algorithms is shown in Figure 3.

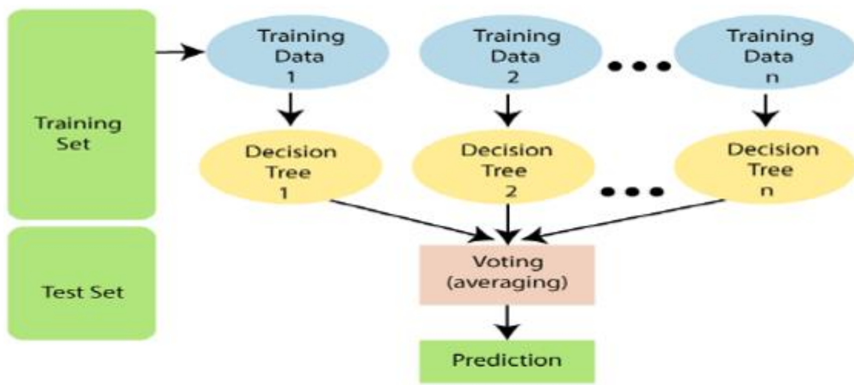

Figure 3. Random Forest Algorithms.

- **SVM**

It is a combination of supervised learning techniques for regression and classification. The SVM method is focused on Structural Risk Minimization (SRM). It plans the input vector to a better-dimensional space in which a maximum dividing hyperplane is built [31]. It utilizes the dividing or separating hyperplane with the expression to view this training information:

$$w.m + b = o \tag{3}$$

In this equation 1, b is a scalar, and A p-dimensional vector, w is perpendicular to the horizontal dividing hyperplane. To optimize the margin's potential, the offset parameter b is used. In the absence of b, the hyperplane must pass through the origin. It is comparable to examining the entire margin for SVM and parallel hyperplanes. A formula is applied to explain the equation of the parallel hyperplane:

$$w.m + b = 1 \tag{4}$$

$$w.m + b = -1 \tag{5}$$

These hyperplanes chose so that there are no points between them If the training data are spread linearly, then the distance between the data points will be a constant value. As seen in Figure 4, an SVM trained on instances from 2 has maximum edge hyperplane classes [32]. To stimulate data points, to confirm that for each I w. li – b ≥ 1 also w. li – b ≤ -1 re-written as.

$$mi\,(w.li - b) \geq 1, 1 \leq i \leq n \tag{6}$$

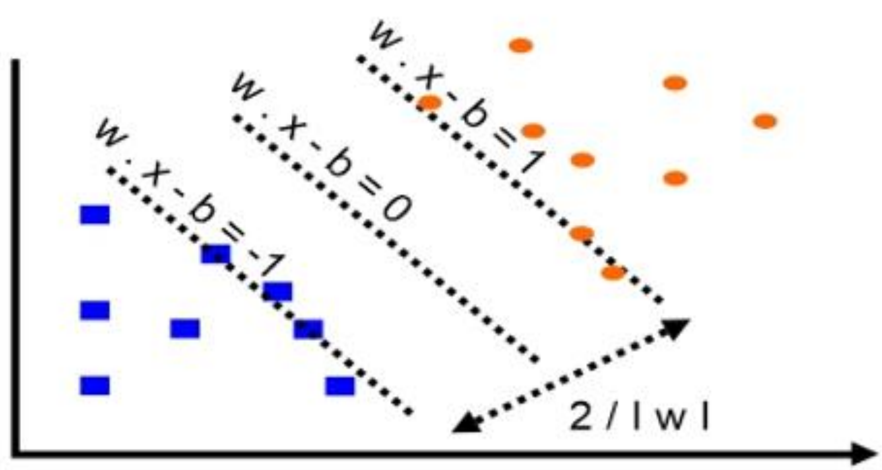

Figure 4. An SVM trained with samples from 2 classes.

- **KNN**

Unlabeled findings are classified using the KNN classifier by assigning them to the class with the most related labeled instances. For both the training and evaluation datasets, characteristics of findings are obtained.

$$\mathrm{D}(\mathrm{p},\mathrm{q}) = \sqrt{(\mathrm{p}_1 - \mathrm{q}_1)^2 + (\mathrm{p}_2 - \mathrm{q}_2)^2 + \cdots + (\mathrm{p}_\mathrm{n} - \mathrm{q}_\mathrm{n})^2} \tag{7}$$

where q and p are the issues to be associated and n are the characteristics to be compared [33].

- **Naive Bayes**

Specified the insurmountable sample difficulty of studying Bayesian classifiers it must seek methods to decrease it. The Naive Bayes classifier does this by assuming conditional freedom which decreases the number of constraints to be calculated when modeling P(X|Y) from 2(2n-1) to only 2n.

Assumed L, M, and N are 3 sets of random variables. If and only if the probability distribution leading L is independent of the value of M provided N, then L is conditionally independent of M given N. i.e.

$$(\forall \mathrm{i},\mathrm{j},\mathrm{k})\mathrm{P}(\mathrm{L} = \mathrm{li}\ |\mathrm{M} = \mathrm{mj}, \mathrm{N} = \mathrm{nk}) = \mathrm{P}(\mathrm{L} = \mathrm{li}|\ \mathrm{N} = \mathrm{nk}) \tag{8}$$

Data is categorized using a looping split in the case space. Directed trees do not have outgoing edges, and the root node of the decision tree is the node from which all other nodes branch out. A single tip has been received by each of the additional nodes [34].

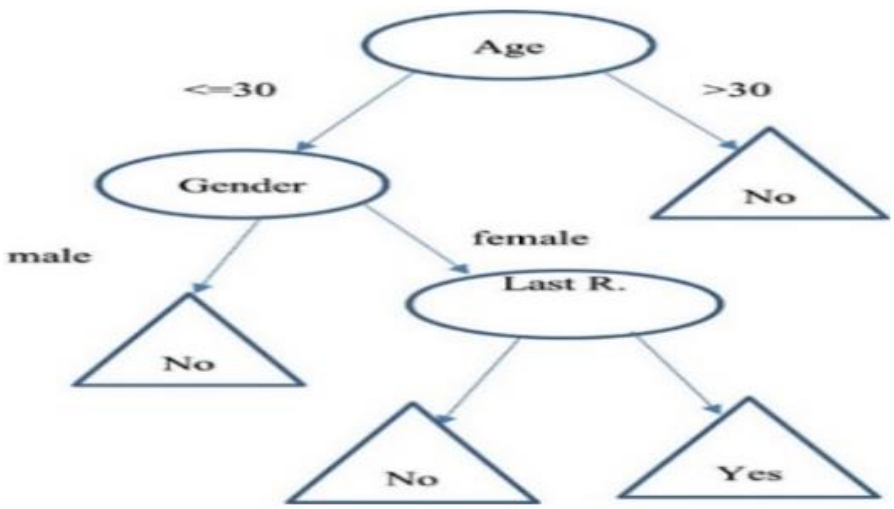

Figure 5. Decision Treed

An internal or evaluation node is a node with outward edges. Figure 5 shows a decision tree for determining if a prospective customer can reply to a direct mailing [35]. The rest nodes are stated as leaves. An internal or evaluation node is a node with outward edges. The rest nodes are mentioned as leaves. It is also called decision nodes or terminals.

$$\mathrm{E}(\mathrm{s}) = \sum_{i=1}^{c} -p_i log_2 p_i \tag{9}$$

Where E is entropy and p is probability.

4. Background Study

The detection of EKD is aided by a ML approach and the EKD data set was obtained from the ML repository of the University of California, Irvine (UCI). Filling in the blanks for each defective sample required the use of KNN imputation, which used a large number of samples with remarkably similar metrics. For a variety of reasons, patients often fail to take the necessary measures in real-life medical situations. An incomplete data set was filled in using six machine learning techniques (LR, RF, SVM, K-Nearest Neighbors, NB Classifier, and FFNN). Following the evaluation of the misjudgments induced by the previous models, it also offered an integrated model combining LR and RF with convolution that could achieve an average accuracy [36].

5. Methodology

The suggested methodology addresses Diagnosing Enduring Kidney Disease using Classifier Ensembling. In the suggested methodology there are two types of Datasets i.e., Dataset 1 and Dataset 2. Both Datasets (Dataset 1 & Dataset 2) give input to the Data preprocessing. This methodology has been described in various steps which have been given below.

Step 1: Dataset

Two types of datasets are taken from UCI for the analysis of EKD. The UCI database from the ML repository is applied by the majority of the researchers.

Step 2: **Pre-processing**

The primary step after obtaining the dataset is to do data pre-processing. Perform different pre-processing methods in this phase, such as data cleaning, data transformation, managing missing values, attribute selection, and many more.

Step 3: Density-based failure selection system

After preprocessing the data, now the information is going to a Density-based failure selection system.

Step 4: Missing value filling using stochastic regression imputation

Using KNN imputation, numerous full samples with similar measurements are selected and used to replace missing data for each incomplete sample.

Step 5: Classifiers

Different classifiers, such as Logistic Regression, Random Forest, SVM, KNN, Nave Bayes, and feed-forward NN, are used in this approach. Figure 6 represents the pictorial representation of the intended methodology.

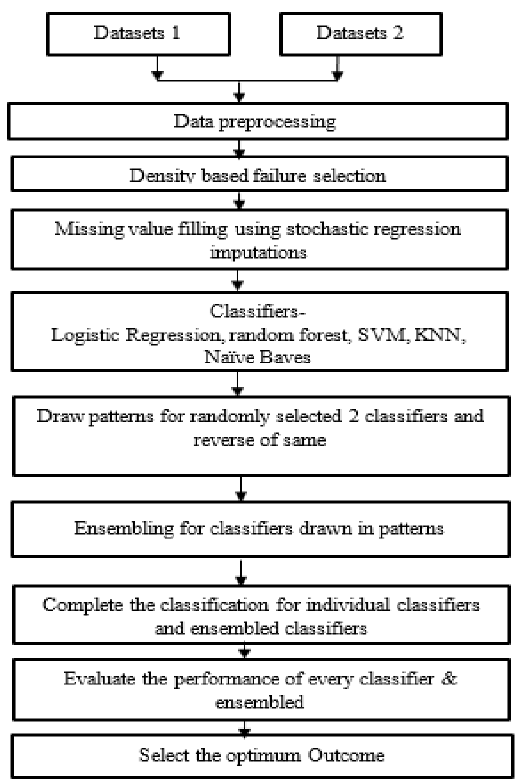

Figure 6.Pictorial representation of Planned Methodology.

Step 6: Draw patterns for randomly selected 2 classifiers and reverse of same

In this step, the pattern is drawn by randomly selecting 2 classifiers and drawing a reverse pattern of the same.

Step 7: Ensembling of Machine Learning

Model outputs improved by integrating various models rather than using just one, using ensemble approaches. The accuracy of the findings is greatly improved by using integrated models. As a consequence, machine learning ensemble approaches are becoming more widely accepted.

Step 8: Complete the classification for individual classifiers and ensembled classifiers

Now, in this step, the classification is done for each classifier and ensemble classifiers.

Step 9: Evaluate the performance of every classifier & ensembled

In this step, the performance of each classifier is computed.

Step 10: Select the Optimum Output

Following an evaluation of classifiers and ensembles, choose the best Output.

6. Conclusion and Future Scope

Millions of people throughout the globe suffer from enduring kidney disease, which is a serious health problem that puts them at risk of financial, social, and medical hardship. EKD could be diagnosed with the use of many automated diagnostic methods. In this research paper, classifier assembling is used to diagnose EKD. Dataset 1 and Dataset 2 are the two kinds of datasets used in the suggested technique. The proposed model is dependent upon the different models, and it shows different accuracy. The accuracy of the SVM, LR classifier, RF, and KNN classifiers are respectively 96.25 percent, 99 percent, 99.87 percent, 97.5 percent, and the accuracy of the Naive Bayes classifier is 95 percent. When these models are coupled in different ways, the results show varying degrees of accuracy. Experimental results also give the mixing accuracy of SVM and LR is 98.75%, RF and SVM is 99.85%, LR and Naïve Bayes is 98.75% and KNN and LR is 99.7%. The whole model is examined, and the accuracy rises to 99.95%. The comparison results show that the suggested give more accurate values than the base paper results. Increasingly complex and representative data will be collected in the future to train the model and improve its generalization ability so that it can detect the severity of the condition. The model is expected to increase in accuracy as more and better data becomes available.

References

[1]. Palma-Mendoza, Raul-Jose, Daniel Rodriguez, and Luis De-Marcos. "Distributed ReliefF-based feature selection in Spark." Knowledge and Information Systems 57, no. 1 (2018): 1-20.

[2]. M. Nassar, H. Safa, A. A. Mutawa, A. Helal, and I. Gaba, “Chi-squared feature selection over Apache spark,” in Proceedings of the 23rd International Database Applications & Engineering Symposium, pp. 1–5, Athens Greece, June 2019.

[3]. Bikbov, Boris, Norberto Perico, and Giuseppe Remuzzi. "Disparities in enduring kidney disease prevalence among males and females in 195 countries: analysis of the global burden of disease 2016 study." Nephron 139 (2018): 313-318.

[4]. Stapleton, Melanie T. "An Expectant Future for Patients with End-Stage Kidney Disease." Physician Assistant Clinics 7, no. 2 (2022): 385-395.

[5]. Harris, David CH, Simon J. Davies, Fredric O. Finkelstein, Vivekanand Jha, Jo-Ann Donner, Georgi Abraham, Aminu K. Bello et al. "Increasing access to integrated ESKD care as part of universal health coverage." Kidney international 95, no. 4 (2019): S1-S33.

[6]. Huang, M. Q., J. Ninić, and Q. B. Zhang. "BIM, machine learning and computer vision techniques in underground construction: Current status and future perspectives." Tunneling and Underground Space Technology 108 (2021): 103677.

[7]. Rácz, Anita, Dávid Bajusz, and Károly Héberger. "Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification." Molecules 26, no. 4 (2021): 1111.

[8]. Rezaoana, Noortaz, Mohammad Shahadat Hossain, and Karl Andersson. "Detection and classification of skin cancer by using a parallel CNN model." In 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 380-386. IEEE, 2020.

[9]. Sultana, Zinnia, Lutfun Nahar, Nanziba Basin, and Mohammad Shahadat Hossain. "Inference and Learning Methodology of Belief Rule Based Expert System to Assess Chikungunya." In International Conference on Applied Intelligence and Informatics, pp. 3-16. Springer, Cham, 2021.

[10]. Revathy, S., B. Bharathi, P. Jeyanthi, and M. Ramesh. "Enduring kidney disease prediction using machine learning models." International Journal of Engineering and Advanced Technology (IJEAT) 9 (2019).

[11]. Jena, Lambodar, Bichitrananda Patra, Soumen Nayak, Sushruta Mishra, and Sushreeta Tripathy. "Risk prediction of kidney disease using machine learning strategies." In Intelligent and cloud computing, pp. 485-494. Springer, Singapore, 2021.

[12]. Besra, Bhagyashree, and Banshidhar Majhi. "An analysis on enduring kidney disease prediction system: Cleaning, preprocessing, and effective classification of data." In Recent findings in intelligent computing techniques, pp. 473-480. Springer, Singapore, 2019.

[13]. Murshid, Golam, Thakor Parvez, Nagani Fazal, Lakhani Azaz, and Mohammad Asif. "Data mining techniques to predict enduring kidney disease." Int. J. Scientific Res. Comput. Sci., Eng. Inf. Technol 5, no. 2 (2019): 1220-6.

[14]. Razia, Shaik, P. Siva Kumar, and A. Srinivasa Rao. "Machine learning techniques for Thyroid disease analysis: A systematic review." Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough (2020): 203-212.

[15]. Siwy, Justyna, Petra Zürbig, Angel Argiles, Joachim Beige, Marion Haubitz, Joachim Jankowski, Bruce A. Julian et al. "Noninvasive analysis of enduring kidney diseases using urinary proteome analysis." Nephrology Dialysis Transplantation 32, no. 12 (2017): 2079-2089.

[16]. Wu, Yuanli, Hong Wang, and Fei Wu. "Automatic classification of pulmonary tuberculosis and sarcoidosis based on random forest." In 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1-5. IEEE, 2017.

[17]. Silveira, Andressa CM da, Álvaro Sobrinho, Leandro Dias da Silva, Evandro de Barros Costa, Maria Eliete Pinheiro, and Angelo Perkusich. "Exploring Early Prediction of Chronic Kidney Disease Using Machine Learning Algorithms for Small and Imbalanced Datasets." Applied Sciences 12, no. 7 (2022): 3673.

[18]. Gokiladevi, M., Sundar Santhoshkumar, and Vijayakumar Varadarajan. "MACHINE LEARNING ALGORITHM SELECTION FOR CHRONIC KIDNEY DISEASE DIAGNOSIS AND CLASSIFICATION." Malaysian Journal of Computer Science (2022): 102-115.

[19]. Suri, Jasjit S., Mrinalini Bhagawati, Sudip Paul, Athanasios D. Protogerou, Petros P. Sfikakis, George D. Kitas, Narendra N. Khanna et al. "A Powerful Paradigm for Cardiovascular Risk Stratification Using Multiclass, Multi-Label, and Ensemble-Based Machine Learning Paradigms: A Narrative Review." Diagnostics 12, no. 3 (2022): 722.

[20]. Yashwanth, Lakkoju Sairam, Chengalvala Anudeep, Gorijilli Venkata Sai Mukesh, and Kompella Subhadra. "Identification Of Chronic Kidney Disease (CKD) Using Artificial Neural Networks Algorithms."

[21]. Singh, Vijendra, Vijayan K. Asari, and Rajkumar Rajasekaran. "A Deep Neural Network for Early Detection and Prediction of Chronic Kidney Disease." Diagnostics 12, no. 1 (2022): 116.

[22]. Moreno-Sanchez, Pedro A. "Development and evaluation of an Explainable Prediction Model for Chronic Kidney Disease Patients based on Ensemble Trees." arXiv preprint arXiv:2105.10368 (2021).

[23]. Jongbo, Olayinka Ayodele, Adebayo Olusola Adetunmbi, Roseline Bosede Ogunrinde, and Bukola Badeji-Ajisafe. "Development of an ensemble approach to chronic kidney disease diagnosis." Scientific African 8 (2020): e00456.

[24]. Ghosh, Pronab, FM Javed Mehedi Shamrat, Shahana Sultana, Saima Afrin, Atqiya Abida Anjum, and Aliza Ahmed Khan. "Optimization of prediction method of chronic kidney disease using a machine learning algorithm." In 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (Isai-NLP), pp. 1-6. IEEE, 2020.

[25]. Elhoseny, Mohamed, K. Shankar, and J. Uthayakumar. "Intelligent diagnostic prediction and classification system for chronic kidney disease." Scientific reports 9, no. 1 (2019): 1-14.

[26]. Koyner, Jay L., Kyle A. Carey, Dana P. Edelson, and Matthew M. Churpek. "The development of a machine learning inpatient acute kidney injury prediction model." Critical care medicine 46, no. 7 (2018): 1070-1077.

[27]. Roy, Pratik, G. S. Mahapatra, and Kashi Nath Dey. "Neuro-genetic approach on logistic model-based software reliability prediction." Expert systems with Applications 42, no. 10 (2015): 4709-4718.

[28]. Ribokaitė, Lina. "Outlier detection in multidimensional streaming data." Ph.D. diss., Vilniaus universitetas, 2021.

[29]. Hengl, Tomislav, Madlene Nussbaum, Marvin N. Wright, Gerard BM Heuvelink, and Benedikt Gräler. "Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables." PeerJ 6 (2018): e5518.

[30]. Li, Ke, Nan Yu, Pengfei Li, Shimin Song, Yalei Wu, Yang Li, and Meng Liu. "Multi-label spacecraft electrical signal classification method based on DBN and random forest." PLoS One 12, no. 5 (2017): e0176614.

[31]. Sathyadevan, Shiju, and Remya R. Nair. "Comparative analysis of decision tree algorithms: ID3, C4. 5 and random forest." In Computational intelligence in data mining-volume 1, pp. 549-562. Springer, New Delhi, 2015.