JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 02
2024
Research Article
https://doi.org/10.56979/702/2024

A Literature Analysis for the Prediction of Chronic Kidney Diseases

Ahmad Abdullah[1], Asif Raza[1*], Qaisar Rasool[1], Umair Rashid[1], Muhammad Minam Aziz[2], and Saad Rasool[1]

[1]Department of Computer Science, Bahuddin Zakariya University, Multan, 60000, Pakistan.
[2]Department of Electrical and Computer Engineering, University of Alabama at Birmingham, Birmingham, AL 35294, United States.
[*]Corresponding Author: Asif Raza. Email: asif79@bzu.edu.pk

Received: July 03, 2024 **Accepted:** August 28, 2024 **Published:** September 01, 2024

Abstract: It is a global health issue that will, nowadays, affects millions of people with severe relief on their quality of life due to the presence of Chronic Kidney Disease (CKD). Many can progress from short-term to long term kidney failure which has complications such as anaemia, osteoporosis, cardiovascular diseases and end stage renal disease. This makes early identification and management of CKD paramount especially for the slow progression of the disease and better results for the patients. The current research aims at focusing on the capability of trained machine learning models to identify community, clinical, and laboratory variables to estimate the likelihood of the initial identification of CKD. Demographic factors which include age, sex, race and economic status are fundamental in determining possibility of contracting CKD. Proteinuria, serum creatinine and estimated glomerular filtration rate (eGFR) are the essential markers of the renal function and CKD development. Algorithms from the machine learning family such as decision trees, random forests as well as the neural networks are used to find out the patterns that may be associated with the asymptomatic high risk patients and in turn help in the prediction of risk towards development of CKD. This holistic approach improves identification and management of risk factors; perhaps keeping off or remitting the early development of CKD.

Keywords: Kidney Chronic Diseases; Systematic Review; CKD Prediction; Machine Learning; Renal Disease Prediction

1. Introduction

A long-term illness named chronic kidney diseases CKD causes the kidneys to progressively lose function. Millions of individuals around the world are impacted by this serious public health problem, with a high illness burden among older adults, people with diabetes and hypertension, and particular ethnic groups. CKD is distinguished by a progressive loss of renal function, which can result in a building of toxins and waste products in the blood, as well as an increased risk of cardiovascular disease [1] and other consequences. We will only have 10% of our kidneys working, and 90% of the time there will be no symptoms. As a result, it has earned the moniker "silent killer." Diabetes, hypertension, smoking, obesity, age, family history, alcohol consumption, and other factors all have a role in kidney disease. The most common causes of CKD are diabetes and high blood pressure, both of which can harm the kidneys' tiny blood capillaries and impair their capacity to filter waste from the blood. Genetic diseases, infections, and certain drugs are some of the other causes of CKD. CKD may lead to Blood in the urine, shortness of breath, nausea and vomiting, irregularities in urinary function, and other symptoms. Since the kidneys are not functioning properly, CKD leads to blood infiltration [2]. Patients in the latter stages of the condition may develop anemia, bone disease, and electrolyte abnormalities, as well as an increased risk of cardiovascular disease. CKD is often diagnosed based on the results of blood and urine tests, as well as imaging examinations such as a computed tomography (CT) scan or magnetic resonance imaging (MRI). Treatment options for CKD include lifestyle modifications like eating a nutritious diet and exercising regularly, as

well as blood pressure and diabetic drugs. Patients in severe stages of the illness may require dialysis or a kidney transplant.

Recent studies have shown that early identification and therapy of CKD can dramatically reduce disease development and improve outcomes. Avoiding and managing risk factors for CKD, such as hypertension and diabetes, is critical for preventing the development and progression of CKD, in addition to early identification and therapy. This involves sustaining healthy lifestyle choices such as regular exercise, a nutritious diet, quitting smoking, and controlling blood pressure and blood sugar levels.

In this analysis of the literature, we conduct a systematically review and compared the performance of the various machine learning and deep learning models across multiple studies. Our contribution lies in the synthesizing these finding to find model effectiveness pattern and trends. A comprehensive understanding of the current state of the art by highlighting the superior performance of the deep learning approaches. This analysis precise these models and recommend optimal research topic.

In summary, millions of people throughout the world suffer from CKD, a serious and progressive illness. Early diagnosis and treatment can significantly slow the progression of the disease and enhance results.

2. Research Objectives

The main goal of this study is to examine how machine learning (ML) and deep learning (DL) methodologies have changed over time, from 2020 to 2024, in terms of methods for predicting CKD. The main goals of this research study are as follows:

- To determine the realistic CKD systems.
- To present an overview of existing research studies based on the benefits of CKD categorization and the direction of future research.
- Determine the most recent research trends and publishing interests based on the categorization of CKD.

2.1. Include & Exclude Criteria

The primary goal of this article is to present a systematic review deep and meta-analysis of relevant research papers for the purpose of this skin disease diagnosis using deep learning approaches. The article's inclusion and exclusion criteria are presented in Table 1 below.

Table 1. Include & Exclude Criteria.

Include Criteria	Exclude Criteria
IC1- Articles presenting concepts of CKD prediction.	EC1- Article that are not focused on CKD
IC2- Articles that are focused on CKD applications and their implementation.	EC2- Articles about micro services-based software and approaches
IC3- Articles that are relevant to machine learning	EC3- Articles presenting general focus on CKD
IC4- Articles that are written in English Language	EC4- Books, thesis and published abstracts
IC5 - Studies from any geographical location.	EC5- Articles that are published before 2020

2.2. Search String

(“kidney Chronic Disease” OR “chronic kidney disease” OR “chronic renal disease” OR “CKD” OR “kidney disease”) AND (“prediction” OR “predicts” OR “diagnose”).

(“last stage kidney disease” OR “kidney failure disease” OR “renal disease” OR “kidney failure”) AND (“Prediction” OR “Predicts” OR “diagnose”).

3. Literature Review

In terms of computing speed and prediction accuracy, Bhaskar and Suchetha's [3] suggested technique beats the conventional methods. Their combined SVM and correlational neural network produced a prediction accuracy of 98.67%. The testing results reveal a 9.85% decrease in overall computing time when compared to the standard CNN technique.

In a supervised learning environment, Islam et al. [4] assessed twelve alternative machine learning-based classifiers for CKD classification. In which the XgBoost classifier outperforms all others with an accuracy of 98.3%.

Nag and Jammula [5] compare the supervised classification methods (decision tree and random forest) with the neural network system. Using decision tree and random forest, they achieve accuracy of around 96.67% and 99.17%, respectively.

The work in [6] used various machine learning algorithms to provide an investigation into early prediction of chronic kidney disease. The classification of topics based on the importance of performance indicators was evaluated with or without normalization, scaling and imputation. After data preprocessing for all classifiers, the random forest classifier has the best accuracy (95.6%).

The main objective of the research in [7] is to develop and evaluate the performance of various unsupervised algorithms to determine the best combination that can improve the accuracy and cost analysis. The researchers developed five unsupervised algorithms using K-Means clustering, DB-Scan, I-Forest, and autoencoders, and then combined them with various custom options. After the integration of the reduction technology and K-Means clustering method, clinical data can be classified into CKD and non-CKD with 99% overall accuracy.

"AdaBoost (ADB), K-Nearest Neighbor (KNN), Extra Tree Classifier (EXT), Extreme Gradient Boosting (XGB), Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF) and Gaussian Naive Bayes (GNB)" are used by researchers [8] to achieve the best results. Researchers have investigated the machine learning algorithms and compared the results before the data. Other methods apply KNN and Extra Tree Classifier to achieve 99% accuracy.

Deep Learning models for the early diagnosis of chronic renal disease are thoroughly examined by the researchers [9]. In addition to standard CNN, ANN, and LSTM models, they also used optimized ANN (OANN), CNN (OCNN), and LSTM (OLSTM) models. The implemented models are compared in order to choose the be most competent model for CKD classification. As a result, they discovered that the optimized models outperformed the traditional models in general. They obtained 92.71% validation accuracy for the classic CNN model, whereas OCNN, OANN, and OLSTM scored 98.75%, 96.25%, and 98.5%, respectively.

An enhanced comparative model for the prediction of CKD is presented by researchers [10]. On a data set comprised of 400 instances, 24 characteristics, and binary classification labels that they got from the UCI repository, they apply various ML classification techniques. On the dataset, the model is evaluated using the 7-fold and 10-fold cross-validation procedures. In order to conduct their experiment, they employed the most well-known ML classification algorithms, including Support Vector Machine (SVM), Naive Bayes (NB), Discriminant Analysis (DA), Random Forest (RF), and K-Nearest Neighbor (KNN). According to the statistical results of all algorithms, RF outperformed KNN, SVM, NB, and DA with accuracy values of 99.75%, 92%, 97%, 98% and 98.25 respectively.

For better CKD prediction, researchers [11] provide a ML strategy using filter-based feature selection. This study developed a method for detecting CKD that combines the information-gain-based feature selection strategy with a cost-sensitive adaptive boosting (AdaBoost) classifier. The suggested methodology was contrasted with recently introduced CKD prediction algorithms and tried and true classifiers. The suggested cost sensitive AdaBoost classifier, with an accuracy of 99.8%, outperformed the other classifiers when trained with the smaller feature set.

Researchers [12] have identified five ML classification methods for predicting CKD. These classification methods include the decision tree, support vector machine, k-nearest neighbors, logistic regression, and random forest. There experiment results indicate that both the decision tree classifier and the logistic regression algorithm achieved the highest level of performance among the classifiers examined, boasting an impressive accuracy rate of 98.75%. Following, the Random Forest method secured the second position with an accuracy score of 97.5%.

Researchers [13] identified early-stage factors for CKD. It can be accomplished through the use of deep learning techniques, which can be created. Through the examination of a number of different features, they make an effort to forecast CKD. In the beginning, the data was preprocessed using a number of methods to account for any missing values and outliers that may have been present. After that, a classification of those with and without CKD was carried out using both a random forest and a deep neural network. When

the results of both methods were compared, it became clear that the DNN model was superior in terms of accuracy for binary classification. Its accuracy was measured at 98.8 percent.

A model for classification and forecasting that incorporates intelligence has been proposed by the researchers [14]. They used a modified version of the Deep Belief Network (DBN) classification algorithm, the SoftMax activation function, and the categorical cross-entropy loss function in order to make a prediction regarding kidney-related illnesses. The assessment of the model that was proposed reveals that it possesses a degree of accuracy of 98.5% when it comes to predicting CKD.

The researchers [15] study objective is to develop a method for predicting chronic renal illness on the basis of symptom onset. The consequences may be predicted with the use of data mining and ML techniques, as described in the article. For the purpose of disease prediction, KNN and SVM Ensemble are the algorithms that are utilized. When it came to data mining, the prediction of the outcome was handled by an SVM equipped with an RBF kernel. In ML, the KNN algorithm combined with a hyperparameter was used to make a prediction of the result. In order to enhance the accuracy of ML, they chose to take an assembly-based approach. The method that is provided has an accuracy of 92% in ML and 87% in Data Mining.

The researchers [16] devised computer-aided systems aimed at predicting diseases associated with CKD and provided recommendations for their practical implementation. The CKD dataset was accessed from the UCL ML repository. These researchers crafted predictive models employing a repertoire of eight diverse ML methodologies, namely, Random Forest, Support Vector Machine (SVM), Naive Bayes, Logistic Regression, K-Nearest Neighbors (KNN), XG Boost, Decision Tree, and AdaBoost. These models' precision was rigorously examined in order to determine which one demonstrated the highest accuracy. The AdaBoost, XG-Boost, Naive Bayes, Decision Tree, and SVM models were found to perform most positively in terms of accuracy after a thorough review. Subsequently, the Random Forest and Logistic Regression models yielded results of lower accuracy, while the KNN classifier model recorded the lowest accuracy rate at 73%.

The researchers [17] investigate the viability of chronic renal disease prediction based on the extensive and crucial CKD dataset characteristics. Three different approaches to the job of picking features have been used: correlation-based feature selection, the Wrapper technique, and LASSO regression. There were seven different classifier algorithms used in the context of this observation. These included a C5.0 artificial neural network, CHAID logistic regression, LSVM, K-Nearest Neighbors (KNN), and RT random tree. All classifier methods performed remarkably well on the set of features selected by LASSO regression, both with and without SMOTE. Out of the five classifiers, SMOTE generated the best results when configured with all of its features. Seven different classifiers in total were used for this experiment.

However, neither Logistic nor KNN were used in SMOTE since they did not yield results that were sufficient. The results suggest that the SMOTE approach is the most efficient tactic for guaranteeing that a dataset is spread equally. It has been shown that SMOTE delivered superior results with selected features using LASSO regression in comparison to the results obtained by the LASSO regression model without the use of SMOTE. In all the studies, LSVM had the greatest accuracy of 98.46% compared to the other classifiers algorithms that were used.

Logistic Regression, Decision Tree, and SVM are the three ML classifiers that we used in order to test the efficacy of the prediction model that researchers [18] developed for chronic renal disease. The constructed model for predicting CKD was trained with the help of categorical and non-categorical variables that were found in the CKD dataset. When compared against the performance of logistic regression and the support vector machine, the decision tree classifier comes out on top. During the second step, we utilized the bagging ensemble technique in order to improve the performance of the fundamental classifiers. They found that the decision tree had the highest accuracy, coming in at 95.92%.

The accuracy of projections for CKD was determined by the researchers [19] by employing two different data mining methods. The result of running the Random Forest Algorithm analysis produced a confidence level of 88.7% and an area under the curve of 99.2% for the receiver's operational characteristics. It appears that bit shear makes a substantial contribution to Random Forest's performance, although the Back Propagation method achieves an accuracy of 98.40%. When it comes to forecasting CKD, the results of the comparison shown above make it abundantly clear that the Back Propagation neural network outperforms the Random Forest Classifier technique.

This paper [20] explores the application of ML and predictive modelling to improve the early diagnosis of CKD, a serious condition that can lead to end-stage renal failure. The authors examine ML methods to improve CKD variable identification via predictive modelling. The study selects the most important features for reliable diagnosis among 25 factors. From 12 classifiers tested, the XgBoost model performed best with 0.983 accuracy, 0.98 precision, 0.98 recall, and 0.98 F1-score. Recent advances in ML and predictive modelling increase CKD diagnosis accuracy and show potential for other medical diagnostic applications, according to the report.

This researcher [21] presents a deep learning method for early CKD identification and classification using normal medical consultation data. Unlike prior studies that employed typical ML or UCI dataset, this research uses a fuzzy deep neural network to improve CKD diagnosis. An optimization method inspired by learning processes examines ordinary clinical data. The suggested model outperforms existing approaches in classification accuracy, precision, F-measure, and sensitivity with 99.23% accuracy. This strong performance shows the model could detect CKD stages without medical intervention in the future. The research shows that advanced artificial intelligence, particularly deep learning, can improve early CKD detection and that AI-driven approaches to medical diagnostics offer a distinct alternative to previous methods that used different datasets and technologies.

This paper [22] uses CT scan pictures to use deep learning to detect CKD early and automatically. It diagnoses kidney stones, cysts, and tumors in CKD. This study uses 12,446 CT scan images categorized into cysts, normal, stones, and tumors, unlike prior studies that used smaller datasets or different imaging methodologies. The method uses a deep learning model to extract deep characteristics from CT scans and create hypergraphs. For representational learning, an HCN is used. The model outperforms state-of-the-art validation approaches with 99.71% accuracy. The three-layer model reduces training and testing time compared to more sophisticated models. With its ability to diagnose kidney disorders, this digital-twin model may help nephrologists detect and prognose early diseases. The article advocates using multimodal imaging to improve kidney disease diagnosis.

This paper [23] presents a robust application of machine learning to the prediction of chronic kidney disease (CKD). Using classifiers like Random Forest, Support Vector Machine and Artificial Neural Networks, on different data patterns, the study is successful in achieving desired accuracy levels. The highest scores can be observed with the Random Forest classifier where the accuracy level reached 92%, which is accompanied by above average results displayed by other measures such as precision, recall, F-score and ROC-AUC. In addition to this, the study will complement these results by combining the output from the various classifiers to get an improved model. The results highlight the applicability of these sophisticated methods in the diagnosis and treatment of CKD at its early stages.

The study [24] focuses on using Machine Learning (ML) techniques to predict Chronic Renal Disease. By analyzing a dataset of 400 samples from the UCI Machine Learning Repository, the study evaluates three ML classifiers: Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM). The Decision Tree classifier was the most accurate with an accuracy level of 95 %, and it was followed by a bagging ensemble method that slightly increase the accuracy to 97%. Their research emphasizes the possibilities of enhancing disease prognosis using ML in early diagnosis and treatment of CKD.

Using a dataset of 201 records and 29 parameters, researchers [25] employed ensemble machine learning techniques to forecast chronic kidney disease (CKD). Ensemble models efficiently capture the range of patterns and interactions included in the data, so offering a valuable tool for early identification and individualized treatment of chronic kidney disease (CKD). This underscores the potential efficacy of ensemble learning in enhancing the quality of medical decision-making in the context of chronic renal illness. They assess the predictive accuracy of several techniques, including RF, DT, SVM, and AdaBoost. SVM and Random Forest achieved the highest accuracy of 98%. Moreover, we identify key characteristics in the dataset that are necessary for predicting chronic kidney disease (CKD).

3.1. Disscusion

We have conducted a thorough analysis of literature studies that use ML classifiers to diagnose CKD. As a result, this section will explore existing limitations and potential future prospects for ML in CKD prediction, as well as the ML approach. Almost all facets of kidney imaging, and blood reports are now covered by ML applications. Different ML approaches have been incorporated for the classification of CKD.

The results of this study demonstrate how ML is widely used in the classification and prediction of kidney chronic disease.

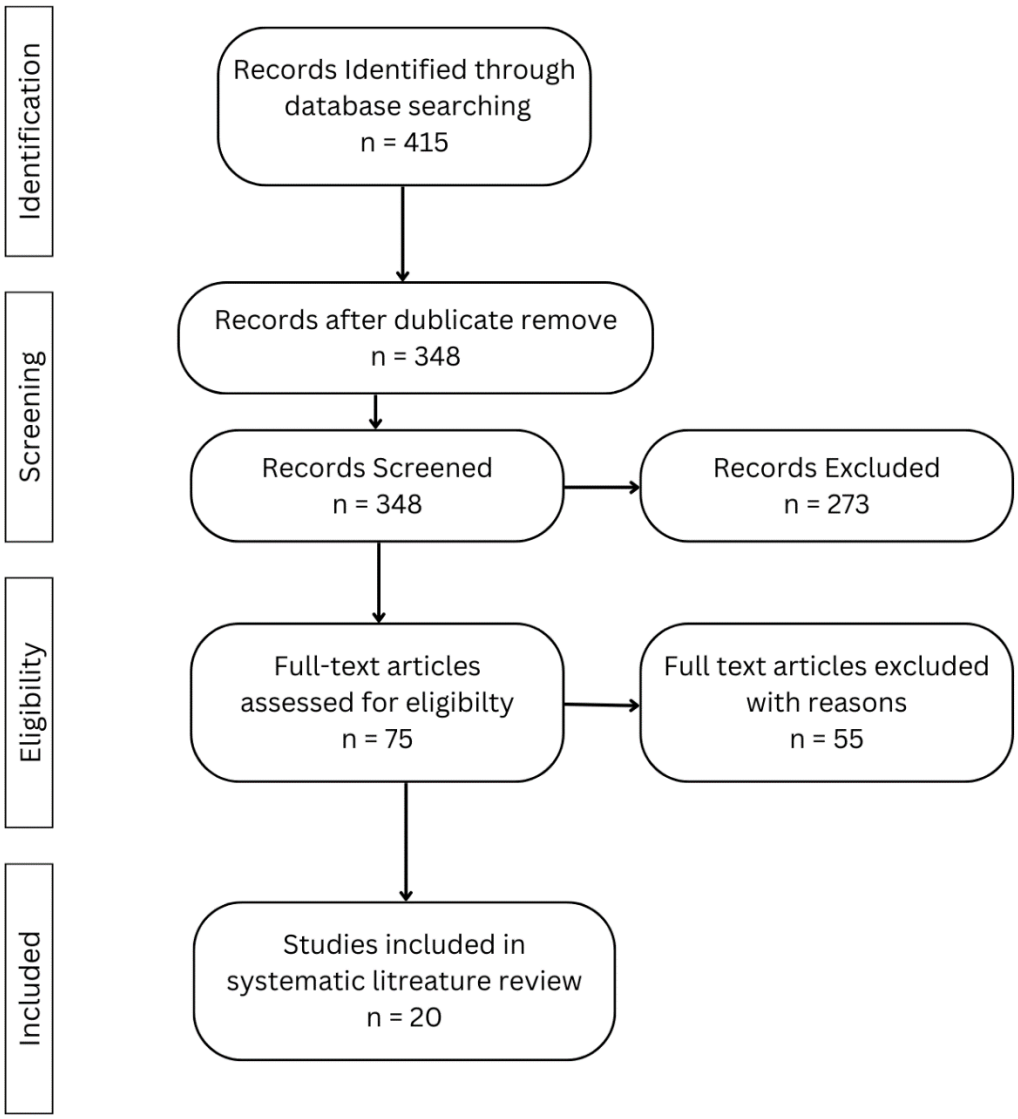

Figure 1. Paper selection Process

Table 2. Comparative Analysis

Reference	Model	Accuracy
[3]	Combined model of SVM & Correlational Neural Network	98.67%
[4]	XgBoost	98.3%
[5]	Neural Network	100%
[6]	Random Forest	95.6
[8]	K-Nearest Neighbor	99%
[7]	K-Mean	99%
[8]	Extra Tree Classifier	99%
[9]	OCNN	98.75%
[10]	Random Forest	99.75%
[11]	AdaBoost Classifier	99.8%
[12]	Logistic Regression	98.75%
[13]	DNN	98.8%
[14]	DBN	98.5%
[15]	KNN	92%

[16]	Random Forest	99%
[16]	Logistic Regression	99%
[17]	Linear SVM	98.46%
[18]	Decision Tree	95.92%
[19]	Back Propagation Neural Network	98.40%
[20]	XGBoost	98.3
[21]	Fuzzy Deep Neural Network	99.23%
[22]	HCNN	99.71%

The above table presents a quick comparison of algorithms that have been used for the classification of kidney chronic disease. The neural network has shown higher accuracy than the other approaches.

In this review, we performed a systematic analysis of various studies, comparing the performances of numerous ML and DL approaches, specifically in terms of their accuracy parameter. Our contribution consists in integrating these results to identify patterns and trends of an effective model. For this, we discuss the state-of-the-art deep learning approaches like Neural Networks, Fuzzy Deep Neural Networks, and Hybrid Convolutional Neural Networks by pointing out their enhanced performance during experimentation. Besides, we also explain the ensemble methods like Random Forest and XGBoost and their reliability for discovering the best model. Considering this, the analysis reveals the benefits of these models and provides ideas about the further development and enhancement of the models in the future studies.

4. Methodology

Machine learning and deep learning methodologies are of great significance as it is widely used when it comes to more effective results. These algorithms can be fine-tuned and can be updated, to mean that the results that are generated are even more accurate. Significant studies on the detection of kidney chronic diseases have incorporated both the machine learning and deep learning techniques. Based on the literature that has been reviewed on CKD classification, the following are some of the most preferred techniques in this field.

4.1. Support Vector Machine

A popular supervised learning technique that may be applied to classification and regression analysis is support vector machine. This discriminative classifier is characterized by a separating hyperplane. This approach can be used to solve both linear and nonlinear issues in the real world. The main goal of the SVM model is to produce a best line or decision border that categorizes n-dimensional space and enables future classification of new data points with ease. The best option is referred to as the hyperplane. The extreme points or vectors selected by the SVM to construct this hyperplane are known as support vectors. There are several potential hyperplanes that can be used to separate the data point classes. However, our primary goal is to find the hyperplane with the maximum margin. Expanding the margin enables more confident classification of future data points and allows for the addition of new data points.

4.2. Linear Support Vector Machine

Linear Support Vector Machines (LSVM) is a modern approach associated with an effort to solve such complex problems as multiclass classification challenges presented to systems dealing with large amounts of data. This technique rests with simple iteration and produces a SVM model in linear CPU time, relative to a dataset in question. LSVM is again seen to be nicely in tune with multi dimensionality and operates well with both sparse and dense data. Significantly, it is useful in response to the numerous ML problems caused by the scale of data, and adeptly manages computational power. The structure of it relies on Support Vector Machine which is a supervised classification algorithm. To deal with the complexities of categorization, it is advisable to use the so-called kernel-based approach. By these kernel adaptations, the LSVM model has the best ability in defining the right boundary between potential output classes. In scenarios defined by nonlinear kernels, which include the Radial Basis Function (RBF) kernel, the conventional SVM method is well used. On the other hand, if one is faced with linear classification problems, then LSVM comes up as a wise decision. It is pertinent to underscore that the LSVM classifier stands as a proficient solution for resolving a spectrum of linear classification challenges.

4.3. K- Nearest Neighbors

KNN is a key classification algorithm in ML. The KNN algorithm belongs to the area of supervised learning and is typically used for data mining, pattern recognition, and intrusion detection. KNN is non-parametric, hence it does not assume anything about the distribution of the data. This is the type of distance-based method that may be adjusted depending on category attributes when all attribute values are continuous. Prior data (also known as training data) is provided, which often divides locations into groups based on certain characteristics. This technique represents the concept of closeness mathematically by calculating the distance between graph nodes. This algorithm is simple to implement, and there is no need to construct a model or make assumptions.

4.4. Artificial Neural Network

AI includes artificial neural networks which are computational models inspired by the human brain's architecture It has the same structure as the human brain. In layers of the network, ANN neurons are related to one another, much like human neurons. There, neurons are called nodes. ANN can tackle problems that were previously insurmountable for humans or statistical methods. The Artificial Neural Network (ANN) comprises three distinct layers: the input layer, the hidden layer, and the output layer. In the initial stage of processing, information is conveyed to the input layer. Subsequently, this data, in conjunction with weight parameters, is transmitted to the hidden layer, where necessary calculations are performed. The primary objective of the hidden layer is to discern concealed structures and patterns within the input data. It is noteworthy that the quantity of hidden layers may be adjusted as necessitated by the specific task at hand. The final phase of the ANN's operation resides within the output layer, where the ultimate computation of the output takes place. Recalculating the weight values, which contrast the expected and actual values, is a necessary step in this process. The network works nonstop to recognize and categorize the classes and learning patterns. This process, known as backpropagation, serves as the fundamental underpinning of the Artificial Neural Network, allowing it to refine its performance and enhance its accuracy over time.

4.5. C5.0

C5.0, as a decision tree variant, constructs a tree structure based on the input data, resulting in numerous branches within the tree. It does this in a manner that easily mimics the linkage of various features with their probable outcomes. During the construction of the tree, at each node of the tree, one attribute of the dataset is selected. The compatibility of C5. It has been found to be zero with both nominal and numeric characteristics, making it versatile in its application. There is one thing that is crucial to remember, and that is C5.0 is an upgrade of the C4. 5 classification approach that takes into account information entropy of text. Entropy can be looked at as a measure of the degree of impurity of the attributes concerning various classes. This in turn gets the in-formation entropy by calculating the individual entropy for both parent and child nodes. This iterative method is followed until there are no more splits possible in the tree like structure.

4.6. Logistic Regression

It is a well-established supervised learning methodology characterized by its statistical modelling framework. Its principal use is to measure the risk of a particular event occurring. In logistic regression, the target attribute is dichotomously classified into two distinct categories: success, denoted as 1, and failure, denoted as 0. The logistic regression equation is formulated as follows:

$$P = 1/(1 + e^{-(b_0 + b_1 x + b_2 x^2)}) \quad (1)$$

Here, 'x' represents an attribute, 'p' signifies the predicted outcome, and 'b_0,' 'b_1,' and 'b_2' denote the respective bias parameters. Logistic regression finds applications across a spectrum of domains, including social sciences and medical ML. Notably, it has been instrumental in tasks such as cancer and diabetes detection, as well as spam filtering. Logistic regression, in essence, represents a more nuanced extension of the linear regression paradigm, as it specifically focuses on modelling the probability of the target variable's outcome.

4.7. Decision Tree

The decision tree methodology also applies as a means of predicting given discrete and continuous attributes, which involve classification and regression. The above framework means that the system can predict discrete attributes within a given format in the context of a given dataset by looking into various columns and relating them. It does this by predicting the states of columns, referred to as "states," that users have claimed can be predicted. The decision tree approach also indicates the input columns that

influence the projected column the most. One strength therefore accompanies the Decision Tree approach, and this is the fact that it is extremely simple and natural since it mimics the thought process that one normally goes through in a decision-making process in the real world. Due to the fact that it resembles human decision-making processes, it is a useful tool for helping in judgments and solving pertinent problems. This means that one should try to consider all possible scenarios when dealing with intricate problems. As such, this approach reduces the need for rigorous data cleaning processes that may otherwise be characteristic of other forms of modeling for prediction.

4.8. Random Forest

The Random Forest (RF) algorithm, which has received considerable popularity, exists within a supervised learning setting. It is useful for a broad range of problems involving ML that require classification as well as regression. This approach relates to the concepts of ensemble learning, a technique of training multiple classifiers to solve a given problem and enhance the performance of the model. In this context, a classifier is constructed using the framework of a random forest, where several decision trees are used on several portions of the entire data set. The outcomes generated from each of these individual decision trees are then combined to strengthen the capability of the data set for factored prediction. Different from the segregated decision tree model, the random forest model arrives at its final decision on the basis of a voting system and is therefore more accurate.

4.9. XGBoost

In the XG Boost algorithm, decision trees are generated in sequence. Weights play an important role in XG Boost. Before each independent variable is fed into the forecasting decision tree, a weight is assigned to it. Variables that were incorrectly predicted by the first tree are given a larger weight before being sent to the second tree. Together, these various classifiers and predictors produce a robust and trustworthy model. It may be used to solve problems including classification, ranking, regression and custom prediction.

4.10. AdaBoost

Boosting techniques combine many weak classifiers to produce strong classifiers in order to enhance classification accuracy. Friedman introduced a new and useful method, adaptive boosting, in 1997 but not until 2000. Logit Boost was shown to have dealt with this situation by utilizing increased generalizations. Boosting algorithms handle several medical issues, including cancer detection, breast cancer diagnosis and protein structure classification.

5. Conclusions

This systematic review includes 20 articles on the automated identification and categorization of chronic kidney disease using machine learning and deep learning algorithms. The article offered a summary of the many approaches for detecting CKD. Before classifying this disease, several procedures were applied to different datasets. The researchers faced numerous obstacles and barriers during their attempt to identify and classify the disease. Researchers who intend to study the same topic may investigate a strategy for further categorization in order to improve the classification and detection outcomes. We conclude that several of the most recent ML algorithms are capable of classifying CKD. When we have a large collection of medical photos and data, it is preferable to use a neural network for classification.

References

1. J. Singh, S. Agarwal, P. Kumar, Kashish, D. Rana, and R. Bajaj, "Prominent Features based Chronic Kidney Disease Prediction Model using Machine Learning," in 3rd International Conference on Electronics and Sustainable Communication Systems, ICESC 2022 - Proceedings, 2022, pp. 1193–1198. doi: 10.1109/ICESC54411.2022.9885524.
2. G. Nandhini and J. Aravinth, "Chronic kidney disease prediction using machine learning techniques," in 2021 6th International Conference on Recent Trends on Electronics, Information, Communication and Technology, RTEICT 2021, Aug. 2021, pp. 227–232. doi: 10.1109/RTEICT52294.2021.9573971.
3. N. Bhaskar and M. Suchetha, "A Computationally Efficient Correlational Neural Network for Automated Prediction of Chronic Kidney Disease," IRBM, vol. 42, no. 4, pp. 268–276, Aug. 2021, doi: 10.1016/j.irbm.2020.07.002.
4. Md. A. Islam, Md. Z. H. Majumder, and Md. A. Hussein, "Chronic kidney disease prediction based on machine learning algorithms," J Pathol Inform, p. 100189, Jan. 2023, doi: 10.1016/j.jpi.2023.100189.
5. S. Nag and N. Jammula, "Comparison of Neural Network and Machine Learning Approaches in Prediction of Chronic Kidney Disease." [Online]. Available: www.JSR.org
6. N. H. Taz, A. Islam, I. Mahmud, E. Haque, and M. R. Rahman, "A Comprehensive Exploration on Impact of Preprocessing for Prediction of Chronic Kidney Disease Using Multiple Machine Learning Approaches," in 2021 International Conference on Science and Contemporary Technologies, ICSCT 2021, 2021. doi: 10.1109/ICSCT53883.2021.9642638.
7. L. Antony et al., "A Comprehensive Unsupervised Framework for Chronic Kidney Disease Prediction," IEEE Access, vol. 9, pp. 126481–126501, 2021, doi: 10.1109/ACCESS.2021.3109168.
8. D. Baidya, U. Umaima, M. N. Islam, F. M. J. M. Shamrat, A. Pramanik, and M. S. Rahman, "A Deep Prediction of Chronic Kidney Disease by Employing Machine Learning Method," in 2022 6th International Conference on Trends in Electronics and Informatics, ICOEI 2022 - Proceedings, 2022, pp. 1305–1310. doi: 10.1109/ICOEI53556.2022.9776876.
9. C. Mondol et al., "Early Prediction of Chronic Kidney Disease: A Comprehensive Performance Analysis of Deep Learning Models," Algorithms, vol. 15, no. 9, Sep. 2022, doi: 10.3390/a15090308.
10. Inayatullah and H. Qayyurn, "An improved comparative model for chronic kidney disease (CKD) prediction," in 2020 14th International Conference on Open Source Systems and Technologies, ICOSST 2020 - Proceedings, Dec. 2020. doi: 10.1109/ICOSST51357.2020.9333097.
11. 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, Aug. 2022, doi: 10.3390/bioengineering9080350.
12. Ms. K. Samatha, Ms. M. R. Reddy, Mr. P. F. Khan, Mr. R. A. Chowdary, and Dr. P. V. R. D. P. Rao, "Chronic Kidney Disease Prediction using Machine Learning Algorithms," International Journal of Preventive Medicine and Health, vol. 1, no. 3, pp. 1–4, Jul. 2021, doi: 10.54105/ijpmh.C1010.071321.
13. K. Jhumka, M. M. Auzine, M. S. Casseem, M. H.-M. Khan, and Z. Mungloo-Dilmohamud, "Chronic Kidney Disease Prediction using Deep Neural Network," Nov. 2022, pp. 1–5. doi: 10.1109/nextcomp55567.2022.9932200.
14. S. M. M. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Early Prediction of Chronic Kidney Disease Using Deep Belief Network," IEEE Access, vol. 9, pp. 135542–135549, 2021, doi: 10.1109/ACCESS.2021.3114306.
15. A. Azmi, A. Hingu, R. Dholaria, and A. Alphonso, "Chronic Kidney Disease Prediction using Data Mining and Machine Learning," International Research Journal of Engineering and Technology, 2020, [Online]. Available: www.irjet.net
16. I. Ahmed, T. E. Chowdhury, B. B. Routh, N. Tasmiya, S. Sakib, and A. A. Chowdhury, "Performance Analysis of Machine Learning Algorithms in Chronic Kidney Disease Prediction," in 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2022, 2022, pp. 417–423. doi: 10.1109/IEMCON56893.2022.9946591.
17. P. Chittora et al., "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," IEEE Access, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 17312–17334, 2021. doi: 10.1109/ACCESS.2021.3053763.
18. M. B. Nirmala, D. K. Priyamvada, P. R. Shetty, and S. Pallavi Singh, "Chronic kidney disease prediction using machine learning techniques," in 12th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2021, 2021, vol. 2021-August, pp. 185–190. doi: 10.1007/s44174-022-00027-y.

19. J. Snegha, V. Tharani, S. D. Preetha, R. Charanya, and S. Bhavani, "Chronic Kidney Disease Prediction Using Data Mining," in International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020, Feb. 2020. doi: 10.1109/ic-ETITE47903.2020.482.
20. Islam, Md Ariful, Md Ziaul Hasan Majumder, and Md Alomgeer Hussein. "Chronic kidney disease prediction based on machine learning algorithms." Journal of pathology informatics 14 (2023): 100189.
21. Kumar, Kailash, M. Pradeepa, Miroslav Mahdal, Shikha Verma, M. V. L. N. RajaRao, and Janjhyam Venkata Naga Ramesh. "A deep learning approach for kidney disease recognition and prediction through image processing." Applied Sciences 13, no. 6 (2023): 3621.
22. Sasikaladevi, N., and A. Revathi. "Digital twin of renal system with CT-radiography for the early diagnosis of chronic kidney diseases." Biomedical Signal Processing and Control 88 (2024): 105632.
23. Pal, S. Prediction for chronic kidney disease by categorical and non_categorical attributes using different machine learning algorithms. *Multimedia Tools and Applications*, *82*(26), 41253-41266 (2023).
24. Kaur, C., Kumar, M. S., Anjum, A., Binda, M. B., Mallu, M. R., & Al Ansari, M. S. (2023). Chronic kidney disease prediction using machine learning. *Journal of Advances in Information Technology*, *14*(2), 384-391.
25. Srikanth, V. Chronic kidney disease prediction using machine learning algorithms. (2023).
26. Imran, A., Li, J., Pei, Y., Akhtar, F., Yang, J.-J., & Wang, Q. (2019). Cataract detection and grading with retinal images using SOM-RBF neural network. In 2019 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 2626-2632). IEEE.
27. Imran, A., Li, J., Pei, Y., Mokbal, F. M. M., Yang, J.-J., & Wang, Q. (2020). Enhanced intelligence using collective data augmentation for CNN based cataract detection. In Frontier Computing: Theory, Technologies and Applications (FC 2019) (Vol. 8, pp. 148-160). Springer Singapore.
28. Khan, M. F., Iftikhar, A., Anwar, H., & Ramay, S. A. (2024). Brain Tumor Segmentation and Classification using Optimized Deep Learning. Journal of Computing & Biomedical Informatics, 7(01), 632-640.
29. Abbas, T., Khan, A. H., Kanwal, K., Daud, A., Irfan, M., Bukhari, A., & Alharbey, R. (2024). IoMT-Based Healthcare Systems: A Review. Computer Systems Science & Engineering, 48(4).
30. Iftikhar, A., Elmagzoub, M. A., Shah, A. M., Al Salem, H. A., ul Hassan, M., Alqahtani, J., & Shaikh, A. (2023). Efficient Energy and Delay Reduction Model for Wireless Sensor Networks. Comput. Syst. Sci. Eng., 46(1), 1153-1168.