International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Prediction of Chronic Kidney Disease Using Machine Learning and Deep Learning Mechanisms: A Survey

Suhaila.K K[1*], Dr.Elamparithi.M[2] and Dr.Anuratha.V[3]
Department of Computer Science, Kamalam College of Arts and Science, Anthiyur,
Bharathiar University, Coimbatore, Tamil Nadu, India.
E-Mail: [1*] kk.suhaila@gmail.com, [2]profelamparithi@gmail.com, [3]profanuratha@gmail.com

Abstract: The ability of the kidneys is gradually reduced by chronic kidney disease (CKD). Early identification and characterization are essential to treating and managing chronic renal disease. Because of its expanding patient population, increased likelihood of progressing to renal failure, and dismal outlook for morbidity and death, CKD is an enormous cost of medical care. While many techniques have been employed to identify CKD, machine learning (ML) and deep learning (DL) algorithms provide more informative outcomes. Therefore, this study looks at several state-of-the-art ML and DL models for CKD detection. In the total corpus of literature, we also notice a few noteworthy problems that merit additional investigation. Lastly, readers and ML and DL researchers will find our study informative on essential aspects of CKD prediction.

Keywords: *CKD detection, Machine Learning, Deep Learning, UCI database, Feature selection*.

1. INTRODUCTION

The kidney remains a vital organ in the human body because it removes waste products from the blood plasma and releases them in urine. In addition, the kidneys generate and discharge hormones that control blood pressure, preserve the body's electrolyte and fluid balance, regulate pH to govern the production of red blood cells and create an active kind of vitamin D that supports healthy and strong bones [1]. A severe chronic illness that affects adults 60 years of age and older, kidney disease is comparable to adult diabetes, hypertension, and hypertension. CKD is a disorder where the kidneys cannot filter blood as well as they should [2]. A 2021 report estimates that over 37 million people in the US alone have CKD. Kidney disease causes about 2.4 million deaths annually (Nikhila, 2021) [3]. It is currently the sixth-leading cause of death globally [4]. The following are the five stages of CKD: normal, mild, moderate, severe, and end-stage [5].

Thus, early detection and screening for people with CKD may lead to interventions that alter the illness's natural course and lower the chance that it will progress to kidney failure in its final stages. Early detection of CKD is the best way to treat it. Computerized assistance examinations are required to help radiologists and doctors make diagnoses due to the rising number of chronic renal patients, the shortage of specialist clinicians, and the high costs of diagnosis and treatment, especially in developing nations. ML has become a potential method for CKD prevention and early kidney disease identification. ML has become a potential method for CKD prevention and early kidney disease identification. To identify CKD early on, however, ML techniques like Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest (RF), Logistic Regression (LR), Naïve Bayes (NB), Decision Tree (DT), etc. are employed. The challenges that come up while evaluating CKD data in the presence of missing values are also covered in this work [5].

However, when the dataset is large, the ML models perform poorly, and they also require human interaction to carry out the feature learning process for renal disease identification. They are not impervious to errors and omissions, especially in constantly changing and complicated situations. Unexpected events, biased training sets, and noisy or faulty data can all lead to inaccurate forecasts. Some of the data pre-processing that is usually done in conjunction with ML is eliminated by DL. These algorithms automate feature extraction, reducing the need for human specialists by ingesting and processing unstructured data, such as text and images. Recent efforts have attempted to use DL models, which work incredibly well, to diagnose CKD. Although they require more effort to set up, DL systems can produce results instantly. DL is still quite successful because of its better performance in CKD prediction than traditional ML algorithms. This motivates us to present a quick overview of the most recent deep learning and machine learning models for early-stage CKD detection.

2.BACKGROUND INFORMATION

A severe chronic illness that affects adults 60 years of age and older, kidney disease is comparable to adult diabetes, high blood pressure, and hypotension. Based on recent findings, there appears to be a rise in the seriousness of this problem.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

CKD is a condition in which the kidneys become so damaged that they can no longer filter blood as effectively as possible. The most effective method for controlling CKD is early detection; nevertheless, if you wait until it is too late, you risk developing renal failure and requiring therapy or a kidney transplant to live life as usual. The steps involved in detecting CKD are generally as follows: data collection, pre-processing, feature selection, classification, and performance analysis. Figure 1 displays the CKD prediction flow chart.

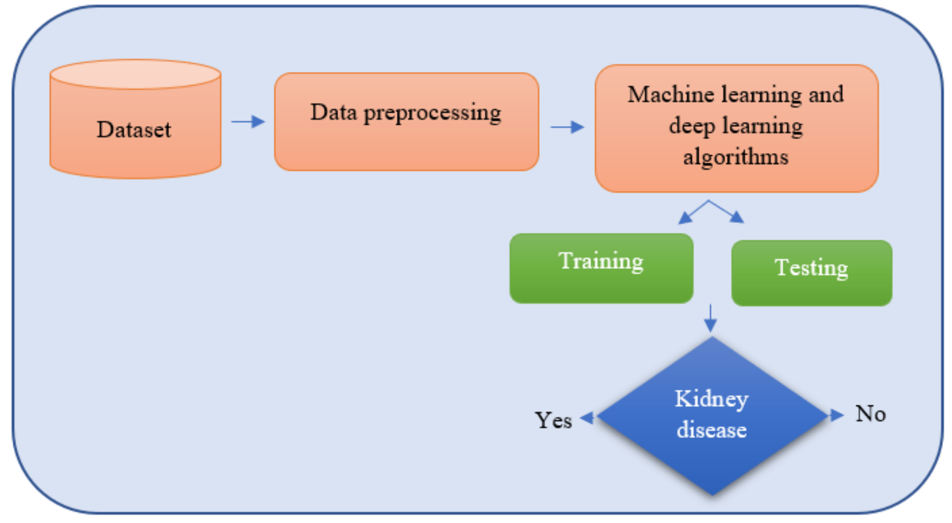

Figure 1: CKD detection system

2.1 Dataset Collection

Since we need to use the gathered data to train our classifier's, gathering data is the initial stage in the analysis process. Using the Kaggle and UCI datasets, numerous researchers can identify CKD. Its objective is to enable knowledge discovery, organization, and deconstruction from unstructured data sources. It also offers a location for information management and organization, facilitating more efficient knowledge discovery.

2.2 Data Preprocessing

The obtained dataset will then go to the stage of data preparation. The process of transforming acquired data into an understandable format is known as data preparation. It employs two pre-processing methods: transforming nominal values to integer values and filling in missing values. The median values are used to fill in the missing numbers.

2.3 Feature Selection

Following pre-processing, feature selection narrows down a feature set by keeping the most sensible features and eliminating drossy ones. Two methods for selecting highly effective features are recursive feature removal, minimum redundancy, and maximum relevance.

2.4 Classification

The final stage in the detection of CKD is classification. The purpose of the classification system is to categorize patient data as either chronic or non-chronic. In order to automatically detect early signs of CKD and prevent it, ML and DL are essential tools. ML is a valuable tool for diagnosing CKD. NB, LR, SVM, RF, KNN, and DT are well-liked machine learning styles. With impressive results, DL models can classify CKD and learn its properties. To alter the neural network's hidden layers, DL uses a multilayer technique. The most popular DL techniques, deep neural networks (DNN), long short-term memory (LSTM), and convolutional neural networks (CNN), yield better results.

3. REVIEW ON METHODS USED FOR CKD PREDICTION

Some experts are now more concerned with early CKD detection due to the recent rise in CKD cases. Much study has been done on the detection of CKD. This analysis offers a thorough look at the approaches used, the results gained, and the restrictions found in each study. By doing this, we hope to provide a thorough and objective knowledge of the advancements and difficulties in CKD research. The most current studies published on ML-based CKD detection are compiled in Table 1.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

Table 1: Summary of methods for CKD detection

Author &Ref No.	Methodology used	Dataset used	Accuracy (%)	Benefits	Drawbacks
Hira Khalid *et al.* [6]	Gaussian NB, Gradient Boosting (GB), DT and RF.	CKD dataset from UCI repository	GB=99, RF=98, DT=96, Hybrid model=100	These classifiers, such as DT, RF, GB, and Gaussian NB, are combined (hybrid model) to predict CKD. Additionally, this solves the overfitting issue and improves accuracy.	ML classifiers such as GB, RF, and DT cause overfitting compared to hybrid models because of their subpar performance on test data.
Debabrata Swain *et al.* [7]	SVMand RF.	UCI CKD Dataset	SVM=99.33, RF=98.67.	The SVM achieved a greater accuracy rate with the best first search engine-based feature selection technique.	The expense of medical diagnostics increases as a result.
Chamandeep Kaur *et al.* [5]	KNN, DTand RF.	UCI dataset	DT=96, RF=97.	By merging other factors and filling in missing values, RF classifiers can reduce the number of features in the prediction algorithm and reach the highest accuracy possible. This could result in fewer medical tests being necessary.	Compared to RF, DT leads to greedy and overfitting strategies because dividing the data requires many nodes.
Md. Ariful Islam*et al.* [2]	KNN, Artificial Neural Network (ANN), adaboost, Xgboost, DT, catboost, RF, NB, GB, Stochastic GB, Light GB Machine, extra tree, SVM, and hybrid.	UCI dataset	KNN=59, DT=97, Adaboost=98, Catboost=97, Xgboost=99, RF=97, NB=88, GB=97, SCB=97, LGBM=98, Extra tree=98, SVM=96, ANN=60, Hybrid=95.	The proposed ML model would obtain optimal performance metrics through PCA and strong classification performance with few features.	The deficiency of abundant data leads to subpar performance for ANN and KNN.
Dibaba Adeba Debaland Tilahun Melak Sitote [4]	RF, SVM and DT with RFECV and UFS feature selection methods.	Data from St. Paulo's Hospital	SVM=RF (with RFECV) =99.8, xgboost=82.56, DT=98.5.	They did not primarily concentrate on predicting the individual stages. This study has used both multiclass and binary classification to predict stages.	It just employs the feature selection method and the supervised ML algorithm to identify CKD. It encounters challenges when utilizing unsupervised ML models and DL.
Muhammad Shoaib Arif 1*et al.* [8]	KNN and NB.	UCI ml repository	KNN=100, NB=97.5.	This model's application demonstrates both its clinical applicability potential and dependability.	However, we must know the possible hazards and ethical issues while applying ML approaches to

					medical diagnosis. In the healthcare industry, complex machine learning models may be difficult to interpret, which raises questions about accountability and trust.

For CKD prediction, **Deema Mohammed Alsek ait *et al.*** [9] provided an ensemble of DL frameworks, including recurrent neural networks (RNN), LSTM, and gated recurrent units (GRU), with various feature selection models. The data was first pre-processed to encode text features. The best feature list was then chosen using various feature selection methods, including mutual information, chi-squared, recursive feature elimination, and tree-based RF. The RNN, LSTM, and GRU stacking ensemble DL models received the chosen features to classify them. The results demonstrated that the combination of ensemble models using the mutual information feature selection model utilizing the UCI database performed best, obtaining 99.69% accuracy, 99.71% precision, 99.69% recall, and a 99.6% F1 score. Due to the numerous mathematical processes involved, the hybrid DL models increase the computing complexity of the system. With the help of the UCI repository, **P.V. Shitole *et al.*** [10] created ML classifiers for the diagnosis of CKD, including SVM, KNN, LR, RF, NB classifier, and neural networks, along with feature selection methods including correlation, Fisher's score, and Chi-square test. The intricate machine learning model reduces overfitting and raises accuracy rates. The outcomes confirmed that the RF achieved the best, obtaining an accuracy of 95.92%. However, this approach necessitates a significant amount of processing power in addition to resources for programming.

The fusion DL model, which combines GNN and tabular data, was introduced by **Patike Kiran Rao *et al.*** [11] as a baseline technique for CKD prediction using the CKD UCI database. The fusion model proved resilient to noise and outliers and adjusted to fluctuations in the dataset by exhibiting consistent performance across various data splits. The results demonstrated that the fusion model, which combined the GNN and tabular data models, performed best, obtaining a 95% accuracy rate. On the other hand, the fusion DL model results in increased time and space complexity and reduced handling of graph edges according to their kinds and relations. Utilizing the Changhua Christian Hospital in Taichung, Taiwan, dataset, Kailash Kumar et al. [12] built the model utilizing the UCI dataset, which combines a fuzzy deep neural network (FDNN) and traditional radioimmunoassay method for the identification of early CKD. FDNN uses fuzzy inference with human-like characteristics to find the best solution. The results showed that, with an accuracy of 99.23%, the FDNN performed the best. Since a smaller dataset was used to test the procedure, a more extensive dataset must be used to generalize the results.

Using a dataset gathered from the General Hospital in the Gashua Local Government Area of Yobe State, **Iliyas Ibrahim Iliya *et al.*** [13] used a DNN to predict the presence or absence of CKD. This model can identify more intricate patterns in the input data, which enhances accuracy and performance. The results demonstrated that the model successfully predicted CKD with a 98% accuracy rate, 100% precision rate, 96% recall rate, 100% ROC score, 98% F1 score, 96% sensitivity, and a specificity of 100%. However, it results in intricacy and difficulties in system tuning and needs more transparency in making choices. Using the UCI dataset, **Khadiime Jhumka *et al.*** [14] created models such as RF and DNN to predict the presence or absence of CKD. Only tabular data is compatible with TheRF. Conversely, a neural network can process various data formats, including text, audio, pictures, and tabular data. In contrast to RF, DNN is more adaptable, has learning capabilities, and can handle complex data. To increase the accuracy rate, it employs more hidden layers. The outcomes demonstrated that the DNN could predict CKD with 98% accuracy. In the absence of appropriate regularization, it could lead to potential overfitting.

Sudip Raj Khadka *et al*. [15] used the shapely additive explanation of shapely additive values (SHAP) and local interpretable model-agnostic explanation (LIME) to develop the four ML algorithms—DT, LR, Multi-layer Perceptron Classifier, and SVM—in conjunction with the explainable AI (XAI) interface for CKD detection. LIME included specific details about how much each case fit into a particular category, but SHAP focused on the degree to which unique characteristics affected the result. The results demonstrated that, when utilizing the UCI ML library, the MLP classifier and SVM performed the best, attaining the most fantastic accuracy of 100%. Unfortunately, this model could not learn the non-linear feature of the input data because of its vast number of learnable model parameters.

4. DISCUSSION

Early diagnosis of CKD can be made by reviewing patient records, and persons with the condition can benefit from appropriate treatment. The UCI CKD database repository's data was gathered and analysed for the literature analysis, which examined CKD detection and preventive strategies that were applied using ML and DL models. The ML approach is frequently used to detect CKD at an early stage. The most popular ML approaches include LR, SVM, RF, NB, KNN, and DT. These are well-liked because they work well with a small set of features and prevent overfitting by obtaining the maximum accuracy rate with the least processing power. However, we must know the possible hazards and ethical issues while applying ML approaches to medical diagnosis. Complex ML algorithms may need to be more interpretable, which raises questions about accountability and trust in the healthcare industry. Regarding CKD detection, DL models have many benefits; the abovementioned literature employed fusion DL models, LSTM, FDNN, and RNN. As a DL model, it automatically extracts the characteristics and improves the performance of the CKD detection system without the need for manual feature extraction using a backpropagation technique. However, it takes a lot of memory and is difficult to apply to other issues. The model’s performance will improve if the data is balanced, noisy, and complete.

5. CONCLUSION

A poor diet, insufficient sleep, and other factors are contributing to the increasing prevalence of CKD in different age groups. Kidney failure can result from CKD (CKD), which first causes a gradual reduction in kidney function. Patients may need to undergo dialysis or a transplant as a result of this. Therefore, early identification of CKD is critical for effective treatment. First, the study's overall flow chart for the CKD detection system was made available. After that, it offered a review of the ML and DL approaches for CKD prediction and presented the information in a tabular format to aid in comprehension and deepen one's grasp of the research. After that, this study evaluates previous research to highlight the benefits and drawbacks of that body of work. It can be enhanced in the future by adding a more significant number of DL algorithms and new data from the healthcare sector.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of Interest: I, Suhaila.K.K, declares no conflicts of Interest to disclose.

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

[1] Samrat Kumar Dey, Khandaker Mohammad Mohi Uddin , Hafiz Md. Hasan Babu , Md. Mahbubur Rahman , Arpita Howlader , K.M. Aslam Uddin. “Chi2 -MI: A hybrid feature selection based machine learning approach in diagnosis of chronic kidney disease.” Intelligent Systems with Applications 16 (2022) 200144

[2] 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.

[3] Hossain, Muhammad Minoar, Reshma Ahmed Swarna, Rafid Mostafiz, Pabon Shaha, Lubna Yasmin Pinky, Mohammad Motiur Rahman, Wahidur Rahman, Md Selim Hossain, Md Elias Hossain, and Md Sadiq Iqbal. "Analysis of the performance of feature optimization techniques for the diagnosis of machine learning-based chronic kidney disease." *Machine Learning with Applications* 9 (2022): 100330.

[4] Debal, Dibaba Adeba, and Tilahun Melak Sitote. "Chronic kidney disease prediction using machine learning techniques." *Journal of Big Data* 9, no. 1 (2022): 1-19.

[5] Kaur, Chamandeep, M. Sunil Kumar, Afsana Anjum, M. B. Binda, Maheswara Reddy Mallu, and Mohammed Saleh Al Ansari. "Chronic Kidney Disease Prediction Using Machine Learning." *Journal of Advances in Information Technology* 14, no. 2 (2023).

[6] Khalid, Hira, Ajab Khan, Muhammad Zahid Khan, Gulzar Mehmood, and Muhammad Shuaib Qureshi. "Machine Learning Hybrid Model for the Prediction of Chronic Kidney Disease." *Computational Intelligence and Neuroscience* 2023 (2023).

[7] Swain, Debabrata, Utsav Mehta, Ayush Bhatt, Hardeep Patel, Kevin Patel, Devanshu Mehta, Biswaranjan Acharya, Vassilis C. Gerogiannis, Andreas Kanavos, and Stella Manika. "A Robust Chronic Kidney Disease Classifier Using Machine Learning." *Electronics* 12, no. 1 (2023): 212.

[8] Arif, Muhammad Shoaib, Aiman Mukheimer, and Daniyal Asif. "Enhancing the early detection of chronic kidney disease: a robust machine learning model." *Big Data and Cognitive Computing* 7, no. 3 (2023): 144.

[9] Alsekait, Deema Mohammed, Hager Saleh, Lubna Abdelkareim Gabralla, Khaled Alnowaiser, Shaker El-Snappish, Radhya Sahal, and Nora El-Rashidy. "Toward Comprehensive Chronic Kidney Disease Prediction Based on Ensemble Deep Learning Models." *Applied Sciences* 13, no. 6 (2023): 3937.

[10] Shitole, P. V., Shivtej Bhosale, Shravani Sonawane, Pranit Wable, and Shloka Togarllu. "Detection of Chronic Kidney Disease Using Machine Learning and Deep Learning."

[11] Rao, Patike Kiran, Subarna Chatterjee, K. Nagaraju, Surbhi B. Khan, Ahlam Almusharraf, and Abdullah I. Alharbi. "Fusion of Graph and Tabular Deep Learning Models for Predicting Chronic Kidney Disease." *Diagnostics* 13, no. 12 (2023): 1981.

[12] 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.

[13] Iliyas, Iliyas Ibrahim, Isah Rambo Saidu, Ali Baba Dauda, and Suleiman Tasiu. "Prediction of chronic kidney disease using deep neural network." *arXiv preprint arXiv: 2012.12089* (2020).

[14] Jhumka, Khadiime, Muhammad Muzzammil Auzine, Mohammad Shoaib Casseem, Maleika Heenaye-Mamode Khan, and Zahra Mungloo-Dilmohamud. "Chronic Kidney Disease Prediction using Deep Neural Network." In *2022 3rd International Conference on Next Generation Computing Applications (NextComp)*, pp. 1-5. IEEE, 2022.

[15] Khadka, Sudip Raj, Saphal Subedi, Bikash Kumar Aidy, Lok Nath Regmi, Ashish Parajuli, and Mohan Bhandari. "Predicting Chronic Kidney Disease using ML algorithms and XAI."