American Journal of Artificial Intelligence
2024, Vol. 8, No. 2, pp. 32-40
https://doi.org/10.11648/j.ajai.20240802.11

Research Article

Feature Selection AI Technique for Predicting Chronic Kidney Disease

Preethi Kolluru Ramanaiah*

Ernest & Young LLP, New York, USA

Abstract

The kidney is a vital organ that plays a crucial role in eliminating waste and excess water from the bloodstream. When renal function is impaired, the filtration process also ceases. This leads to an elevation of harmful molecules in the body, a condition referred to as chronic kidney disease (CKD). Early-stage chronic kidney disease often lacks noticeable symptoms, making it challenging to detect in its early stages. Diagnosing chronic kidney disease (CKD) typically involves advanced blood and urine tests, but unfortunately, by the time these tests are conducted, the disease may already be life-threatening. Our research focuses on the early prediction of chronic kidney disease (CKD) using machine learning (ML) and deep learning (DL) techniques. Utilized a dataset from the machine learning repository at the University of California, Irvine (UCI) to train various machine learning algorithms in conjunction with a Convolutional Neural Network (CNN) model. The algorithms encompassed in this set are Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB). Based on the experimental results, the CNN model achieves a prediction accuracy of precisely 97% after feature selection, the highest among all models tested. Hence, the objective of this project is to develop a deep learning-based prediction model to aid healthcare professionals in the timely identification of chronic kidney disease (CKD), potentially leading to life-saving interventions for patients.

Keywords

Chronic Kidney Disease, Convolutional Neural Network, Feature Selection, Gradient Boost, Extra Trees Classifier

1. Introduction

Chronic kidney disease is a common and significant health issue affecting millions of individuals worldwide. The kidneys play a crucial role in maintaining overall health by removing waste materials and excess fluids from the bloodstream through filtration. The presence of chronic kidney disease (CKD) hinders the filtering process, leading to the buildup of harmful substances in the body. Early detection of chronic kidney disease (CKD) is crucial because prompt medical attention can prevent the condition from progressing to more severe stages and reduce the likelihood of cardiovascular complications. But people with early-stage CKD often don't have any symptoms, and standard ways of diagnosing the disease depend on finding problems in tests such as urine and blood tests, which might not show up until the disease is far along. This delay in identification can make things worse for patients. The latest technology and computer

*Corresponding author: Preethiram4@gmail.com (Preethi Kolluru Ramanaiah), Preethi.kolluru.ramanaiah@ey.com (Preethi Kolluru Ramanaiah)

Received: 31 May 2024; **Accepted:** 21 June 2024; **Published:** 8 July 2024

methods, like machine learning (ML) and deep learning (DL), have provided promise in healthcare to forecast different diseases early on.

Joseph et al [11] provided a way to classify chronic kidney disease Using a dataset from the UCI machine learning repository, this work investigated the use of ML and DL techniques to forecast chronic kidney disease (CKD). Compared the performance of a Convolutional Neural Network (CNN) model to that of several ML algorithms, including Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB). The results of our study indicate that the CNN model exhibits the most precise predictions, thereby emphasizing the promising prospects of DL methods in the realm of early disease detection. The primary objective of this study is to construct a dependable prognostic model that may assist medical practitioners in the prompt detection and treatment of chronic kidney disease (CKD), thereby enhancing patient outcomes.

Main Contributions:

1. Acquire a trained dataset pertaining to chronic kidney disease (CKD) and conduct a descriptive examination of the dataset to identify the top-K features that are most significant in predicting CKD.
2. Train different classification algorithms using machine learning (ML) and deep learning (DL) methods. Check how well these models work by looking at things like accuracy, precision, recall, and F1-score. Determine the optimal categorization model based on these performance indicators.
3. Utilize the most effective model to make predictions about Chronic Kidney Disease (CKD) using newly obtained data points. This approach aims to facilitate the prompt identification of CKD by assessing the testing data and generating predictions using the trained model.

The report format of this research study should be structured as follows: The previous studies are summarized in Section 2. The methodology of the system, which makes use of ML and DL techniques, is introduced in Section 3. The results of deploying the system are examined in Section 4. Section 6 wraps up the anticipated work, while section 5 delves into the thorough evaluation of the system. The artificial neural network (ANN) and LR, the machine learning algorithms are used in healthcare industry to minimize the diagnosis cost [11, 12]. Both parametric and nonparametric models are used to solve classification problems. Classification accuracy in ANN can be enhanced by rearranging the weights of neurons [13]. ANN and LR with decision support play a very important role in improving the quality of service in health care or in medical centers. Thus, ANN model-based diagnoses have significantly helped health care by providing effective diagnosis [14].

2. Literature Survey

Ifraz et al. [4] evaluated three different machine learning approaches by conducting a comparison study using the CKD dataset. These methods were Logistic Regression (LR), Decision Tree (DT), and k-Nearest Neighbours (KNN). Both methods were evaluated. According to the findings of their investigation, the Logistic Regression algorithm reached the highest level of accuracy when it came to predicting kidney illness.

In their study, Hamsagayathri and Vigneshwaran [3] introduced a machine learning (ML) algorithm that can accurately forecast various diseases by analyzing symptoms. A dataset was compiled that included a wide range of disorders and their related symptoms. The researchers employed various machine learning classifiers for analysis, such as Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM). Out of all the models tested, the Random Forest model exhibited the highest level of accuracy in their studies. Consequently, they opted for the Random Forest model to forecast various diseases by utilizing symptoms as input parameters.

A predictive model for the detection of chronic kidney disease (CKD) was proposed by Ogunleye et al. [8]. The chronic kidney disease (CKD) affects around 10% of the world's population and 15% of the people in South Africa. For the identification of chronic kidney disease (CKD), they chose extreme gradient boosting (XGBoost) after testing a few different AI methods. In terms of accuracy, sensitivity, and specificity, the proposed technique achieved a perfect score of 1.000 for each of these aspects. These findings can be of assistance to nephrologists in the process of diagnosis, which allows for a reduction in both time and cost. Giraddi et al. [15] proposed a system based on BPNN for the detection of diabetic retinopathy (DR). DR is a complication in which vision of a person gets affected by long-term diabetes. The author explored various architectures to find best model for the detection of diabetic retinopathy.

Nazin Ahmed et al. [5] investigated the prediction of diabetes using machine learning techniques. For comparative analysis, they utilized several different datasets. The study employed various machine learning approaches, including Decision Tree (DT), Naive Bayes (NB), k-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR), and Support Vector Machine (SVM), to predict diabetic illness. To enhance prediction accuracy, they applied the chi-square feature selection method to extract the most relevant information.

Fu et al. [2], Nishanth et al [6] and Pal et al [9] used different machine learning algorithms to develop a system for early chronic kidney disease (CKD) prediction. Models were thoroughly tested and validated using input attributes from the CKD dataset. Based on the design, developed support vector classifier, random forest, and machine learning decision tree models for chronic kidney disease detection. Accuracy of the prediction model dictated the performance of the model. Compared to decision trees and support vector classifiers, random forests perform better for predicting chronic kidney

disease (CKD).

Critical Observation on Previous Works

Table 1 below shows critical observations made by researchers on early detection of chronic disease using advanced Machine learning algorithms.

Table 1. *Literature survey observations.*

Author/s	Observations
Ifraz et al. [4]	Assessed and contrasted the efficacy of three distinct machine learning methodologies in forecasting chronic kidney disease (CKD). Three machine learning algorithms were used to predict Chronic Kidney Disease (CKD), and Logistic Regression was shown to be the most precise.
Hamsagayathri and Vigneshwaran [3]	The text explicitly outlines the objective of the study, which is to predict different diseases by examining symptoms through the utilization of machine learning algorithms.
Ogunleye et al. [8]	The researchers investigated several different artificial intelligence (AI) strategies, and after doing an in-depth analysis, they decided to go with extreme gradient boosting (XGBoost) due to its remarkable results. On the other hand, the specific artificial intelligence methods that were compared are not revealed.
Nazin Ahmed et al. [5]	There is a lack of clarity about dataset characteristics, model comparison, and other aspects of machine learning, even though the article discusses a number of different machine learning methodologies and makes use of feature selection methods.
Fu et al. [2]	The research provides a breakdown of the process of developing machine learning models for the diagnosis of chronic kidney disease (CKD). These models include decision tree classifiers, random forest classifiers, and support vector classifiers. One can see that this exhibits a methodical approach to the building of models using ML.
Desai et al [1]	In this research paper, two classification models: back-propagation neural network (BPNN) and logistic regression (LR), are used for the study. The developed classification model will assist domain experts to take effective diagnostic decision. 10-fold cross validation method is used to measure the unbiased estimate of these classification models to diagnose cardiovascular disease
Nithya et al [7]	Uses a combination of clustering and classification approach detection and segmentation of kidney disease. This is achieved used Neural Network along with multi-kernel k-means clustering algorithm. This method seems to achieve 99.61% accuracy in detecting kidney disease as well as differentiating tumor vs non-tumor stones

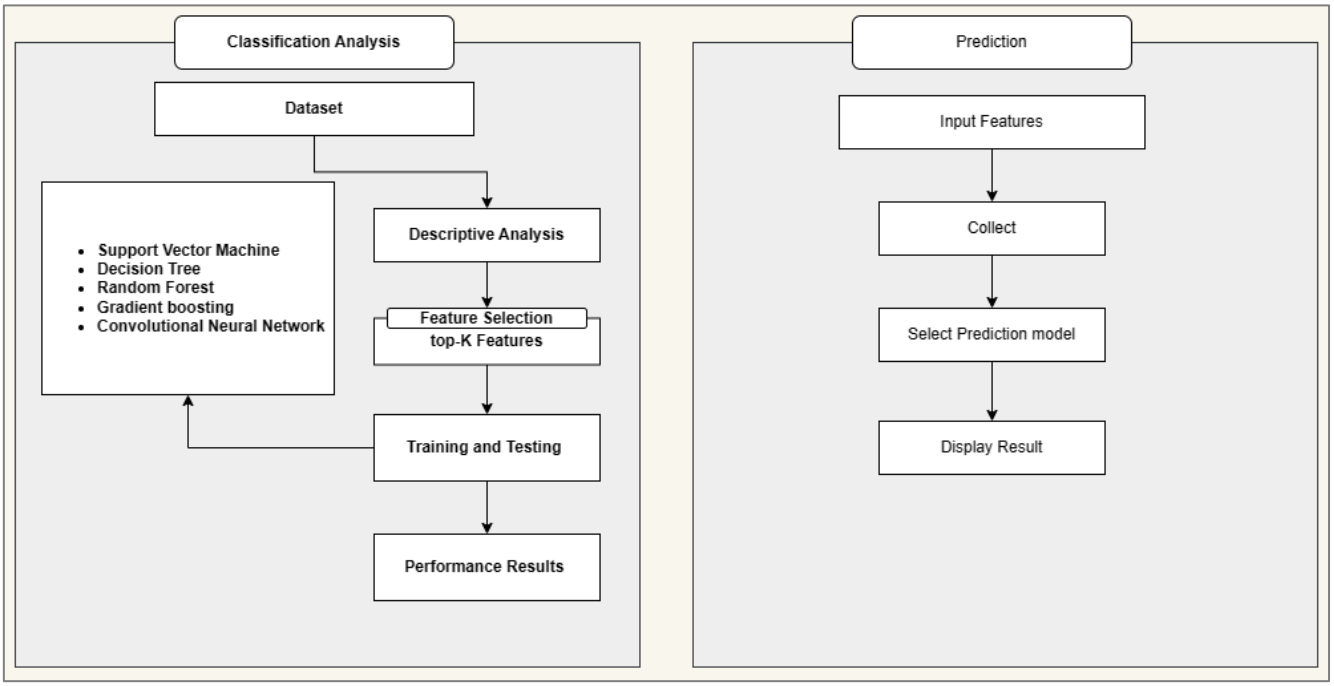

Figure 1. *Proposed Methodology of CKD.*

3. Methodology

The recommended approach consists of two primary components: the evaluation and forecasting modules. Figure 1 depicts the progression of the planned project. Prior studies have utilized machine learning algorithms to precisely detect different kidney illnesses by analyzing input features. This study examines the utilization of machine learning and deep learning models to identify chronic renal disease. A total of five classification algorithms, including a Convolutional Neural Network (CNN) Deep Learning model, were utilized. The objective is to determine the optimal categorization model for accurately predicting chronic renal disease using examination findings. The section that follows outlines the subsequent steps outlined in the proposed system methodology.

3.1. Dataset Collection

In this work, the chronic renal failure dataset from the UCI repository was employed published by Rubini and Eswaran [10]. Figure 2 represents the dataset on chronic renal sickness consisting of 24 characteristics that are annotated with class data, specifically "yes" and "no". These characteristics possess the capacity to discern a patient afflicted with chronic renal disease.

Column	Non-Null Count	Dtype
age	400 non-null	float64
blood_pressure	400 non-null	float64
specific_gravity	400 non-null	float64
albumin	400 non-null	float64
sugar	400 non-null	float64
red_blood_cells	400 non-null	int64
pus_cell	400 non-null	int64
pus_cell_clumps	400 non-null	int64
bacteria	400 non-null	int64
blood_glucose_random	400 non-null	float64
blood_urea	400 non-null	float64
serum_creatinine	400 non-null	float64
sodium	400 non-null	float64
potassium	400 non-null	float64
haemoglobin	400 non-null	float64
packed_cell_volume	400 non-null	float64
white_blood_cell_count	400 non-null	float64
red_blood_cell_count	400 non-null	float64
hypertension	400 non-null	int64
diabetes_mellitus	400 non-null	int64
coronary_artery_disease	400 non-null	int64
appetite	400 non-null	int64
peda_edema	400 non-null	int64
aanemia	400 non-null	int64
class	400 non-null	int64

Figure 2. *Description of the dataset.*

3.2. Data Pre-processing and Feature Selection

The process of data preprocessing is an essential stage in the machine learning pipeline. Its purpose is to guarantee that the data is in a format that is suitable for modelling. While this procedure was being carried out, null values and duplicate entries were eliminated. The dataset has 24 features. To decrease the number of features, this study employed the ExtraTreesClassifier technique to determine the importance of each feature. The given method is an ensemble learning technique referred to as an Extremely Randomized Trees classifier.

This technique relates to the ensemble learning family, wherein many models are trained to address the same problem, and their predictions are aggregated to enhance performance.

The ExtraTreesClassifier algorithm constructs several decision trees by randomly selecting subsets of characteristics and applying random thresholds to separate nodes. This stochasticity aids in mitigating overfitting.

Figure 4 represents the outcomes of this technique on the dataset for all the characteristics, providing information on the importance of each characteristic. This technique is valuable for feature selection and gaining insights into the data. In this study, the top-10 features were chosen for classification and prediction, and they are represented Figure 3.

Feature	Score
specific_gravity	0.194122
diabetes_mellitus	0.13243
hypertension	0.111859
haemoglobin	0.098245
albumin	0.085076
packed_cell_volume	0.080988
appetite	0.047786
red_blood_cell_count	0.031146
peda_edema	0.031005
pus_cell	0.0288

Figure 3. *Top-10 features.*

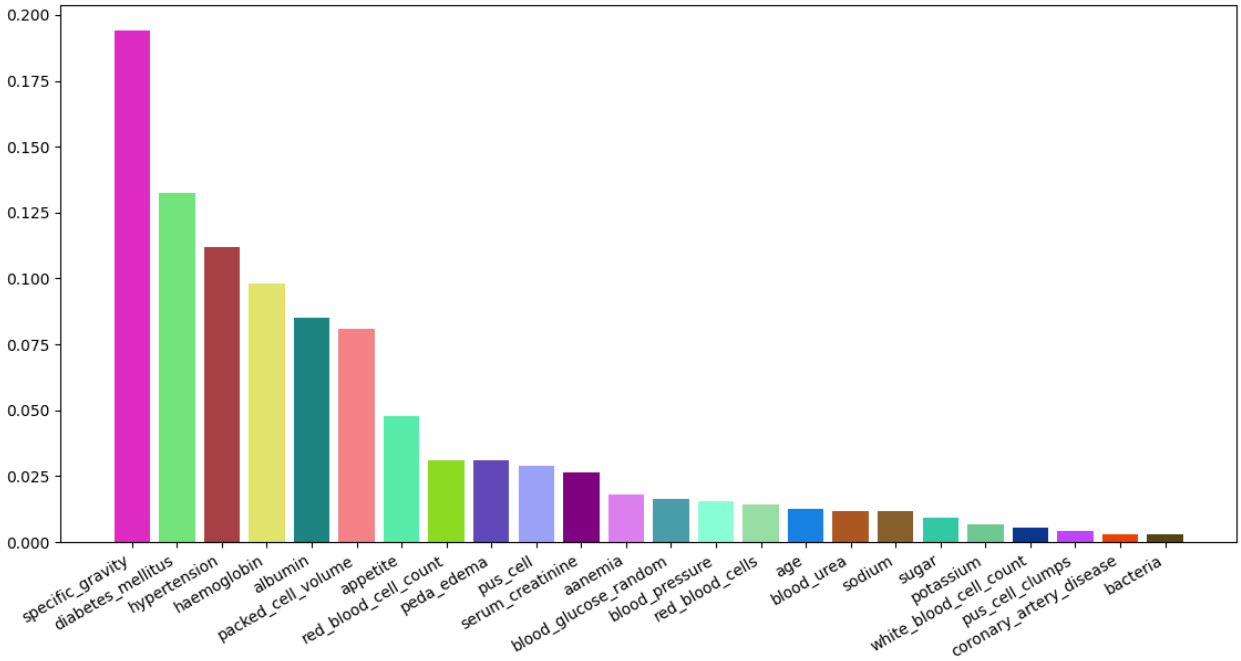

Figure 4. *Feature scores of the dataset.*

3.3. Training and Testing

The primary purpose of this research is to identify the classification method that is responsible for the best accurate prediction of chronic renal illness. Previous research has mostly focused on the use of data mining and traditional machine learning models for the purpose of performing classification analysis and making predictions for renal illnesses in general. However, even though they were aware of the advantages of Deep Learning models, the researchers also considered the possibility of utilizing a Convolutional Neural Network (CNN) model in conjunction with traditional machine learning techniques.

Machine Learning Models:

Within the scope of this work, the following conventional machine learning models are taken into consideration:

1. Random Forest (RF) is a set of ensemble models that use the combination of numerous decision trees to enhance

accuracy.

2. A decision tree, sometimes known as a DT, is a tree-based model that divides data into chunks according to feature thresholds.
3. A technique known as Gradient Boosting (GB) is an ensemble method that creates weak learners (often decision trees) in a sequential manner to produce a robust model.
4. A Support Vector Machine, often known as an SVM, is a model that determines the best hyperplane to divide groups of data into distinct categories.

Deep Learning Model:

Additional components of this research include the CNN model, which is an example of a Deep Learning architecture. The CNN is composed of several layers, which give it its distinctive characteristics. These levels include the convolutional layer, the pooling layer, the linear unit of correction layer (ReLU), then the layer that is completely connected.

3.4. Evaluation

All the models, including RF, DT, GB, SVM, and CNN, will be analyzed with the testing dataset once the training phase is complete. For evaluating their performance, evaluation criteria including as accuracy, precision, recall, and F1-score will be utilized. It is expected that the model that achieves the best results on these criteria will be regarded as the most efficient classification algorithm for the forecasting of chronic kidney disease.

3.5. Prediction

This information will be gathered by the application whenever a user fills out a prediction form and provides input parameters. These parameters may include patient demographics, laboratory findings, and medical records. Once the most appropriate model has been chosen, these input parameters will be used as the features of the model. A prediction result will be generated by the model after it has processed the features that were input. When it comes to chronic renal disease, the outcome of the forecast is going to be one of the instances that follows:

1. Positive: This shows that the model can accurately forecast the possibility of chronic kidney disease for the patient.
2. Negative: The presence of persistent kidney failure is something that the model does not predict, which is a negative indicator.

4. Results

In this work, a variety of various algorithms for machine learning are presented together with a deep learning model of CNN, with the goal of achieving an earlier diagnosis of chronic kidney disease (CKD). The models that are developed with the help of people with chronic kidney disease are then trained and verified with the help of the parameters used for input that were described earlier. To reduce the total number of attributes and eliminate unnecessary data, research has been conducted on the associations that exist between the various parameters. When an ExtraTree feature selection approach was used to select the top-k characteristics, it was found that things like hemoglobin, albumin, specific gravity, and other related variables had the most significant influence on the ability to predict chronic kidney disease (CKD). Sciket-learn and TensorFlow, which are open-source APIs for the ML and DL implementation in Python, have been utilized in a variety of various ways throughout the course of undertaking this project. For this investigation, the evaluation criteria that have been considered were recall, precision, accuracy, and F1-score.

All algorithms have shown exceptional performance for CNN in accurately identifying chronic kidney disease (CKD), with an accuracy rate over 97%, as revealed by the final evaluation results. The work utilized the top 10 characteristics from the dataset, as depicted in Figure 3. These features encompassed variables such as hemoglobin, specific gravity, blood pressure, and hypertension, among others. The Random Forest, Gradient Boosting, and Decision Tree models all achieved scores above 86%, indicating that they were successful at identifying persons who were not affected by the illness or were in good health. The only exception was the SVM model. This can be demonstrated by placing emphasis on precision in addition to other measurements. By employing this methodology, it is possible to produce a precise forecast regarding the existence of chronic renal illness. This development has the potential to aid the medical community in expanding their understanding of biological science. The evaluation results comparing the performance of the model scores and the top-k characteristics data are displayed in Table 2 and Figure 5.

Table 2. *Performance comparison table.*

Algorithms	Accuracy	Precision	Recall	F1-score
SVM	60	30	50	37.5
DT	84.16667	83.64876	84.44444	83.98238

Algorithms	Accuracy	Precision	Recall	F1-score
RF	87.5	87.56631	87.22222	87.38657
GB	86.66667	86.52778	86.52778	86.52778
CNN	97.16667	97.24324	97.1831	97.17241

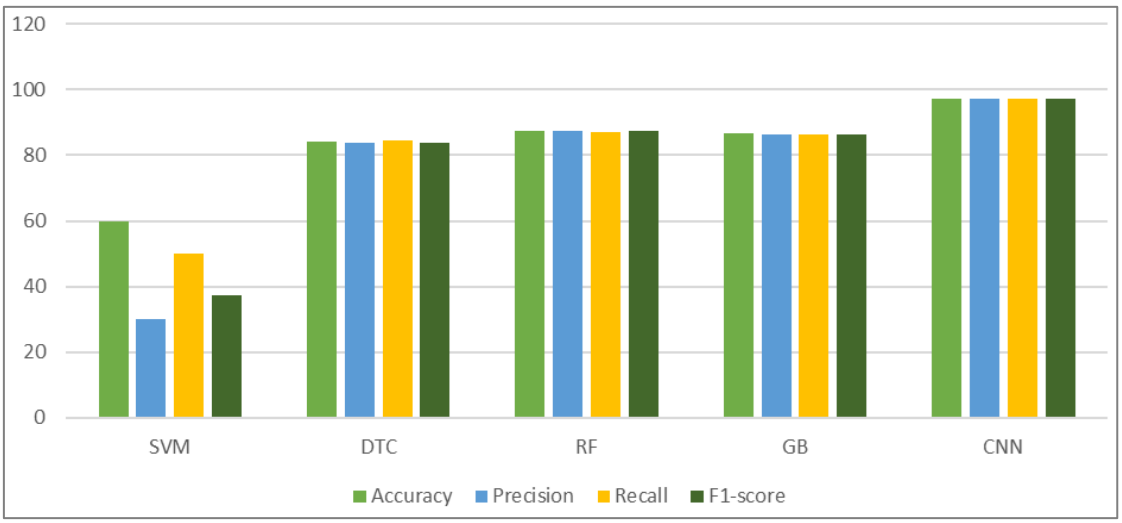

Figure 5. *Performance comparison graph.*

5. Critical Analysis

The primary objective and research question of this study are to determine the most effective classifier for predicting chronic renal disease using machine learning and deep learning techniques. The work evaluates four classification models: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB). Additionally, a Convolutional Neural Network (CNN) is also assessed. Performance metrics such as accuracy, precision, recall, and F1-score are used to compare these classifiers, with thirty percent of the test dataset.

For reducing the overall number of attributes and getting rid of data that was not necessary, carried out research on the associations that exist between the various criteria. The top k (ten) features were chosen by the application of an ExtraTreeClassifier feature selection strategy. The results of our investigation showed that the ExtraTreeClassifier performed exceptionally well in two different experiments: the first experiment used the complete dataset's 25 features, while the second experiment used the ExtraTreeClassifier to choose the best 10 features. Through the process of comparing the results, discovered that all the algorithms, except for SVM, had improved performance scores, with a minimum improvement of 0.9. A comparison of the performance scores obtained with and without the selection of features is shown in Table 3 and Figure 6, respectively.

Table 3. *Performance comparison table between the total dataset and top-10 features.*

	Accuracy	Precision	Recall	F1-score
SVM (Top-10 Features)	60	30	50	37.5
SVM (Total Dataset)	60	30	50	37.5
DT (Top-10 Features)	84.16667	84.64876	84.44444	84.98238
DT (Total Dataset)	83.33333	83.64865	83.91667	83.25124
RF (Top-10 Features)	87.5	87.56631	87.22222	87.38657
RF (Total Dataset)	86.33333	86.64865	86.91667	86.25124

	Accuracy	Precision	Recall	F1-score
GB (Top-10 Features)	88.33333	88.64865	87.91667	88.25124
GB (Total Dataset)	86.66667	86.52778	86.52778	86.52778
CNN (Top-10 Features)	97.16667	97.24324	97.1831	97.17241
CNN (Total Dataset)	96.33333	96.80952	96.80952	96.80952
RF (Top-10 Features)	87.5	87.56631	87.22222	87.38657
RF (Total Dataset)	86.33333	86.64865	86.91667	86.25124

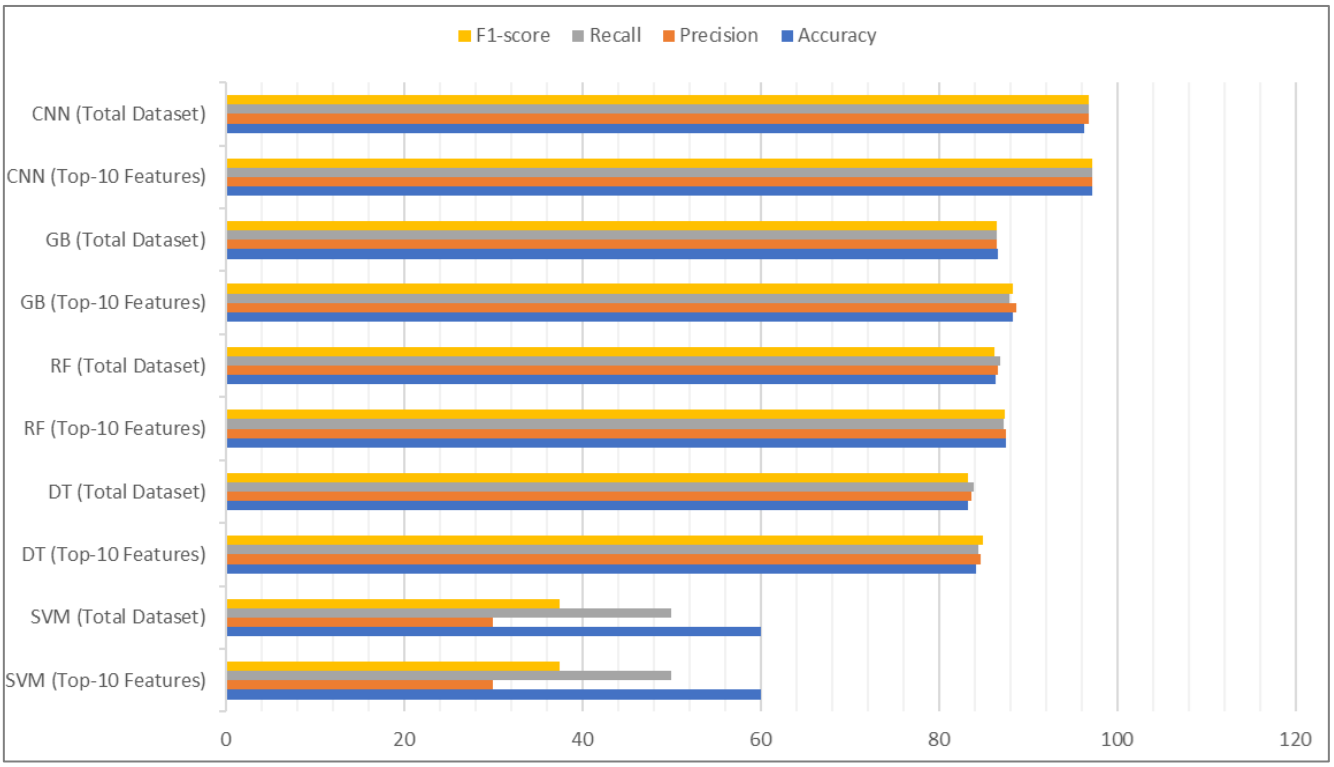

Figure 6. *Performance comparison graph between the total dataset and top-10 features.*

6. Conclusions

Ultimately, this study emphasizes the crucial need of promptly identifying and addressing chronic kidney disease (CKD), a condition that can have severe and potentially fatal outcomes. Given the kidneys' crucial function in maintaining body equilibrium, any decline in renal function can have significant consequences for general health. The difficulty lies in detecting chronic kidney disease (CKD) during its initial phases when symptoms are frequently not present, which requires the creation of inventive diagnostic instruments. By utilizing machine learning (ML) and deep learning (DL) methods, our research provides a viable approach for accurately predicting chronic kidney disease (CKD) at an early stage. Using a dataset from the University of California, Irvine (UCI) and advanced algorithms such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), and Convolutional Neural Network (CNN), have shown that DL-based methods can achieve a prediction accuracy of 97%. This advancement has great potential to provide healthcare practitioners with tools to quickly identify CKD, allowing for timely interventions that might save lives. In future work, a substantial quantity of increasingly sophisticated and complete data will be gathered with the aim of developing the predictive tool. This will improve its capacity for generalization and allow it to evaluate the seriousness of the illness.

Abbreviations

AI Artificial Intelligence

BPNN	Back-Propagation Neural Network
CKD	Chronic Kidney Disease
CNN	Convolutional Neural Network
DT	Decision Tree
GB	Gradient Boosting
KNN	k-Nearest Neighbours
LR	Logistic Regression
ML	Machine Learning
NB	Naive Bayes
RF	Random Forest
SVM	Support Vector Machine
UCI	University of California, Irvine

Author Contributions

Preethi Kolluru Ramanaiah is the sole author. The author read and approved the final manuscript.

Funding

This work is not supported by any external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Desai. S. D, S. Giraddi, P. Narayankar, N. R. Pudakalakatti, S. Sulegaon, Back Propagation neural network versus logistic regression in heart disease classification in advanced computing and communication technologies (Springer, Singapore, 2019). https://doi.org/10.1007/978-981-13-0680-8_13

[2] G.-S. Fu, Y. Levin-Schwartz, Q.-H. Lin, and D. Zhang, "Machine learning for medical imaging," Journal of healthcare engineering, vol. 2019, pp. 1-2, 2019. https://doi.org/10.1155%2F2019%2F9874591

[3] Hamsagayathri. P, Vigneshwaran. V, "Symptoms Based Disease Prediction Using Machine Learning Techniques", 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 31 March 2021. https://doi.org/10.1109/ICICV50876.2021.9388603

[4] Ifraz G. M., M. H. Rashid, T. Tazin, S. Bourouis, M. M. Khan, Comparative analysis for prediction of kidney disease using intelligent machine learning methods. Comput. Math. Methods Med. (2021). https://doi.org/10.1155/2021/6141470

[5] Nazin Ahmed, Rayhan Ahammed, Md. Manowarul Islam, Md. Ashraf Uddin, Arnisha Akhter, Md. Alamin Talukder, Bikash Kumar Paul, "Machine learning based diabetes prediction and development of smart web application", International Journal of Cognitive Computing in Engineering, Volume 2, Pages 229-241, June 2021. http://dx.doi.org/10.1016/j.ijcce.2021.12.001

[6] Nishanth. A, T. Thiruvaran, Identifying important attributes for early detection of chronic kidney disease. IEEE Rev. Biomed. Eng. 11, 208–216 (2018). https://doi.org/10.1109/rbme.2017.2787480

[7] Nithya. A, A. Appathurai, N. Venkatadri, D. R. Ramji, C. A. Palagan, Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images. Measurement (2020). https://doi.org/10.1016/j.measurement.2019.106952

[8] Ogunleye. A, Q.-G. Wang, XGBoost model for chronic kidney disease diagnosis. IEEE/ACM Trans. Comput. Biol. Bioinform. 17, 2131–2140 (2019). http://dx.doi.org/10.1109/TCBB.2019.2911071

[9] Pal, S. Chronic Kidney Disease Prediction Using Machine Learning Techniques. Biomedical Materials & Devices (2022). https://doi.org/10.1007/s44174-022-00027-y

[10] Rubini, L., Soundarapandian, P., and Eswaran, P.. (2015). Chronic Kidney Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C5G020

[11] Bhatla, N., Jyoti, K.: An analysis of heart disease prediction using different data mining techniques. Int. J. Eng. Res. Technol. 1(8) (2012). ISSN: 2278-0181.

[12] Artificial Neural Networks Architecture and Training Process: https://doi.org/10.1007/978-3-319-43162-8_2

[13] Vijayarani, S., Sudha, S.: Comparative Analysis of Classification FunctionTechniques for Heart Disease Prediction. Int. J. Innov. Res. Comput. Commun. Eng. 1(3) (2013). http://ijircce.com/admin/main/storage/app/pdf/piBcBYWcQ2A51R5e37eswoIWnupB8syG5cQLpUnm.pdf

[14] Murthy, H. N., Meenakshi, M.: ANN model to predict coronary heart disease based on riskfactors. Bonfring Int. J. Man Mach. Interface 3(2) (2013) http://dx.doi.org/10.9756/BIJMMI.4473

[15] Giraddi, S., Pujari, J., Seeri, S.: Role of GLCM Features in Identifying Abnormalities in the Retinal Images. IJIGSP 7(6), 45–51 (2015). https://doi.org/10.5815/ijigsp.2015.06.06

Biography

Preethi Kolluru Ramanaiah senior cloud architect and leading AI initiative team on Ernst and Young. Her passion for numbers is unlimited. She would like to play with numbers to identify the hidden truth and concepts around it and this passion made me to work on creating Advanced data mining and machine learning algorithms since 2009. Also, she has been architecting platforms for clients on Banking, Healthcare and Defense systems. Experienced on wide range of legacy as well as modern and cloud computing technologies. Been tutoring young generation on AI and blockchain from 2017 and my works were posted on Udemy, and my research papers were published on different journals from 2009.