IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101249

SURVEY ON CHRONIC KIDNEY DISEASE PREDICTION SYSTEM

Deepak K.N[1], Adhwaidh P.S[2], Akshay P.D[3], Athira K.S[4], Jisna Jayan[5]

Assistant Professor, Department of Computer Science And Engineering, Universal Engineering College, Vallivattom, Thrissur, India.[1]

B.Tech Student, Department of Computer Science And Engineering, Universal Engineering College, Vallivattom, Thrissur, India.[2,3,4,5]

Abstract: Chronic Kidney Disease (CKD) is a slow diminishing in renal capacity over a time of a while or on the other hand years. Diabetes and hypertension are the most well-known reasons for persistent kidney illness. The manifestations of this infection can't be recognized in the beginning phase. Truth be told, exceptionally lesser individuals know about this infection and can foresee the side effects at the prior stage. With the accessibility of organized clinical information, specialists have drawn in scores to concentrate on clinical illness discovery mechanization with machine learning and data mining. The machine takes in designs from the current dataset, and afterward applies them to an obscure dataset to foresee the result. CKD has been such a field of study for a long while presently. In this manner, the framework means to analyze kidney illness utilizing various machine learning techniques and to choose the best one to evaluate the degree of CKD patients. By utilizing information of CKD patients with 21 attributes and 400 records we use different machine learning methods like DT, SVM, DNN. The attributes are inputted naturally utilizing image processing and letter recognition. To construct a model with the most extreme exactness of anticipating whether or not CKD and in the event that indeed, its Severity.

Keywords: CKD, image processing, machine learning, letter recognition, DT, SVM, DNN

I. INTRODUCTION

Kidney infection is a significant general wellbeing challenge around the world. Particularly in our country life is so costly and month to month or daily incomes of people are low so that patients were not going to hospitals and clinics; they can simply take painkillers as well as most hospitals and clinics are not ready to handle Kidney problems early. As shown by WHO, “Tolerating that hazard Factors are perceived early, uncommon kidney injury and reliable kidney issue can be ruined and, assuming kidney infection is examined early, debilitating of kidney breaking point can be moved back or prevent by humble intervention”. Ongoing kidney illnesses incorporate conditions that hurt kidneys and diminish their capacity to keep us fit. Assuming kidney issue deteriorates, the waste can develop at pinnacle levels in our blood and can cause challenges, for example, hypertension, powerless bones, weakness, terrible sustenance, and nerve harm. Besides, kidney illness raises the chance of heart and vascular sicknesses. CKD can be brought about by diabetes, hypertension, coronary illness, lupus, weakness, microorganisms and egg whites in the urine, difficulties of certain medications, sodium, and potassium insufficiency in the blood, and family ancestry, and numerous others. Early exposure and clinical consideration can for the most part keep away from the deteriorating of ongoing kidney sickness. Assuming that it can deteriorate it can result in kidney disappointment requiring dialysis. It can likewise result in kidney transplantation to support living.

In this paper, we aim to implement a machine learning prediction system to develop a prediction model which to recognize the beginning phase and movement of chronic kidney disease which helps support in decision-making of wellbeing experts to avoid the CKD issue.

II. THEORY

A. Support Vector Machine

This algorithm is the highest well-known and important supervised machine-learning algorithm that is working on classification and regression problems however mostly used for classification problems. SVM utilized kernel function to isolate marked information One of the benefits related with involving portions in SVM is that SVM applies kernel dentitions to non-vector sources of info and Kernels can be characterized dependent on a blend of various information

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101249

types. In the SVM algorithm data classify into two data points in which hyper-plane lies between two branches. SVM makes a discrete hyperplane in the signifier space of the preparation information and mixtures are arranged depending on the side of the hyperplane found.
SVM has been used in a few fields, including bioinformatics and handwritten recognition. SVM is additionally utilized in different applications, including clinical conclusion, climate forecast, securities exchange examination, and picture handling. Like any remaining machine learning procedures, SVM is a computational algorithm that gains as a matter of fact and guides to dispense marks to objects.

B. Decision Tree

Decision tree algorithm belongs to the supervised machine learning algorithm and is used to solve both classification and regression problems, but mostly used for classification problems. Decision Tree algorithm resolves the classification problem by converting the dataset into a tree representation by sorting them by feature values. In decision tree every node indicates features in an instance to be classified and every leaf node indicates a class label the instance belongs to.

C. Deep Neural Network

Deep neural network (DNN) is a type of machine learning algorithms in which the input and output layers are separated by numerous layers. The main aim is to imitate the data handling of the brain. DNN shave more than one secret layer arranged between the input and output layers. One of the key reasons deep learning is more remarkable than classical machine learning is that it makes adaptable arrangements. A model can be worked with a solitary layer of neurons, and adding layers allows the computer to make an ever increasing number of explicit elements that lead to a more perplexing last result.

D. Image Processing

Image processing is a strategy for performing a specific procedure on an image in order to improve it or extract useful data from it. It is a type of sign processing in which the input is a picture and the output could be a picture or qualities/features related to that picture.

III.RELATED WORK

Here are some papers based on the technologies used in the CKD prediction.

Paper [1], proposed a machine learning methodology for diagnosing CKD. Dataset obtained from university of California Irvine (UCI) machine learning repository which has large number of missing values, KNN imputation was used to fill in the missing values, which select several complete samples with the most similar measurement to process the missing data for each incomplete dataset, they use six machine learning algorithms such as logistic regression, random forest,support vectors machine, K- nearest neighbour, naive bays classification and feed forward neural network were used to establish model. Among these machine learning models random forest achieve the best performance with 99.75% diagnostic accuracy and they analyse the misjudgement generated by the establish model to propose an integrated model that combines logistic regression with random forest by using perception and achieve average accuracy of 99.83%.

The aim of the paper [2], is to predict CKD using machine learning technique based on dataset from UCI machine learning dataset warehouse. CKD is detected using the apriori association techniques for 400 instance of chronic kidney patients with 10-fold-cross-validation test and results are compared across a number of classification algorithms including ZeroR, OneR, naïve Bayes, J48 and IBK (K-nearest Neighbour). The dataset pre-processed by completing and normalizing missing data. They select the most relevant feature from the dataset for improved accuracy and reduce training time. Detection accuracy for CKD based on Apriori is 99%.

This research [3], examines data from CKD patients and proposes a system from which it will be possible to predict the risk of CKD. This system should have used 456 patient's data which is collected from online UCI machine learning repository and real time dataset from Khulna city medical college. Python high level interpreted programming language used to develop this system. This system is trained with the data using 10-fold CV and applied random forest and ANN. The accuracy achieved by random forest algorithm is 97.12% and ANN is 94.5%.

This study [4], makes use of a lab dataset of 361 patients with chronic kidney disease. It computes the duration of chronic kidney disease using PNN, SVM, and MLP algorithms. The outcomes of algorithms such as Probabilistic Neural Networks (PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Radial Basis Function

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 12, December 2021
DOI: 10.17148/IJARCCE.2021.101249

(RBF) have been compared. This investigation suggests that the probabilistic neural organisation calculation is the best performing calculation that doctors can use to eliminate demonstrative and treatment errors.

In this work [5], a feature selection method is used to reduce the attributes and select only the most important ones. Four classifiers are used to classify data: JRip, SMO, Naive Bayes, and IBK. This system can predict the correct and best classifiers by comparing the results of the reduced attribute dataset and the original dataset using these four classifiers. The most important part of the process is classification, which is carried out using data mining techniques based on machine learning. The classification can be used to predict data instance group membership.

This paper [6], is based on deep neural network which predicts the presence or absence of chronic kidney disease with an accuracy of 97%. Compared to other available algorithms, this model shows better results which is implemented using the cross-validation technique to keep the model safe from overfitting. This automatic chronic kidney disease treatment helps reduce the kidney damage progression, but for this chronic kidney disease detection at initial stage is necessary.

The main motive of this work [7], is to detect chronic kidney disease by imposing various classification algorithms on the patient's medical record. Based on the classification report and performance factors, this research work is primarily focused on determining the best suitable classification algorithm that can be used for the diagnosis of CKD. Empirical work is done on various algorithms such as Support Vector Machine, Random Forest, XGBoost, Logistic Regression, Neural Networks, and Naive Bayes Classifier. The experimental results show that Random Forest and XGBoost outperform other classification algorithms, achieving 99.29 percent accuracy.

In this paper [8], chronic kidney disease (CKD) is detected using machine learning algorithms. Random forest, support vector machine (SVM), linear regression (LR), decision tree, and naive Bayes classifier are the supervised machine learning algorithms used in this paper. After comparing the results obtained by using these methods, random forest technique is the best suited technique for detection. In this paper 99.3% of detection accuracy is obtained using random forest.

In this work [9], classification models have been built with different classification algorithms, Wrappersubset attribute evaluator and bestfirst search method to predict and classify the CKD and non CKD patients. These models have applied on recently collected CKD dataset downloaded from the UCI repository. The models have shown better performance in classifying CKD and non CKD cases. Results of different models are compared. From the comparison it has been observed that classifiers performed better on reduced dataset than the original dataset.

This paper [10], proposes a methodology using inspired optimization model and learning procedure to classify CKD. The proposed method uses the Ant Lion Optimization (ALO) technique to select applicable features of kidney data for the classification process. The CKD data is then sorted based on selected features using a Deep Neural Network (DNN). When compared to other data mining classifiers, performance comparison shows that this proposed model achieves higher classification accuracy, precision, F-measure, and sensitivity measures.

In this work [11], main focus is on predicting the patient's status of CKD or non CKD. To predict the value in machine learning classification algorithms have been used. Classification models have been built with different classification algorithms will predict the CKD and non CKD status of the patient. These models have applied on recently collected CKD dataset downloaded from the UCI repository with 400 data records and 25 attributes. Results of different models are compared. From the comparison it has been observed that the model with Multiclass Decision forest algorithm performed best with an accuracy of 99.1% for the reduced dataset with the 14 attributes.

The main objective of this work [12], is to determine kidney function failure by applying the classification algorithm on the test result obtained from the patient medical report. The aim of this work is to use classification algorithms to reduce diagnosis time and improve diagnosis accuracy. The proposed work is concerned with the severity classification of different stages of chronic kidney disease. The experiment is performed on different algorithms like Back Propagation Neural Network, Radial Basis Function and Random Forest. The experimental results show that the Radial basis function algorithm gives better result than the other classification algorithms and produces 85.3% accuracy.

The main intent of this paper [13] is to analyze the various data mining techniques in medical domain and some of the algorithms used to predict kidney diseases eventually. From this survey, it is proven that results may vary for different stages of kidney disease diagnosis based on the tools and techniques used. Data mining provides better results in disease diagnosis when appropriate techniques used. Thus, data mining is the significant field for healthcare predictions.

In this paper [14], they use novel approach to detect CKD using machine learning algorithm. They get result on dataset which having 400 records and 25 attributes which gives result of patient having CKD or not CKD. They use k-nearest neighbours, random forest and neural network to get results. For feature reduction they use wrapper method which detect CKD with high accuracy.

The main objective of this proposed work [16], is to search the effect of class imbalance in training data when developing neural network classifier for medical decision making on chronic kidney disease. Neural networks are widely used in a number of applications including data mining and decision systems. Back propagation networks are a popular type of neural networks that can be trained to recognize different patterns. The importance of these networks

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 12, December 2021
DOI: 10.17148/IJARCCE.2021.101249

was considered and a comparative study of some sampling algorithms was performed based on multilayer perceptron with different learning rate values for the prediction of chronic kidney disease. This study reveals that sampling algorithms can improve the performance of classification algorithms and learning rate is a crucial parameter which can significantly effect on multilayer perceptron.

In this work [16], they assessed 12 different classification algorithm on dataset which having 400 records and 24 attributes. They had compared their calculated results with actual results for calculating the accuracy of prediction results. They used assessment metrics like accuracy, sensitivity, precision, and specificity. They find that the decision tree technique gives accuracy up to 98.6%, sensitivity of 0.9720, and precision of 1 and specificity of 1.

This research work [17], is based on big data in healthcare which has been developed by using data mining techniques. For analysis and prediction of Chronic Kidney Disease (CKD) Data mining techniques has been used. Expectation Maximization [EM] is the clustering algorithm which is used to cluster similar type of person into one group. Artificial Neural Network [ANN] and C4.5 are classification algorithm which is used for prediction of the disease.

The main objective of this research work [18], is to predict kidney diseases using classification algorithms such as Naïve Bayes and Support Vector Machine. This research work mainly focused on finding the best classification algorithm based on the classification accuracy and execution time performance factors. From the results, it can be concluded that the SVM achieves increased classification performance, yields results that are accurate, hence it is considered as best classifier when compared with Naïve Bayes classifier algorithm. Perhaps, Naïve Bayes classifier classifies the data with minimum execution time.

In this research [19], K-Nearest Neighbor, J48, Artificial Neural Network, Naïve Bayes, and Support Vector Machine classification techniques were used to diagnose Chronic Kidney Disease. To predict chronic kidney disease, build two important models. Namely, feature selection method and ensemble model. To build chronic kidney disease prediction, used Info gain attributes evaluator with ranker search engine and wrapper subset evaluator with the best first engine was used. The result showed that the K-nearest neighbor classifier by using Wrapper Sub set Evaluator with Best first search engine feature selection method has 99% accuracy, J48 with Info Gain Attribute Evaluator with ranker search engine has 98.75, Artificial Neural Network with Wrapper Sub set Evaluator with Best first search engine has 99.5% accuracy, Naïve Bayes with Wrapper Sub set Evaluator with Best first search engine has 99% accuracy, Support Vector Machine with Info Gain Attribute Evaluator with ranker has 98.25% accuracy in prediction of chronic kidney disease compared to other with and without feature section method. The second model building method ensemble model by combing the five heterogeneous classifiers based on a voting algorithm. The effectiveness of the proposed ensemble model was examined by comparison of the base classifier. The experimental result showed that the proposed ensemble model achieved 99% accuracy.

The objective of this work [20], is to develop a clinical decision support system using machine learning techniques. In this paper first the classification techniques like neural network based back propagation (BPN), probability based Naive Bayes, LDA classifier, lazy learner K Nearest Neighbor (KNN), tree based decision tree, and Random subspace classification algorithms are analyzed. The accuracy of each algorithm found is 81.5%, 78%, 76%, 90%, 93% and 94% respectively on a dataset collected from UCI repository which contains 25 attributes and 400 instances. From the results obtained, the algorithm which gave better result was used for the developing the Clinical Decision Support System.

IV. CONCLUSION

This paper describes a clinical area application that uses machine learning algorithms to assist clinical professionals in anticipating CKD based on CKD parameters. Support Vector Machine, Decision Tree, and Deep Neural Network are three machine learning algorithms that have been considered for predicting chronic kidney disease. The dataset is obtained from the University of California, Irvine's Machine Learning Repository. The dataset contains 400 records and 26 features. From 26 features, we only use 21 of them. The features are automatically entered using image processing and letter recognition. This system will identify chronic kidney disease in early stages. This assists specialist with diagnosing and recommending the treatment at the beginning phase. Patients will be able to learn about their health condition at an earlier stage and follow the necessary diet and prescriptions to improve the progression of this condition.

ACKNOWLEDGMENT

We utilize this opportunity to convey our gratitude towards all those who have helped us directly or indirectly for the completion of our work. We deeply and wholeheartedly thank Dr. Sreeraj R -HOD, Computer Science and Engineering for his extreme valuable advice and encouragement. We especially thankful to our guide and supervisor Mr. Deepak K N -Assistant Professor, Computer Science and Engineering for giving me valuable suggestions and critical inputs in the preparation of this paper. We would like to extend our sincere gratitude to all faculty of Computer

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101249

Science and Engineering department for the support and suggestions that helped us in the development of our work to what it is now. We thank our parents and friends for the mental support provided during the course of our work at the times when our energies were the lowest.

REFERENCES

[1]. Y. Liu, C. Liu, J. Qin, L. Chen, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease," IEEE Access, vol. 8, pp. 20991–21002, 2020, doi: 10.1109/ACCESS.2019.2963053

[2]. Z. Wang, J. Won Chung, X. Jiang, Y. Cui, M. Wang, and A. Zheng, "Machine Learning-Based Prediction System For Chronic Kidney Disease Using Associative Classification Technique," Int. J. Eng. Technol., vol. 7, no. 4.36, p. 1161, 2018, doi: 10.14419/ijet.v7i4.36.25377.

[3]. S. Y. Yashfi et al., "Risk Prediction of Chronic Kidney Disease Using Machine Learning Algorithms," 2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020, 2020, doi: 10.1109/ICCCNT49239.2020.9225548.

[4]. A. S. Anwar and E. H. A. Rady , "Prediction of kidney disease stages using data mining algorithms," Informatics Med. Unlocked, vol. 15, no. March, p. 100178, 2019, doi: 10.1016/j.imu.2019.100178.

[5]. E. Perumal , P. Arulanthu , "Predicting the Chronic Kidney Disease using Various Classifiers," 4th Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2019, pp. 70–75, 2019, doi: 10.1109/ICEECCOT46775.2019.9114653

[6]. H. Kriplani, B. Patel, and S. Roy, "Prediction of chronic kidney diseases using deep artificial neural network technique," vol. 31. Springer International Publishing, 2019.

[7]. N. V. Ganapathi Raju, K. Prasanna Lakshmi, K. G. Praharshitha, and C. Likhitha, "Prediction of chronic kidney disease (CKD) using Data Science," 2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019, no. Iciccs, pp. 642–647, 2019, doi: 10.1109/ICCS45141.2019.9065309.

[8]. 9.P. Kotturu, V. V. S. Sasank, G. Supriya, C. S. Manoj, and M. V. Maheshwarredy, "Prediction of chronic kidney disease using machine learning techniques," Int. J. Adv. Sci. Technol., vol. 28, no. 16, pp. 1436–1443, 2019, doi: 10.17148/IJARCCE.2018.71021.

[9]. S. D. Sudarsan,N. Chetty, K. S. Vaisla, "Role of attributes selection in classification of Chronic Kidney Disease patients," pp. 1–6, 2016, doi: 10.1109/cccs.2015.7374193.

[10]. K. Shankar, G. Devika, P. Manickam, M. Ilayaraja, "Optimal Feature Selection for Chronic Kidney Disease Classification using Deep Learning Classifier," 2018 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2018, pp. 1–5, 2018, doi: 10.1109/ICCIC.2018.8782340.

[11]. Gunarathne W.H.S.D,Perera K.D.M, Kahandawaarachchi K.A.D.C.P, "Performance Evaluation on Machine Learning Classification Techniques for Disease Classification and Forecasting through Data Analytics for Chronic Kidney Disease (CKD)",2017 IEEE 17th International Conference on Bioinformatics and Bioengineering

[12]. S.Ramya, Dr. N.Radha, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms," Proc. International Journal of Innovative Research in Computer and Communication Engineering,Vol. 4, Issue 1, January 2016.

[13]. S.Dilli Arasu and Dr. R.Thirumalaiselvi, "Review of Chronic Kidney Disease based on Data Mining Techniques",International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 (2017) pp. 13498-13505

[14]. Asif Salekin, John Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes," Proc. IEEE International Conference on Healthcare Informatics (ICHI), IEEE, Oct. 2016, doi:10.1109/ICHI.2016.36.

[15]. Pinar Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron: Chronic Kidney Disease Prediction," Proc. 41st IEEE International Conference on Computer Software and Applications (COMPSAC), IEEE, Jul. 2017, doi: 10.1109/COMPSAC.2017.84

[16]. Sahil Sharma, Vinod Sharma, Atul Sharma, "Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis," July18, 2016.

[17]. Tabassum S, [2] Mamatha Bai BG, [3] Jharna Majumdar," Analysis and Prediction of Chronic Kidney Disease using Data Mining Techniques," International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 4, Issue 9, September 2017.

[18]. Dr.S. Vijayarani1, Mr.S.Dhayanand2," DATA MINING CLASSIFICATION ALGORITHMS FOR KIDNEY DISEASE PREDICTION," International Journal on Cybernetics & Informatics (IJCI) Vol. 4, No. 4, August 2015.

[19]. S. Zeynu and S. Patil, "Prediction of Chronic Kidney Disease using Data Mining Feature Selection and Ensemble Method," International Journal of Data Mining in Genomics and Proteomics, 9(1), pp. 1-9, 2018.

[20]. R. Ani, S. Greeshma, S. Resmi, and O. Deepa, "Decision Support System for Diagnosis and Prediction of Chronic Renal Failure Using Random SubSpace Classification," International. Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, pp. 1-6, 2016.