DIAGNOSIS OF CHRONIC KIDNEY DISEASE USING DEEP LEARNING

Fathima S.K[1], Priyadharshini T[2], Poorna Chandra K[3], MahaSwetha S[4]

[1, 2,3,4] *Department of Computer Science and Engineering , Sona College of Technology, Tamilnadu, India*

Fathima.sk@sonatech.ac.in, dharshinithangaraju01@gmail.com, kpoornachandra1@gmail.com, mahaswetha37@gmail.com

Abstract:

Disease prediction is one of the most important procedures in saving a person's life. Machine learning and deep learning plays a key role in disease prediction, and research is ongoing to improve accuracy. Kidney illness is one of the primary diseases that poses a threat to our health. Ignoring a kidney problem can lead to chronic renal disease and death. Chronic Kidney Disease and its indications are sometimes gentle and progressive, going undiagnosed for years before being discovered. Unfortunately, there has yet to be found a technological cure for the disease. Bade General Hospital provided us with a dataset of 400 patients with ten characteristics. DNN is now being used in medical image processing to diagnose diseases like cancer and diabetes. The model has a 98 percent accuracy rate. We also selected and underlined the relevance of Factors to produce a rating of the features utilized in CKD prediction. The results showed that 2 factors, Creatinine and Urea, have the most impact onCKD prognosis.

Keywords: *Chronic Kidney Disease (CKD),Deep Learning (DL), disease prediction, symptoms and accuracy*.

I.INTRODUCTION

The kidneys are two bean-shaped organs that are around the size of fists [1]. One on either side of the spine, they are positioned immediately below the rib cage. The kidneys filter 120 to 150 quarts of blood every day to generate 1 to 2 quarts of urine. The kidneys' primary role is to eliminate waste materials and overflow fluid from the body through urine. Urine is made up of a series of very complicated excretion and restoration stages. This mechanism is required to keep the body's chemical equilibrium steady. The kidneys regulate the body's salt, potassium, and acid levels, as well as producing hormones that impact the operation of additional organs. A hormone generated by the kidneys, for example, drives red blood cell synthesis, controls blood pressure, and regulates calcium metabolism, among other things.

Early detection of renal disease may aid with rectification, which is not always achievable. To avert major harm, we must first have a better awareness of a few renal disease signs. Deep learning classifiers are used to predict the class, target, labels, and categories of a data item. Classification is a type of supervised learning in which the objectives are given input data [2]. A few of the uses include medical diagnosis, spam detection, and targeted marketing. They achieve this by converting discrete input variables (X) into discrete output variables using a mapping function (f) (Y).

There are two main diseases of CKD:

1. Diabetes and
2. High blood pressure

As a result, preventing CKD requires managing these two disorders. CKD usually does not manifest itself until the kidneys have been severely damaged. According to research, hospitalisation cases are increasing at a rate of 6.23 percent per year, yet the worldwide fatality rate stays constant.

There are few diagnostic tests to check the condition of CKD:

1. estimated glomerular filtration rate(eGFR)
2. urine test
3. Blood pressure

Determine which groups may benefit from more systematic CKD screening and offer an overview of screening and diagnostic procedures.

Outline therapy options for CKD patients to slow the course of renal impairment and reduce morbidity and death.

Emphasize the significance of proactive therapy of prevalent co-morbid disorders such as cardiovascular disease and diabetes to potentially reduce morbidity and mortality among patients with CKD.

2. RELATED WORK

Nikhila (2021) explains how four basic calculations are used to evaluate whether a patient has Chronic Kidney Disease in its initial stages. Accuracy, sensitivity, specificity,and F1-Score, are among the six execution metrics used to assessDL models. The Mathew Correlation Coefficient and Area Under the Bend scores for AdaBoost and Random Forest were both 100 percent. By allowing clinical doctors to examine the condition at an early stage, the DL model suggested on this work will provide an effective ways for preventing Disease [3].

G Nandhini; J Aravinth (2021), depicts CKD is an infirmity which will in general harm the kidney and influence their viable working of discharging waste and adjusting body liquids. A portion of the confusions included are hypertension, iron deficiency (low blood count), mineral bone issue, poor dietary wellbeing, corrosive base irregularities, and neurological complexities. Early and blunder free recognition of CKD can be useful in deflecting further crumbling of patient's wellbeing [4]. These persistent illnesses are visualized utilizing different sorts of information mining grouping approaches and (DL) calculations.

Md. MohsinSarkerRaihanet.al (2021), talks about an earlier guess of CKD can feed the personal satisfaction to a higher reach in such conditions and can upgrade the quality of life to a bigger territory. Sara Alshakraniet.al (2021), states one of the significant medical conditions is persistent kidney illness (CKD). It is expanding consistently because of unfortunate dietary patterns, deficient water utilization, and absence of wellbeing mindfulness. Innovative progressions, for example, Deep Learning (DL), affect the wellbeing area by giving more precise recognition and fruitful therapy of numerous persistent illnesses. This examination paper investigates an assortment of DL strategies to foresee kidney infection [6].

Pronab Ghosh et.al (2021), talks about he in general review has been carried out in view of four dependable methodologies, for example, Support Vector Machine (from this time forward SVM), AdaBoost (hence AB), Linear Discriminant Analysis (consequently LDA), and Gradient Boosting (hence GB) to obtain exceptionally precise aftereffects of expectation. These calculations are executed on an online dataset of UCI AI store .Afterward, unique execution assessment measurements have additionally been shown to show fitting results [7].

3. PROPOSED METHODOLOGY

The suggested approach may estimate the risk of CKD for public health using generally available health data. When we evaluate actual data, the suggested model will fail to categorise, and the classification result will be unacceptable. In order to predict CKD, they recommended comparing Decision Tree and Support Vector Machine (SVM) techniques. During the classification stage, it was discovered that Decision Tree recognition outperformed SVM recognition in terms of proper classification.

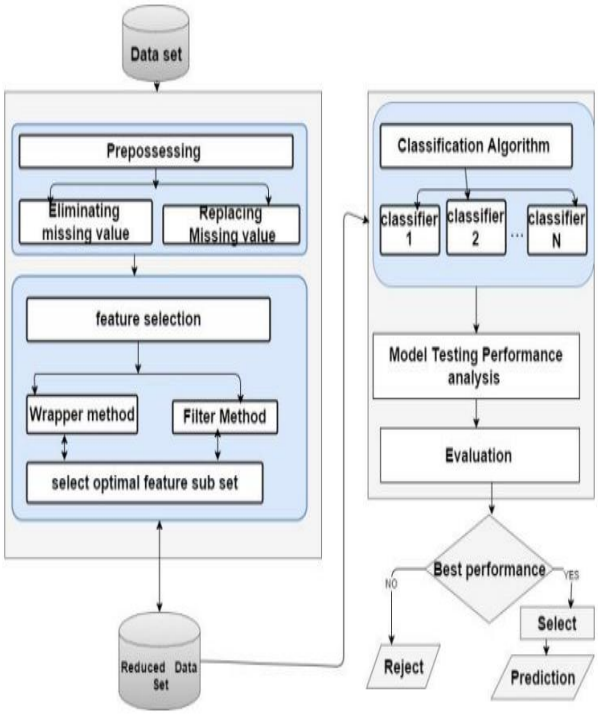


Figure 1.1: proposed system architecture

A Data Feature Selection (DFS) model is suggested in this work to pick and filter important characteristics from the medical dataset for early CKD prediction, using CKD clinical data with 12 numerical and 12 nominal features. Separate feature selection methods on numerical and nominal data are used in the proposed approach to determine the most prevalent qualities that might help in early CKD prediction.

Data Collection:

The source of data for this study is a patient's renal disease record. This data was gathered at the General Hospital in Yobe State's Gashua Local Government Area. It comprises 400 patient records with 11 attributes/parameters, which includes a goal variable defined as CKD or non-CKD illness.

Attribute	Attribute Description
Sex	Gender
Age	Age
Sod	Sodium
Pot	Potassium
Chl	Chloride
Bica	Bicarbonate
Urea	Urea
Cre	Creatinine
UA	Urea Acid
Alb	Albumin
Class	{Kidney Disease, NoKidney Disease}

Figure 2: Dataset Attributes

Data Pre-processing

Pre-processing is the majority significant work in data mining techniques; it entails cleaning, taking out, and transformation of information into a machine-readable layout. Raw data holds missing information, improper layout, and incorrect data, which leads to catastrophe in machine learning calculation. Because string values are not understood by the machine, they had few unfilled cells, which were restored using attribution.

Target variable

The CKD dataset has a wealth of valuable factors that are critical for identifying the illness in patients. We based our factors on the test and procedure used by the hospital in detecting the existence of renal disease, which includes ten blood tests. Names were also included in the exam, but we opted to delete them because they had no bearing on our test and the patients' privacy must be protected.

Prediction using DNN

DNN is a subset of ANN that simulates the structure and functionality of biological neural networks, which include an input, weights, and activation function. DNN contains an input, hidden layer, and output structure. The output in DNN is denoted by a, where Wi and Xi are the weight and input, respectively. This is expressed as follows: The accepted model contains ten (10) inputs, (2) hidden layers, and (1) output.

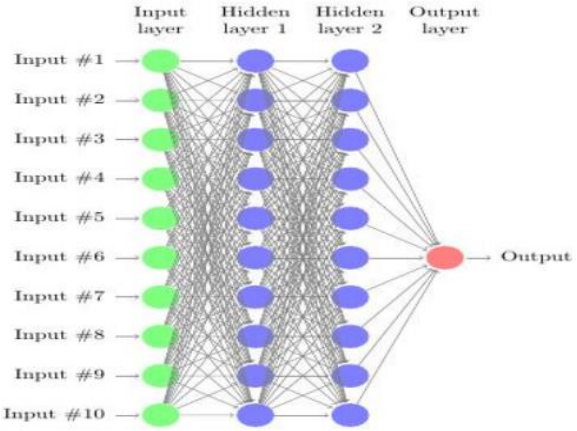


Figure 3: Deep Neural Network

The key in layer is made up of neurons that process the real incoming information. The number of neurons may be determined based on the data supplied. Three layers were utilized to train the system. The inauguration function "Relu" is utilized for key in and the secreted layer. Relu produces a 0 or 1 output. With Sigmoid as the activation function, the result layer has just one output result: KD or NKD. The optimizer for the model was Stochastic Gradient. The Python programming language was used to create the experiments.

4. EXPERIMENTAL RESULT

Confusion Matrix: A confusion matrix is a table structure that is especially used for algorithm presentation. It demonstrates the statistical adequacy of the system and its compatibility with the dataset.

Classification		Observation	
		Negative	Positive
Observations	Negative	True Negative(TN)	False Positive(FP)
	Positive	False Negative(FN)	True Positive(TP)

Classification		Prediction	
		Absence	Presence
Observation	Absence	29	0
	Presence	1	30

Accuracy- It's used to count how many data points were successfully predicted out of all the data points. The work is to calculate by using dividing the to entire number of right guesses by the entire number of forecasts.

Sensitivity: It's the number of positive events that were predicted as positive. The ratio is true positive to the sum of true positive and false negative. The ability is to correctly identify persons who have the ailment in medical diagnosis is known as test sensitivity (recall).

Specificity: It's distinct as the percentage of real negatives that were anticipated as negatives. It's determined by the method of dividing the entire number of negatives by the total number of valid negative predictions. True negative rate is another name for it.

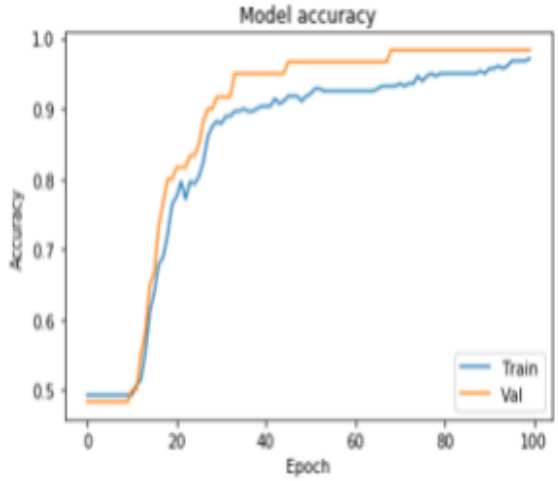
Cohen Kappa: This is a comparison of the presentation of two sets of categorizedinformation. The Kappa output values range from 0 to 1. With higher kappa values, the results become more relevant. It assesses how well the ML classifier's classifications coordinated the data labelled as the positive.

Precision: It's the proportion of retrieved cases with relevant occurrences. This is a correlation between correctly classified modules and also the overall faultprone modules.

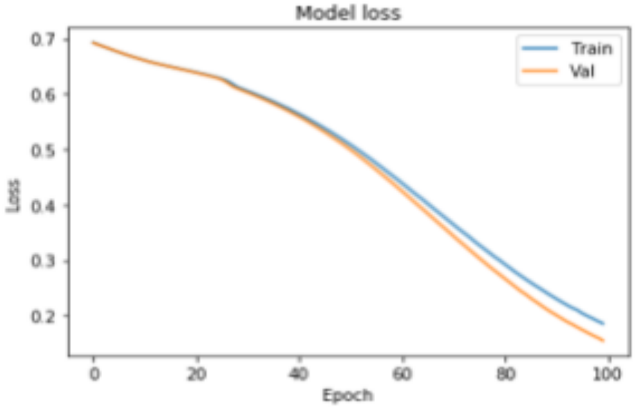
Recall/ Sensitivity: The number of positive identifications of a disease produced out of all potential positive predictions is known as recall.

F1 Score: It is to define the harmonic mean of accuracy, specificity and recall. The ranges from(0 to 1) It describes the model's precision (how many records it can properly classify) and resilience (if it misses a significant number of data).

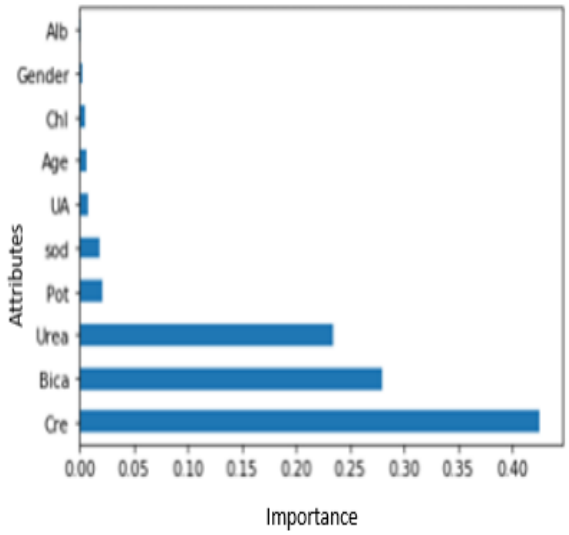
Performance evaluation for Deep Neural Networks Model



Graph of training against theValidation Accuracy



Graph of Training against the Validation Loss



Importance of the feature

For testing and training, we partitioned our dataset into 70 percent (280) and 30 percent (60). The confusion matrix revealed the Deep neural network model's performance on the given dataset; this model successfully identified 34 non-CKD samples and 25 CKD samples, while one CKD sample was classified wrongly. The DNN model worked on our datasets, with 59 trained samples correctly identified, and only one sample out of the sample picked for tested was wrongly classified. When the model was tested with unknown data, it correctly classified renal patients' records as CKD or non-CKD with an overall prediction accuracy of 98 percent. The algorithm stated that it is highly intelligent to generate seamless predictions into CKD and non-CKD based on its accuracy of 98 percent.

5. CONCLUSION

The primary purpose of this study is to develop a DNN model that can accurately and easily to predict the renal illness.. The DNN model that was used proven to be effective and appropriate for predicting renal disease. The research also emphasized the significance of the variables employed in renal disease prediction. This demonstrated that, out of the ten qualities, Creatinine and Bicarbonate had the very important role on CKD, contrary to Ahmed and Alshebly's results that Creatinine and Urea had the most impact on CKD.

REFERENCES

[1] S.Revathy, B.Bharathi, P.Jeyanthi, M.Ramesh, “Chronic Kidney Disease Prediction using Machine Learning Models” International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019.

[2] Gazi Mohammed IfrazMuhammadHasnath Rashid, TahiaTazin, Sami Bourouis, and Mohammad Monirujjaman Khan, “Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods” Knowledge-Based Intelligent Systems in E-Health and Medical Communication Services Volume 2021 |Article ID 6141470.

[3] Nikhila, "Chronic Kidney Disease Prediction using Machine Learning Ensemble Algorithm," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2021, pp. 476-480, doi: 10.1109/ICCCIS51004.2021.9397144.

[4] G. Nandhini and J. Aravinth, "Chronic kidney disease prediction using machine learning techniques," 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2021, pp. 227-232, doi: 10.1109/RTEICT52294.2021.9573971.

[5] M. M. S. Raihan et al., "Chronic Renal Disease Prediction using Clinical Data and Different Machine Learning Techniques," 2021 2nd International Informatics and Software Engineering Conference (IISEC), 2021, pp. 1-5, doi: 10.1109/IISEC54230.2021.9672365.

[6] S. Alshakrani, R. Taha and N. Hewahi, "Chronic Kidney Disease Classification Using Machine Learning Classifiers," 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), 2021, pp. 516-519, doi: 10.1109/3ICT53449.2021.9581345.

[7] P. Ghosh, F. M. JavedMehediShamrat, S. Shultana, S. Afrin, A. A. Anjum and A. A. Khan, "Optimization of Prediction Method of Chronic Kidney Disease Using Machine Learning Algorithm," 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), 2020, pp. 1-6, doi: 10.1109/iSAI-NLP51646.2020.9376787.

[8] Deshpande, A. (2016). A beginner's guide to understanding convolutional neural networks. Retrieved March, 31(2017).

[9] FRANÇOIS CHOLLET. (2018). Deep Learning Using Python. In 2018 21st International Conference on Information Fusion, FUSION 2018.

[10] Ge, Y., Wang, Q., Wang, L., Wu, H., Peng, C., Wang, J. & Yi, Y. (2019). Predicting post-stroke pneumonia using deep neural network approaches. International Journal of Medical Informatics, 132, 103986.