ISSN PRINT 2319 1775 Online 2320 7876

Research paper Volume 11, Iss 10, 2022

A Neural Approach to Predicting Chronic Kidney Disease

Rahamtulla Baig[1], Mukesh Kumar Vidam [2], Santhi Sri. T[3]
[1] Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India
[2] Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India
[3] Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AP, India

Corresponding Author Email: sri_santhi2003@yahoo.com
DOI : 10.48047/IJFANS/V11/ISS10/419

ABSTRACT
Chronic kidney illnesses are currently affecting a large number of individuals globally. Because of the numerous risk factors, including the diet, environment, and quality of living .Many individuals get abrupt illness without knowing why. Many people throughout the world are today affected by Chronic Kidney diseases. The diagnosis of chronic renal disease is usually invasive, costly, time-consuming, and risky. Early sickness detection techniques are still essential, especially in developing countries where infections are sometimes only found after they are too progressed. The need to address the problems and circumvent obstacles made this inquiry necessary in the state of them. The diagnosis when dealing with CKD is often invasive, expensive, time-consuming, and frequently dangerous. That is the reason why many people leave it untreated till late stages, especially in those nations with little resources. Consequently, the early detection method for the disease is predicted in the proposed method of this paper. In order to predict we have taken the help of machine learning. We have used the Artificial Neural Networks to predict the patient kidney disease present or not.

Keywords:
Chronic kidney disease, Machine Learning, Artificial Neural Networks, Sequential model

1. INTRODUCTION

CKD illustrates us about the kidney diseases which has been caused by high B.P levels. Due to which it can strain the small vessels in blood within the kidney which can make working of kidney to get stopped once it is much strained. Diabetes - having too much aldohexose in your blood might harm your kidneys extremely small filters. A build up of fatty deposits inside the blood vessels supplying your kidneys as a result of high sterol levels may make it difficult for them to function correctly. Renal disease such as glomerulonephritis is known as inflammation of the kidneys. Polycystic kidney disease is a hereditary disorder that creates abnormal rates of growth in the kidneys. [11]

In India, One of the most common diseases, CKD is brought on when the kidneys extract excess fluid and waste from the bloodstream and excrete it in the urine. As the condition progresses, waste products build up in the blood, causing high BP, anaemia, weak bones, malnutrition, and nerve damage. With the help of machine learning, it is possible to identify undiagnosed CKD, forecast the risk that patients will develop chronic diseases, and give patient-specific preventative treatments. Dietary management is based on the disease severity and current Glomerular Filtration Rate (GFR). A crucial diet is necessary to maintain the balance of electrolytes and water in the body as well as to improve the kidneys and avoid future damage.[12]

It is a typical upset that is oftentimes joined to ageing. Though it's going to have an effect on anybody, those of African or South Asian descent appear to expertise it a lot of times. Though it's uncommon, CKD can worsen over time and likely cause the kidneys not engaging at all. many CKD patients can lead long lives despite their pathological state. CKD testing Blood and body waste tests are often used to confirm the presence of CKD. [1]

These exams check for elevated amounts of certain compounds in your blood and urine, which are symptoms of a functioning kidney. Regular testing to monitor for CKD may be indicated if you have a high chance of developing kidney disease (for instance, if you have a known risk factor like high blood pressure or diabetes). This will help to detect the illness at an early stage. You can determine the stage of your kidney illness based on the results of your blood and urine tests. A greater value indicates more significant CKD, and it measures how severely your kidneys have been damaged. The remedies for CKD cannot be cured, however medication can lessen symptoms and prevent further progression. The extent of your disease will determine how you are treated.

Machine learning may be used to predict a positive CKD status and the phases of CKD. [19] When it comes to making predictions based on historical data using classification and regression techniques, machine learning captures a significant portion of artificial intelligence. Based on various data sets, the use of machine learning techniques to predict CKD has been investigated. A benchmark dataset among them is the dataset from the UCI repository [7] (referred to as UCI dataset henceforth) [10,18]. This research considers the aforementioned benchmark dataset, much like the majority of similar studies.2]

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 10, 2022

1.1. Machine Learning

The study of computer algorithms through self-learning through data is considered a component of AI. What the ML algorithms do is generally start by collecting the data. This step is not considered as this is done before starting with Machine Learning Model. The initial step is pre-processing the data where cleaning the data, which is a process where we correct mismatched values or "NA" values, will be replaced by using some cleaning techniques. In machine learning, we train models on collected samples of information, also referred to as training models. We use the trained models there to make predictions for making decisions. Because there are two kinds of learning. Most of them are based on statistics and probability, and we use mostly predefined and self-defined algorithms to train and process the data.

It's considered to be a part of AI. Without an expression to teach, which machine learning algorithms will choose to build a model using the collected sample, we generate a hypothesis or prediction. We can use this in a variety of development applications, including computer vision applications and inbox filtering of inbox mail. As normal and natural algorithms have a hard time coping with the impracticality of a situation, we use these machine learning techniques. We have a houseful of predicting algorithms which are closely connected among themselves. Because machine learning is not only a part of statistics, but also image processing, it can identify more solutions. We have two types of machine learning algorithms, which are supervised learning for structured data where we use statistics , and unsupervised learning for unstructured data where we focus on data analysis and complete different studies on data mining. As from the above information we can clearly see how we can use machine learning aspects to improve our proposed method in terms of predicting chronic kidney disease.

It is feasible to build algorithms that instruct the device on how to carry out all of the procedures necessary to tackle the situation at hand; no computer learning is required. It may be difficult for a human to manually create the requisite algorithms for increasingly complex jobs. Supporting the computer in constructing its algorithm may be more useful than having human programmers explain each and every necessary step. Machine learning incorporates a variety of ways to teach computers to perform jobs for which there is no ideal solution. When there are several accurate answers, one option is to declare some of them valid. We have a large number of algorithms to predict using machine learning the method we have chosen is ANN as this paper aims in using neural approach we have studied about the ANN and increasing the accuracy based on the dataset.

1.2. KNN IMPUTATION

SKITKIT LEARN KNN The method of using an imputer to fill in the gaps in data is common. It frequently replaces traditional imputation techniques. In the modern world, data is gathered from many sources and used for study, insight discovery, theory testing, and other goals. Frequently, the data that was compiled from numerous sources may have some information missing. This can be the outcome of a human error that affected the data extraction or collection process. The pre-processing of data must therefore include a critical step for handling these missing values.

In this paper, we present a method for imputing missing values in a dataset using the values of nearby data points' observations. The imputation technique that one chooses is important since it might have a big influence on their study. The cause of missing values and solutions to the problem are topics covered in certain statistics literature. The ideal approach is to substitute an estimated value for these missing observations.

We employ the well-liked KNN Imputer by scikit-learn K-Nearest Neighbours Algorithm for this. Any data scientist may find missing values in a dataset to be a real hornet's nest. Since there is no simple solution to handle variables with missing values, they might provide a challenging situation. In most cases, we may simply eliminate such observations from the data if the fraction of missing observations is low in comparison to the total number of observations. However, this is not always the situation. Rows with missing values might be lost if they are deleted, along with any important data or patterns.[8]

Researchers attempted to predict the Long-Term Kidney Transplantation Outcome in a different study [4]. Ten training and validation datasets were used in the study to predict renal rejection in kidney transplant recipients. According to the experimental findings, ANN may be regarded as a helpful auxiliary algo in the process of prediction of the specified problem. In conclusion, the sensitivity of kidney rejection prediction was 39% for LR and 63% for ANN. Compared to ANN's 84% accuracy, LR's 65% specificity for predicting no-rejection.

When choosing the median of related variables in K full samples, it is preferable to pick an odd number for K since, in this situation, when the values of K are represented in order of their numeric values, the middle number is naturally the median. The choice of K shouldn't be either too big or too little. As we have seen that missing values leads to wrong direction in terms of prediction. The missing values can manipulate the win or loss of prediction. We have used KNN to fill the missing values of our dataset in the most valuable form to have a good or accurate performance of our methos.

The unobtrusive mode may go unnoticed if K is set too high, which might be crucial. In contrast, an overly low K value results in noise and irregular data has a significant negative impact on filling in the missing values. K was therefore chosen to have the values 3, 5, 7, 9, and 11 in this study.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 10, 2022

2. METHODS AND MATERIALS

A dataset and the study technique will be discussed in this section.

S.no	Attribute	Desciption
1	Age(numerical)	age in years
2	Blood Pressure(numerical)	bp in mm/Hg
3	Specific Gravity(nominal)	sg - (1.005 -1.025)
4	Albumin(nominal)	al - (0-5)
5	Sugar(nominal)	su - (0-5)
6	Red Blood Cells(nominal)	rbc - (normal , abnormal)
7	Pus Cell (nominal)	pc - (normal , abnormal)
8	Pus Cell clumps(nominal)	pcc - (present,not present)
9	Bacteria(nominal)	ba - (present ,not present)
10	Blood Glucose Random(numerical)	bgr in mgs/dl
11	Blood Urea(numerical)	bu in mgs/dl
12	Serum Creatinine(numerical)	sc in mgs/dl
13	Sodium(numerical)	sod in mEq/L
14	Potassium(numerical)	pot in mEq/L
15	Hemoglobin(numerical)	hemo in gms
16	Packed Cell Volume(numerical)	% of rbc im circulating blood
17	White Blood Cell Count(numerical)	wc in cells/cumm
18	Red Blood Cell Count(numerical)	rc in millions/cmm
19	Hypertension(nominal)	htn - (yes , no)
20	Diabetes Mellitus(nominal)	dm - (yes, no)
21	Coronary Artery Disease(nominal)	cad - (yes ,no)
22	Appetite(nominal)	ppet - (good , poor)
23	Pedal Edema(nominal)	pe - (yes,no)
24	Anemia(nominal)	ane - (yes , no)
25	Class (nominal)	class - (ckd , not ckd)

Table 1. **Characteristics in the Dataset**

2.1. DATASET

Data was collected in India over a two month period using 25 features (e.g., RBC count, WBC count, etc.).This dataset has a column with two conclusions like CKD or NOTCKD ,with more than 450 rows .

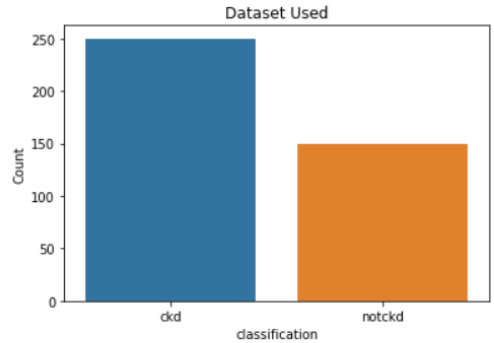

Figure 1. Distribution of target value in the Dataset

The Machine Learning Repository on the UCI website makes this dataset accessible. This dataset has 400 occurrences with 25 attributes, including 1 target attribute. Two classes on the target property are labelled to denote CKD or non CKD. Additionally, there is a value gap. Each of the 24 traits is listed in table 1 above. We have also taken the dataset from the Kaggle repository. In the dataset we used contains multivariate characteristics and we use 24 attributes and a class which totals to 25 in which there are 11 numeric representation and 14 nominal values are present. [18]

Each nominal (categorical) variable was coded to make computer processing easier. 1 and 0 were classified as Normal and abnormal, respectively, for the rbc levels and pc. Present and not present were recorded as 1 and 0, respectively, for the values of pcc and ba. Yes and no were denoted by the codes 1 and 0 respectively for dm , pe , htn, ane , and cad.

Good and bad were recorded as 1 and 0, respectively, for the value of applet. Although three variables—sg, al, and su—are categorical in nature according to the original data description, their values are still dependent on numbers, hence these variables

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 10, 2022

were handled as numbers. Each and every category variable was changed into a factor. Because it contains NaNs and needs to be converted to floating, the data needs to be cleaned. Basically, we were told to remove each row containing even one NaN, which translates to ALL ROWS with Nans, with no cut-off. The reference page will provide a link to the Kaggle page where the dataset was obtained. The dataset for chronic kidney disease was used for this study. Numerous researchers had used this data set[10,13].

The distribution has adequately covered the data of whole domain of CKD, however general characteristics like starvation, anaemia, and oedema are skewed in favour of CKD. With this data set, we can make accurate forecast, but in a wider context, it can result in false, as seen in the of Table 1. Furthermore, it was difficult to attain 100% accuracy without filling in the missing values from a collaborative imputer rather than a constant since they were entirely overlooked at random. Given the relative relevance of the traits, some of them have a stronger correlation than others based on the stage at which the patient manifests them.
The training for the models is affected by the training process. This was overcome by using KNN algorithms to fill the missing values in the data set,

2.2. METHODOLOGY

In this paper we used machine learning algorithms to develop a ML model to check CKD or not with the patient information. We have used feature scaling methods with wrapping and filtering which allows to select the important features. In this we have used classifiers like ANN classifier algorithms for training the ML models. This process includes the above stages.
(i) Dataset pre-processing(ii) Feature Selection
(iii) Proposed methodology (iv)Performing analysis

2.2.1 Data Pre-processing

Unclean data can be turned into a clean dataset using a technique called data preparation. Training is a key component of every machine learning classifier algorithm. This technique finishes duties including handling missing values, downscaling, binary data conversion, and standardising the dataset. When a group of attributes had a range of scales, rescaling was employed to scale the dataset. The binary transformation has changed the value into 0 and 1. There are only two sides that are considered for each value of the attributes either 1 or 0 with mean 0 and Deviation of 1.[3]

Data preparation is a technique that may be used to transform unclean data into a clean dataset. Every machine learning classifier algorithm requires training as a fundamental step. This method completes tasks including handling missing values, downscaling the dataset, converting it to binary data, and standardising the dataset. Rescaling is used to scale the dataset when the set of attributes had scales that varied. The value has been transformed into 0 and 1 using the binary transformation. Every attribute's value is regarded as either 1 for values above the threshold or 0 for values below the threshold. Each attribute must have a mean with a value equal to Zero and a standard deviation with a value equal to One according to the standardised technique.[8]

2.2.2 Feature Selection

Each machine learning classifier that has been trained has to employ feature selection since the results might be impacted if extraneous attributes are not removed from the dataset. A classifier method that incorporates feature selection performs better and speeds up model execution. Three distinct feature selection techniques were employed in this approach. You can see the features from the dataset section in the first page of this paper or visit Table 1.[4]

The objective of boosting is to identify new classifiers that are better equipped to forecast situations for which the current ensemble performs badly. The classifier utilises one of two methods: (a) choosing a selection of examples based on their probabilities, or (b) utilising all instances and dividing each example's error by its likelihood. The advantage of the latter strategy is that each sample is included in the training set.
Furthermore, this type of Ada-Boosting has been shown by Friedman, J. et al. to be an example of additive modelling for improving a logistic loss function.[6]

2.2.3 Proposed methodology Artificial Neural Networks

Using the framework of Artificial Neural Networks with a fold strategic validation approach is used as the classification to predict patients with CKD. Using ANN, we aim to predict if a patient has CKD or not by analysing the data of patients who have CKD. Neural networks recognise underlying relationships between the data and operate similarly to how human brains do. Many layers with both linear and non-linear functions of activation are merged to form a network model.

Combining these multiple layers with linear and non-linear activating functions, we form a resultant neural network to solve a problem. [4] These activation functions may involve large and complex problems in order to produce a result that falls within a certain range. There are two common activation functions called "relu" and "sigmoid" that are two examples of common activation mechanisms that are described below. to forecast the long-term outcome of kidney transplantation.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 10, 2022

The artificial neural network is a supervised machine learning approach (ANN). The structure of its three layers input, hidden, and output is identical to that of the human brain. In the levels of the network, ANN neurons are connected to one another, much as how all neurons in humans are interconnected.

A comparative study has been done between the ANN and LR algorithms. A comparison study has been put into place on the basis of performance indicators including accuracy, sensitivity, and specificity. Ten training and validation datasets were used in the study to predict renal disease in kidney transplant recipients. According to the experimental findings, ANN may be regarded as a helpful auxiliary algorithm used in prediction for a specific problem.

We can use this in a variety of development applications, including computer vision applications and inbox filtering of inbox mail. As normal and natural algorithms have a hard time coping with the impracticality of a situation, we use these machine learning techniques.

2.2.4 Filter Method

The filter is one tool you may use to pick the right feature. Without using any learning classifier method, it chooses the features based on their core characteristics. This approach provides results more quickly than the wrapper approach. The technique bases each attribute's score on the statistical correlation between those qualities.

2.3. Flow Chart

This flow chart here shows the procedure followed to get the desired result from stages of learning how each stage is interdependent on the other internally. This flowchart is a clear picture of the idea behind working with machine learning perspective.[9]

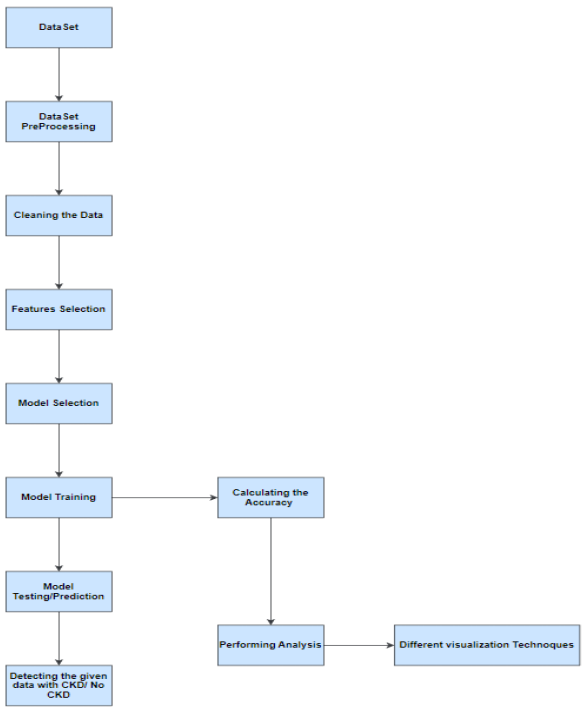

Figure 2. Proposed Method Flow to predict CKD

We can see that the above flow chart starts by collecting the dataset we have collected the dataset and started the next step as data cleaning where we have take care of the null values , but the missing values are not just removed we have used the KNN algorithm to this missing values to get the best and accurate values for the prediction using ANN. Then this follows with the feature selection which includes classifiers followed by the neural approach of training by building the model. After building the process of training takes place with
Original prediction with our trained model (see figure 6) and the analysis of accuracy will follow.

2.4. Model Algorithm

The categorization strategy is a crucial part of supervised learning. After learning from the training dataset, classifiers utilise the testing dataset to identify the desired attribute. The study's classification techniques are mentioned below.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 10, 2022

Artificial intelligence includes artificial neural networks. This kind of machine learning is supervised. Similar to the human brain, it has the same structure. The neurons in ANN are connected to one another in layers of the network, much way all of the neurons in humans are interconnected. There, neurons are referred to as nodes. An issue that has eluded human or statistical criteria can now be resolved by ANN. The input, hidden, and output layers make up an ANN. Input and weight are sent from the input layer to the hidden layer, which performs calculations and looks for hidden structures and patterns. When further layers are needed, they can be added. The output is calculated in the output layer. You can see more at reference [1].

3. LITERATURE REVIEW

The steady decrease of kidney function is referred to as chronic kidney disease, often known as chronic kidney failure. You might not have many symptoms or indicators when chronic kidney disease is first developing. It is possible that chronic kidney disease won't show symptoms until your kidney function is seriously compromised. The goal of treating chronic kidney disease is to delay the damage to the kidneys from progressing, usually by addressing the underlying cause.

This is focused on four classifying algorithms that provide predictive models for chronic renal failure, according to Anusorn Charleonnan et al. [14]. Here, the aim was to pick the most accurate classifier among the four alternatives—LR, support vector machines, decision trees, and KNN.
The predictive model was constructed using the dataset for chronic kidney disease, and performance evaluations were then carried out to identify which classifier was most efficient in accurately predicting chronic kidney disease.

M.P.N.M. Wickramasinghe et al. [15] published a research study that will offer CKD patients with an appropriate diet plan by conducting research from patient medical records and using algorithms to these records. Dr. S. Vijayarani et al discussion of a comparison of Support Vector Machines and Artificial Neural Networks, two categorization systems, is found in [16]. The objective of predicting CKD was attained based on their individual accuracies and timings. The one with better timing and precision was picked.

Ms. Astha Ameta et al [17] major focus was on data mining methods and how they may be used to forecast chronic renal illnesses. They demonstrated that data mining was a more effective method for forecasting chronic renal illnesses as a result.

Applying machine learning techniques to address issues in Health care is not a new industry. Several scientists have attempted that. Various strategies and approaches have been employed for the patient's state for CKD categorization and prediction. A decision support system for diagnosis has been developed. You use the random data to forecast chronic kidney failure categorization of subspaces. They employed strategies like K-nearest Nave Bayes and K nearest neighbour (KNN) for predicting the existence of CRF [5].

Performance on an individual basis and group learning used to anticipate the development of chronic kidney disease they are targeted and work at predicting renal disease with machine learning algorithms are used. In a study report, scientists have examined three distinct prediction methods that had The best models with 10 parameters are continuous and categorical. eight predictors and a category model that is more straightforward predictive tools.[7]

Progressive decrease of kidney function over time is a feature of chronic renal disease. Since the majority of victims show no symptoms, it is a quiet illness. The medical community has severe difficulty in the early detection and treatment of CKD, and they turn to machine learning theory to provide an effective answer.

4. TRAINING AND ANALYSING SEQUENTIAL MODEL

4.1 PERFORMANCE EVALUATION MEASURE

The ML algorithm used for predicting the CKD over the given information is ANN. After the construction of neural networks, the next step is to train the model we have created with the data we have. The objective of enrolling our model to train is to minimise the number of corrections or mistakes and improve accuracy. In this state, we split our training population with an "80%-20%" testing size.

We have used multiple evaluation matrices in order to identify the performance of the proposed model, hence we got the below confusion matrix. It generally is two x two matrix which is because of the dataset we used. This confusion matrix includes two positive outcomes and two negative outcomes. The Confusion matrix is presented in above figure 5.

4.1.1 CONFUSION MATRIX DESCRIPTION

TP: True Positive indicates that the projected outcome was accurately categorised.

True Negative (TN) output signifies that the projected outcome is accurately categorised.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 10, 2022

FP: False Positive indicates that the projected result was erroneously categorised.

FN: False Negative indicates that the projected outcome was improperly categorised.

4.1.2 CLASSIFICATION RELIABILITY

Classification accuracy demonstrates the success rate of predictions. The confusion matrix is used to compute it. Equation 1 yields the detection accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (1)$$

4.1.3 CLASSIFICATION ERROR

Classification error displays the percentage of predictions that are inaccurate. From the confusion matrix, it computes. Equation 2 determines the categorization error.

$$\text{Error} = \frac{FP+FN}{TP+TN+FP+FN} * 100 \quad (2)$$

4.1.4 PRECISION

A key component of the model performance evaluation matrix is precision. The percentage of linked occurrences among all retrieved instances is what this term refers to. It is a forecasted value that is positive. Equation 3 calculates the accuracy as follows.

$$\text{Precision} = \frac{TP}{TP+FP} * 100 \quad (3)$$

4.1.5 RECALL

Another crucial component of the model performance assessment matrix is recall. It represents the proportion of related occurrences among all retrieved instances. Equation 4 calculates the recall as follows.

$$\text{Recall} = \frac{TP}{TP+FN} * 100 \quad (4)$$

4.1.6 F-MEASURE

It also goes by the name F Score. In order to gauge test accuracy, the F-measure is calculated. Equation 5 calculates it based on the accuracy and recall.

$$\text{F-Measure} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

Every time the entire training set is used to update the weights, it is referred to as an epoch. Using this method, we can predict the CKD of the patients in the supplied dataset with an accuracy of 96%. The term "training the network" refers to the process of refreshing or updating the weights of the nodes.

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 10, 2022

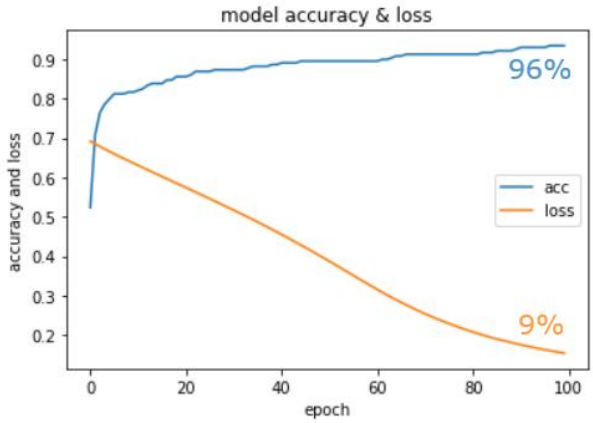

Figure 3. Output of Sequential MODEL

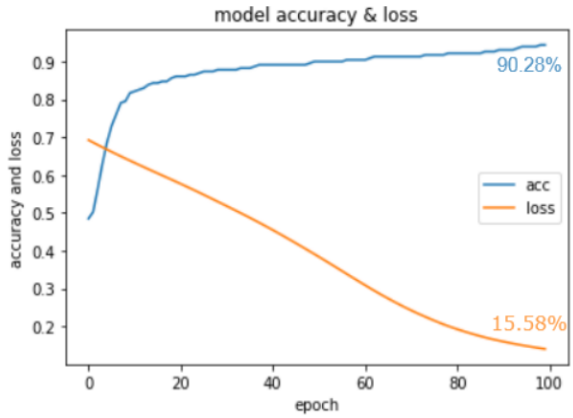

Figure 4. Output of Standard ANN Model

The above graphs are the comparison between the standard ANN Model with the enhanced Sequential model that we have used to get higher accuracy i.e., when tested with 100 epochs to achieve an accuracy of 96% with a loss of 9% is the result of our model. The result of ANN model was 90% accuracy with 15% loss. The above results were accurate with the feature selection and training we did as a team to better the performance.

	precision	recall	f1-score	support
0	0.96	1.00	0.98	25
1	1.00	0.98	0.99	55
accuracy			0.99	80
macro avg	0.98	0.99	0.99	80
weighted avg	0.99	0.99	0.99	80

Accuracy of Sequential model is :0.9655

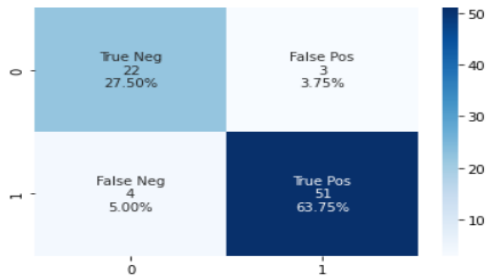

Figure 5. Classification Report for the Sequential model

This report is based on 100 epochs that we have used to attain this above classification.

5. EXPERIMENTAL SETUP AND RESULTS

For the predictions we made to examine the CKD, the trained ANN referenced CKD model has an accuracy of 100% and a loss of 0.09%. We, as a team, employed Keras and the Tensor flow libraries for ANN to build the method depicted in Fig 2.To boost accuracy, the ANN we utilise here is trained over 2000 epochs. This method aims to offer a neural network-based answer to the

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 11, Iss 10, 2022

machine learning problem. We handled this challenge as a team utilising data and produced accurate outcomes with minimal loss. We have used the Google Collab and Jupiter notebook to establish the environment setup for the study of this experiment.

```
Model file:  ckd.model
2/2 [==============================] - 0s 7ms/step
2/2 [==============================] - 0s 6ms/step - loss: 0.0090 - accuracy: 1.0000

Original  : 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0,

Predicted : 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0,

Scores    : loss =  0.009027427993714 81  acc =  1.0
------------------------------------------------------------
```

Figure 6. Testing and Prediction of Sequential Model

6. CONCLUSION

With the research method we have proposed in a specific way to provide a prediction for CKD by utilising the concepts of artificial neural networks, we have achieved high accuracy and minimal loss. So, after doing this research, we conclude that with ANN we can get much more accurate results and without having to worry about the loss by having the right machine learning algorithm like ANN. In comparison to previous research using the same dataset, filling in missing values based on their distribution and the collocation of other variables using K Nearest Neighbours'-impute results in prediction models that are more accurate. Additionally, compared to other models, the additional trees classifier and the random forest classifier are superior algorithms for making predictions for CKD because they have 100% overall accuracy and little bias toward certain features. In order to forecast CKD status, this study proposes a novel approach that includes data preparation, managing missing values, and feature selection.

REFERENCES

[1] Chittora, P., Chaurasia, S., Chakrabarti, P., Kumawat, G., Chakrabarti, T., Leonowicz, Z., … Bolshev, V. (2021). Prediction of Chronic Kidney Disease - A Machine Learning Perspective. IEEE Access, 9, 17312– 17334. doi:10.1109/access.2021.3053763

[2] Ekanayake, I. U., & Herath, D. (2020). Chronic Kidney Disease Prediction Using Machine Learning Methods. 2020 Moratuwa Engineering Research Conference(MERCon).doi:10.1109/mercon50084.202 0.9185249 10.1109/mercon50084.2020.9185249

[3] Charleonnan,A.,Fufaung,T.Niyomwong,T.Chokchueypattanakit,W.Suwannawach,S&Ninchawee(2016). Predictive analytics for chronic kidney disease using machine learning techniques. 2016 Management and Innovation Technology International Conference (MITicon). doi:10.1109/miticon.2016.8025242

[4] A.Wosiak and D. Zakrzewska, ''Integrating correlation-based feature selection and clustering for improved cardiovascular disease diagnosis,'' Complexity, vol. 2018, Oct. 2018, Art. no. 2520706

[5] A. R, G. Sasi, R. Sankar, and O. . Deepa, "Decision Support system for diagnosis and prediction of Chronic Renal Failure using Random Subspace Classification," 2016.

[6] D. S. Sisodia and A. Verma, "Prediction Performance of Individual and Ensemble learners for Chronic Kidney Disease," 2017.

[7] A. V Kshirsagar et al., "A Simple Algorithm to Predict Incident Kidney Disease," ARCH Intern Med, vol. 168, no. 22, pp. 2466– 2473, 2008.

[8] Gunarathne, W. H. S. ., Perera, K. D. ., & Kahandawaarachchi, K. A. D. C.(2017). 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 (BIBE). doi:10.1109/bibe.2017.00-39

[9] R. Kei Chiu, R. Y. Chen, S. Wang and S. Jian, "Intelligent systems on the cloud for the early detection of chronic kidney disease", Machine Learning and Cybernetics IEEE, pp. 1737-1742, July 2012.

[10] Akash Maurya, Rahul Wable, Rasika Shinde , Sebin John , Rahul Jadhav, Dakshayani.R, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet plan by using Machine Learning" IEEE 10.1109/ICNTE44896.2019.8946029

[11] M. S. Gharibdousti, K. Azimi, S. Hathikal, and D. H. Won, ''Prediction of chronic kidney disease using data mining techniques,'' in Proc. Ind. Syst. Eng. Conf., K. Coperich, E. Cudney, H. Nembhard, Eds., 2017, pp. 2135–2140.

[12] T. F. T. N. W. C. Anusorn Charleonnan, "Predictive Analytics for Chronic Kidney Disease," in The 2016 Management and Innovation Technology International Conference (MITiCON-2016), 2016.

[13] D. P. a. K. K. M.P.N.M. Wickramasinghe, "Dietary prediction for patients with Chronic Kidney Disease (CKD) by considering blood potassium level using machine learning algorithms," in 2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, Australia, 2017.

[14] M. Dr. S. Vijayarani, "data mining classification algorithma for kidney disease prediction," International Journal on Cybernetics & Informatics (IJCI), vol. 4, no. 4, p. 13, August 2015.

[15] M. K. J. Ms. Astha Ameta, "Data Mining Techniques for the Prediction of Kidney Diseases and Treatment: A Review," International Journal Of Engineering And Computer Science , vol. 4, no. 2, p. 20376, february 2017.

[16] J. Xiao, R. Ding, X. Xu, H. Guan, X. Feng, T. Sun, S. Zhu, and Z. Ye, "Comparison and development of machine learning tools in the prediction of chronic kidney disease progression," Journal of translational medicine, vol. 17, no. 1, p. 119, 2019.