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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal ⁜ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121001

A Deep Belief Neural Network Model for Predicting the Early Phases of Chronic Kidney Disease

Aloy-Okwelle C[1], O.T. Olise [2], O.P Nweke[3]

Department of Computer Science, Rivers State University, Port Harcourt, Nigeria[1,2,3]

Abstract: Chronic Kidney Diseases are undoubtedly one of many fatal diseases that are very difficult to diagnose. With early diagnosis and treatment, it is possible to slow or stop the progression of chronic kidney disease (CDK). However, the most basic kidney screening tests available, such as the Blood Urea Nitrogen (BUN)/creatinine ratio test, cost $25 or more. This disease remains a leading cause of end-stage kidney disease thus requiring renal replacement therapy with dialysis or kidney transplantation. In this work we developed a model to predict the various stages of kidney diseases using deep belief neural network using clinical data. The model was trained using dataset containing 400 patient records. The attributes used for building our model includes; age, blood pressure, albumin, red blood cells, pus cells, pus cells clumps, hemoglobin, White blood cell count, Red blood cell count, etc. Object-oriented design methodology was used for system development. Localization model was used to identify quality kidney images. Deep belief neural network was used to analyze the image of the kidney and predict the status of kidney health. System was implemented in python programming language. The system successfully classified kidney disease dataset into CKD and non-CKD with 98% overall accuracy when the model was tested with a set of data that were not used during the training process.

Keywords: Deep neural networks, chronic kidney disease, supervised machine learning.

1. INTRODUCTION

Chronic kidney disease (CKD) is a type of kidney disease in which there is a gradual loss of kidney function over a period of months to years. Initially there are generally no symptoms; at a later stage, the symptoms may include swelling of the legs, fatigue, nauseating, and loss of appetite. Further complications may include an increased risk of heart disease, high blood pressure, bone disease, and anemia (Chase, 2014).

Causes of chronic kidney disease include diabetes, high blood pressure, glomerulonephritis, and polycystic kidney disease. Diagnosis of CKD is by blood tests which is done to measure the estimated Glomerular Filtration Rate (eGFR), and a urine test to measure albumin. Ultrasound or kidney biopsy may be performed to determine the underlying cause. Several severity-based staging systems are in use (Chew & Pai, 2019). Purusothaman & Krishnakumari, (2015) indicated that kidney failure falls among several classes of disease such as heart disease and blindness. Dialysis and transplant are the only method to keep the kidneys function artificially and it is also a painful and expensive process. According Luyckx, & Stanifer, (2018) kidney disease increased globally from 19 million in 1990 to 33 million in 2013 in 2016 and in 2010 2.62 million people received dialysis worldwide and the need for dialysis is projected to double by the year 2030.

Between 8 and 10 percent of the world's adult population have some form of kidney damage, and every year millions die prematurely of complications related to chronic kidney diseases (CKD) (KDSUS & UDIC, 2012). With early diagnosis and treatment, it is possible to slow or stop the progression of chronic kidney disease. However, the most basic kidney screening tests available, such as the Blood Urea Nitrogen (BUN)/creatinine ratio test, cost $25 or more (FLT, 2016). The high cost affects the frequency of testing among the lower income population, making early detection less likely. Patients get tested only after they start experiencing symptoms and this can result in late diagnosis.

In Nigeria, the situation is such that Kidney Disease (KD) represent about 8% to 10% of hospital admissions (Ulasi & Ijoma, 2010). By combining machine learning and statistical analysis, useful information can be drawn from medical databases. Levey et al. (2007) carried out a Modification of diet in renal disease (MDRD) equation, which is used to determine kidney health indicator called estimated glomerular filtration rate. The MDRD equation accounts for factors such as age, gender and ethnicity, which results in a better estimate of effective renal function. The equation gives a value in the range of 0–150, where 60 and above indicate a healthy kidney, and patients with values between 0–15 require immediate dialysis or kidney transplant. This study focuses on developing a deep neural network model that can predict the early stages of kidney disease using clinical data. The data considers age, blood sugar red blood cell count among others. Deep belief neural network will be used to analyze the image of the kidney and predict the status of kidney health.

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal ⁜ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121001

II. RELATED WORK

Kriplani et al. (2019) studied 224 records of chronic kidney disease available on the UC Irvine (UCI) machine learning repository named chronic kidney diseases dating back to 2015 and proposed an algorithm. Their proposed method 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, the model built shows better results which is implemented using the cross-validation technique to keep the model safe from over fitting.Chimwayi et al. (2017) applied neuro-fuzzy algorithm to determine the risk of CKD patients, using a dataset which had 25 attributes (14 nominal and 11 numeric) describing early stages of CKD in Indians, with accuracy of 100%, sensitivity of 100% and specificity of 97%, they suggested that the work should be added to the domain of healthcare and can be used for providing suggestions in the domain by making it easy for healthcare professionals in diagnosis and treatment of patients as well as for identifying relationships within diseases suffer by patients.

Arafat et al. (2018) studied an automated detection of CKD with clinical data using Random Forest (RF), Naïve Bayes (ND) and Decision Tree based on a comparative study on UCI dataset, they computed the weight of each attribute used in the dataset. Their result shows that RF has higher accuracy of 97.5% and 96% for Naïve Bayes and Decision Tree in the prediction of CKD.

Aljaaf (2018) in their study examined four machine learning (ML) models for early prediction of CKD, which were: support vector machine (SVM), classification and regression tree (CART), logistic regression (LR), and multilayer perceptron neural network (MLP). By using the CKD dataset from UCI and seven features out of 24, they compared the performance of these ML models. The results showed that the MLP model had the highest AUC and sensitivity. It was also noticeable that logistic regression almost had the same performance as MLP but with the advantage of the simplicity of the LR algorithm.

Shafi et al. (2020) Proposed a machine learning based solution to avoid cleft in the mother's womb with Deep Leaning method and other four methods, on 1000 pregnant female samples from 3 different hospitals in Lahore, Punjab. The authors performed data cleaning, scaling and feature selection method and compared the accuracy for all the algorithms with Random Forest(RF) algorithm:85.77%, Decision Tree(DT):88.14%, K-Nearest Neighbor (KNN):89.72%, Support vector Machine(SVM):90.69% and Multilayer perception(MLP) which is a Deep Neural Network:92.6% and indicated that MLP yield a better accuracy.

Polat et al. (2017) used SVM algorithm along with two feature selection methods: filter and wrapper to reduce the dimensionality of the CKD dataset with two different evaluations for each method. For the wrapper approach, the ClassifierSubsetEval with the Greedy Stepwise search engine and WrapperSubsetEval with the Best First search engine were used. For the Filter approach, CfsSubsetEval with the Greedy Stepwise search engine and FilterSubsetEval with the Best First search engine were used. However, the best accuracy was 98.5% with 13 features using FilterSubsetEval with the Best First search engine using the SVM algorithm without mentioning which features were used.

Yildirim (2017) studied the effect of sampling algorithms in predicting chronic kidney disease. The experiment was done by comparing the effect of the three sampling algorithms: Resample, SMOTE, and Spread Sup Sample on the prediction by multilayer perceptron classification algorithm. The study showed that sampling algorithms could improve the classification algorithm performance, and the resample method has a higher accuracy among the sampling algorithms. On the other hand, Spread Sub Sample was better in terms of execution time.Xiao, J. (2019) compared nine ML models, including LR, Elastic Net, ridge regression lasso regression SVM, RF, XGBoost, k-nearest neighbor and neural network to predict the progression of CKD. They used available clinical features from 551 CKD follow-up patients. They conclude that linear models have the overall predictive power with an average AUC above 0.87 and precision above 0.8 and 0.8, respectively.

Wang & Chung (2018) downloaded the data from the UCI machine learning repository for diagnosing CKD. Correlation-based feature selection is used to select initial features. Different classifiers like Adaboost, SVM, K-NN and Navie Bayes were used. Four parameters where used to predict the accuracy of each classifier namely accuracy, precision, recall and f-measure. K-NN achieved the best result with a 98.1% accuracy rate.

In the study carried out by Gunarathne et al. (2017), Microsoft Azore has been used to predict the patient status of CKD. By considering 14 attributes out of 25, they compared four different algorithms, which were Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Decision Regression, and Multiclass Neural Network. After comparison, they found that Multiclass Decision Forest performed the best with 99.1% accuracy.

DOI: 10.17148/IJARCCE.2023.121001

Arif-UI-Islam (2019) proposed data mining techniques like boosting and rule extraction for analyzing kidney disease. Adaboost and LogitBoost algorithms are used for comparing the performance of classification and Ant-Miner is used to decline the rules. The proposed work has two aims: used a boosting algorithm for analyzing the performance to detect the CKD and deriving rules.

Chew & Pai (2019) predicted chronic kidney disease using ultrasound imaging techniques. This technique is invasive, low-cost. The dataset was collected from Keelung Chang Memorial Hospital in Taiwan. The preprocessing technique called image in painting and median filter was applied to remove noise. The feature extraction technique was applied to take out important features from the image and finally, it is classified using the SVM model.

III. METHODOLOGY

The system uses Term Frequency-Inverse Document Frequency (Tf-Idf) to extract text feature in big data and Deep Belief Neural Network (DBNN) to train and categorize the phases of kidney disease. It can learn continually which is very desirable in classifying and estimate missing data. Figure 1 represents architectural design of deep belief neural network system

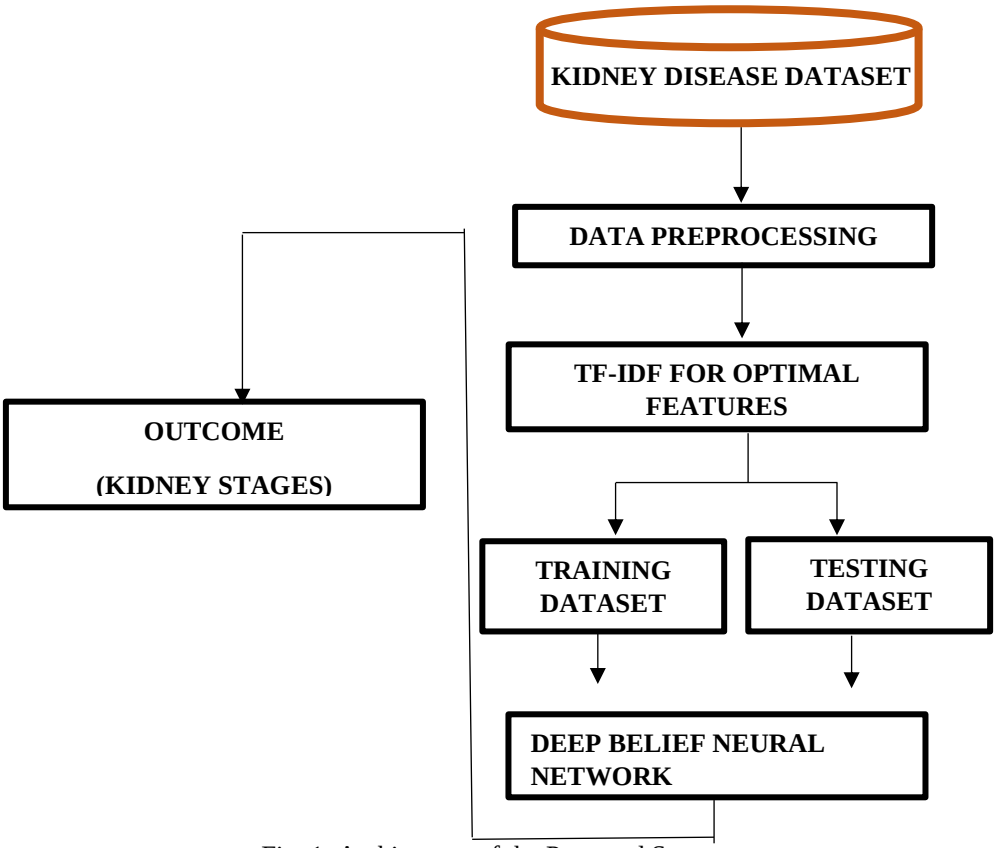

Fig. 1: Architecture of the Proposed System

In data preprocessing, the raw data was transformed to a clean collection of data. In the extraction of text features, the TF-IDF checks how many times a word appears in the document; in order to arrange multiple symptoms. The Deep Belief Neural Network trains and classifies kidney disease into categories.

A. Dataset

The dataset was gotten from UCI repository. The dataset contains 400 patient records. The data set contains a total of 25 attributes. However, 14 attributes was used for building our model. The attributes include Age, Blood pressure, Albumin, Red blood cells, Pus cell, Pus cells clumps, Serum creatinine, Hemoglobin, White blood cell count, Red blood cell count, Anemia, Classification, Appetite, Packed cell volume all this 14 attributes are used to build model.

B. Data Preprocessing

The data gotten from UCI repository did not contain the name of the attribute so we first assigned the names to the attribute. Missing values in the dataset like NA's or blank values where removed and replaced with the mean values of that attribute.

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal ⁜ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121001

Table 1: Attribute of Kidney Disease Dataset

S|N	Attribute	Value Used
1	Age	Discrete Integer Values
2	Blood Pressure	Discrete Integer Values
3	Albumin	Nominal Values
4	Red Blood cells	Nominal Values(Normal, Abnormal)
5	Pus cell	Nominal Values(Normal, Abnormal)
6	Pus cells clumps	Nominal Values(Present, Not-Present)
7	Serum creatinine	Numeric Values
8	Haemoglobin	Numeric Values
9	White blood cell count	Discrete Integer Values
10	Red blood cell count	Numeric Values
11	Anaemia	Nominal Values(Yes, No)
12	Classification	Nominal Values(CKD, Not CKD)
13	Appetite	Nominal Values(Good, Poor)
14	Packed cell volume	Discrete Integer Values

C. Feature Extraction

Text feature extraction was done by taking out the list of words from the text data and then transforming them into a feature set which is usable by a classifier.

D. Term Frequency-Inverse Document Frequency (TF-IDF)

In TF-IDF, the frequency of the words is rescaled by considering how frequently the words occur in all the documents. Due to this, the scores for frequent words are also frequent among all the documents are reduced. This way of scoring is known as Term Frequency – Inverse Document Frequency (Waykole and Thakare, 2018). Term Frequency (TF) is the frequency of the word in the current document. Inverse Document Frequency (IDF) is the score of the words among all the documents.

These scores can highlight the words that are unique that is the words that represent needful information in a specified document. Therefore, the IDF of an infrequent term is high, and the IDF of a frequent term is low. Suppose we have a document (or a collection of documents), and we want to summarize it using a few keywords. In the end, we want some method to compute the importance of each word. Term Frequency-Inverse Document Frequency (TF-IDF) is given as:

$$tf - idf(t, D) = tf(t, d) \,.\, idf(t, D) \qquad (3.3)$$

$$tf(t, d) = f_{t|d} = \frac{number\ of\ time\ t\ appears\ in\ a\ document}{total\ number\ of\ terms\ in\ the\ document}$$

$$idf(t, D) = \log\left(\frac{N}{number\ of\ documents\ with\ t\ in\ it}\right)$$

Where,
N is the total number of documents

Considering two documents D_1 and D_2:
D_1 contains “Appetite is good”.
D_2 contains “Appetite is bad”.
Here,
N = |D|= 2

Table 2: TF-IDF for Text Extraction

Text	tf		idf	tf-idt	
	A	**B**		**A**	**B**
Appetite	1	1	Log(2/2)	0	0
Is	1	1	Log(2/2)	0	0
good	1	1	Log(2/2)	0	0
bad	0	1	Log(2/1)	0	Log(2)

The only thing that differentiates D1 and D2 is the word “bad”. The TF-IDF reflects that “bad” is really important in terms of IDF given as log(2) for the word “bad”.

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal ⁜ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121001

E. Training and Testing Dataset

The dataset was divided into two sub datasets both containing 14 attributes.

Training data: training dataset is derived from main dataset and it contains 300 out of 400 records in main dataset of CKD.

Testing data: testing dataset is of 100 out of 400 records from main CKD dataset.

F. Deep Belief Network (DBN) for Classification

A Deep-Belief Network (DBN) is a generative graphical model or alternatively deep-neural networking groups consisting of several layers of latent variables (the "hidden units"). The two top levels of DBN, which are associative memories, are symmetrically connected. The input data is obtained in the lowest layer or the observable units. Integer or actual input details should be. The intra layer connections like the RBM are not provided. Hidden units are software which gathers correlations in the data. The symmetrical weight matrix W connects two layers. In each layer each device in each next layer is connected to another level as shown in Figure 2.

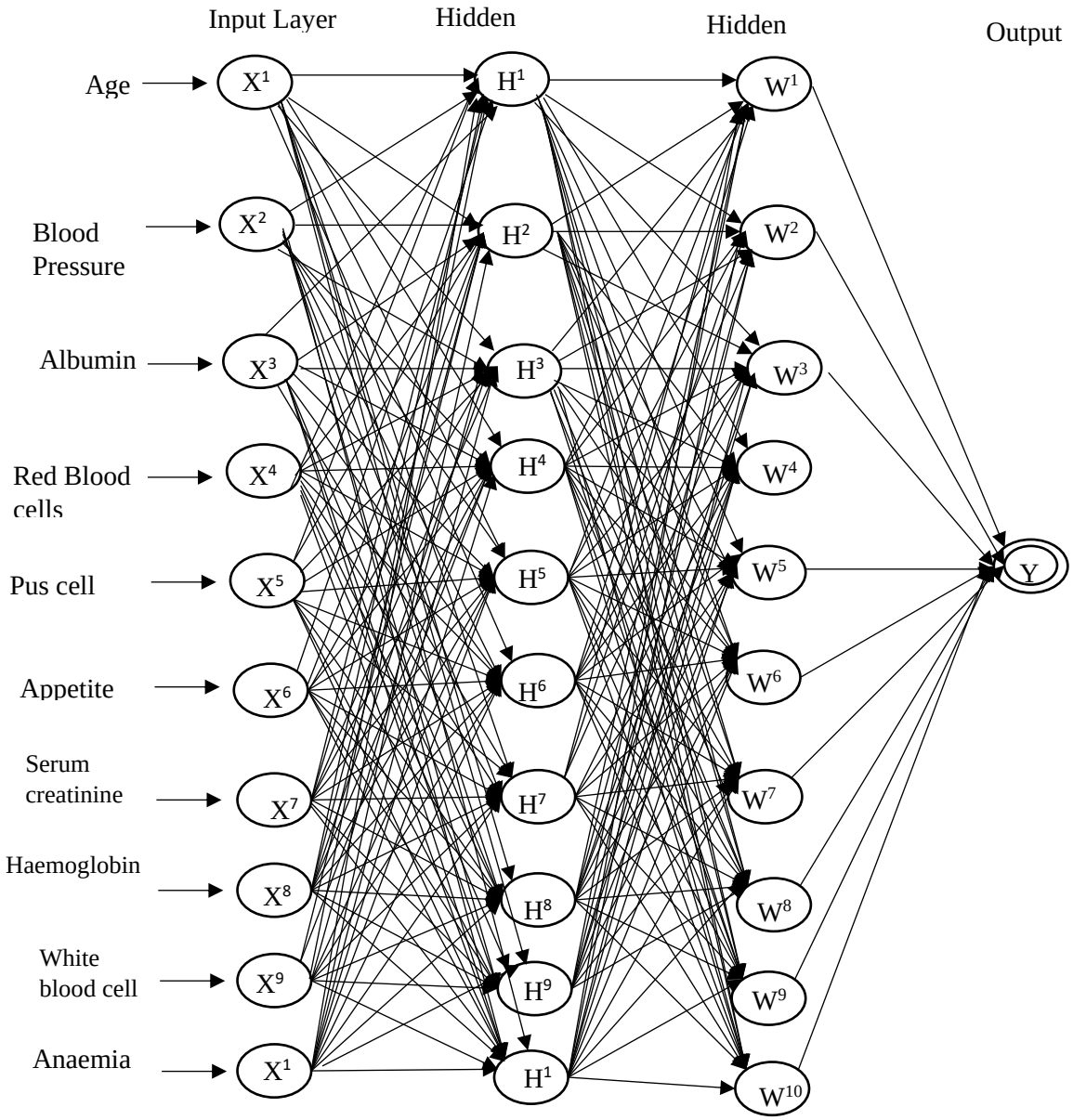

Fig. 2: Model architecture of Deep Belief Network

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal ⁜ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121001

DNN is a subset of Artificial Neural network which simulate the structure and functionalities of biological neural network consisting of an input, weights and activation function, the structure of DNN has an input, hidden layer and an output.

In the model, Xi is the input, Wi and Hi are the weight and Y referred to the output. Input layer is set of neurons which take the actual input data for processing. The number of neurons can be decided in accordance with the input data. To train the model, three layers where used. For input and the hidden layer, "Relu" is used as activation function. Relu outputs in 0 or 1. The output layer has only one output result either kidney disease or not kidney disease with Sigmoid as activation function.

IV. RESULT

The main goal of this project is to discover an effective indicator which can assist physicians to identify different stages of CKD. The dataset contains 24 features including the class (CKD or NOT CKD) and stages (severity, moderate and mild). Some features used in this study are age, blood pressure, albumin, sugar, senum creatinine, red blood cells, blood glucose random, blood urea, pus cell, pus cell clumps, white blood cell count, haemoglobin, sodium, potassium, diabetes mellitus, coronary artery disease, appetite.

The features in the dataset that has more influence in CKD prediction from the 24 attributes are white blood cell count, senum creatinine, BP, blood urea and blood glucose as shown in table 3.

The patients with earlier stages of CKD possess higher values of white-to-black transition indicator than the patients at his/her ending stages.

Table 3: Stages of CKD

BP	Blood Urea	Blood Glucose	Senum Creatinine	White blood cell count	Stages
100	40	252	3.2	26400	severity
90	89	139	3	9700	moderate
70	31	75	1.2	7800	mild
80	16	85	1.1	5800	none

The BP for ckd is more variant and lie between 60-100 compared to non ckd as shown in Figure 4.1.
The albumin content of non ckd lies less than 1 whereas the albumin content of ckd is more dispersed upto 5 as shown in Figure 4.2.
The blood glucose level for non ckd lies 50 to 150 whereas less ckd suffers are having level in this normal range as shown in Figure 4.3. Ckd sufferers are having more dispersed data.
Similarly, for blood urea, only few ckd sufferers are having blood urea in normal range as non ckd sufferers as shown in Figure 4.4.
Maximum ckd sufferers are having serum creatinine in range of non ckd sufferers.
The sodium and potassium levels of ckd and non ckd are in approximate same range
The haemoglobin levels for non ckd is more in range of 12.5-17.5 than ckd in range of 5-17.5 having dispersed data.
The packed cell volume of ckd lies low in normal range i.e. 40-60.
The white blood count of non ckd and ckd are in approximately same range with approximately same dispersed data.
The red blood count of non ckd is more in range of 4-7 whereas ckd sufferers are having less red blood count in range of 3-6.

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal ⁜ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121001

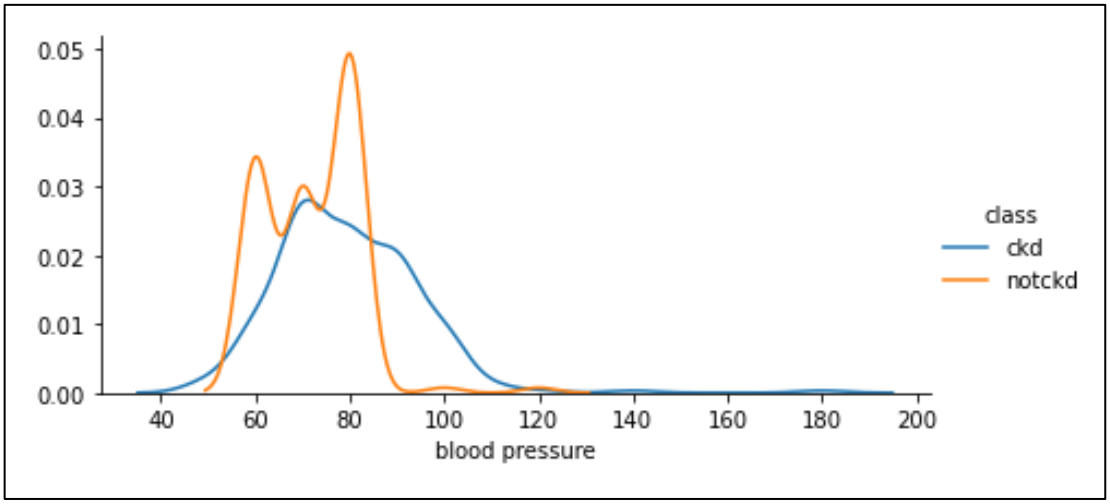

Fig. 3: Blood Pressure associated with CKD or NotCKD

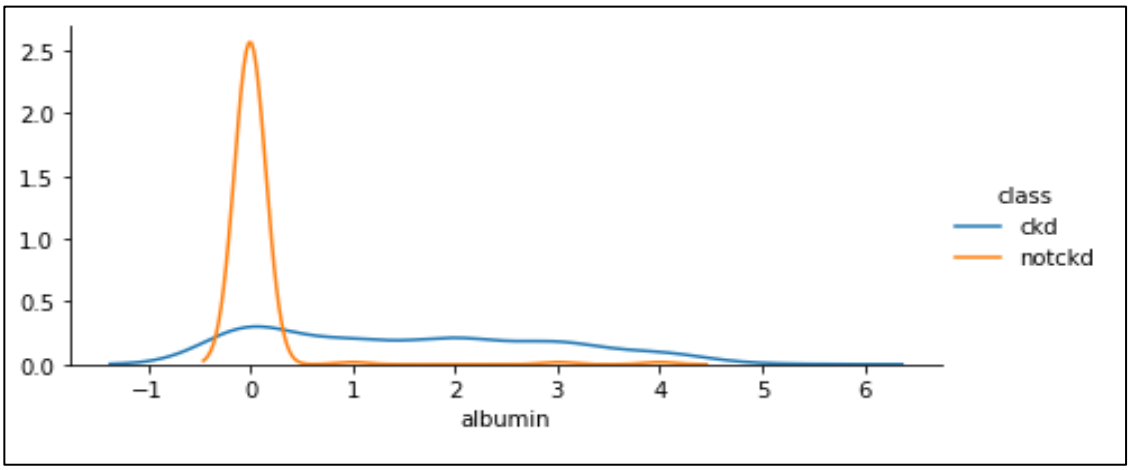

Fig 4: Albumin associated with CKD or NotCKD

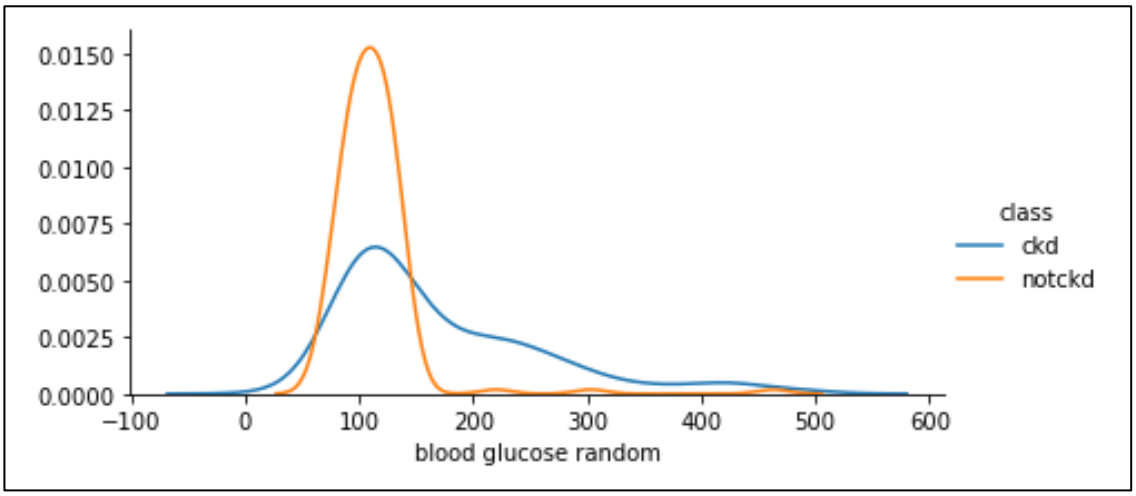

Fig 5: Blood Glucose associated with CKD or NotCKD

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal ⁜ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121001

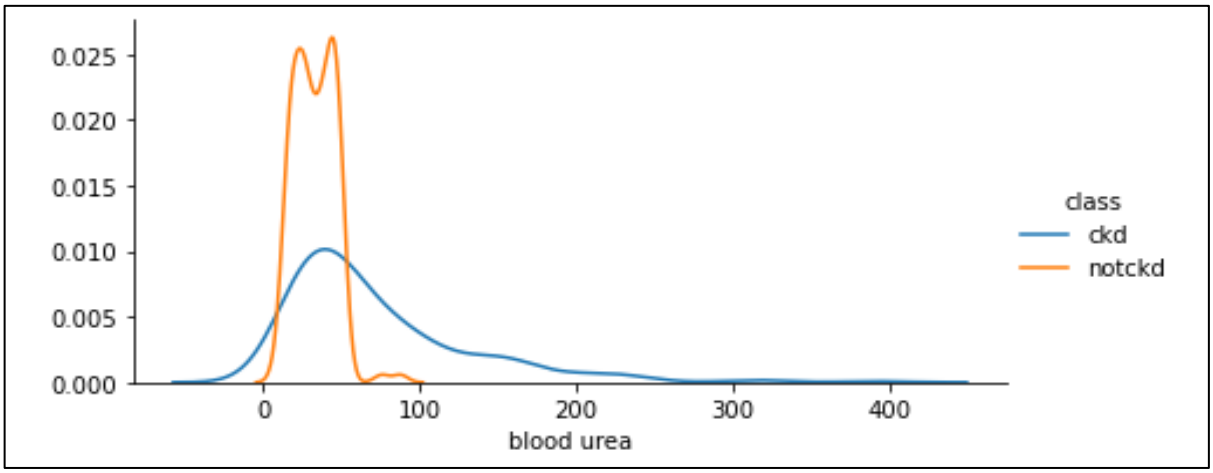

Fig 6: Blood Urea associated with CKD or NotCKD

V. DISCUSSION OF RESULT

Data pre-processing was done which involved the changing the names and the columns, data cleaning involved removing the nan values, data visualization was done to study the skewness, outliers, balance or imbalance data, correlation and distribution of data. After that, feature selection was done using SelectKBest after applying the label encoding for categorical data. This lead to selection of only top 10 columns based on their feature scores. Function file was made separately which is called later in main file. Using the deep neural network algorithm, a prediction of model was built.

The Deep Neural Network (DNN) can outperform the SVC and Naïve Bayes Classifier in the task of diagnosing chronic kidney disease. The DNN was able to achieve a highly acceptable testing score as compared to the Naïve Bayes Classifier is 100% and the Logistic Regression and SVM attained a score of 96% and 82% respectively. The generated confusion matrix illustrates the efficient classification performance of the generated model. A Deep Neural Network model is highly robust to the fluttering environment which makes it immune to noise. The precision score for the model came out to be 1.0 which implies that the model is highly efficient in determining the percentage of the Positive predicted cases for patients suffering from CKD out of the total positive result. For 'notckd' class we obtained a value of 1.0 which was the most optimal. It illustrates the True Negative rate of the model. To evaluate the model's performance keeping in mind the precision and the recall trade-off, the F1 score was evaluated which gave the harmonic mean of the precision and recall values. F1 Score for the proposed model is 100% and the Final testing accuracy received from the same is 100%.

VI. CONCLUSION AND RECOMMENDATION

Conclusion

Diagnosing CKD is a cumbersome task as there are no major symptom that can be used as a benchmark to detect the disease. In cases when diagnosis persists, some results may be interpreted incorrectly. This paper proposes a Multi-Layered Perceptron Classifier that used deep neural network to predict if a patient has CKD or not, and the stages of kidney disease. The model was trained on a dataset of about four hundred patients. The attributes of the dataset includes; blood pressure, age, sugar level, red blood cell count, etc. The experimental results display that the proposed model can perform classification with the testing accuracy of 98 %. The most encouraging aspect of this system is its evident capability to resolve the invariant problems despite all the adversities and difficulties.

Recommendation

The application of deep learning approaches for CKD prediction is still subject to further research and analysis, despite the potential contribution made in this study. Incorporating more data to have a larger dataset will provide more accuracy and efficiency; hence more dataset is needed to accommodate enough samples. Having enough sample will make the prediction wider to capture and identify regions and areas with CKD vulnerability.

Contribution to Knowledge

The system successfully classified kidney disease dataset into CKD and non-CKD with 98% overall accuracy when the model was tested with a set of data that were not used during the training process. The adopted DNN model proved to be efficient and suitable for the prediction of kidney disease. The study also highlighted the importance of the features used

DOI: 10.17148/IJARCCE.2023.121001

in the prediction of kidney disease. This revealed that from the 10 attributes, Creatinine and Bicarbonate are the attributes with highest influence on CKD.

REFERENCES

[1] Aljaaf, A. J. (2018). Early prediction of chronic kidney disease using machine learning supported by predictive analytics. In *IEEE Congress on Evolutionary Computation (CEC).*

[2] Ani, R. & Sasi, G. (2016). Decision Support System for Diagnosis and Prediction of Chronic Renal Failure using Random Subspace Classification. *International Conference on Advances in Computing, Communication and Informatics*, 21-24.

[3] Arafat, F., Fatema, K. & Islam, S. (2018). Classification of chronic kidney disease (ckd) using data mining techniques. *Doctoral dissertation, Daffodil International University.*

[4] Arif-UI-Islam, S. H. R. (2019). Rule Induction and Prediction of Chronic Kidney Disease using Boosting classifiers, Ant- Miner and J48 Decision Tree. *International Conference on Electrical, Computer, and Communication Engineering*, 7-9.

[5] Chase, H. S. (2014). Presence of early CKD-related metabolic complications predict progression of stage 3 CKD: a case-controlled study. *BMC Nephrology,* 15(1), 187.

[6] Chew, C-J. & Pai, J-W. (2019). Prediction of Chronic Kidney Disease stages by Renal ultrasound imaging. *Enterprise Information Systems*, 26, ISSN 1751-7575.

[7] Chimwayi, K. B., Haris, N., Caytiles, R. D. & Iyengar, N. C. S. N. (2017). Risk Level Prediction of Chronic Kidney Disease Using Neuro- Fuzzy and Hierarchical Clustering Algorithm (s). *International Journal of Multimedia and Ubiquitous Engineering,* 12(8), 23-36.

[8] 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.

[9] Gunarathne, W., Perera, K. & Kahandawaarachchi, K. (2017). Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD). In *Bioinformatics and Bioengineering (BIBE), IEEE 17th International Conference*.

[10] Han, J., Pei, J. & Kamber, M. (2012). Data Mining: Concepts and Techniques. *Waltham, MA:*

[11] Kidney disease statistics for the United States and U.S. Department of Health and Human Services. (2012). National Institute of Diabetes and Digestive and Kidney Diseases, 3 Kidney Information Way, Bethesda, MD 20892, *Tech. Rep*., 6, 12-3895.

[12] Kriplani, H., Patel, B., & Roy, S. (2019). Prediction of Chronic Kidney Diseases Using Deep Artificial Neural Network Technique. In *Computer Aided Intervention and Diagnostics in Clinical and Medical Images Springer, Cham. Chicago*, 179-187.

[13] Krishna, K. D. & Akkala, V. (2016). Computer-Aided Abnormality Detection for Kidney on FPGA Based IOT Enabled Portable Ultrasound Imaging System. *Science Direct*, 189-197.

[14] Levey, A. S., Coresh, J., Balk, E., Kausz, A. T., Levin, A. & Steffes, M. W. (2003). National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Annual International Medical*, 139(2), 137–47.

[15] Levey, A. S., Coresh, J., Greene, T., Marsh, J., Stevens, L. A., Kusek, J. W. & Lente, F. V. (2007). Expressing the modification of diet in renal disease study equation for estimating glomerular filtration rate with standardized serum creatinine values, Clinical Chemistry, 53(4), 766–772. [Online*]. http://clinchem.aaccjnls.org/content/53/4/766*

[16] Luyckx, V. A., Tonelli, M. & Stanifer, J. W. (2018). The global burden of kidney disease and the sustainable development goals. *Bulletin of the World Health Organization*, 96(6), 414.

[17] Polat, H., Mehr, H. D. & Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. *Journal Medical System*, 41(4), 55.

[18] Purusothaman, G. & Krishnakumari, P. (2015). A Survey of Data Mining Techniques on risk prediction: Heart disease. *Indian Journal of Science and Technology*, 8(12), 1.

[19] Shafi, N., Bukhari, F., Iqbal, W., Almustafa, K. M., Asif, M., & Nawaz, Z. (2020). Cleft prediction before birth using deep neural network. *Health Informatics Journal*,1(18) 1460458220911789.

[20] Stanifer, J. W., Jing, B., Tolan, S., Helmke, N., Mukerjee, R. & Naicker, S. (2014). The epidemiology of chronic kidney disease in sub-Saharan Africa: a systematic review and meta-analysis. *Lancet Glob Health*, 2(3), 174–81.

[21] Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal Clinic Epidemiol*, 49(11), 1225–31.

[22] Ulasi, I. I. & Ijoma, C. K. (2010). The enormity of chronic kidney disease in Nigeria: The situation in a teaching hospital in south-east Nigeria. *Journal of Tropical Medicine.*

[23] Wang, Z. & Chung, J. W. (2018). Machine Learning Based Prediction System for Chronic Kidney Disease Using Association Classification Technique. *International Journal of Engineering and Technology*, 1161-1167.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed journal ⁜ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121001

[24] Xiao, J. (2019). Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. *Journal of Translational Medicine*, 17(1), 119.

[25] Yildirim, P. (2017). Chronic kidney disease prediction on imbalanced data by multilayer perceptron: Chronic kidney disease prediction. In *Computer Software and Applications Conference (COMPSAC), IEEE 41st Annual*.

