Chronic Kidney Disease Detection Using Deep Learning Algorithm

Ganga.R[1], Jayasheeli. A[2], Pandiya Rajan. G[3]

[1,2] Student, Department of Information Technology, National Engineering College, Kovilpatti

[4] Assistant Professor, Department of Information Technology, National Engineering College, Kovilpatti

Abstract. Chronic Kidney Disease (CKD) refers to damaged kidneys that do not filter your blood properly. As a kidney's primary function is to filter your blood of extra water and waste, CKD occurs when wastes accumulate in the body due to overflow of the kidneys. A large number of people with CKD are increasing, implying that effective measurements are needed to ensure early diagnosis of CKD. CKD is among the top causes of death worldwide, affecting nearly one by tenth of the world adult population. Researchers use machine learning techniques to improve early detection of CKD in order to optimize prevention. The purpose of this project is to develop a Deep Neural Network model to assess whether the patient is affected by chronic kidney disease. Deep Neural Network improves accuracy by increasing the number of hidden layers in the model. It is evaluated on twenty four feature patient dataset containing RBC count, blood pressure, blood sugar, etc. Therefore, machine learning techniques can be extremely useful in the early detection of kidney disease. They assist experts and doctors in detecting kidney disease at an early stage so that kidney failure can be avoided.

Keywords. Chronic Kidney Disease (CKD), Deep Neural Network (DNN), Linear Support Vector Machine (LSVM) and K Nearest Neighbors (KNN)

1. INTRODUCTION

A person with Chronic Kidney Disease has an abundance of wastes accumulated in the body because the kidneys are responsible for filtering extra water and waste from the blood to produce urine. The damage in these cases occur gradually over a long period of time, thus making them chronic. Worldwide, it is common. Some health problems may be experienced due to CKD. In addition to diabetes and high blood pressure, CKD can also be caused by heart disease. These critical diseases as well as age and gender contribute to the disease. It is possible to experience any or many of these symptoms when your kidney is not functioning, including abdominal pain, back pain, diarrhea, fever, nosebleeds, rash, and vomiting.

The purpose of machine learning is to make healthcare smarter. However, a number of obstacles impede its deeper integration into healthcare. For training state of the art machine learning models, it is particularly challenging to obtain patient data sets that qualify as having the appropriate size and quality. In the medical field, It can be difficult to collect, share, and distribute patient data due to strict privacy and security rules. In addition to these challenges, the format and quality of the data pose significant challenges to machine learning analysis and require a significant amount of effort to prepare and clean up.

As a result of analyzing a training data set, a machine learning model learns how to

make decisions. Understanding the prediction process provides us with knowledge of these rules and how they apply new data as well as what certain characteristics are considered most relevant to predicting the outcome. In the prediction explanation, each of the input values is assigned a measure of importance, and if you sum them up, they will equal 100%.
Several deep learning architectures have been developed, including deep neural networks, deep belief networks, recurrent neural networks, and convolution neural networks. In fields such as computer vision, machine vision, speech recognition, natural language processing, social network filtering, audio recognition, machine translation, bioinformatics, and material inspection, machines can sometimes be better than humans in certain fields.

As a result of using multiple layers in the network, deep learning acquires the adjective "deep". In early work, the linear perception was proved to be incapable of acting as a universal classifier, but a network with a non-polynomial activation function and one hidden layer with indefinite width might fulfill that role. Basically, deep learning refers to an unbounded number of layers of bounded size, which enables practical application while maintaining theoretical universality under mild conditions. The layers of deep learning can also be heterogeneous and diverge greatly from biologically informed connectionist models in the interests of efficiencies, trainability, and understandability, hence the "structured" parts.

Similar to neurons in the brain, nodes are tiny nodes within a system. Whenever these nodes receive a stimulus, a process occurs within them. Whenever a task is performed, the system processes layers of data between input and output.

When a network is considered deeper, more layers have to be processed to get the result. Credit Assignment Path (CAP) refers to how many layers are required for a system to complete a task. When the CAP index exceeds two, the neural network is deep. Deep neural networks make it possible to replace human labor with autonomous work without compromising its efficiency. These networks are used in a variety of real-life situations.

The training process involves the processing of sets of examples with known inputs and outputs, forming probability-weighted associations between them, and then storing these associations in the neural network data structure. To train a neural network based on given examples, it is typically done by determining the difference between the output of the processing (often a prediction) and the target output. By adjusting the neural network repeatedly, it will produce outputs that are increasingly similar to the target outputs. Depending on certain criteria, a training can be terminated after a sufficient number of these adjustments have been made. This is known as supervised learning.

Unlike programmed systems, these systems learn how to do a task by analyzing examples rather than programming rules specific to that task. During the processing of examples, they automatically generate identifying characteristics.

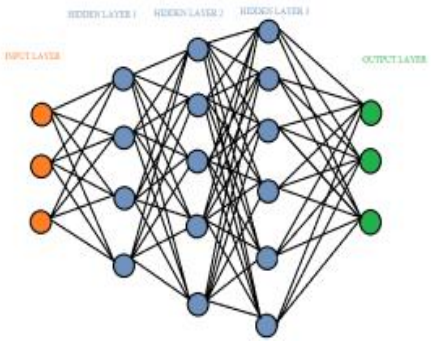


Fig. 1: Deep Neural Network

2. PROBLEM STATEMENT

Various diseases are being encountered due to changes in the environment and lifestyles of people and people are also being affected by them due to these changes. One such widely identified disease is Chronic Kidney Disease. Most often, this results in lethal health conditions or the requirement for kidney transplants. These cases are handled by doctors, researchers, and other technicians who are involved in controlling and managing the disease. In these cases, the main challenge is to accurately detect and characterize the disease in order to effectively manage and control it. Thus the implementation of modern techniques like machine learning, were the algorithms are implemented as Classifiers. When used in conjunction with the classifier algorithms like LSVM (Linear Support Vector Machine), CHAID (Chi-Square Automatic Interaction Detection), and KNN (K-Nearest Neighbors), the accuracy results are expected to be much higher. Thus, the use of neural network concepts not only increases detection accuracy but also facilitates handling of large datasets. Deep Neural Network's accuracy can be achieved by increasing the number of hidden layers in the model. Then the models are trained with the testing data set thus building the required prediction neural network model. Thus, machine learning techniques are crucial in the early detection of chronic kidney disease. Experts and doctors use these techniques to diagnose kidney failure early and prevent it from progressing.

3. EXISTING SYSTEM

The existing system machine learning algorithms were used as a form of classifier for predicting chronic kidney disease. When predicting chronic kidney disease, around 25 attributes (24 attributes and 1 target attribute) were considered and found out which was the quickest and most accurate. CKD or non-CKD was designated as two classes as the target attribute. Diagnostics factors to check the condition of CKD like estimated glomerular filtration rate (eGFR), urine test, blood pressure were considered as major factors. Using machine learning classification techniques such as:

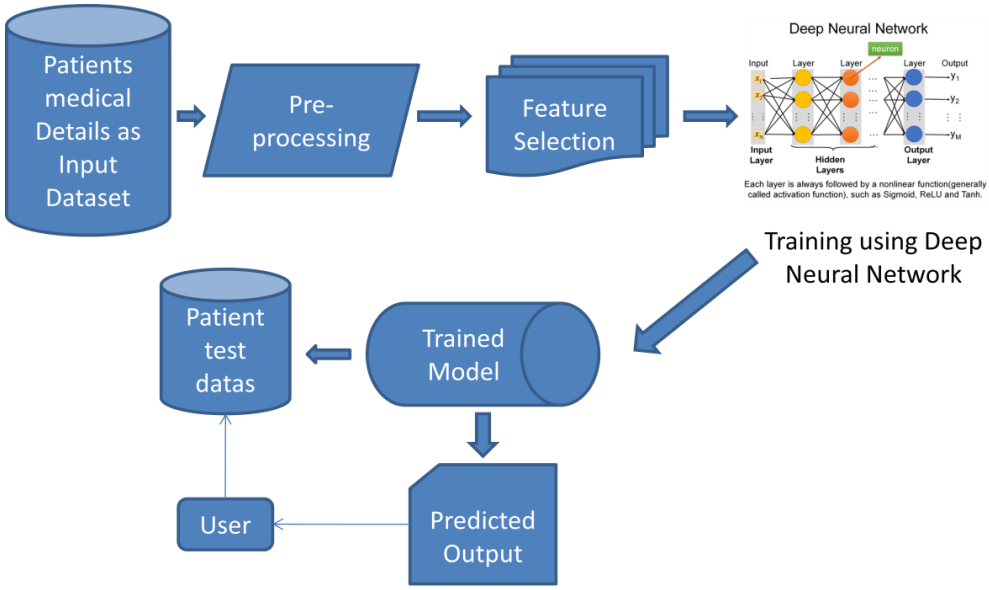
- C5.0
- Logistic regression
- Linear support vector machine,
- K- nearest neighbors

- Random trees

They developed a model to predict CKD disease from patient data. The collected data and preprocessed them in accordance with the requirement and then trained the model using classification techniques using the given dataset. A second training dataset was then used for testing the model by applying classification techniques. When comparing the accuracy of the above algorithms, LSVM achieved the highest level of 95.12%, while KNN had the lowest level of 53.17%.

4. PROPOSED SYSTEM

In the proposed system, a deep learning model was employed that is Deep Neural Network (The method of machine learning when high-level functions are derived from input data by using multiple layers of nodes and it is based on transforming the data into something more creative and abstract). In recent times, the implementation of deep neural network achieves greater accuracy. The prediction of chronic kidney disease solely depends on the dataset considered and the accuracy of the trained model. By using DNN, we can predict the chronic kidney disease with more than 95% of accuracy. In the DNN, we have more hidden layers and hence its accuracy also high.



.

Fig. 2 Proposed System Network

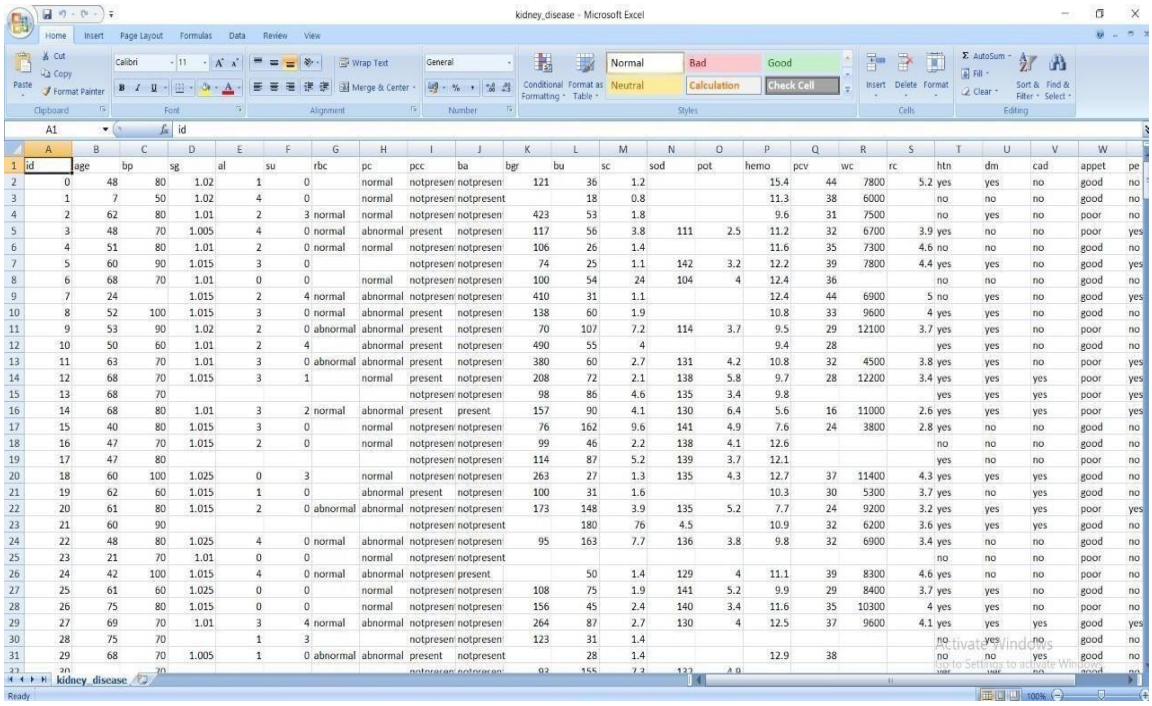
4.1 Data Collection:

The dataset of the proposed system is obtained from Kaggle. The data set consists of 25 features (eg: red blood cell count, white blood cell count, appetite, heart disease etc). The target attribute here is the 'classification', which is either 'ckd' or 'not ckd' - ckd=chronic kidney disease. The Data are retrieved from the input dataset by using pandas library. Retrieval of rows from the data frame is easily done as pandas provide unique method.

dataframe.loc[] method is a method that takes only index labels and returns row or data frame if the index label exists in the caller data frame.

4.2 Featured Attributes:

S.No	Attribute Name	Description
1	Age	Patient age (It is in years)
2	Bp	Patient blood pressure (It is in mm/HG)
3	Sg	Patient urine specific gravity
4	Al	Patient albumin ranges from 0-5
5	Su	Patient sugar ranges from 0-5
6	Rbc	Patient red blood cells two value normal and abnormal
7	Pc	Patient pus cell two value normal and abnormal
8	Pcc	Patient pus cell clumps two values present and not present
9	Bgr	Patient blood glucose ransom in my/dl
10	Bu	Patient blood urea in mg/dl
11	Sc	Patient serum creatinine
12	Sod	Patient sodium
13	Pot	Patient potassium
14	Hemo	Patient hemoglobin (protein molecule in red blood cells)
15	Pcv	Patient packed cell volume
16	Wc	Patient white blood cell count in per microliter
17	Htn	Patient hypertension two values Yes and No
18	Dm	Patient diabetes mellitus two values Yes and No
19	Cad	Patient coronary artery diseases two values Yes and No
20	Appet	Patient appetite two values good and poor
21	Pe	Patient pedal edema two values Yes and No
22	Ane	Patient anemia two values Yes and No
23	Ba	Patient bacteria two values present and not present
24	Class	Target Variable (CKD or Not)
25	Rc	Patient red blood cell count in per microliter



id	age	bp	sg	al	su	rbc	pc	pcc	ba	bgr	bu	sc	sod	pot	hemo	pcv	wc	rc	htn	dm	cad	appet	pe
0	48	80	1.02	1	0		normal	notpresen	notpresen	121	36	1.2			15.4	44	7800	5.2	yes	yes	no	good	no
1	7	50	1.02	4	0		normal	notpresen	notpresent		18	0.8			11.3	38	6000		no	no	no	good	no
2	62	80	1.01	2	3	normal	normal	notpresen	notpresen	423	53	1.8			9.6	31	7500		no	yes	no	poor	no
3	48	70	1.005	4	0	normal	abnormal	present	notpresen	117	56	3.8	111	2.5	11.2	32	6700	3.9	yes	no	no	poor	yes
4	51	80	1.01	2	0	normal	normal	notpresen	notpresen	106	26	1.4			11.6	35	7300	4.6	no	no	no	good	no
5	60	90	1.015	3	0			notpresen	notpresen	74	25	1.1	142	3.2	12.2	39	7800	4.4	yes	yes	no	good	yes
6	68	70	1.01	0	0		normal	notpresen	notpresen	100	54	24	104	4	12.4	36			no	no	no	good	no
7	24		1.015	2	4	normal	abnormal	notpresen	notpresen	410	31	1.1			12.4	44	6900	5	no	yes	no	good	yes
8	52	100	1.015	3	0	normal	abnormal	present	notpresen	138	60	1.9			10.8	33	9600	4	yes	yes	no	good	no
9	53	90	1.02	2	0	abnormal	abnormal	present	notpresen	70	107	7.2	114	3.7	9.5	29	12100	3.7	yes	yes	no	poor	no
10	50	60	1.01	2	4		abnormal	present	notpresen	490	55	4			9.4	28			yes	yes	no	good	no
11	63	70	1.01	3	0	abnormal	abnormal	present	notpresen	380	60	2.7	131	4.2	10.8	32	4500	3.8	yes	yes	no	poor	yes
12	68	70	1.015	3	1		normal	present	notpresen	208	72	2.1	138	5.8	9.7	28	12200	3.4	yes	yes	yes	poor	yes
13	68	70						notpresen	notpresen	98	86	4.6	135	3.4	9.8				yes	yes	yes	poor	yes
14	68	80	1.01	3	2	normal	abnormal	present	present	157	90	4.1	130	6.4	5.6	16	11000	2.6	yes	yes	yes	poor	yes
15	40	80	1.015	3	0		normal	notpresen	notpresen	76	162	9.6	141	4.9	7.6	24	3800	2.8	yes	no	no	good	no
16	47	70	1.015	2	0		normal	notpresen	notpresen	99	46	2.2	138	4.1	12.6				no	no	no	good	no
17	47	80						notpresen	notpresen	114	87	5.2	139	3.7	12.1				yes	no	no	poor	no
18	60	100	1.025	0	3		normal	notpresen	notpresen	263	27	1.3	135	4.3	12.7	37	11400	4.3	yes	yes	yes	good	no
19	62	60	1.015	1	0		abnormal	present	notpresen	100	31	1.6			10.3	30	5300	3.7	yes	no	yes	good	no
20	61	80	1.015	2	0	abnormal	abnormal	notpresen	notpresen	173	148	3.9	135	5.2	7.7	24	9200	3.2	yes	yes	yes	poor	yes
21	60	90						notpresen	notpresent		180	76	4.5		10.9	32	6200	3.6	yes	yes	yes	good	no
22	48	80	1.025	4	0	normal	abnormal	notpresen	notpresen	95	163	7.7	136	3.8	9.8	32	6900	3.4	yes	no	no	good	no
23	21	70	1.01	0	0		normal	notpresen	notpresent										no	no	no	poor	no
24	42	100	1.015	4	0	normal	abnormal	notpresen	present		50	1.4	129	4	11.1	39	8300	4.6	yes	no	no	poor	no
25	61	60	1.025	0	0		normal	notpresen	notpresen	108	75	1.9	141	5.2	9.9	29	8400	3.7	yes	yes	no	good	no
26	75	80	1.015	0	0		normal	notpresen	notpresen	156	45	2.4	140	3.4	11.6	35	10300	4	yes	yes	no	poor	no
27	69	70	1.01	3	4	normal	abnormal	notpresen	notpresen	264	87	2.7	130	4	12.5	37	9600	4.1	yes	yes	yes	good	yes
28	75	70		1	3			notpresen	notpresen	123	31	1.4							no	yes	no	good	no
29	68	70	1.005	1	0	abnormal	abnormal	present	notpresent		28	1.4			12.9	38			no	no	yes	good	no
30		70						notpresen	notpresen	93	155	7.2	137	4.9					yes	no	no	good	no

Fig. 3 Data set

4.3 Data Pre-processing:

Preprocessing involves preparing the raw data for machine learning by making it suitable for the model. This dataset contains lots of null values. So replace all the null values into 0. This is because this Deep learning model cannot process null values or else it will result in some error or invalid result. Also here Conversion of String values to integers and float values are also done.

4.4 Feature Selection:

After the data preprocessing phase, feature selection phase is done. In feature selection, the attributes are classified as inputs and outputs. In this case all the attributes are considered inputs except the classification attribute which is considered as output. Then segregation of training and testing of data is done. Here consider 70% of the data is considered for training and other 30% is considered for testing the model.

4.5 Module Description:

This model contains three hidden layers that is the CAP index (Credit Assignment Path). The neural network is deep if the CAP index is more than two.

Training

The extracted features are inserted into the designed DNN model. Then this model gets trained according to the given input data.

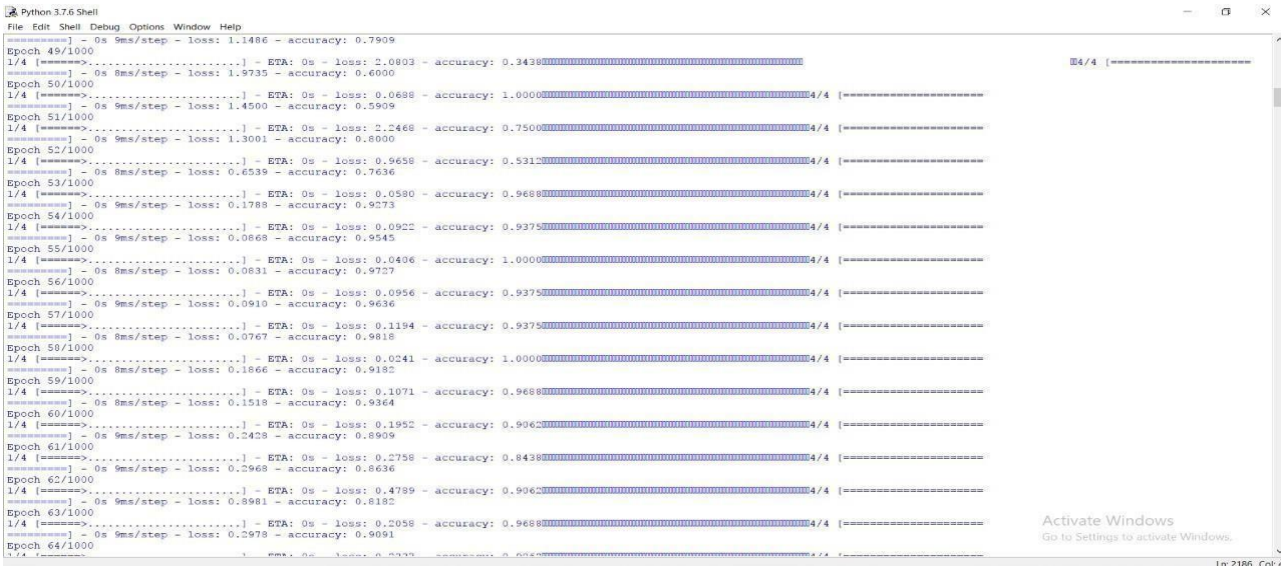


Fig. 4 Training of the dataset

Working

A pandas library package is used to extract the initial data from the input dataset. Then the pre-processing the data involves replacing null values with zeroes and conversion to numeric values. Then the neural network model is created with input layer, hidden layers and output layer. Our model contains 1 input layer with 64 nodes, 3 hidden layers, 1output layer with 1 node. First hidden layer contains 128 nodes, second hidden layer contains 256 nodes and third hidden layer contains 512 nodes. Once the creation of model and adding of nodes is done training of the t model is done. The trained model is tested with the testing data the result is compared with the confusion matrix to get the better and accurate result. With this technique we achieve 98% accuracy in the prediction.

5. RESULTS

Fig. 5 CKD title page

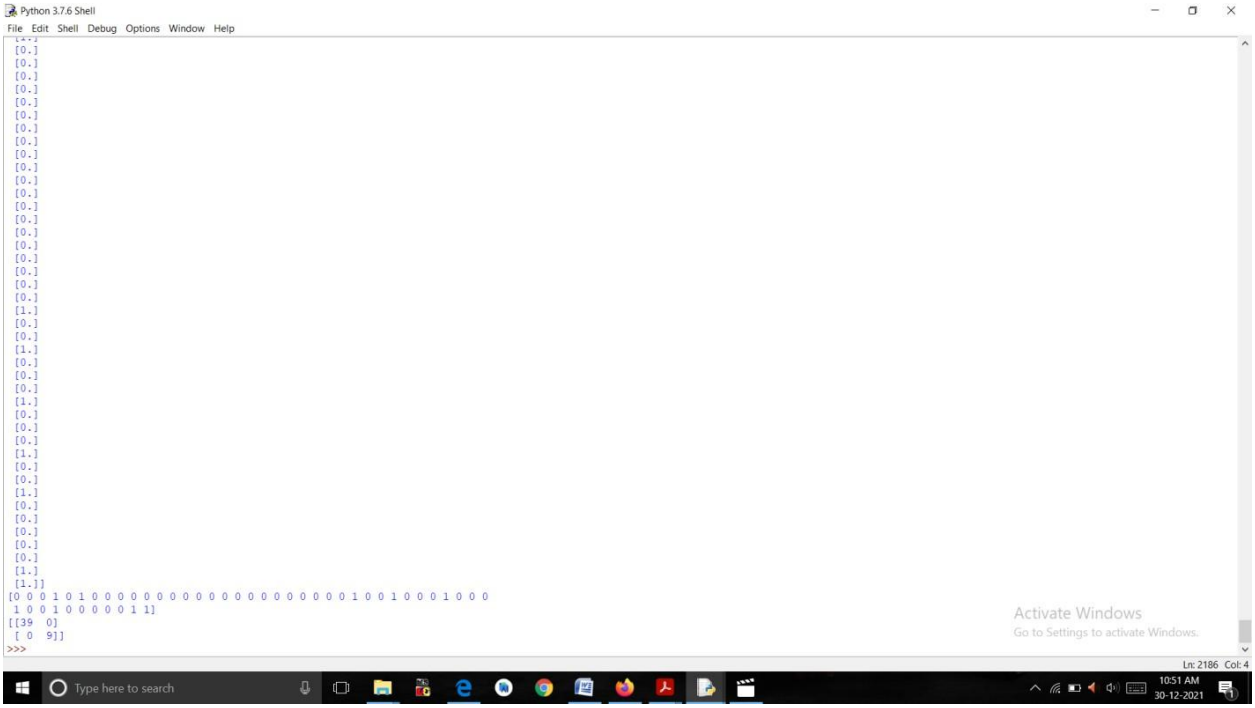


Fig. 6 Confusion matrix of the data

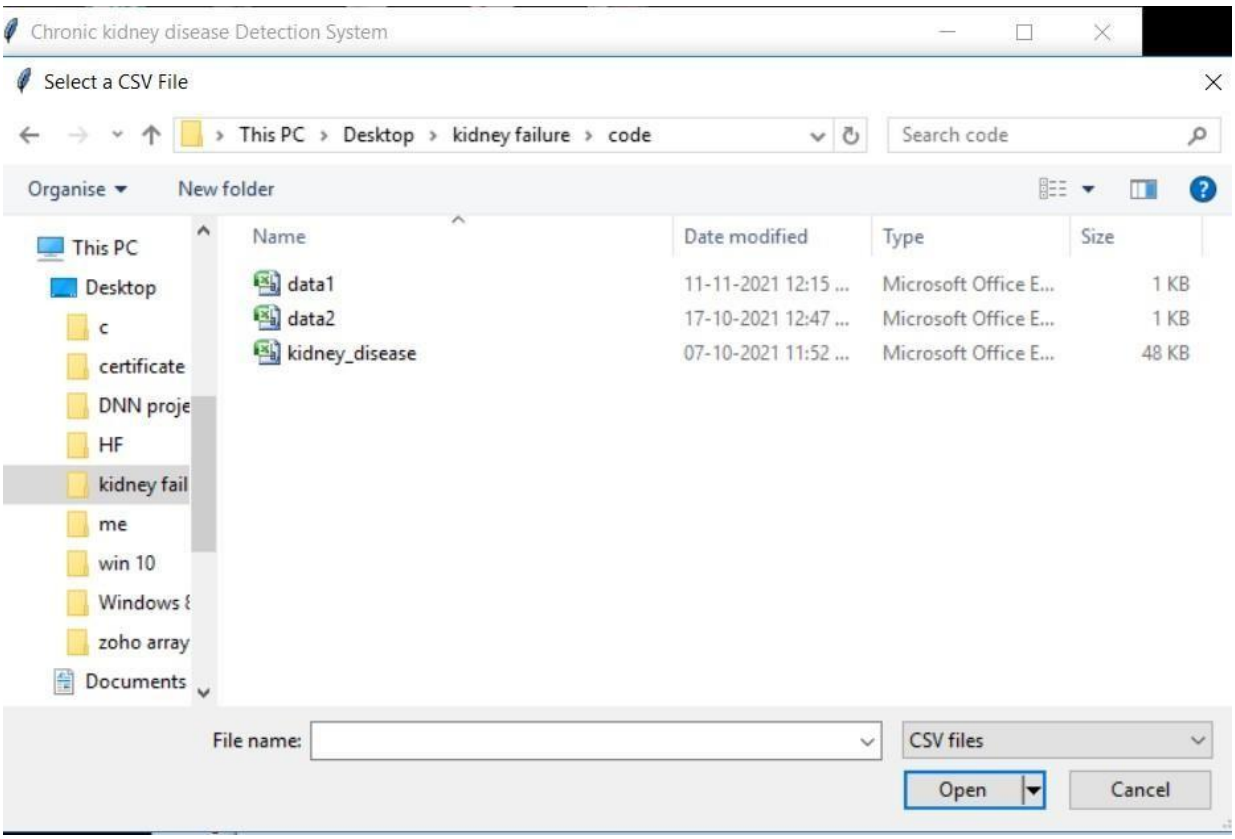


Fig. 7 Browsing the data

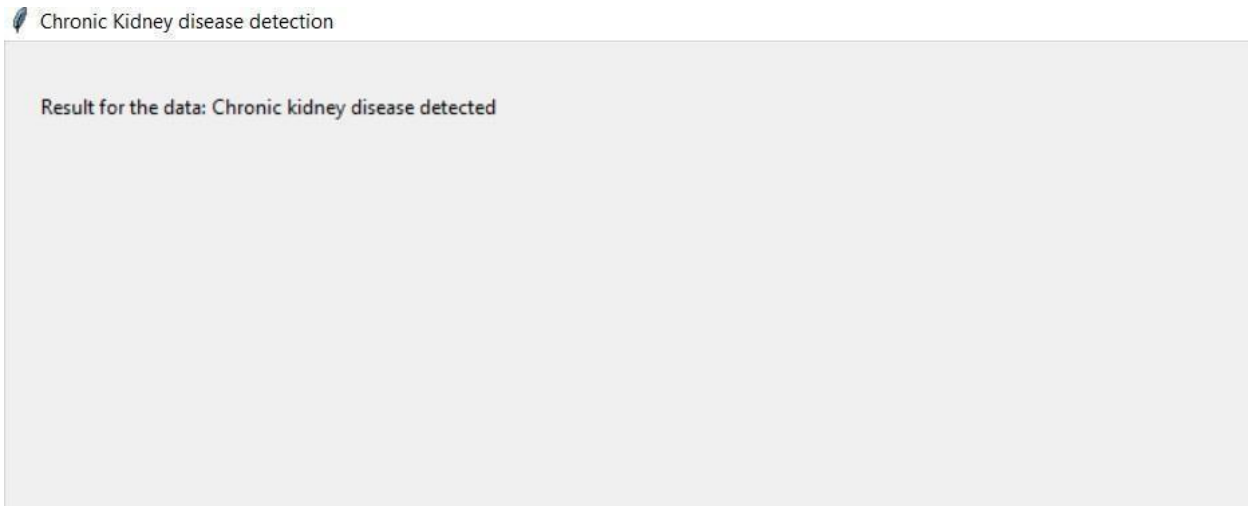


Fig. 8 Result of Dataset 1

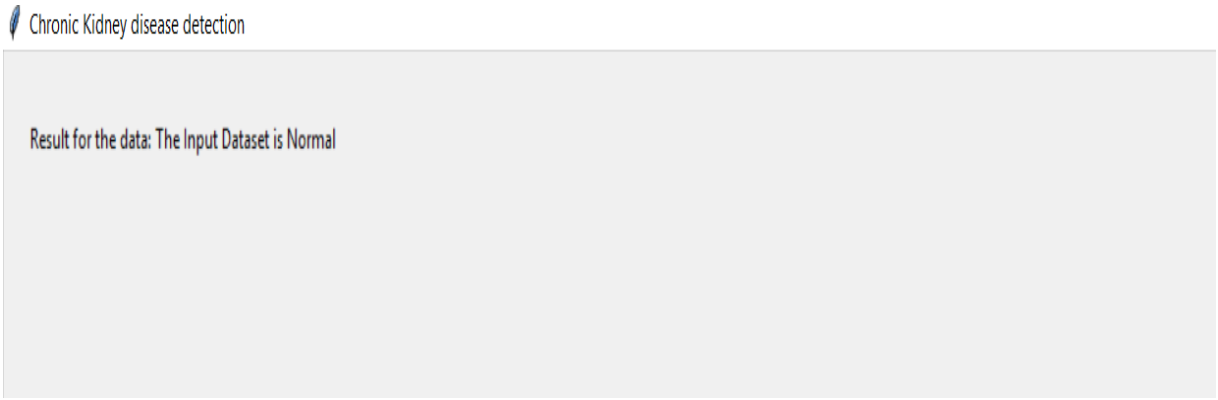


Fig. 9 Result of Dataset 2

6. CONCLUSION

The objective of this project was to predict chronic kidney disease using Deep Neural Network. This is a deep learning technique that can be used to train and predict kidney disease based upon input data. We used chronic kidney disease datasets that hold over 24 different features for training purpose. When the model is trained using the training dataset, it is tested with the testing dataset. Our model has an accuracy of more than 98% when tested and trained. Our DNN model predicts accurate and stable results, its patterns are also matched to the existing datasets and, therefore, it can predict kidney disease with high accuracy. Our next step will be to use hybrid approaches to improve the accuracy.

7. REFERENCES

[1] Q.-L. Zhang and D. Rothenbacher, ‘‘Prevalence of chronic kidney disease in population-based studies: Systematic review,’’ BMC Public Health, vol. 8, no. 1, p. 117, Dec.2008.

[2] W.M.McClellan,D.G.Warnock,S.Judd,P.Muntner,R.Kewalramani,M.Cushman,L.A. McClure, B. B. Newsome, and G. Howard, ‘‘Albumin- uria and racial disparities in the risk for ESRD,’’ J. Amer. Soc. Nephrol., vol. 22, no. 9, pp. 1721–1728, Aug. 2011.

[3] M. K. Haroun, ‘‘Risk factors for chronic kidney disease: A prospective study of 23,534 men and women in Washington County, Maryland,’’ J. Amer. Soc. Nephrol., vol. 14, no. 11, pp. 2934–2941, Nov.2003.

[4] W. D. Souza, L. C. D. Abreu, L. G. D. SilvaI, and I. M. P. Bezerra, ‘‘Incidence of chronic kidney disease hospitalisations and mortality in Espírito Santo between 1996 to 2017,’’ WisitCheungpasitporn, Univ. Mississippi Medical Center, Rochester, MN, USA, Tech. Rep., 2019, doi:10.1371/journal.pone.0224889.

[5] W. Mula-Abed, K. A. Rasadi, and D. Al-Riyami, ‘‘Estimated glomerular filtration rate (eGFR): A serum creatinine-based test for the detection of chronic kidney disease and its impact on clinical practice,’’ Oman Med. J., vol. 27, no. 4, pp. 339–340,2012.

[6] A. S. Levey, D. Cattran, A. Friedman, W. G. Miller,B. Kasiske, and T. Hostette, ‘‘Proteinuria as a surrogate outcome in CKD: Report of a scientific workshop sponsored by the national kidney founda- tion and the US food and drug administration,’’ Amer. J. Kidney Diseases, vol. 54, no. 2, pp. 205–226, Aug.2009.

[7] S. Gerogianni, ‘‘Concerns of patients on dialysis: A research study,’’ Health Sci. J., vol. 8, no. 4, pp. 423–437,2014.

[8] K. R. A. Padmanaban and G. Parthiban, ‘‘Applying machine learning techniques for predicting the risk of chronic kidney disease,’’ Indian J. Sci. Technol., vol. 9, no. 29, Aug. 2016.

[9] W. Gunarathne, K. D. M Perera, and K. A. D. C. P Kahandawaarachchi, ‘‘Performance evaluation on machine learning classification techniques for disease classification,forecasting through data analytics for chronic kidney disease (CKD),’’ in Proc. IEEE 17th Int. Conf. Bioinf. Bioeng. (BIBE), Oct. 2017, pp.291–296.

[10] S. Drall, G. S. Drall, S. Singh, and B. B. Naib, ‘‘Chronic kidney disease prediction using machine learning: A new approach,’’ Int. J. Manage., Technol. Eng., vol. 8, pp. 278–287, May 2018.