International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

Prediction of Chronic Kidney Disease using SVM and CNN

M Purushotham Reddy*[1], K Pavan Kumar[2], Y Suresh[3], T Vara Lakshmi[4]
[1]Professor, Department of Information Technology,
Institute of Aeronautical Engineering, Hyderabad, Telangana,500043
purushotham.mps@gmail.com
[2]Assistant Professor, Department of Information Technology,
Prasad V. Potluri Siddartha Institute of Technology, Vijayawada, AP – 520007
pavanpvpsit@gmail.com
[3]Associate Professor, Department of Information Technology,
Prasad V. Potluri Siddartha Institute of Technology, Vijayawada, AP – 520007
sureshyadlapati@gmail.com
[4]Professor, Department of MBA,
Institute of Aeronautical Engineering, Hyderabad, Telangana,500043
mvaralu2011@gmail.com

Abstract— Chronic kidney disease is one of the deadliest diseases today and it is vital to have a good diagnosis as soon as possible. In medical treatment, machine learning has been reported to be effective. A doctor can diagnose the disease early by using machine learning classifier algorithms. This study investigated the chronic disease prognosis of this concept. Disease data was taken from the University of California, Irvine. Other measurement algorithms used in this study include C5.0, Chi-square automatic interaction detector, line extraction, SVM line with L1 and L2 flap, and neural network random tree. The database was also submitted to a feature selection program that merited the database. Scores are computer generated for each category segment using the following methods: Full Version, (ii) Link-Based Feature Selection, (iii) Folder Feature Selection, (iv) Minimal Collapse and Selected Optional Retrospective Features, (v) integrated small oversampling method with very small reduction features and selected bias on the selected operator, and (vi) how to do multiple samples combined with full functions. In the full multi-sample processing process, the findings show that L2-loaded LSVM has a very high accuracy of 96.86 percent. The graph shows the results of different methods, as well as precision, precision, recall, F-score, area under the curve, and GINI coefficient. The minimum absolute reduction and selection regression operation selected features using the synthetic minority oversampling approach produced the best results after using the synthetic minority oversampling method with full features. The support vector machine achieved a high accuracy of 96.46 percent in the process of making very large samples with very small turndowns and selected operator features. Machine learning methods used with convolutional neural networks and SVM classifier models on the same database, with 96.7 percent of high-definition support machine models and networks are used.

Keywords- Chronic Kidney Disease, Machine Learning, Deep Learning, Sklearn,prediction.

I. INTRODUCTION

Chronic kidney disease is called CKD. Kidney disease (CKD) is a disorder in which the kidneys are damaged and cannot filter blood as efficiently as they should. As a result, extra fluid and waste from the bloodstream remain in the body, which can lead to various health problems, including heart disease and stroke. Other health effects of CKD include: (i) Anemia (low red blood cell count), (ii) Increased risk of infection, (iii) Low levels of calcium, potassium and phosphorus in the blood, (iv) Loss of appetite (eating less) and (v) depression (low quality of life). CKD can be mild, moderate, or severe. Although treatment has been shown to slow progression, it usually worsens over time. If neglected, CKD can lead to kidney failure and death. Kidney failure treated with dialysis or a kidney transplant is known as end-stage renal disease (ESRD). Kidney failure usually does not occur in patients with kidney disease. Manage your CKD risk factors, get checked every year, make lifestyle adjustments, take medication if needed, and monitor your progress. See your healthcare team regularly to help prevent CKD and minimize the chance of kidney failure.

If your kidneys are not working properly, you may experience abdominal pain, back pain, diarrhea, fever, nosebleeds, rash, and vomiting. Diabetes and high blood pressure are the two most common causes of CKD. As a result, CKD prevention requires addressing these two diseases. CKD usually does not appear until the kidneys are severely damaged. According to studies, hospitalizations are increasing by 6.23% per year, but the global mortality rate remains stable. There are several tests that can be used to determine the severity of CKD: blood pressure measurement, urinalysis, and estimation of glomerular filtration rate (eGFR) are all required.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

A. The EGFR gene

The estimated glomerular filtration rate (eGFR) is a measure of how successfully your kidneys filter your blood.Your kidneys are in good shape if your eGFR is above 90. If your eGFR is below 60, you have chronic kidney disease.

B. URINE EXAMINATION

Since the kidneys produce urine, the doctor also requires a urine test to determine kidney function. Your kidney is not working properly if your urine contains blood and protein.

C. BLOOD PRESSURE

A doctor measures blood pressure because it reveals how well your heart is pumping blood. A patient has end-stage renal disease if their eGFR is less than 15. Currently, only two therapies are available: (i) dialysis and (ii) kidney transplant. Age, sex, frequency and duration of dialysis, physical mobility of the body, and mental health are elements that affect a patient's life after dialysis. If dialysis is not an option, the doctor has only one option: a kidney transplant. However, this is prohibitively expensive.

Therefore, it is vital to pay attention to early detection, monitoring and treatment of the disease. Due to the dynamic and hidden nature of CKD in its early stages, it is extremely important to accurately predict its progression, as well as patient abnormalities. The stage of CKD determines the pharmacological treatment. In addition, it is of the utmost importance to characterize the organization of the infection, since it provides several indicators. This is the basis for the provision of basic intercessions and medicines.

Kidney disease is now a serious public health problem. It is a condition in which the kidneys are damaged and cannot filter toxic waste from the body. To predict life-threatening diseases such as chronic kidney disease, we apply classification algorithms such as the naive Bayes classifier, neural network, SVM, and CNN.

II. LITERATURE SURVEY

The article Vasquez-Morales et al [1] worked on a data set of 40,000 specimens, created a neural network to predict the risk of developing chronic kidney disease, and the accuracy of their model was 95%.

The report by Chen et al [2] worked on the data set provided by UCI, three models were implemented. They calculate patient risk using KNN, SVM, and SIMCA (smooth independent modeling of class analogy) classifiers. The SVM and KNN models have the highest accuracy of 99.7%, and the SVM model has the highest ability to withstand noise disturbances.

Amirgaliyev's paper [3] The SVM machine learning classifier technique achieved an experimental result of 93 percent accuracy. Because CKD is invasive and expensive, many people reach the final stages of the disease without receiving therapy. As a result, it is vital to detect this disease early.

They used decision trees, random forests, and SVMs with linear, polynomial, sigmoid, and RBF functions in their research by de Almeida et al [4]. They carried out their study using the MIMIC-II database. They concluded that the random forest and the decision tree gave the best results, with a prediction accuracy of 80% and 87%, respectively.

Gunarathne et al [5] built a model of different machine learning classification methods and evaluated which strategy was best for the data set. They used the UCI data set, which had 400 events and 14 functions. With 99.1% accuracy, the multi-class decision forest strategy was found to be the correct combination for the CKD data set.

The SVM process of Polat et al [6] was used to predict CKD. They focus on one important factor to achieve an accurate result. They used a bidirectional convolution and an SVM-type filter to identify the appropriate feature. Wrapper includes both a step-by-step greedy search engine sub-segment search tool and an excellent Wrapper sub-segment search feature. There was a greedy finder of the subset element of the link element, as well as a first finder of the subset filter element. When all method findings were tested, the SVM was found to have very high accuracy with a 98.5 percent low-pass filter.

Almasoud and Ward [7] used a 400-item ERC database with 25 features in their study.

They used the filter feature selection approach to search for feature attributes in the CKD dataset and found that hemoglobin, albumin, and specific gravity were feature attributes. Use the 10-step verification to verify the database by following the feature selection. With a 99.1 percent accuracy rate, the gradient-boosting strategy was the most accurate.

The article by Vijayarani and Dhayanand [8] collected data from health laboratories, research centers and hospitals for kidney function tests (KFT). There were 584 instances and 6 attributes in the data set, and two classification techniques were used: artificial neural network (ANN) and support vector machine (SVM). The ANN was found to have a high accuracy of 87.7 percent.

Ma and colleagues [9] proposed a comprehensive survey strategy to predict the occurrence of kidney disease. The deep neural network is built using the heterogeneous modified artificial neural network algorithm. The model was made using ultrasound images. Findings were compared using three different classifiers: SVM, ANN, and multilayer perceptron.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

UI Haq et al [10] proposed a machine-readable method for the early diagnosis of diabetes. They realized that machine learning could be useful in the field of healthcare.

The article Amin et al [11] proposed a machine learning model to predict early-stage Parkinson's disease. They used the SVM phase to create a model. The following feature options are used to extract important features: Feature Help to remove ACO features.

Dodge et al. [12] see how image quality deviations (blur, noise, brightness, compression in JPEG and JPEG 2000) affect deep neural networks (VGG-16, VGG-CNN-S, GoogleNet), but not correction from image.

The article by Dejean et al [13] shows how CNN's classification performance is affected by compression. They also state that for JPEG and JPEG2000 compression, without compromising classification, the image can be reduced by a number of 7, 16, or 40.

The article Krizhevsky et al [14], 2012 created a CNN model that was able to significantly reduce the error rate in the ILSVRC Krizhev competition.

III. SYSTEM DESIGN

A. System Architecture

System architecture is shown in Fig 1.

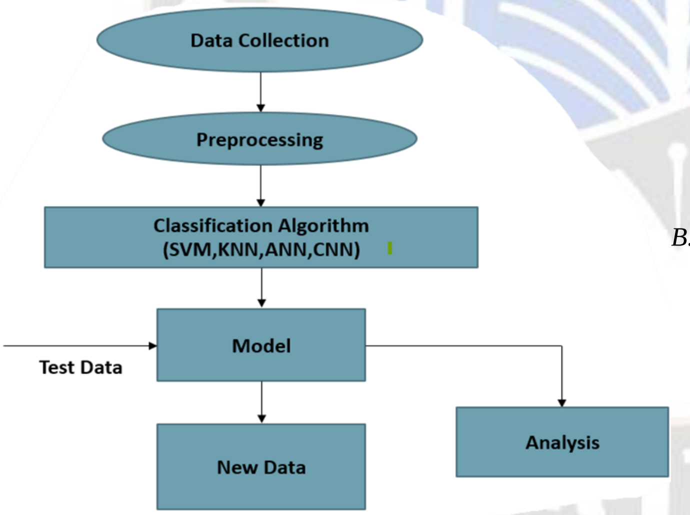

Fig 1: System Architecture

IV. METHODOLOGY

A. SVM (SUPPORT VECTOR MECHINE)

SVMs (Support Vector Machines) are a new way of separating linear and indirect data as shown in fig 2. In this case, the Support Vector Machine (SVM) is a solution algorithm in this sense. Only some training data is converted to high quality using an indirect mapping method. Looking for a straight division a hyperplane or "decision boundary" that separates tuples from one section to another surrounded by this new depth. A well-designed hyperplane without a line of the required maximum size can divide the data into two classes. SVM uses support vectors and genes to connect the hyperplane. Despite the fact that even the fastest SVMs have a long training period, they are remarkably accurate in modeling complex nonlinear decision bounds. They are less likely to work better than other strategies. The first support vectors also serve as a brief summary of the studied model. SVMs can be used to predict future splits. Some of the applications include handwriting attention, object recognition, speaker recognition, and timeline series prediction.

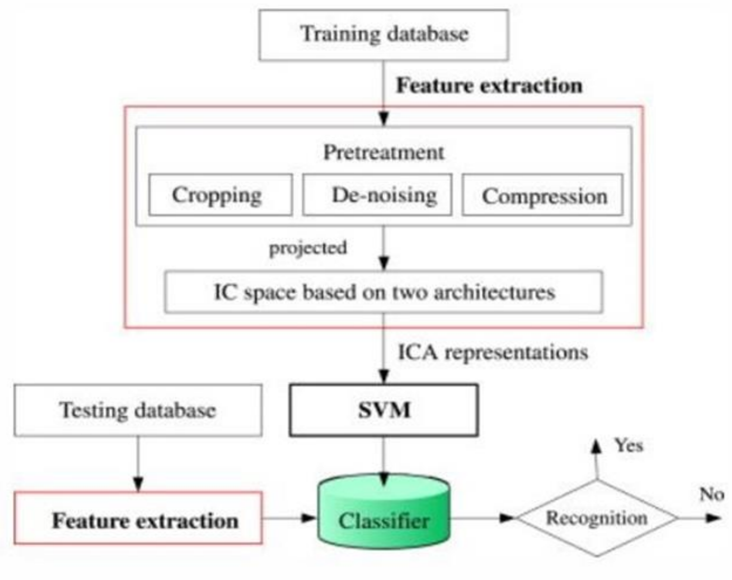

Fig 2: Support Vector Machine

B. CNN (CONVOLUTIONAL NEURAL NETWORK)

Deep learning or deep neural networks are artificial neural networks (ANN) with multiple layers. It has been hailed as one of the most useful tools in recent years, and its ability to control massive amounts of data has made it a hot topic in academia. In some fields, such as pattern recognition, the desire for deeper hidden layers has only recently begun to overtake traditional methods. Convolutional Neural Networks (CNNs) are a well-known type of Deep Neural Networks (CNNs) as shown in Fig 3. The convolution operation, which is a logical linear process connecting two matrices, gives it its name. A CNN consists of multiple layers, including a flexible layer, a fixed layer, a composite layer, and a fully integrated layer. Fully connected and flexible layers have limitations, but nonlinear and pooled layers have no limitations. When it comes to machine learning, CNN performs admirably. Particularly impressive were the applications that work with image data, such as the world's largest collection of image classification data (Image Net). Computer vision and natural language processing

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

(NLP) were used with the collected data. In this document, we will describe and outline all the critical aspects and issues related to CNN and how these elements work. We'll also look at parameters that affect CNN's performance. Machine learning and realized neural networks are assumed to be familiar to readers of this article.

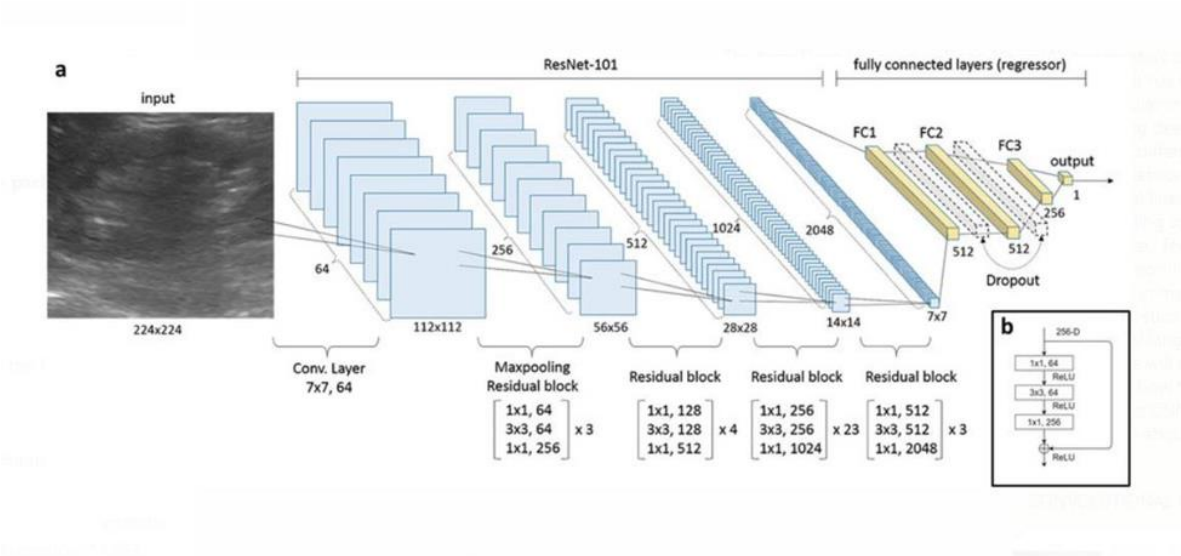

Fig 3: Convolution Neural Networks

V. EXPERIMENTAL SETUP

Clinical data for the 400 records used in the study come from the UCI machine learning repository. The data comes to a total of 220 after cleaning and removing the missing numbers. Rapid Miner was used to collect the data. The data set contains a total of 25 records. Age, blood pressure, random blood sugar, blood urea, serum creatinine, potassium, hemoglobin, whole cell volume, white blood cell count and red blood cell count among the digital indicators. Albumin, sugar, red blood cells, pus cells, clumps of pus cells, viruses, high blood pressure, diabetes, cardiovascular disease, hunger, anemia, and social status are some of the common factors.

VI. RESULT

SVM and CNN test comparisons are made based on performance vectors. It is a mathematical performance test of division functions and contains a to-do list. It is shown in the following figures 4 – 13.

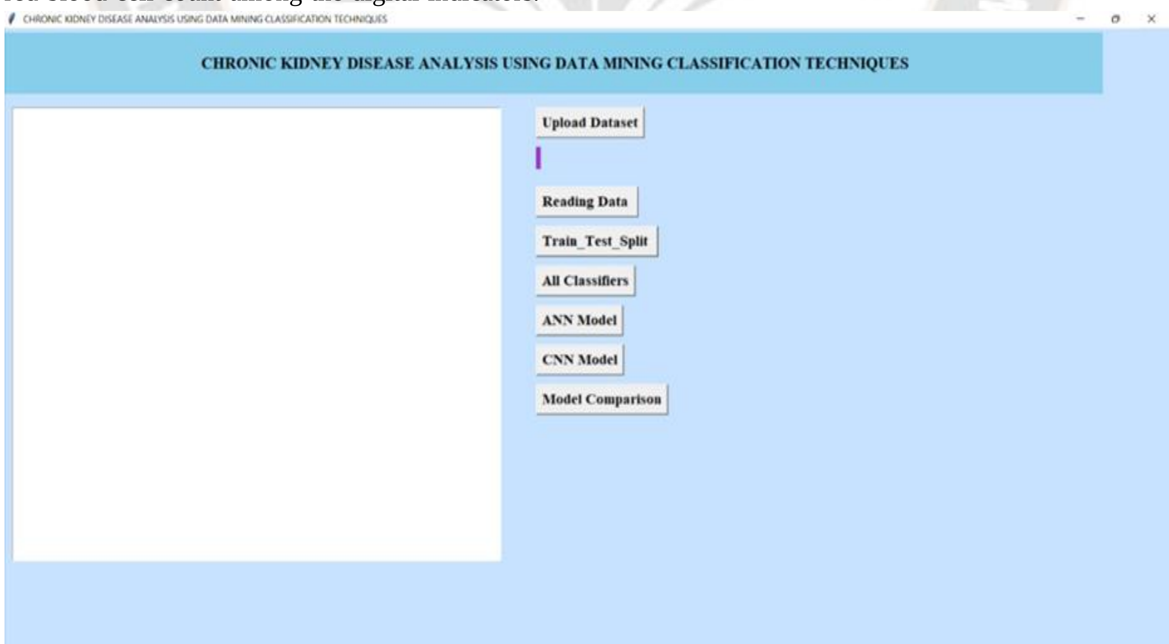

Fig 4: Classification Techniques

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

To upload the data, click 'upload dataset'.

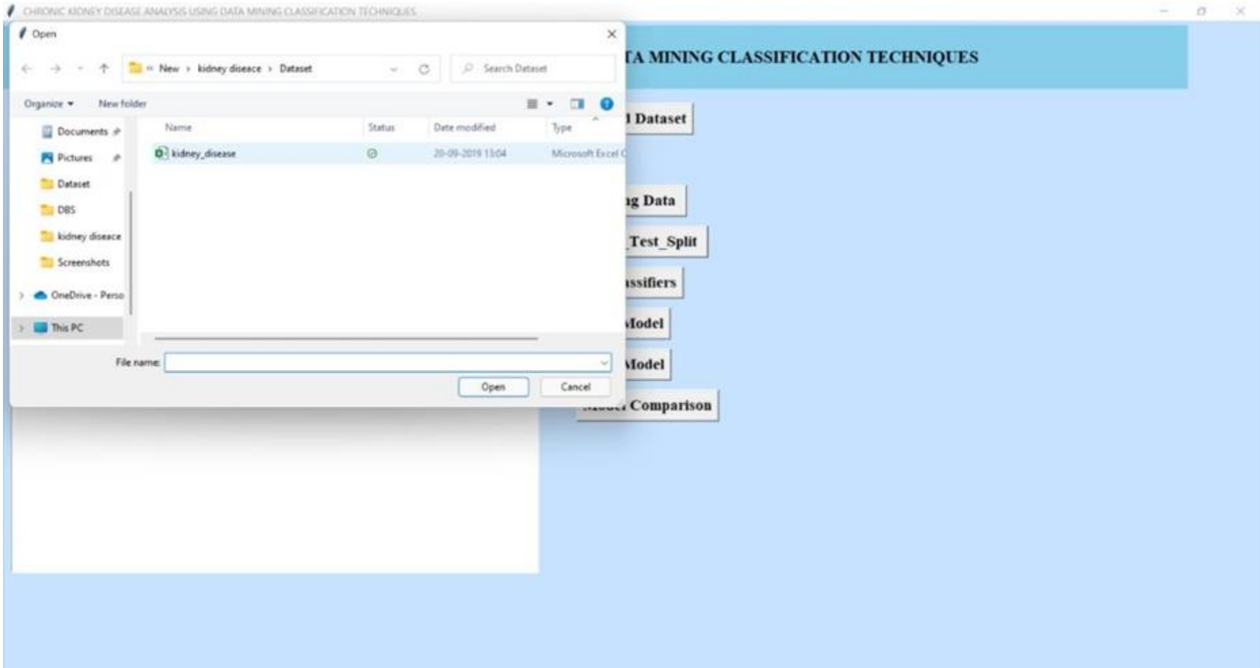

Fig 5: Upload of data set

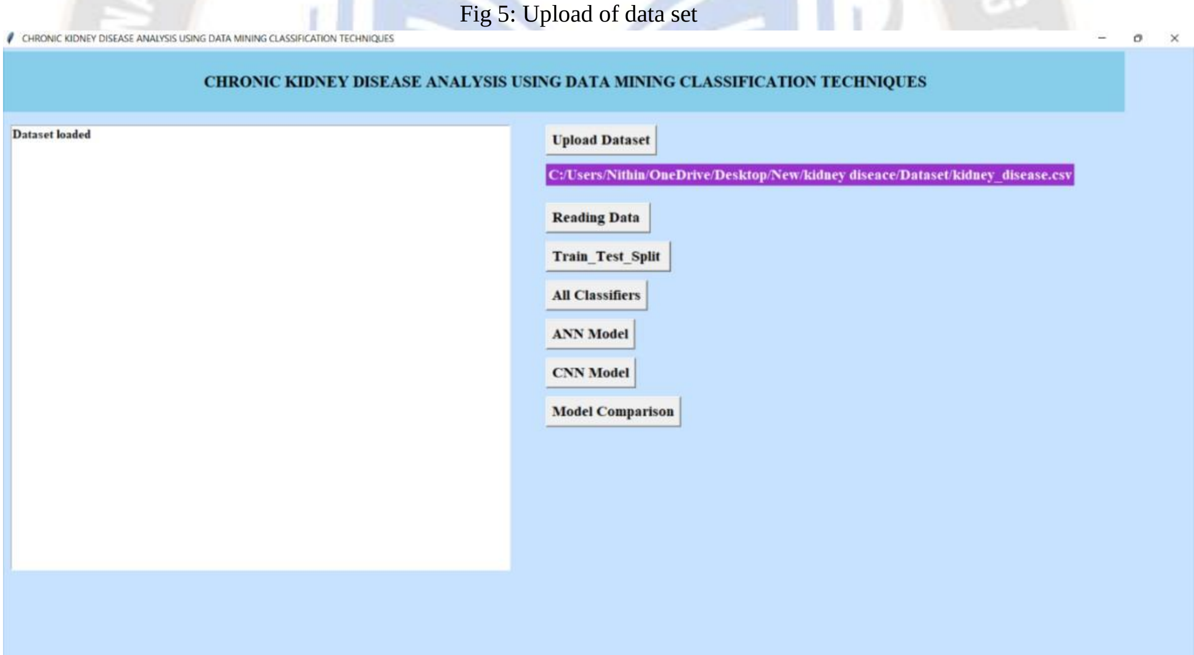

Fig 6: Read Data

Now select 'read data' from the drop-down menu. It reads the data.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

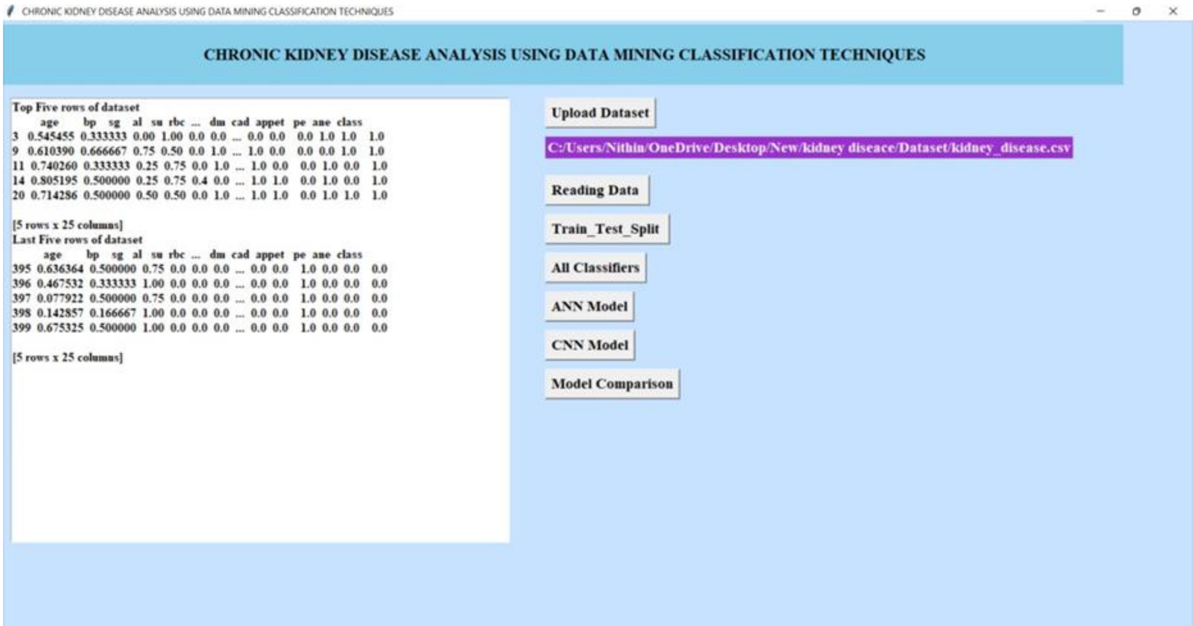

Fig 7: Train and Test Split

To split the data into training and testing, click 'Train Test split'.

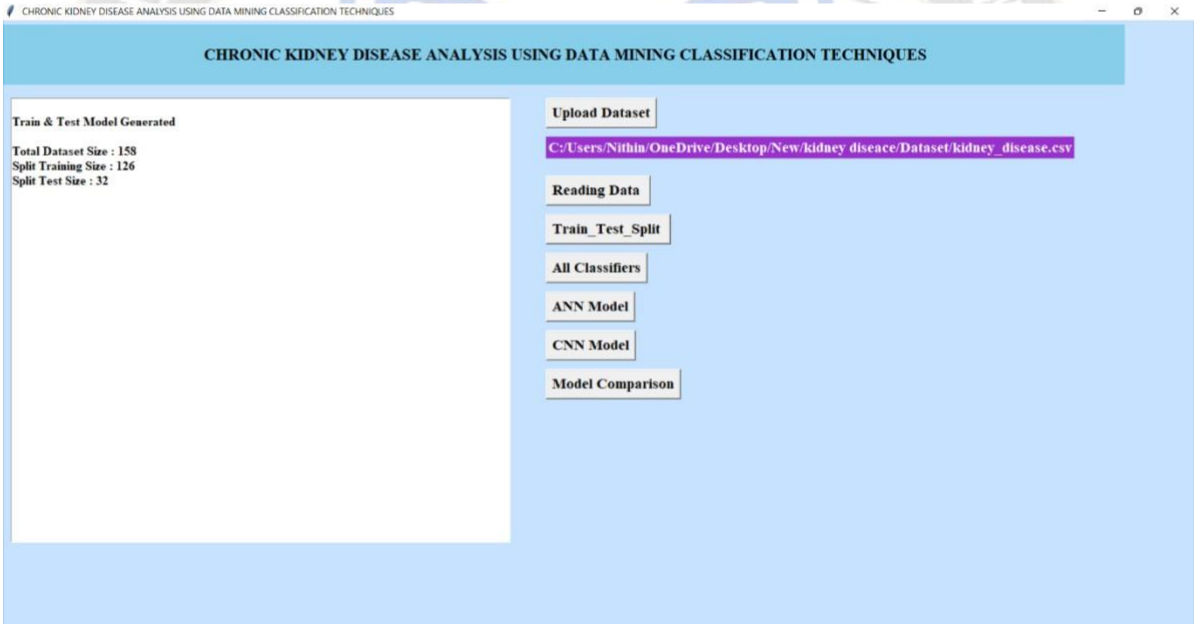

Fig 8: Train Data

To categorize the model, select 'All classifiers' from the drop-down menu.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

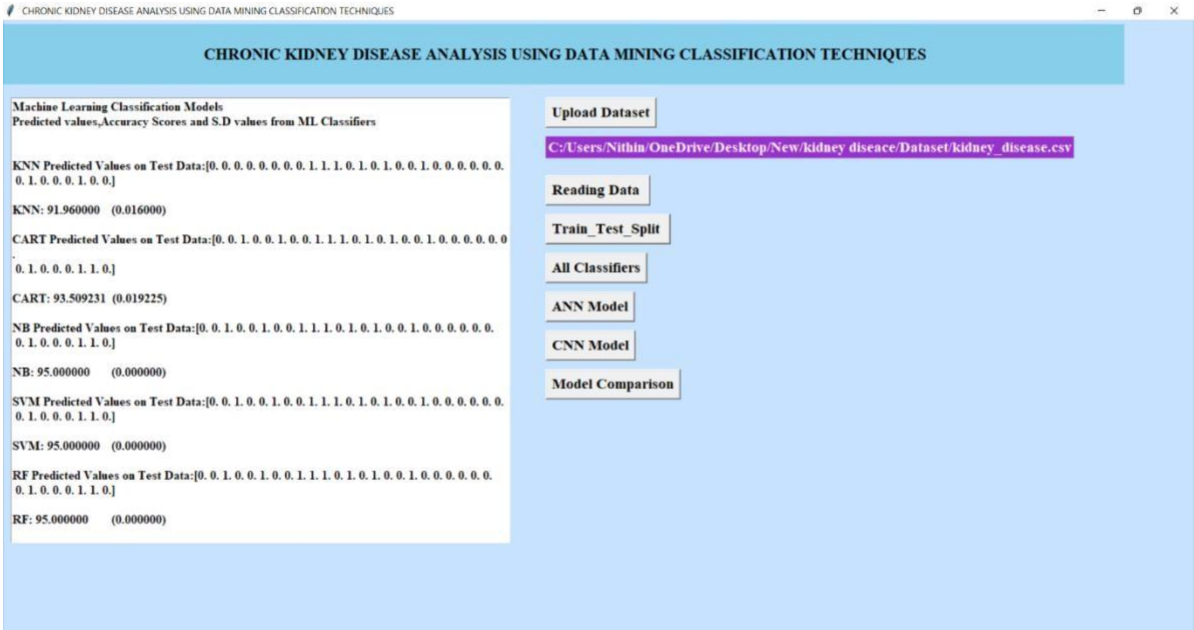

Fig 9: Test Data

KNN Predicted Values on Test Data is 90.22% CART Predicted Values on Test Data is 91.93% SVM Predicted Values on Test Data is 95%

RF Predicted Values on Test Data is 90.24% NB Predicted Values on Test Data is 95%

Now click on 'ANN Model'

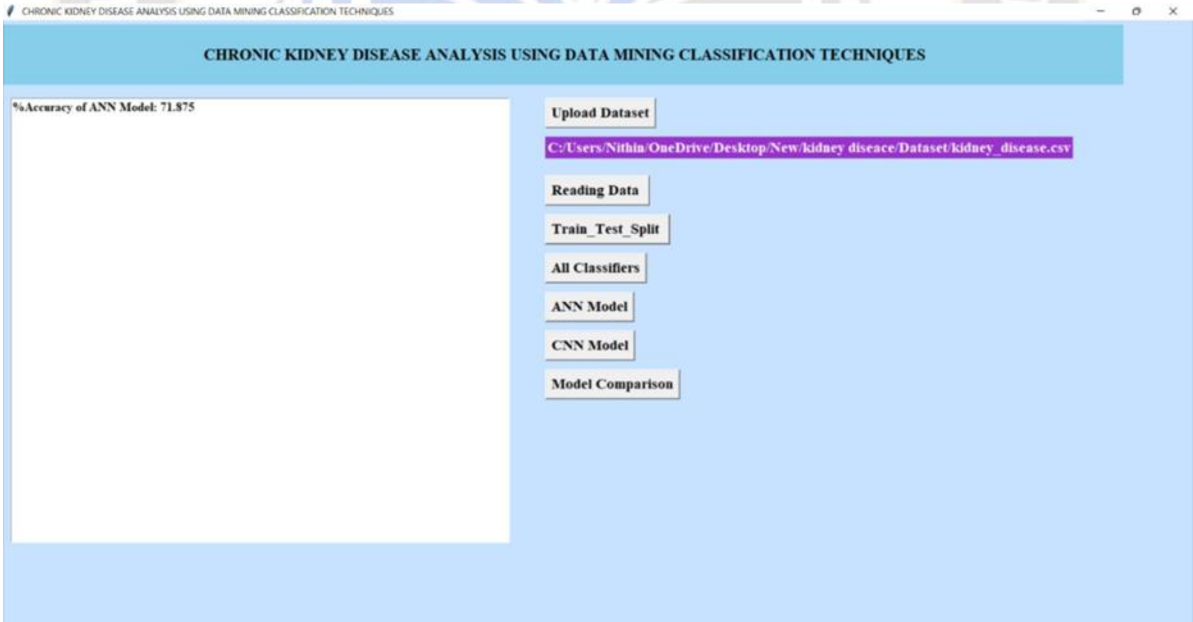

Fig 10: ANN Model Selection

Now select 'CNN Model' from the drop-down menu.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

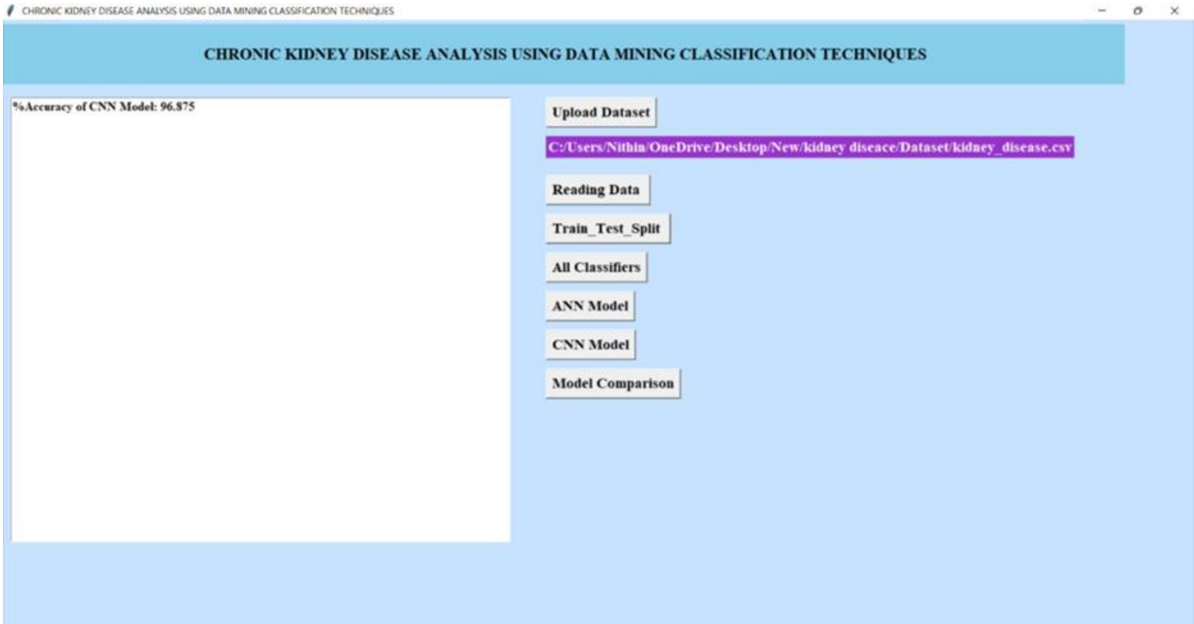

Fig 11: CNN Model Selection

Now select 'Model comparison' to see a comparison of the models.

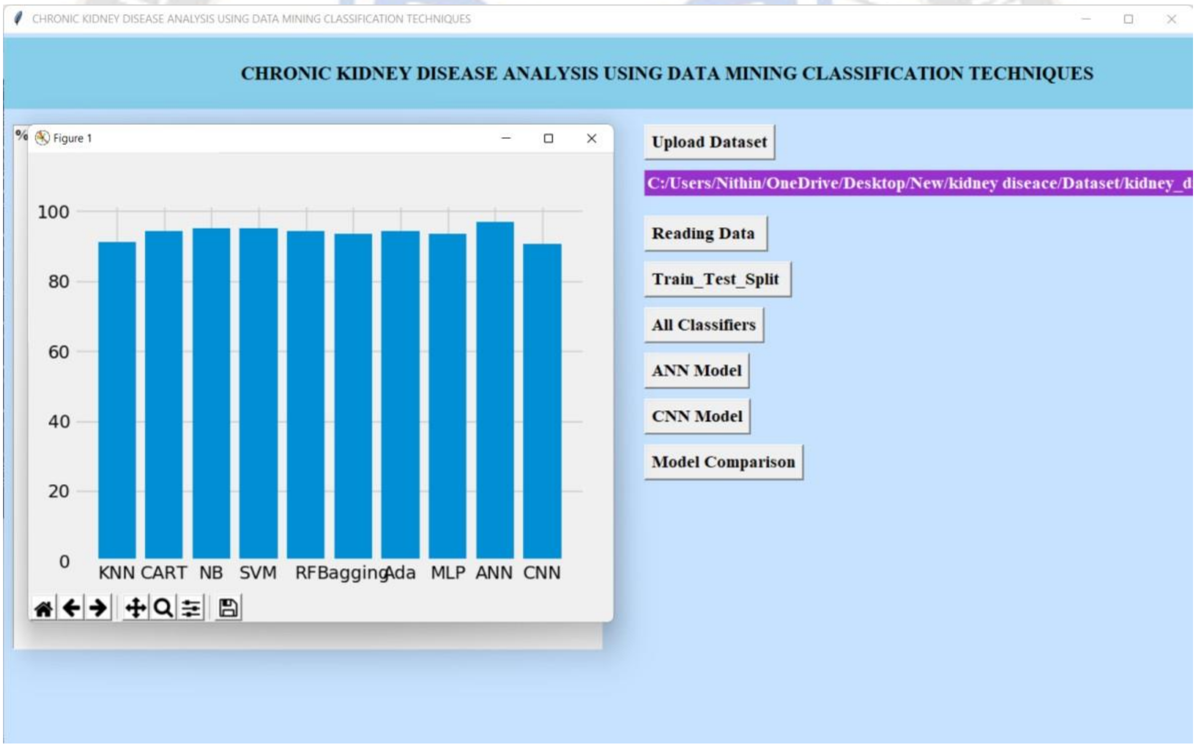

Fig 12: Different Models Comparison

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6632
***Article Received**: 22 February 2023 **Revised**: 28 March 2023 **Accepted**: 16 April 2023*

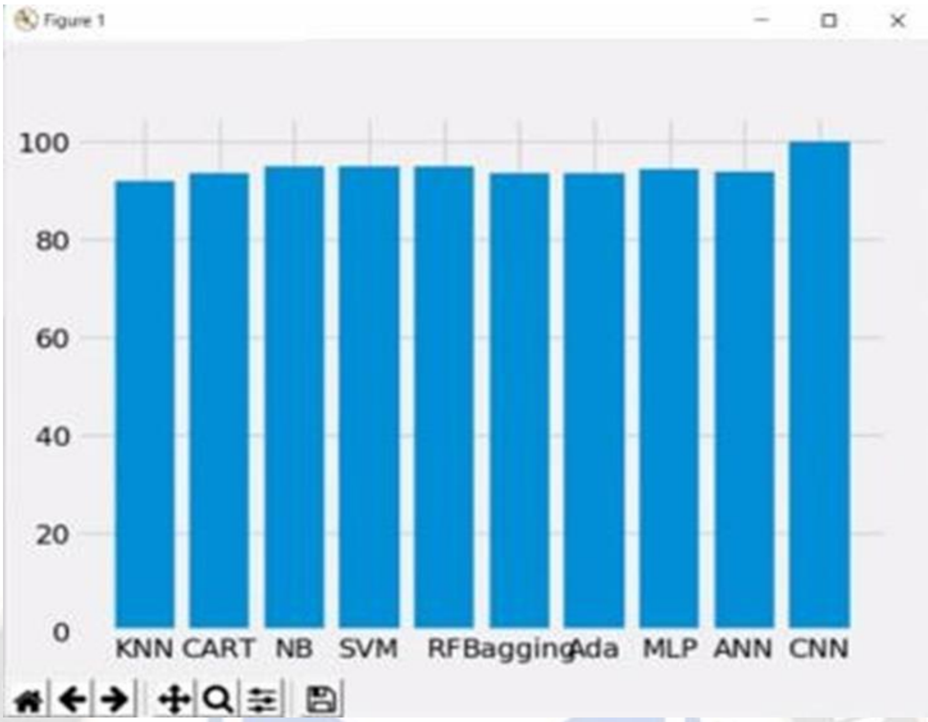

Fig 13: Comparison of Models

VII. CONCLUSION

Data mining classifiers such as ANN and Naive Bayes have been used to predict and diagnose chronic kidney disease. To evaluate the results of these methods, RapidMiner tools are busy. The results obtained show that CNN has an accurate classifier with an accuracy of 96.8% compared to ANN that has an accuracy of 71.73%. In this research study, some of the factors considered are age, diabetes, blood pressure, red blood cell count, etc. The work can be expanded taking into account other parameters such as the type of food, the work environment, living conditions, the availability of clean water, environmental factors, etc. to detect kidney disease. Further research can be done using other classifiers such as fuzzy logic, KNN, CNN, and ANN.

REFERENCES

[1] G. R. Vasquez-Morales, S. M. Martinez-Monterrubio, P. Moreno-Gier, and J. A. Recio- Garcia, Interpretive Predictability of Renal Disease in Colombian People Using Networks of Emotional and Cognitive Networks.

[2] Z. Chen, X. Zhang, and Z. Zhang. "Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models. Int. Urol Nephrol., vol. 48, no. 12, pp. 2069-2075, Jun 2016.

[3] Y Amirgaliyev, S. Shamiluulu, and A. Serek, "Analysis of chronic kidney disease dataset by applying machine learning methods," in Proc. IEEE 12th Int. Conf. Appl Inf. Commun. Technol(AFCT), Oct.2018, pp. 1-4.

[4] L. Kilvia De Almeida, L. Lessa, A. Peixoto, R. Gomes, and J Celestino, "Kidney failure detection using machine learning techniques," in Poe St Int. Workshop ADVANCES ICT Infrastructures Services, 2020, pp. 1-8.

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

[6] H. Polat, H. D. Mehr, and A. Cetin, "Huseyin polat & homay danari mehrl & aydin cetin. J. Med. Syst., vol. 41, no. 4, pp. 1-11, Apr 2017, doi: 10.1007/s10916-017-0703-x.

[7] M. Almasoud and T. E. Ward, "Detection of chronic kidney disease using machine learning algorithms with least number of predictors. Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 8, pp. 89-95, 2019.

[8] S. Vijayaranil and S. Dhayanand. "Kidney disease prediction using SVM and ANN algorithms, Int. J. Comput. Bus. Res., vol. 6, no. 2, pp. 1-12, Mar. 2015.

[9] F. Ma, T Sun, L. Liu, and H. Jing. Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artifi cial neural network, Future Gener. Comput. Syst., vol. 111. pp. 17-26. Oct. 2020.

[10] A. U Haq, I. P. Li, J. Khan, M. H. Memon, S. Nazir, S. Ahmad, G. A. Khan, and A. Aliss. Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data," Sensors, vol. 20, no. 9. p. 2649, May 2020.

[11] UH. Amin, J. Li, Z. Ali, M. H. Memon, M. Abbas, and S. Nazir, Recognition of the Parkinson's disease using a hybrid feature selection. approach, J. Iell. Fuzzy Syst, Jul. 2020.

[12] Dodge, Samuel, and Lina Karam. "Understanding how image quality affects deep neural networks.", Quality

Multimedia Experience (QoMEX), Eighth 2016 International Conference forTEEE, 2016.

[13] Dejean-Servieres. Mathieu, et al. "Study of the Impact of Standard Image Compression Techniques on Performance of Image Classification with a Convolutional Neural Network, Diss. INSA Rennes Univ Rennes; IETR; Institut Pascal, 2017.

[14] A. Krizhevsky, et al., "ImageNet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[15] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," The Journal of physiology, vol. 160, pp. 106-154, 1962.

[16] K. Fukushima and S. Miyake, "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition," in Competition and cooperation in neural nets, ed: Springer, 1982, pp. 267-285.

[17] Y. LeCun, et al., "Handwritten digit recognition with a back propagation network," in Advances in neural information processing systems, 1990, pp. 396-404.

[18] Y. LeCun, et al., "Gradient-based learning applied to document recognition," Proceedings ofthe IEEE, vol. 86, pp. 2278-2324, 1998.

[19] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in Neural networks for perception, ed: Elsevier, 1992, pp. 65-93.