International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 12 Issue 8 ǁ August 2024 ǁ PP. 75-82

Chronic Kidney Disease Stage Identification in HIV Infected Patients Using Machine Learning.

Dr. T.R. Arunkumar[1], Dr. Raju Hiremath[2], Mr. Sunil Jadhav[3]
[1] *Assistant Professor, Department of Computer Science, Rani Channamma University, Dr.P.G.Halakatti Post Graduate Center "Vachana Sangam", Vijayapura*
[2]*Teaching Assistant, Department of Computer Science, Rani Channamma University, Dr.P.G.Halakatti Post Graduate Center "Vachana Sangam", Vijayapura.*
[*3]*PG scholar, Department of Computer Science, Rani Channamma University, Dr.P.G.Halakatti Post Graduate Center"Vachana Sangam", Vijayapura.*

Abstract
Chronic Kidney Disease (CKD) is one of worldwide medical challenges with high morbidity and death rate. Since there is no symptom during the early stages of CKD, patients often fail to diagnose the disease. Patients with HIV have more chances to be affected with CKD in critical condition. Early detection of CKD helps patients to obtain prompt care ald delays the further progression of disease. With the availability of pathology data, the use of machine-learning techniques in health care for classification and prediction disease has become more common. This paper presents the classification of CKD using machine learning models. Based on the glomerular filtration rate, the CKD stages are also calculated for patients diagnosed with CKD. DNN model outperforms with 99% of accuracy in classifying CKD patients with HIV.

Keywords: *Chronic kidney disease ; CKD stage identification; Machine Learning, Support Vector Machine, KNN.*

Date of Submission: 09-08-2024 Date of acceptance: 21-08-2024

I. INTRODUCTION

CKD is an incurable condition of kidney associated with higher risk of many other diseases such as heart failure, anemia, and bone disease. Kidneys are very adaptable. However symptoms will reveal kidney damage slowly. In many cases, patients do not feel symptoms until disease is in last stage. Fig- 1 shows the common symptoms that is over lapped with other disease. Some forms of kidney disease are treatable by avoiding symptoms. In the patients to keep the disease from getting worse by restoring few kidney functions. Especially in case CKD, dailies is and kidney transplant are two major treatment options for end-stage kidney disease. Due to high treatment cost, only 10% of people receive delays is or kidney transplant worldwide [2]. Each year, more than one million individuals from 112 low-earning countries suffer and die due to kidney failure [5]. Patients having Acquired Immunodeficiency Syndrome(AIDS) have more complication in kidney disease due to deficiency of glomeruli filters also known as nephrons. The medication used for Human Immunodeficiency Viruses (HIV) can also infect the cells in kidney. It is very import ant to detect, control, progression of CKD in early stage. Increasing interest in automated diagnosis and rapid development in machine learning methods has played an important role in health care. Although many researches have used machine learning techniques to classify CKD in multiple stages. However, a few researcher has identified relation of CKD with HIV. In this paper, we have explored ML techniques and done experimental analysis to classify stages of CKD based on Glomerular Filtration Rate (eGFR)

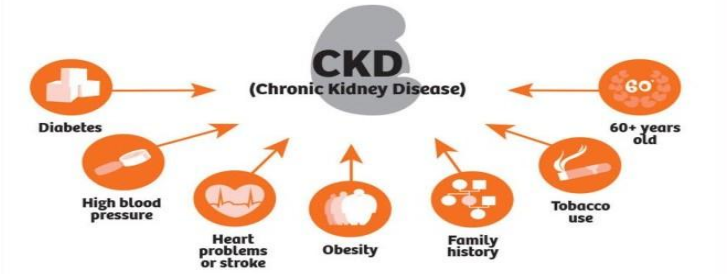

Fig. 1. Causes of Chronic Kidney Disease

Automated computer aided diagnose for CKD is a process of getting stage information using patient data such as age, blood pressure, blood test reports. Yu et al. [2] has utilized the Support Vector Machine (SVM) algorithm to recognize and anticipated and pre-diabetic patients. The outcome show that SVM is able to distinguishes patients with common diseases. E.Perumaletal. [6] has used the decision tree algorithm to predict the occurrence of heart disease, Naïve Bayes algorithm, and Probabilistic Neural Network (PNN) algorithm. It provides better results compared to other cardio vascular prediction algorithms. R. Shinde et al. [8] The Multilayered Perceptron(MLP) separator was used to predict HBV-induced hepatic cirrhoses is, and the findings indicate that the MLP se par at or provide sex cellist predictive results for liver disease, particularly in HBV-related patients with liver failure.

II. RELATED WORKS

Corinne Isnard Bagnis, Jack Edward Heron, David M. Gracey et al. [1] conducted are port on Chronic Kidney Disease and its connection to more deplorable out comes It shows that controlling blood pressure with angiotens in converting enzyme inhibitors and angiotens in receptor blockers slows the progression of CKD in HIV patients, particularly when protein uria is present. Y. Liu, J. Qin, C. Feng, L. Chen, C. Liu, and B. Chen et al. [2] Reveals that data imputation and sample diagnosis are possible with CKD. The integrated model presented in this paper can achieve sufficient accuracy using the KNN algorithm. Since the dataset contains two classes, Chronicle Kidney Infection and Not Chronic Kidney Disease, the model cannot investigate the stages of chronic kidney disease .A. S. Anwar and E. H. A. Rady et al. [3] The uses of lab data set of 361 persistent kidney sickness patients. It uses PNN, SVM, and MLP algorithms to calculate period of chronic kidney sickness. This examination suggests that the probabilistic neural organization calculation is best performing calculation that can be utilized by doctors to kill demonstrative and treatment mistakes. M. N. Amin, A. Al Imran and F. T. Johora et al. [4] analyze model performance on real (imbalanced) data and model performance on over sampled (balanced)data using logistic regression and feed forward neural networks. Feed for ward neural networks showed the best results for both real and over sampled data, with 0.99 Recall, 0.97 Precision, 0.99 F1-Score and 0.99 AUC score. K. S. Vaisl a, N. Chetty and S. D. Sudarsan et al. [5] recommended On the CKD dataset, attribute assessment and classification models were used. The attribute evaluator model performed better by decreasing the number of attributes from 25 to 6, 12, and 7. P.Arulanthu and E. Perumal et al. [6] utilizes JRip, SMO, Naive Bayes, algorithms and analyses that JRip generate best performance. P. Manickam, K. Shankar, M. Ilayaraja and G. Devika et al. [7] Uses Ant Lion Optimization (ALO) technique to choose ideal features for classification. This optimization results in better classification accuracy for deep neural network. R. Shinde, Maurya, R. Wable, S. John, R. Dakshayani and R.Jadhav,et al. [8] To slow the progression of CKD and to follow the recommended diet plans, use the potassium zone, which is computed using blood potassium levels. R. Yadav and S. C. Jatetal. [9] Investigate the relation of various methods of selection and dimensionality reduction to the performance of chronic disease classification and prediction.

III. MODEL SELECTION

There are several machine learning algorithms used in literature for CKD classification. In this paper, we have built 6 ML models using, KNN, SVM, random forest, decision tree, ada-boost and xg-boost algorithms, along with a simple deep neural network to classify weather a patient has CKD or not. The flow of the proposed experimental setup is depicted in Fig-2 For binary classification situations, A SVM (Support Vector Machine) is a classification-based supervised machine learning model. K-nearest neighbors (KNN) algorithm utilizes feature comparing to predict a value according on how closely it is similar in the training dataset. A decision tree is used to visually represent decisions of classification. Often, a single decision tree is not sufficient for producing effective classification accuracy. Random Forest algorithm solves this problem by leveraging multiple decision trees. AdaBoost algorithm, also called adaptive boosting, is a boosting technique

used as an ensemble method in machine learning. It aims to convert a set of weak classifiers into a strong one by reassigning the weights to each instance. XG Boost (eXtreme Gradient Boosting) is another boosting algorithm that uses a gradient boosting frame work. Other than machine learning, Many researcher have utilized feature based deep neural network (DNN) for better classification results. Deep Neural Networks are capable of detecting crucial diseases in they use several layers of nodes to accomplish high-level functions from input data. Before applying classification algorithm, we have eliminated few features using feature selection method.

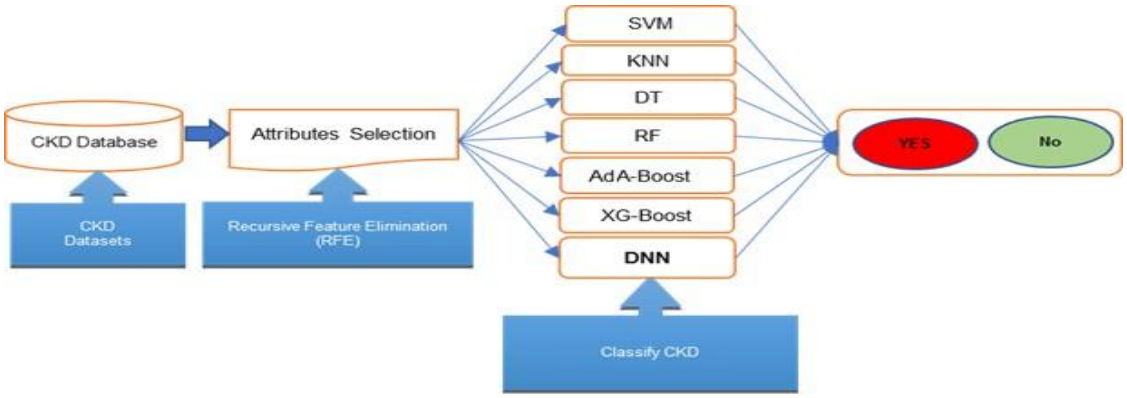

Fig.2. Proposed System

A. CKD Dataset

The dataset is obtained from the University of California, Irvine's Machine Learning Repository. There are 26 features in the dataset, 20 of which are numeric and 6 of which are categorical. This data set contained 158 unique cases. Following Table 1, list out the attributes of dataset we used for experiment.

TABLE I. ATTRIBUTES OF HIV CKD PATIENT DATASET

Sr no	Data	Type
1	Age	Numerical
2	Gender	Categorical
3	Ethnicity	Numerical
4	Blood Pressure	Numerical
5	Specific Gravity	Numerical
6	Albumin	Numerical
7	Sugar	Numerical
8	Red Blood Cells	Numerical
9	Pus Cell	Numerical
10	Pus Cell clumps	Numerical
11	Bacteria	Numerical
12	Blood Glucose Random	Numerical
13	Blood Urea	Numerical
14	Serum Cretonne	Numerical
15	Sodium	Numerical
16	Potassium	Numerical
17	Hemoglobin	Numerical
18	Packed Cell Volume	Numerical
19	White Blood Cell Count	Numerical
20	Red Blood Cell Count	Numerical
21	Hypertension	Numerical
22	Diabetes Mellitus	Categorical
23	Coronary Artery Disease	Categorical
24	Appetite	Categorical
25	Pedal Edema	Categorical

26	Anemia	Categorical
27	Class	Categorical

B. Attributes Selection

RFE, or Recursive Feature Elimination, is a wide spread attributes selection algorithm that selects the features (columns) in a training dataset that are more or more important in predicting the target variable. When using RFE, there are two essential configuration options: the number of features to select from and the algorithm used it to help choose features. Both of these hyper parameters can be examined, although the accuracy of the method is not strongly dependent on these hyper parameters being configured. In our experiment the feature set is reduced to 14 attributes

C. CKD classification

For classification of HIV patients with CKD or not KNN, SVM, random forest, decision tree, xg-boost, ada-boost algorithms and a deep neural network has been implemented

D. CKD Stage identification

After classification of CKD patients in to two class CKD and non-CKD, stage identification is done for patient having CKD. There are 6 stages of CKD based on eGFR as shown in Table

TABLE II. CHRONIC KIDNEY DISEASE STAGE

Stages	Explanation	GFR
One	Normal damage of kidney function	>90%
Two	Minor damage of kidney job	89-60%
Three(A)	Minor to Modes damage	59-45%
Three(B)	Modes to simple damage	44-30%
Four	Simple damage of kidney meaning	29-15%
Five	Kidney Stop Working.	<15%

The phases of progressive kidney illness are assessed using glomerular filtration rate

eGFR = 175 x (Creatinine/88.4)^1.154 x (Age) ^0.203 x (0.742iffemale)x(1.210 if dark) **(1)**

According to above equation (1), creatinine, sex, and age of patients are used to calculate eGFR. Alteration of Diet in Renal For example we have value of Creatinine=4.1, Age=68, ethnicity=black, Gender= Male, then the value of eGFR= 175x (4.1/88.4) ^1.154x (68) ^0.203x1.210= 14.41. As the value of eGFR lies in the range of 0 to 15, so the patient will be considered in 5th stage of CKD.

IV. RESULTS AND ANALYSIS

In this section, we have reported the results of each model separately and shown the comparative analysis based on parameters such as model accuracy, precision, recall confusion matrix.

$$Accuracy = \frac{TN+TP}{FP+TP+TN+FN} \quad \textbf{(2)}$$

$$Precision = \frac{TP}{FP+TP} \quad \textbf{(3)}$$

$$Recall = \frac{TP}{TP+FN} \quad \textbf{(4)}$$

Where, True Negative(TN), True Positive(TP),False Negative(FN) and False Positive (FP).

Confusion Matrix is one of the tool for evaluating the behavior a Binary Classifier. For better visualization of results, we have used heat maps for each model as shown in Fig-3 to 9.

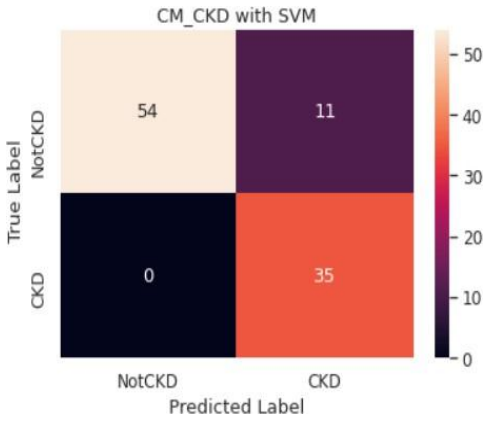

Fig .3 .SVM Classifier Result

As visible in Fig- 3 SVM is tested on 100-sample among 54-sample classified as non-CKD, 35-sample as CKD and remain 11-classify as False.

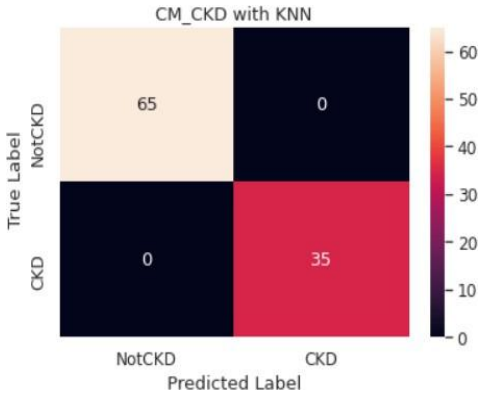

Fig.4. KNN Classifier Result

As visible in Fig- 4 KNN is tested on 100-sample among 65-sampleclassified as non-ckd, 35-sampleas CKD.

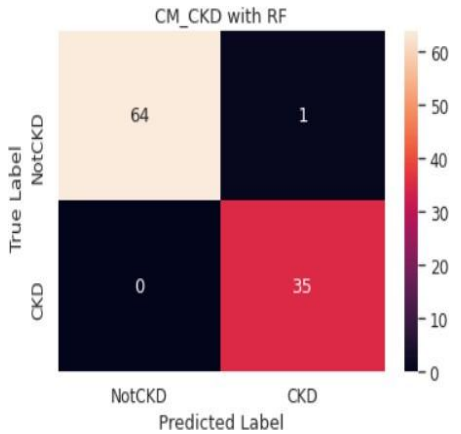

Fig.5 .Random Forest Classifier Result

As visible in Fig-5 RF is tested on 100- sample among 64 –sample classified as non-ckd, 35-sample as CKD and remain 1-classify as false.

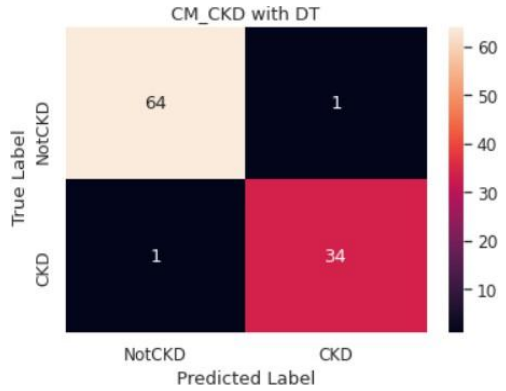

Fig.6.Decision Tree Classifier Result

As visible in Fig-6 DT is tested on 100-sample among 64-sample classified as non-CKD, 34-sample as CKD and remain 2-classify as false.

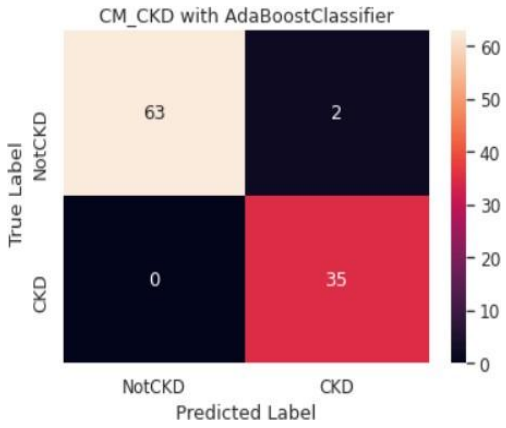

Fig.7.Ada-boost Classifier Result

As visible in Fig- 7 Ada-Boost is tested on 100-sample among 63-sample classified as non-ckd, 35-sample as CKD and remain2-classifyas false.

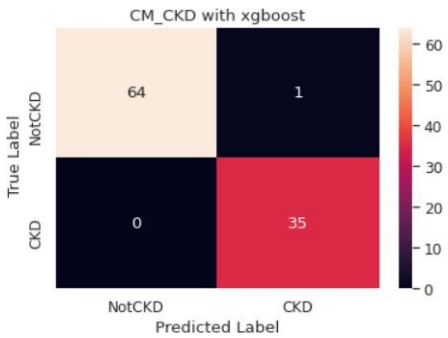

Fig.8. XG-Boost Classifier Result

As visible in Fig- 8 XG-Boost is tested on 100-sample among 64-sample classified as non-CKD, 35-sample as CKD and remain1- classify as false.

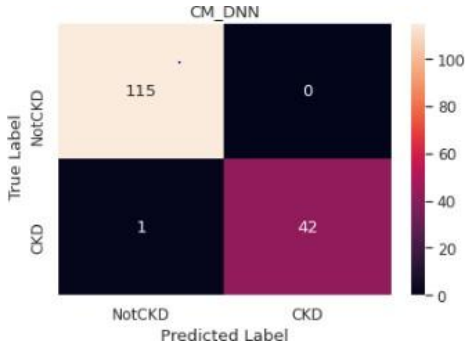

Fig.9. Deep Neural Network(DNN) Result

For DNN model we have used all 24 attributes for training. The heat map shown in Fig- 9 shows the classes of total 158 instances. The results indicate that DNN has achieved high performance, even though it uses 24 attributes in comparison to other techniques that uses only 14 attributes. SVM has not performed well in comparison, other Machine Learning Algorithms.

TABLE III. ACCURACY, PRECISION AND RECALL COMPARISON

Classifier	Attributes	Accuracy(%_)	Precision(%)	Recal(%)
SVM	14	93	91	92
KNN	14	97	95	96
DT	14	97	96	96
RF	14	95	95	94
AdaBoost	14	97	96	97
XgBoost	14	97	95	96
DNN	24	99	99	98

Above Table III shows comparative study of different classifier with respect to accuracy, precision and recall as parameters. Fig- 10 shows the graph of above comparison table .DNN gives almost 99% accuracy for stage classification.

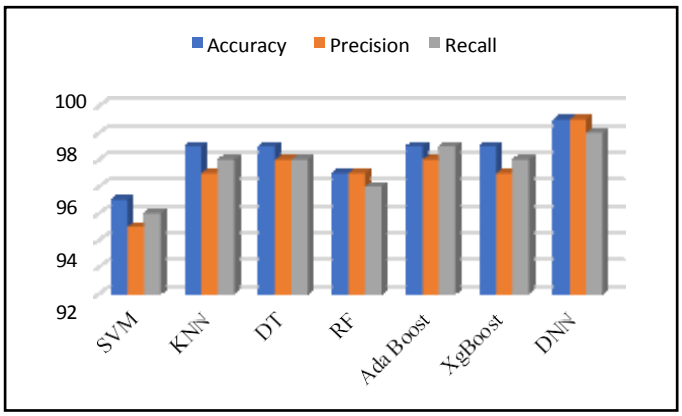

Fig.10.Analysis graph of ML and DL methods

After successful classification of CKD patient's stage of disease is calculated using eGFR. Total 44 patients out of 158 instances have CKD. The CKD progression of all 44 patients has identified in 6 stages as given in Table 4.

TABLE IV. STAGE CLASSIFICATION USING E-GFR

Stage	Stage 1	Stage 2	Stage 3A	Stage 3B	Stage 4	Stage 5
Patient	1	1	2	5	10	25

V. CONCLUSION

Classification of Chronic Kidney diseases stage in HIV infected patient are extremely useful to patients as well as doctor for timely and accurate clinical decisions. In this paper we have compared the performance of state of art machine learning algorithms along with DNN for classification of CKD for patients having HIV. Our study indicates that DNN has outperformed in CKD classification. We have also shown the use of eGFR formula to identify stages of disease. In future, features based DNN can be combined with medical image analysis to support diagnosis based on different imaging modalities.

REFERENCE

[1] I. Bagnis, J. E. Heron, and D. M. Gracey, "Contemporary issues and new challenges in chronic kidney disease amongst people living with HIV," AIDS Res. T her., vol. 17, no. 1, pp. 1 –13, 2020, doi:10.1186/s12981-020-

[2] Y. Liu, C. Liu, J. Qin, L. Chen, C. Feng, and B. Chen, " A machine learning met homology for diagnosing chronic kidney disease," IEEE Access, vol. 8, pp. 20991–21002, 2020, doi:10.1109/ACCESS.2019.2963053.S. Anwar and E. H. A. Rady , " Predict ion of kidney disease stages using data mining algorithms," Informatics Med. Unlocked, vol. 15, no. March, p. 100178, 2019, doi: 10.1016/j.imu.2019.100178.

[3] M. N. Amin, A. Al Imran, and F. T. Johora, "Classification of Chronic Kidney Disease using Logistic Regression, Feed forward Neural Network and Wide Deep Learning," 2018 Int . Conf. Innov. Eng. Technol. ICIET 2018, pp. 1–6, 2019, doi: 10.1109/CIET.2018.8660844.

[4] S. D. Sudarsan, N. Chetty, K. S. Vaisla, "Role of attributes selection in classification of Chronic Kidney Disease patients," pp. 1 –6, 2016, doi:10.1109/cccs.2015.7374193.

[5] Perumal , P. Arulanthu , "Predicting the Chronic Kidney Disease using Various Classifiers," 4t h Int. Conf. Electr. Electron.Commun.Comput. Technol. Optim. Tech. ICEECCOT 2019, pp. 70–75, 2019, doi:10.1109/ICEECCOT46775.2019.9114653.

[6] K. Shankar, G. Devika, P. Manickam, M. Ilayaraja, "Optimal Feature Selection for Chronic Kidney Disease Classification using Deep Learning Classifier," 2018 IEEE Int. Conf. Comput. Intell.Comput . Res. ICCIC 2018, pp. 1–5, 2018, doi: 10.1109/ICCIC.2018.8782340.

[7] R. Shinde, S. John, R. Jadhav, A. Maurya, R. Wable, and R.Dakshayani, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet P lan by using Machine Learning," 2019 Int . Conf.Nascent Technol. Eng. ICNTE 2019 - Proc., no.Icnte, pp. 1–4, 2019,doi: 10.1109/ICNTE44896.2019.8946029.

[8] S. C. Jat . R. Yadav, "Feature selection and dimensionality reduction methods for chronic disease prediction," Int. J. Sci. Technol. Res., vol.9, no. 4, pp. 2912–2918, 2020.

[9] S. saraswat, S. Vashisth, I. dhall, CKD diagnosis using Multilayer Perceptron Classifier' 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)" 2020.

[10] K. Shankar, M. Elhoseny, J. Uthayakumar, "Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease," Sci.Rep., vol. 9, no. 1, pp. 1–14, 2019, doi: 10.1038/s41598-019-46074-2.

[11] M. H. AbdWahab, A. A. Johari, A. Must apha, "Two-Class Classification: Comparative Experiments for Chronic Kidney Disease,"2019 4th Int. Conf. Inf. Syst .Comput. Networks, ISCON 2019, pp. 789–792, 2019, doi: 10.1109/ISCON47742.2019.9036306.

[12] G. Saini, B. Pandey, A. Khamparia, S. Tiwari, D. Gupta, and A.Khanna, " KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked auto encoder network,"Multimed. Tools Appl., 2019, doi: 10.1007/s11042-019-07839-z.

[13] S. Maheshwari, S. Singhal, and M. Meena, Recent Findings in Intelligent Computing Techniques, vol. 707. Springer Singapore, 2019.

[14] V. Sharma, S. Sharma, A. Sharma, "Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis," 2016, [Online]. Available: http://arxiv.org/abs/1606.09581.

[15] Learn More About Kidney Disease: Types, Symptoms, and Causes,Siemens-healthineers.com, 2021. [Online]. Available:https://www.siemens-healthineers.com/en-us/clinicalspecialities/kidney-disease/about-kidney-disease.

[16] Estimated Glomerular Filt rat ion Rate (eGFR), National Kidney Foundation, 2021. [Online]. Available:https://www.kidney.org/atoz/content/gfr. [Accessed: 04- Mar- 2021]

[17] The dawn of a new era in chronic kidney disease, Astrazeneca.com,2021. [Online]. Available: https://www.ast razeneca.com/what-sciencecan-do/stories/the-dawn-of-a-new-era-in-chronic-kidney-disease.html.[Accessed: 04- Mar- 2021]

[18] Applications of Deep Learning in Business - Algorithm-X Lab, Algorithm-X Lab, 2021. [Online]. Available :https://algorithmxlab.com/blog/10-applicat ions-deep-learning-in business/.[Accessed: 04- Mar- 2021]

[19] Check out this amazing template to make your presentations look awesome at, Slide team.net , 2021. [Online]. Available:https://www.slideteam.net/examples-of-deep-learning-applications-pptpowerpoint-presentation-introduction.html. [Accessed: 04- Mar- 2021]

[20] Wang, Huafeng et al. 'A Hybrid CNN Feature Model for Pulmonary Nodule Malignancy Risk Differentiation'. 1 Jan. 2018 : 171 – 187.

[21] How Deep Learning Can Help Prevent Financial Fraud, Investopedia,2021. [Online]. Available: https://www.investopedia.com/terms/d/deeplearning.asp. [Accessed: 04- Mar- 2021]