Estimation of Accuracy through Deep Learning Based Classifiers Algorithms for CKD Prediction

Vipin Rai [1], Dr. P.K. Bharti [2]
Ph.D. Research Scholar [1], Professor & Vice Chancellor [2]
School of Engineering & Technology [1]
Shri Venkateshwara University, Gajraula, UP [1, 2]

Abstract: A deep learning computer plays an important role in predicting the presence or absence of movement disorders and kidney disease. The resting part of the body as compared to the kidneys, is the largest and most concentrated organ in the human body. Data analysis helps in predicting kidney disease in the medical field is an important task. Data analysis helps predict from a lot of information and helps medical institutions to predict various diseases. The main purpose of this research paper is to compare the performance of different classification algorithms for the diagnosis of kidney disease. Large amounts of patient-related data are stored on a monthly basis. Databases can be used to predict the occurrence of a future disease. It is curious to guess the disease using the latest diagnostic report. In the world of science there are various tools or learning methods already tested in different types of data sets. This research study examines the best learning algorithm for predicting kidney disease using various learning tools. This paper uses three different data sets to determine the accuracy of the prediction by the accuracy rate. The data sets were taken from kagrag and the UCI study machine with more than 310 databases each. This paper investigates the accuracy, accuracy and F1-value in each data set using various learning algorithms. This paper investigates eleven highly variable algorithms namely Logistic Regression, KNN Prediction, Tree Decision, Random Tree, SVM Prediction, Gaussian NB, Linear Discriminant Analysis, Ada Boost Classifier Gradient Boosting Classifier, -Quadratic Discriminant Analysis with MLP Classifier. And this have compared each result from the data set and it is present in visual form.

Keywords: K-Neighbor Classifier, Vector Classifier, Tree Decision Classifier, Random Forest Classifier, Deep Learning, Kidney Disease Prediction.

I. Introduction

Chronic kidney disease (CKD) (Salekin et. al., 2016) investigation based on the patient data base is usually cover under data mining and analysis (Xia et al., 2020). Recently various scientist work to predict the disease using deep learning techniques (Ma et. al., 2020). On the collective diagnosis report of large number of patient it is possible to predict their behaviors. Medical centers around the world collect information about various health problems. It can use various methods of deep learning to enter this data for useful information. However, the data collected is very large and in many cases the data can be very noisy. These confusing details can be easily explored using different techniques of deep learning. Therefore, these algorithms have recently been very helpful in accurately predicting the presence or absence of kidney-related diseases.

A. Prediction of kidney disease

Deep computing algorithms primarily use the dimension reduction method to reduce the data set. It is the first step and the most basic procedure used to filter important data sets that have the greatest impact on disease. This prediction includes the next step in which the data is already made for prediction (Ju et al., 2008).

- To summarize the most important data
- Missing value treatment (Replace the shapes or median values of the blank spaces)
- Divide data sets into two parts
- One as a test data and the second as a train data set
- Use good order over data sets
- Get more precision
- Find the best accuracy algorithm.

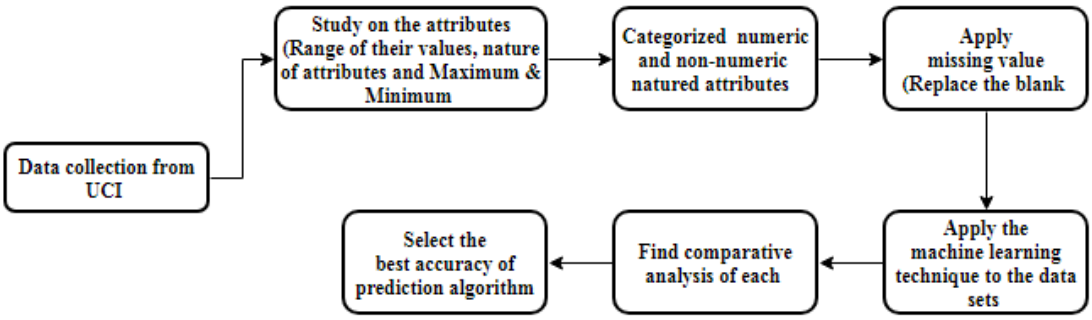

Fig 1: Process of prediction and analysis

B. Depreciation

Data analysis of CKD involves selecting a mathematical representation so that most but not all the distribution of specific data is combined to include the most important information. Information that is deemed to be a work or problem may contain many or many dimensions, but not all of the relevant aspects. A large number of symptoms or functions may affect the complexity of the complex or cause excess, with serious consequences. Therefore, reducing the size of given CKD data is a very important step that needs to be considered when building a model.

C. Release Factor

In this case, the new feature set is based on the original feature set. Feature extraction involves feature conversion. This change is often unrepentant. This change no useful information is lost in the process. Principal component analysis (PCA) was used for feature extraction. The principal component analysis is a general transformation algorithm. In the feature field, find the index with the largest variance and find the corresponding coordinates. This is a global algorithm that provides the best multiplication.

D. Feature Selection

In this case, the bottom set of the original feature is selected. The features are selected using a combination of CFS (Correlation based Feature Selection) and a good size reduction method. The chi-square test is used to select the most important factors.

II. Research Background

Tekale et al. (2018), in this paper they have studied different machine learning algorithms. They have analyzed 14 attributes related to CKD patients and predicted the accuracy of different machine learning algorithms such as Decision of the graduate and support machine. From the analysis of the results, it can be seen that the decision tree algorithms provide 91.75% accuracy and the SVM gives 96.75% accuracy. To create a machine learning model that identifies chronic kidney disease with an overall accuracy of 99.99%, it would require millions of records with zero non-zero.

Alshebly & Ahmed (2019), the purpose of this study is to compare the performance of Artificial Neural Networks (ANNs) and Logistic Regression (LR) categories on the basis of the following criteria: Accuracy, Sensitivity, Specification, Positioning, and Areas under the curve (ROC) in the prediction of CKD. From the experimental results, it is considered that the performance of the ANNs classifier is better than the Logistic Regression model. With an accuracy of 84.44%, sensitivity of 84.21%, specificity of 84.61% and AUCROC of 84.41%. Also, with the final integrated models used, the most important factors that have a clear impact on chronic kidney disease patients are creatinine and urea.

Norouzi et al. (2016), proposed the prediction analysis for kidney failure. This model may accurately describe more than 95% accuracy of prediction. This paper actually limitation work considering the GFR value at much extent in time interval.

Arasu & Thirumalaiselvi (2017, they proposed one dimensional predictive analysis. Basically this model based on the regression and random forest. For big data application there are already RF, Classified and Regression Tree and C4.5 works well. In the exploration of kidney prediction they proposed a new hybrid algorithm which enhance the prediction graph. This model assign the importance of each attribute to a dataset and to carry the priorities of the classification process.

Ravindra et al. (2018), in this paper, the effectiveness of the proposed scheme is evaluated in terms of the sensitivity, specificity and accuracy of the phases. The results reveal the accuracy of the total separation of 94.44% obtained by combining the 6 attributes. It can be concluded that the SVM-based method found may be a member of CKD and NCKD.

Rahman et al. (2019), this paper introduces the concept of detecting the presence of kidney disease using a computerized systematic study model, considering the patient's ECG signal. In their research, they found an accuracy rate of 97.6% which is an average using both the QT and RR intervals, comparing the accuracy obtained when using one of the signals.

Hore et al. (2018), in this article, it has been proposed that in the medical network a neural-trained algorithm to detect early kidney disease (CKD). The results revealed that NN-GA was more active than others in the past and was able to detect CKD more effectively. Future research could focus on studying some such teaching methods for training NNs to improve the performance of NNs in real applications.

Khamparia et al. (2019), this research paper presents a framework for the number of deep neural pathology of kidney disease (CKD) accident using focal auto encoder installation. The model is built using embedded auto codes that contain two auto codes arranged in a fashionable form with one deep max component.

Sisodia and Verma (2017), in this paper, the comparative performance of specific and integrative students is analyzed in the CKD data set of the UCI reading library. The performance of CKD using the predictive analysis need data collection, filtering of data, missing value treatment. They model give accuracy, recall, f-rate, and the ROC-Curve as metrics to compare classroom performance. They are also developing the Wrapper and hybrid method to select the desire feature.

Subasi et al. (2017), in this study, different machine learning methods are used to diagnose CKD. This study has shown that RF isolation results in significantly higher performance during the partition operation.

Zhang et al. (2018), in this work, they build two virtual neural networks, one of which is classical MLPs, and the other combines the selection of the LASSO feature and is highly functional with a small but deep structure. The results indicate that the performance of each model is the same, both of which obtain high accuracy.

III. Architecture of the Experiments

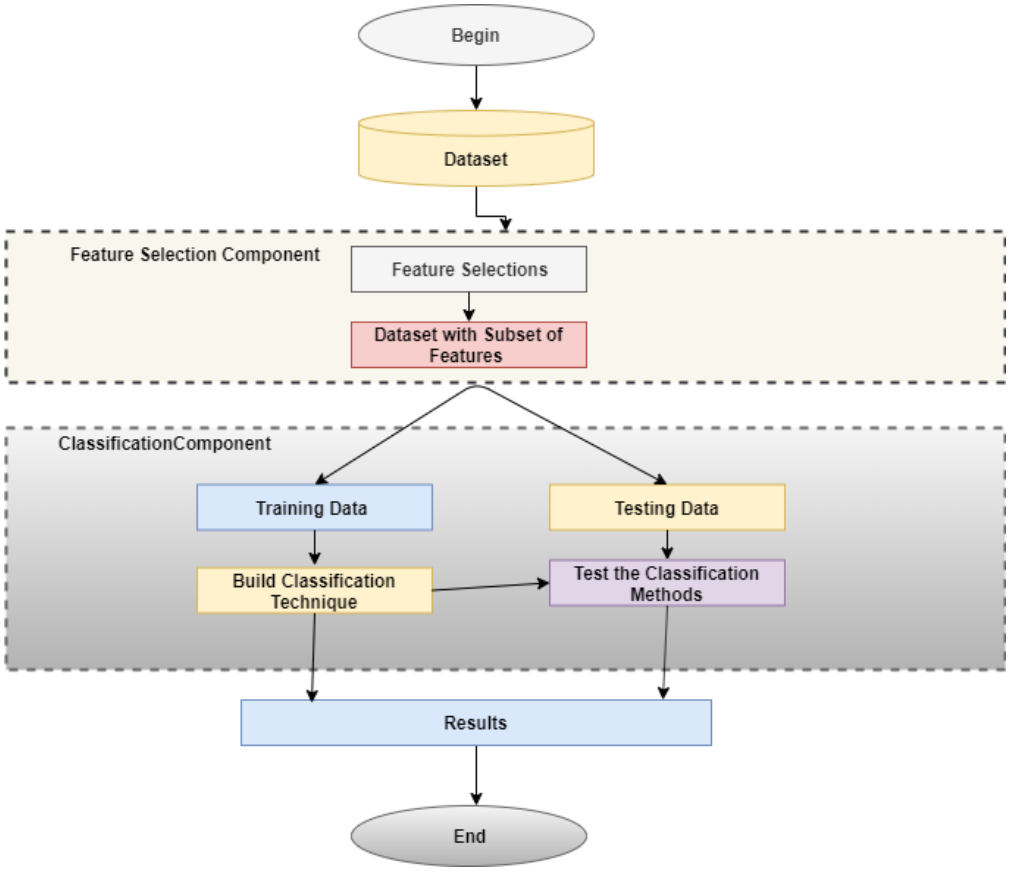

Fig 2: Architecture of the Deep learning experiments

IV. Deep Learning Algorithms for Investigation of CKD Data

A. Curvilinear Regression

Name retrieval can be defined as measuring and analyzing the relationships between independent or dependent variables. Dissolution can be divided into two categories: Direct undo and direct undo. The normalization of normal thinking is reproduced with a straight line. It is primarily used to measure the relative or multidimensional variables (Kahraman et al., 2004).

The translation of rational insight can begin with the interpretation of general cognitive functions. A logical function is a sigmoid function that takes a real value between 0 and 1. It is defined as

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

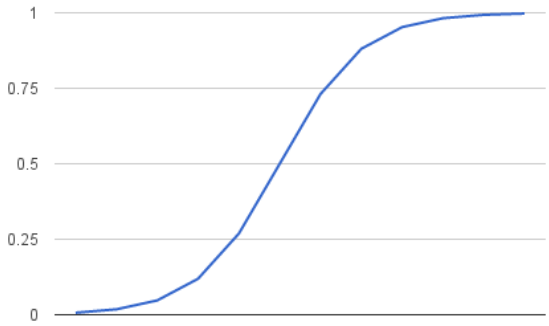

Fig 3: Representing the non-linear curvilinear regression

Let's take a look at the functionality of a line in an unequal return model.

$$t = \beta_0 + \beta_1 x$$

So the non-linear curvilinear regression Equation becomes

$$p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

B. Support Vector Machine

Support vector machines are a very popular way of studying the monitored machine (using pre-programmed target objects) that can be used as processors and predictors.

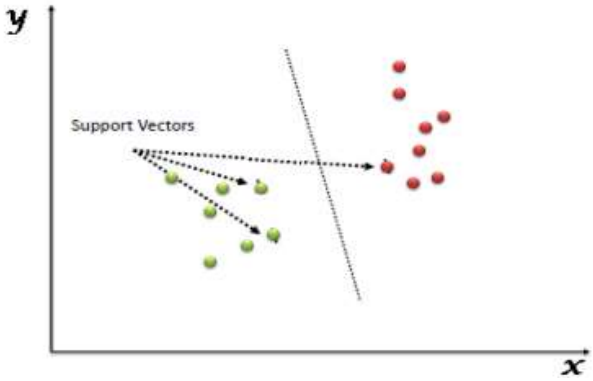

Fig 4: A presentation of SVM

It is the extension of the linear regression with involvement of clusters. A main trend line iterate and continue till the optimization reach (Suthaharan, 2016)

C. K - Nearest Neighbor

K-Near method is one of the best ways to differentiate data. It is used for non-swearing classification functions about data and little or no information about the distribution of data. The algorithm finds the nearest k points closest to the data points

not found in the training set and the amount of data points obtained from it. In classification settings, the nearest neighbor algorithm basically creates the most votes among the most similar K conditions for some "intangible" recognition.

$$d(x, x') = \sqrt{(x_1 - x_1')^2 + \cdots (x_n - x_n')^2}$$

However, there are some metrics that are appropriate for specific settings, such as Manhattan, Chebyshev, and the Hamming distance.

D. Decision Tree

This method is used mainly for separation problems. Make it easy with continuous signs and paragraphs. This algorithm divides the population into two or more similar sets based on the most important predictors. The algorithm tree first lists the entries for each attribute. The data set is segmented by a variable or predictor with the largest or shortest gain. These two steps are performed in duplicate with the remaining structures

$$Entropy(S) = \sum_{i=1}^{c} -p_i \, log_2 p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

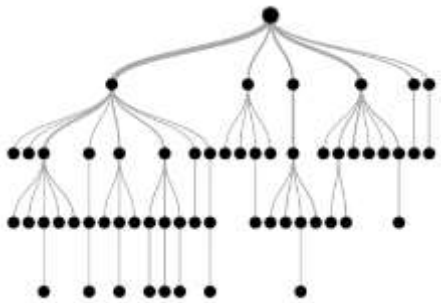

Fig 5: Shows the Decision Tree

E. Random Forests

Random Forest is also a supervised learning algorithm. This method can be used for recording and classification functions, but generally improves the performance of classification functions. As the name implies, a random forest path looks at many decision trees before making a result.

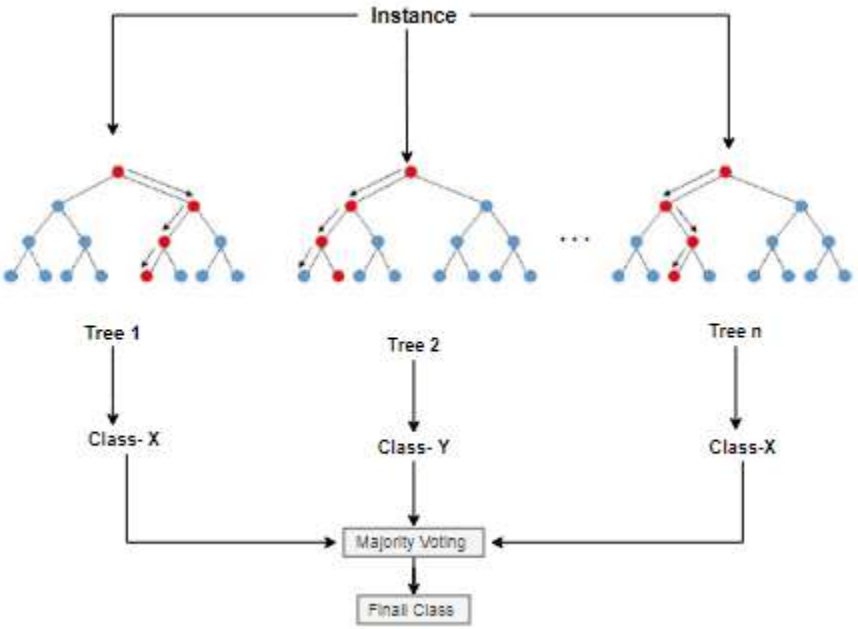

Fig 6: Random Forest

Random Forest (RF) creates many decision trees during training. Summarize all tree predictions with the final prediction; Partition model or average forecast when making final decisions using a set of results, a combination of techniques is used.

F. Gaussian NB

Gaussian Naive Bayes is simply understand as simple distribution with control curve. This algorithm based on the conditional probability based on the occurrence of data under Gaussian plane.

G. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is the most widely used methods for reducing size and separation. With given set of data their feature has been estimated and centroid and covariance extract the reducible sets of data (Tharwat et. al., 2017).

LDA makes simple assumptions about your data:

1. That your data is Gaussian, that each variable is formed as a bell curve when plotted.
2. For each attribute having the same value, those values of each variable differ by the mean of the same value on average.

H. Ada Boost Classifier

AdaBoost is short for Adaptive Boosting. Basically, Ada Boosting was the first successful add-on algorithm designed for binary classification.

AdaBoost could be apply even in an unequal group. It has good properties for general practice that can be proven to increase accuracy of prediction.

I. Gradient Boosting Classifier

Gradient intensification is one of the most competitive algorithms that works with the goal of strengthening weak learne rs by shifting focus from problematic observations that were difficult to predict in the past and to meeting weak readers, general decision trees. It builds the model in a way that is almost identical to other lifting methods, but proves them by a llowing good performance for mysterious losses. For starters, we fit the model in the model that produces 75% accuracy a nd the remaining undisclosed difference is taken in the error name. After that we fit another model in the error set to pull a n additional descriptive component and add it to the original model, which should improve the overall accuracy:

$$\text{Error} = G(x) + \text{Error2}$$

$$f_0(x) = \frac{\arg min}{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$$

J. Quadratic Discrimination Analysis

Quadratic discriminant analysis is performed directly as in the direct discrimination analysis except that we use the followi ng functions based on the matrices of each category.

$$D_i(X) = -1/2LN(|S_i| - 1/2)(X - Y)S_i^{-1}(X - Y)$$

$$S_i(X) = d_i(X) + LN(\pi)$$

K. MLP Classifier

The MLP can be viewed as a classical regression classifier where the input is first modified using a non-linear regression l earned \ Phi. This modification puts the input data into a space where it is directly split. This middle layer is called the hid den layer. One hidden layer is enough to make MLPs a universal standard. An MLP (Ram et al., 2016) (or Artificial Neura l Network - ANN) with a single hidden layer can be represented in bold as follows:

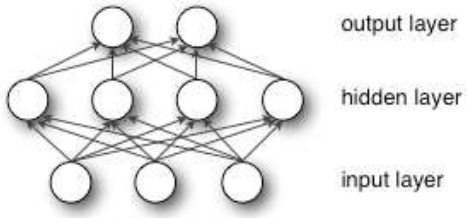

Fig 7: Neural network

V. Proposed Data Measurement Measures

STEP 1: find the essential attributes of kidney data sets. An attribute with low and high data is selected for statistical analysis.
STEP 2: Estimate the accuracy of the data in the statistical analysis.
STEP 3: Measure the methods and methods of treatment that are missing.

STEP 4: Fill in the missing values by means of and between data sets.
STEP 5: Divide train test data and analysis by 70:30.
STEP 6: Perform a deep learning algorithm on train data sets
STEP 7: Measure accuracy with test data sets.

VI. Methodology

Step 1: Data recovery
{
Outline of data
Identify and complete the output
Identify and manage missing information
Appropriate strategies are used
Replace the adjective and spelling
}

Step 2: Model Selection
{
Checking the amount of data (classes)
An in-depth study of Algorithm Sorting's
}

Step 3: Implementation Example using Python
{
Import Data Entry all models together using Python
}

Step 4: Performance classification
{
Calculate Accuracy using the "Operation" operator to analyze the result by the calibration process
}
Step 5: Comparison Result
{
Comparing accuracy between all models by comparing the result with all the proposed algorithms for learning
Calculate the final output of each proposed algorithm
Find the best of all.
}

VII. Pseudo code

Let a1 = {a1, q2, a3,... an} be the given data
A = {}, a set of Algorithms programmers
M = mean and Median {c1, c2, c3, ... cn}, a set of
Z = Means, the mediator of M.
For (i = vacant, i = 0, i ++);
{
For (j = vacant, j = 0, j ++);
}
Apply in-depth Algorithm
f = Depth (Mod: Data);
Let D = {d1, d2, d3,... Dn} be the given data
E = {E1, E2, E3,... En}, a set of ensemble classifiers
C = {c1, c2, c3, ... cn}, a set of arrays
X = training set, X D
Y = test set, Y D
K = class of Meta levelifier
L = n (D)
Because I = 1 to L do

M (i) = A model trained to use E (i) in X
The following i
M = M K
Result = Y divided by M

VIII. Result and Discussion

Below the table present the accuracy, precision and F1 score of the python simulation. The result has been drawn on table from most of the deep learning algorithms.

Table 1: Presents the accuracy, precision and F1 score

Algorithms	LR	LDA	KNN	NB	SVM	RF	MLP	QDA	GBC	ADB	DT
Accuracy	0.72	0.71	0.68	0.69	0.71	0.72	0.72	0.67	0.72	0.71	0.72
Precision	0.72	0.72	0.70	0.73	0.72	0.70	0.72	0.73	0.70	0.70	0.70
f1-score	0.84	0.84	0.79	0.78	0.84	0.79	0.84	0.78	0.79	0.79	0.79

As find from the above table it is LR and MLP of Deep learning algorithm predict better as compare to other learning algorithms used in CKD data analysis for more than 300 data set. It could be varies as per the data sets and attributes considering for the simulation.

IX. Conclusion & Future Scope

This paper mainly explore major deep learning algorithm for the prediction of the CKD analysis. The prediction in terms of accuracy and precision give the better understanding of the deep learning algorithms which could be further use in the real time data analysis and prediction using the graphical user interface. The main purpose of this research paper is to analyze the efficiency of various classification algorithms for the diagnosis of kidney disease. Databases can be used to forecast the incidence of a possible illness. So in future a graphical use interface will have do the same work as doctor does for a specific disease.

References

1. Tekale, S., Shingavi, P., Wandhekar, S., & Chatorikar, A. (2018). Prediction of chronic kidney disease using machine learning algorithm. *Int. J. Adv. Res. Comput. Commun. Eng*, *7*, 92-96.
2. Alshebly, O. Q., & Ahmed, R. M. (2019). Prediction and Factors Affecting of Chronic Kidney Disease Diagnosis using Artificial Neural Networks Model and Logistic Regression Model. *Iraqi Journal of Statistical Science*, *28*, 1-19.
3. Norouzi, J., Yadollahpour, A., Mirbagheri, S. A., Mazdeh, M. M., & Hosseini, S. A. (2016). Predicting renal failure progression in chronic kidney disease using integrated intelligent fuzzy expert system. *Computational and mathematical methods in medicine*, *2016*.
4. Arasu, S. D., & Thirumalaiselvi, R. (2017, February). A novel imputation method for effective prediction of coronary Kidney disease. In *2017 2nd International Conference on Computing and Communications Technologies (ICCCT)* (pp. 127-136). IEEE.
5. Salekin, A., & Stankovic, J. (2016, October). Detection of chronic kidney disease and selecting important predictive attributes. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 262-270). IEEE.
6. Xia, P., Gao, K., Xie, J., Sun, W., Shi, M., Li, W., & Wang, X. (2020). Data Mining-Based Analysis of Chinese Medicinal Herb Formulae in Chronic Kidney Disease Treatment. *Evidence-Based Complementary and Alternative Medicine*, *2020*.
7. Ma, F., Sun, T., Liu, L., & Jing, H. (2020). Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network. *Future Generation Computer Systems*.
8. Ju, W., Eichinger, F., Bitzer, M., Oh, J., McWeeney, S., Berthier, C. C., & Kopp, J. B. (2009). Renal gene and protein expression signatures for prediction of kidney disease progression. *The American journal of pathology*, *174*(6), 2073-2085.
9. Kahraman, S., Fener, M., & Gunaydin, O. (2004). Predicting the sawability of carbonate rocks using multiple curvilinear regression analysis. *International journal of rock mechanics and mining sciences*, *41*(7), 1123-1131.
10. Suthaharan, S. (2016). Support vector machine. In *Machine learning models and algorithms for big data classification* (pp. 207-235). Springer, Boston, MA.

11. Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI communications*, *30*(2), 169-190.
12. Ravindra, B. V., Sriraam, N., & Geetha, M. (2018). Classification of non-chronic and chronic kidney disease using SVM neural networks. *International Journal of Engineering & Technology, 7*(1.3), 191-194.
13. Rahman, T. M., Siddiqua, S., Rabby, S. E., Hasan, N., & Imam, M. H. (2019, January). Early Detection of Kidney Disease Using ECG Signals Through Machine Learning Based Modelling. In *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 319-323). IEEE.
14. Hore, S., Chatterjee, S., Shaw, R. K., Dey, N., & Virmani, J. (2018). Detection of chronic kidney disease: A NN-GA-based approach. In *Nature Inspired Computing* (pp. 109-115). Springer, Singapore.
15. Khamparia, A., Saini, G., Pandey, B., Tiwari, S., Gupta, D., & Khanna, A. (2019). KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network. *Multimedia Tools and Applications*, 1-16.
16. Sisodia, D. S., & Verma, A. (2017, November). Prediction performance of individual and ensemble learners for chronic kidney disease. In *2017 International Conference on Inventive Computing and Informatics (ICICI)* (pp. 1027-1031). IEEE.
17. Subasi, A., Alickovic, E., & Kevric, J. (2017). Diagnosis of chronic kidney disease by using random forest. In *CMBEBIH 2017* (pp. 589-594). Springer, Singapore.
18. Zhang, H., Hung, C. L., Chu, W. C. C., Chiu, P. F., & Tang, C. Y. (2018, December). Chronic Kidney Disease Survival Prediction with Artificial Neural Networks. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1351-1356). IEEE.
19. Ram, R., Palo, H. K., & Mohanty, M. N. (2016, March). Recognition of fear from speech using adaptive algorithm with mlp classifier. In *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)* (pp. 1-5). IEEE.