Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 04, April 2021: 2321-2335

Research Article

Prediction of Early Symptoms for Chronic Kidney Disease by Ensembled classifiers

Sumitha T[1], Charulatha S[2], Dr.S.Selvi[3]

[1]Assistant Professor, Department of CSE, RMK Engineering College, Chennai, India
[2]PG student, Department of CSE, R.M.K. Engineering college, Chennai, India
[3]Professor, Department of CSE,RMK Engineering College,Chennai, India

Email: sat.cse@rmkec.ac.in[1], char19102.pcs@rmkec.ac.in[2], ssi.cse@rmkec.ac.in[3]

Abstract

Chronic Kidney Disease is an important health issue with higher death rate. since, there is no indications at the beginning stage of chronic kidney disease (CKD). Patients doesn't aware of that illness. The early prediction of CKD can help the patients to recover from their infections, here the AI can implement this scope for accurate and quick process. In this paper, we purpose AI diagnoses of chronic kidney disease. The Dataset of CDK was taken from the University of California Irvine (UCI). The missing information of patients are left due to some reasons, that missing data are filled with the KNN. After completion of incomplete qualities, six AI algorithms (Random forest, linear regression, support vector machine, Decision tree and naive Bayes classifier) were utilized to setup models. Among these AI models, support vector machine accomplished the best execution with 96.25% determination precision. By investigating the miscalculation produced by the setup models, a combined model is proposed that consolidates Random forest and naïve bayes along with utilizing perceptron, which could accomplish a normal precision of 98.75% after multiple times of reproduction. Consequently, we estimated this procedure may be material to more complex clinical information for disease finding.

Keywords: *Chronic Kidney Disease, Classifiers, Linear regression, Random forest, prediction*

Introduction

Data mining is the extraction of data from the large data sets. It is used in various fields, now a days its familiar in health fields too. In this paper we analyse the required data from large collected data sets of chronic health disease with the help of machine learning algorithms. Machine learning algorithm Very much useful for large amount of dataset, here also we have taken large dataset to generate model. ML considerably improves efficiency and accuracy Respect to the amounts of data that are processed. Chronic health disease infection is a worldwide public medical condition[12] [15], influencing roughly 10% of the total population,[18] and the scope of commonness of CDK in china is 10.7%, and in the united states the scope of commonness is 10%-15%. CKD occurs slowly over a period of time and continues for a long time afterwards [13]. Also, it may affect other components of the body [16]. The infection of CDK doesn't show any side effects in the initial Stage so we can't predict in the early stage[11]. Chronic kidney disease identified only after the 25% kidney loses of its capacity.

In existing system, the level of accuracy is relatively low. Compared to existing system the proposed system is improved in terms of accuracy. 1) we utilized KNN ascription to fill in the missing data in the informational collection, which could be applied to the informational index with the analytic classes are obscure [14]. 2) Logistic regression(LOG), RF, SVM, KNN, naive Bayes classifier (NB) and feed forward neural network(FNN) were utilized to build up CKD analytic models on the total CKD dataset. The models with better execution were obtained for analysis, therefore in this paper we can say proposed system improvement in accuracy level with zero probability of wrong prediction.

In order to Identify the disease infection in early stage we collect 25 attribute of chronic health disease, none of the existing system taken these much of attributes. Some data may have missing values that can be filled with one of the machine learning approach that is KNN, and also many methods are used over here like SVM, naive bayes, logistic regression. With the help of this algorithms generate the predicted model, finally pass the test data to this generated predicted model and get the results.

Literature Survey

In recent years, there has been a lot of research activity in the areas of Data analytics. Specially in the field of medical application with large amount of data set data analytics play a vital role to predict the expected model. Many of the existing work concentrate on this.

The paper titled "Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers" doing the same work with fuzzy classifiers. In this paper the author proposes the model with the terms of the practicality of two in-house fuzzy classifiers, fuzzy principle building master framework and fuzzy ideal affiliated adaptive padding, for determination of patients with ongoing kidney infection. For examination a classifier, halfway least squares discriminant investigation was utilized. Proposed classifiers are valuable methods for the analysis of CKD patients with palatable strength, and can likewise be utilized for different sorts of patients[1].

The paper title “Diagnosis of Chronic Kidney Disease by Using Random Forest” doing the same diagnosis with the help of random forest algorithm. This paper examines how Chronic Kidney Disease can be analysed by utilizing AI (ML) strategies. ML calculations have been a main impetus in recognition of anomalies in various physiological information, and are, with an extraordinary achievement, utilized in various grouping assignments. This paper mainly focusses on the accuracy of the system, compared to other classifiers Random forest gives better performance in terms of time also[2].

The paper titled “Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration” discuss about the predictive model. The proposed results show that consolidating time value data in patient clinical history can prompt better forecast of loss of kidney functionality. Specifically results show that the overall significance of various indicators fluctuates over the long haul, particularly this paper center around worldly data.[3].

The paper titled “A Method to Predict Diagnostic Codes for Chronic Diseases using Machine Learning Techniques” propose a method to predict Chronic disease using Machine Learning algorithm. This paper centers around clinical and claims information for examining 11 persistent illnesses like kidney sickness, osteoporosis, joint pain and so on proposed paper by and large stands up ongoing disease[4].

The paper title “Explainable Prediction of Chronic Renal Disease in the Colombian Population using Neural Networks and Case-Based Reasoning” talks about chronic disease with the help of Neural networks algorithm. This paper presents one classifier which is used to find out whether the patient is affected by particular disease or not. The model created via preparing segment information and clinical consideration data of various populace gatherings. This paper at last discover of 3,494,516 individuals were distinguished as being in danger of creating CKD in Colombia, or 7% of the complete population.[5].

The paper titled “A Deep Learning-based System for Automated Sensing of Chronic Kidney Disease” use the classification and feature extraction methodology to predict the model. This paper builds up another detecting strategy for the computerized discovery of kidney infection. The salivary urea focus is checked to recognize the illness. With the assistance of this we can detect the urea levels in the salivation test. Additionally played out the factual examination to decide the proposed detecting strategy which gives better execution on assessment or not[6].

The paper titled “An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy” discuss about diagnosing the CKD[21] in early stage. Numerous methods and models have been created to analyse the CKD in beginning phase, among all strategies Machine Learning (ML) procedures assume a huge part in the early estimating of various types infirmities. The result of this paper gives minimum minimal error rates compared to other discussed work.[7].

The paper titled "XG Boost Model for Chronic Kidney Disease Diagnosis" use Recursive Feature Elimination Algorithm (RFE) of machine learning technique to evaluate the disease. In this paper, Extreme slope boosting model is utilized for improved execution thus a few ordinary and ongoing AI calculations are utilized with regards to Chronic Kidney Disease. These produced models combined with the study can help decrease the expense and time to analyse a Chronic Kidney Disease patient. This paper is noticed that a few tests for CKD could create pictures as crude information. We can use this image as a dataset for future work, and so it will improve the accuracy.[8]

The paper titled "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform" did the same work which we proposed but use the different methodology. This paper uses the Convolutional neural network to predict the chronic kidney disease.[9].

The paper titled "Ensemble Feature Ranking for Cost-Based Non-Overlapping Groups: A Case Study of Chronic Kidney Disease Diagnosis" use the ensemble feature ranking method to study about CKD. This paper adopts a more reasonable strategy to bunch based component determination, and utilize two competitor arrangements are proposed for bunch based element choice to meet diverse objectives[10].

One of the existing works carried out to find the relation between air pollution data and the chronic kidney disease with the help of deep learning. [17].

One of the existing papers propose the learning machine to predict the chronic kidney disease. The learning machine developed by the concept of kernel based extreme learning. [19].

System for Predicting Chronic Kidney disease symptoms

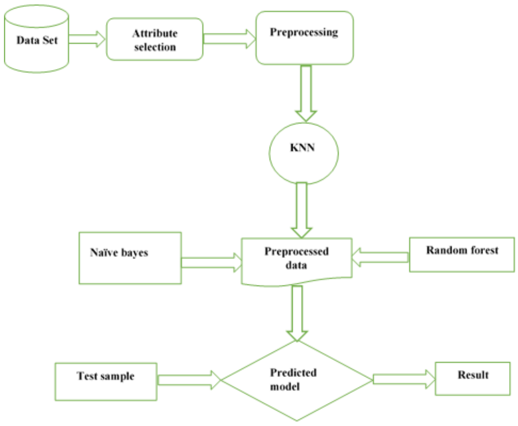


Figure 1 CKD symptoms prediction system

3.1 Data set:

The dataset is the collection of large sets of data from patients for the further analysis of process. The collected data will be preprocessed with various levels of algorithm in machine learning. In this paper we have taken ckd data set from UCI, it contains 400 instances with 25 attributes like su, sg, bp, rbc, pc etc... these attributes play a vital role in predicting the health condition. [22].

Dataset contain more than one database table, each column in that table represents a different variable and each row represents different record. Dataset may contain missed data, outliers etc... The basic algorithm used here are KNN to fill the missed data.

3.2 Attribute selection:

Choosing a subset of the originally available attributes to use for model construction early on is part of the attribute selection task. General-purpose attribute selection algorithms may be applied to a wide variety of target algorithms and, in some cases, various target tasks.

The various types of attributes are taken over here to show the high level of accuracy we take 25 attributes for preprocessing. the important attributes such as blood pressure, sugar level, blood glucose random, white blood cell count, red blood cell count, these all the basic quality values collected from the patients as a data sets to find the health status of patients. Main purpose of attribute selection is improving the efficiency.

3.3 Preprocessing:

Data preprocessing is the procedure for processing relevant data for use in a machine learning model. It's the first and most important step in developing a machine learning model. It is not always the event that we run across clean and structured data when operating on a data science project. Processing the collected data sets, that is the attributes. some missing quality values of attributes are left due to some reasons. from this the unfilled attributes are predicted using the KNN algorithm.

3.4 KNN (K- Nearest Neighbors):

K-Nearest Neighbor calculation relies upon the Supervised Learning strategy and is among the most essential Machine Learning calculations. The K-NN calculation expects that the new case/information and current cases are indistinguishable and places the new case in the classification that is generally like the current classifications. The K-NN calculation stores all accessible information and characterizes another information point dependent on its similitude to the current information. The K-NN calculation can be utilized for one or the other characterization and relapse issues, yet it is most usually utilized for order assignments.

The K-NN calculation is a non-parametric calculation, which implies it makes no presumptions about the basic information. It's otherwise called a sluggish student calculation since it doesn't gain from the preparation set straight away; all things considered, it saves the dataset and utilizations it to arrange the outcomes.

During the preparation interaction, the KNN calculation just stores the dataset, and when it gets new information, it arranges it into a bunch that is extremely near the new information. Think about the accompanying situation: Let's say we have an image of an animal that resembles a feline or a canine, however we couldn't say whether it's a feline or a canine. We can utilize the KNN calculation for this distinguishing proof since it depends on a similitude measure. Our KNN model will analyze the likenesses between the new informational collection and the felines and canines' pictures, and classifications it as either a feline or a canine dependent on the most comparable highlights.

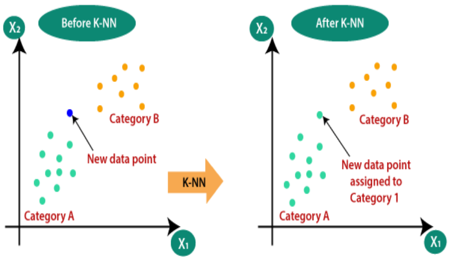


Figure 2

This algorithm can be implemented for both the regression problem and classification. most probably it’s used to solve problem of classification in data science industry. It is a common algorithm that keeps track of all available occurrence and classifies any new occurrence by accepting the majority of votes from its k neighbors.

$$d(\mathbf{p},\mathbf{q}) = d(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2}$$
$$= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.$$

This algorithm runs this formula to process the distance between test data and the data points. then highest possibilities of data are shared.

3.5 Preprocessed data:

Data preprocessing is a data mining technique that requires translating raw data into a format that can be understood. Data from the real world is often incomplete, unreliable, deficient in certain habits or patterns, and prone to numerous errors. Preprocessing data is a tried and tested way of addressing such problems. Preprocessing raw data prepares it for further processing. Database-driven systems, such as customer relationship management and rule-based applications, use data preprocessing (like neural networks). Data preprocessing is important in Machine Learning (ML) processes because it encodes the dataset in a way that the algorithm can understand and parse. By the missing attributes the data are predicted using KNN algorithm and for the further exact accuracy linear regression and random forest algorithm are used.

3.6 Naive Bayes:

The Naive Bayes calculation is an administered learning calculation that utilizes the Bayes hypothesis to tackle order issues. It is essentially utilized in text order undertakings that require a huge preparing dataset [11]. The Naive Bayes Classifier is a straightforward and successful grouping calculation that guides in the advancement of quick AI models equipped for making snappy forecasts. It's a probabilistic classifier, which implies it makes forecasts dependent on an item's probability. Spam filtration, feeling investigation, and article characterization are generally normal employments of the Naive Bayes Algorithm.

For multi-class prediction problems, Naive Bayes is a good option. If its assumption of feature independence holds true, it can outperform other models while requiring much fewer training data. Categorical input variables are better suited to Naive Bayes than numerical input variables. The Naive Bayes model is simple to construct and is particularly useful for large data sets. Naive Bayes is considered to outperform even the most advanced classification methods due to its simplicity.

The Bayes theorem allows you to calculate posterior likelihood P(c|x) from P(c), P(x), and P(x|c) using P(c), P(x), and P(x|c).

Consider the following equation:

$$\underbrace{P(c \mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x \mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

The posterior probability of class (c, target) given a predictor is P(c|x) (x, attributes). The prior probability of class is P(c). The likelihood is P(x|c), which is the probability of a predictor given a class. P(x) is the predictor's prior likelihood

3.7 Random Forest:

A Random Forest is a classifier that joins various choice trees on various subsets of a dataset and midpoints the outcomes to build the dataset's prescient exactness. Rather than relying upon a solitary choice tree, the irregular woods take the expectations from each tree and figures the last presentation dependent on the larger part votes of forecasts. The arbitrary woods are the assortment of choice tree. To comprehend the working standards of arbitrary woods calculation with these means.

Step 1: start by selecting random samples from a dataset

Step 2: For every samples the algorithm will develop a decision tree. Then from that decision tree the result will be predicted for each.

Step 3: In the following step, for all the predicted result the voting will be generated.

Step 4: finally, most voted prediction is selected as final result.

3.8 Predicted model:

Predictive modelling is a mathematical technique for predicting future actions[24] that is widely used. Predictive modelling solutions are a form of data-mining technology that analyses historical and current data to create a model that can be used to forecast future outcomes.

By using the KNN,[23] linear regression, and the random forest the missing data are predicted with zero probability of wrong record. once the missing data are predicted is made to match with the test samples.

3.9 Test samples*:*

An example is a more modest, simpler to-oversee subset of a bigger bunch. It's a subset of a more extensive populace with comparative qualities. On the off chance that the populace size is excessively enormous for the test to contain every single likely member or discoveries, tests are utilized in factual exploration.

Chronic kidney disease patient’s test samples are collected and that are tested with the predicted result and it verifies whether the patients are affected are not.

3.10 Result:

A result (also known as an upshot) is the qualitative or quantitative expression of the end result of a series of acts or events. Advantage, disadvantage, benefit, damage, loss, value, and victory are examples of possible outcomes... in general, the outcome of any type of study, action, or phenomenon. Finally risk analysis of patients are predicted with high level of accuracy. the output is represented with the graphical representation with percentage of normal and abnormal.

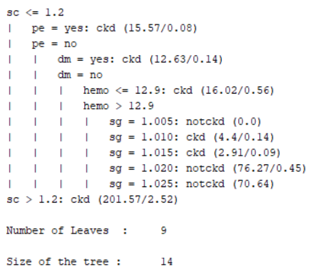
Experiments and Results

Machine Learning algorithm gives better performance for large dataset in terms of accuracy and time. There were 400 patient details collected from UCI Repository with 25 attributes.

The collected data may have some missed values, incorrect format, and some may have outliers. For that we need to preprocess the data, preprocessing done with the KNN algorithm. KNN gives better results than other algorithms. Then the pre-processed data are classified with different mining algorithms like decision tree, Naïve bayes, Random forest and its performance measures are evaluated. Among these algorithms Random forest gave more accuracy.

The classifiers are chosen in the aspect of improving the classifier accuracy. The classifiers are selected individually, so that the performance of the prediction models re observed. In this work, we have ensembled the probabilistic based naïve bayes model and tree evaluation based random forest model. The following are the observations of the prediction models.

4.1 J48 decision tree based prediction:



```
sc <= 1.2
|   pe = yes: ckd (15.57/0.08)
|   pe = no
|   |   dm = yes: ckd (12.63/0.14)
|   |   dm = no
|   |   |   hemo <= 12.9: ckd (16.02/0.56)
|   |   |   hemo > 12.9
|   |   |   |   sg = 1.005: notckd (0.0)
|   |   |   |   sg = 1.010: ckd (4.4/0.14)
|   |   |   |   sg = 1.015: ckd (2.91/0.09)
|   |   |   |   sg = 1.020: notckd (76.27/0.45)
|   |   |   |   sg = 1.025: notckd (70.64)
sc > 1.2: ckd (201.57/2.52)

Number of Leaves  :     9

Size of the tree :      14
```

Figure 3 J48 decision tree-based prediction

The above diagram shows the tree construction for the taken dataset by the decision tree algorithm, and this algorithm gives 94% accuracy. Accuracy depends on many elements like Precision, Recall etc.

Accuracy =TP + TN / TP + FN + FP + TN

Recall = TP / TP + FN

Precision =TP / TP + FP

When the system correctly predicts the positive data, that is called true positive. Similarly, The anticipated model effectively predicts the negative class called genuine negative. Inverse to this idea the bogus negative is, the created model inaccurately predicts the negative class.

Table 1: Detailed accuracy table of decision tree algorithm

TP Rate	FP Rate	Prec ision	Recall	F-Meas Ure	MCC	Roc Area	PRC Area	Class
0.996	0.020	0.988	0.996	0.996	0.979	0.999	1.000	Ckd
0.980	0.004	0.993	0.980	0.980	0.979	0.999	0.999	notckd
0.990	0.014	0.990	0.990	0.990	0.979	0.999	0.999	

Table 1 Shows the detailed accuracy information by decision tree algorithm for the dataset taken from UCI. All 400 data are classified into two classes one is ckd and another one is not ckd. With the help of this predicted model, we can predict sample test data. The confusion matrix generated by decision tree is shown below in Fig 4.2. By this model, among 400 patient's data 249 patients are classified as ckd and 147 records under the classification of not ckd. The Remining 4 only the wrong prediction by this decision tree algorithm.

```
=== Confusion Matrix ===

   a   b   <-- classified as
 249   1 |   a = ckd
   3 147 |   b = notckd
```

Figure 4 Confusion matrix of J48 decision tree model

4.2 Naïve Bayes based prediction:

The input data is evaluated with the naïve bayes classifier based prediction. The model is built based on the training performed on the input data. The naïve bayes algorithm is implemented using WEKA classifiers.[20]. This algorithm works based on bayes theorem and also it tells strong independence among them. It is also called probabilistic classifiers. There are totally 380 instances correctly classified out of 400 instances. Incorrectly classified instances are 20. The accuracy of the naïve bayes classifier for the correctly classified instances are 95%. The error rates are likewise noticed utilizing different estimates like mean total mistake, root mean squared blunder, relative total blunder and root relative squared mistake. The error rate is less and the CKD is predicted to the maximum for the taken dataset. Naïve bayes algorithm takes 0.01 seconds to build the model. The prediction model builds summary and its evaluation is shown in figure 4.3.

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         380               95      %
Incorrectly Classified Instances        20                5      %
Kappa statistic                          0.8961
Mean absolute error                      0.0479
Root mean squared error                  0.2046
Relative absolute error                 10.2125 %
Root relative squared error             42.2526 %
Total Number of Instances              400
```

Figure 4 Prediction model build summary and its evaluation

Table 2: Class wise detailed Accuracy of the prediction model

TP Rate	FP Rate	Prec Ision	Rec all	F-Meas ure	M CC	Roc Area	PRC Area	Class
0.920	0.000	1.000	0.920	0.958	0.901	1.000	1.000	Ckd
1.000	0.080	0.882	1.000	0.938	0.901	1.000	1.000	notckd
0.950	0.030	0.956	0.950	0.951	0.901	1.000	1.000	

The class wise detailed accuracy of the prediction model is shown in table 4.2. The true positive and false positive rates are observed and the precision, recall and f-measure are also measured. It indicates that better accuracy values are observed in the prediction model evaluation.

```
=== Confusion Matrix ===

   a   b   <-- classified as
 230  20 |   a = ckd
   0 150 |   b = notckd
```

Figure 5 Confusion matrix

The confusion matrix of the naïve bayes prediction model is shown in figure 4.4. There are 230 instances are correctly predicted as chronic kidney disease and 20 instances are wrongly predicted as chronic kidney disease. Further, there are 150 instances are correctly classified as not having chronic kidney disease.

4.3 Random forest based prediction:

With the help of WEKA classifiers Random forest algorithm is implemented. It is one of best algorithm under tree-based structure. For this ckd dataset also it gives the best predicted model and high accuracy in prediction. Among 400 instance 385 instances are correctly identified, so it gives 96% accuracy. The performance of Random forest algorithm somewhat improved compared to other algorithms which we have discussed in previous study in terms of accuracy and time also. It takes 0.29 seconds to build the model. Confusion matrix of random forest algorithm shown below.

```
  a      b      < - - classified as
235     15 |    a = ckd
  0    150 |    b = notckd
```

Figure 6. Confusion matrix of random forest

4.4 Proposed model

Finally Compare the above models, in that got high accuracy in random forest algorithm and the second high accuracy is Naïve bayes algorithm. So, try to combine both these two models and improve the accuracy percentage. This is proposed system, with the help of this we can predict the chronic disease much earlier. Naïve bayes is the probabilistic algorithm and Random forest is one of the decision tree algorithms After combining these two, the integrated model is generated here also same dataset is used to train as well as test. Among 400 instances 395 instances are correctly identified, so the performance of Proposed model improved compared to previous studied models that is the accuracy is 98%. Confusion matrix of integrated model is given below.

```
     a      b        < - - classified as

   245      5 |      a = ckd
     0    150 |      b = notckd
```

Figure 7. Confusion matrix of proposed model

4.5 Performance Comparison of models:

Table 3: Performance comparison

Models	**Accuracy**
Decision tree	94%
Naïve bayes	95%
Random forest	96%
Proposed model (Naïve + Random)	**98%**

Table 5 infer that the proposed model that is ensemble classifier gives higher accuracy in classification-based prediction. This work helpful to find out the early-stage symptoms of ckd with good accuracy results.

4.6 Visualization of predicted model:

Visualization of predicted model shown below in the figure 4.7 and 4.8. This figure describes the sample predicted result model which gives the classification based on some limited or selected attributes. Figure 4.7 give the relation between the age and hypertension of the patient record. The attribute age denoted by numerical value and hypertension denoted by the value yes or no. The model predicts the ckd affected patient with respect to the age and hypertension. Blue color is used to indicate the ckd affected patient and the red color denotes the notckd patients.

From this figure we can infer that the patient who have the age above 40 and also having hypertension then there may be possibility to have chronic kidney disease.

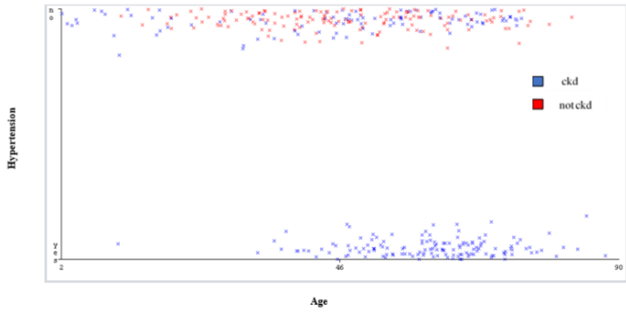


Figure 8 predicted result based on age and hypertension

Figure 8 shows the relation between diabetes and anemia. Both are not a numerical value and denoted by yes or no. X-axis represents the diabetes and the Y-axis represents the anemia. Based on these two, the predicted model classify the class whether the patient having ckd or not. [25] From this graph we can infer that the patient who have diabetes as well as animia then there may lots of possibility to have chronic kidney disease.

Figure.9 Predicted result based on diabetes and anemia

Conclusion

In relations of data attribution and sample analysis, the proposed CKD diagnostic approach is feasible. The integrated model could achieve adequate accuracy after unsupervised attribution of absent values in information set by means of KNN attribution. As a result, we believe that extending this approach to the real-world analysis of CKD will have a positive outcome.

Furthermore, this approach may be applied to clinical evidence from other disorders in real-world therapeutic analysis. Though, in the method of establishing the prototypical, because of the limits of the situations, the existing information examples are moderately limited, counting only 400 examples. As a result, the model's generality routine may be partial. In adding, because around only two groups (ckd and notckd) of information examples in the information set, the model cannot identify the sternness of CKD..

References

[1] Zewei Chen, Zhuoyong Zhang, Ruohua Zhu , Yuhong Xiang , Peter B. Harrington b "Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers" Chemometrics and Intelligent Laboratory Systems, Volume 153, Pages 140-145, 2016.

[2] Subasi A., Alickovic E., Kevric J "Diagnosis of Chronic Kidney Disease by Using Random Forest" IFMBE Proceedings, vol 62. Springer, 2018.

[3] Anima Singh a, Girish Nadkarni b , Omri Gottesman b , Stephen B. Ellis b , Erwin P. Bottinger b, John V. Guttag, "Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration", Journal of Biomedicl Infomtics, pp. 220-228 2015.

[4] D. Gupta, S. Khare and A. Aggarwal "A Method to Predict Diagnostic Codes for Chronic Diseases using Machine Learning Techniques" International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, pp. 281-287, 2016.

[5] G. R. Vásquez-Morales, S. M. Martínez-Monterrubio, P. Moreno-Ger and J. A. Recio-García "Explainable Prediction of Chronic Renal Disease in the Colombian Population using Neural Networks and Case-Based Reasoning" n IEEE Access, vol. 7, pp. 152900-152910, 2019.

[6] N. Bhaskar and S. Manikandan, "A Deep Learning-based System for Automated Sensing of Chronic Kidney Disease" in IEEE Sensors Letters, vol. 3, no. 10, pp. 1-4, Oct. 2019.

[7] B. Khan, R. Naseem, F. Muhammad, G. Abbas and S. Kim, "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy" in IEEE Access, vol. 8, pp. 55012-55022, 2020.

[8] A. Ogunleye and Q. -G. Wang, "XG Boost Model for Chronic Kidney Disease Diagnosis." IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 17, no. 6, pp. 2131-2140, 1 Nov.-Dec. 2020.

[9] G. Chen et al., "Prediction of Chronic Kidney Disease Using Adaptive Hybridized Deep Convolutional Neural Network on the Internet of Medical Things Platform." IEEE Access, vol. 8, pp. 100497-100508, 2020.

[10] S. I. Ali et al., "Ensemble Feature Ranking for Cost-Based Non-Overlapping Groups: A Case Study of Chronic Kidney Disease Diagnosis" IEEE Access, vol. 8, pp. 215623-215648, 2020.

[11] A. Maurya, R. Wable, R. Shinde, S. John, R. Jadhav and R. Dakshayani, "Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning," International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, pp. 1-4, 2019.

[12] Y. Amirgaliyev, S. Shamiluulu and A. Serek, "Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods," IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, pp. 1-4, 2018.

[13] R. Gupta, N. Koli, N. Mahor and N. Tejashri, "Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease," International Conference for Emerging Technology (INCET), Belgaum, India, pp. 1-4, 2020.
[14] F. Kayaalp, M. S. Basarslan and K. Polat, "A hybrid classification example in describing chronic kidney disease," Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, Turkey, pp. 1-4, 2018.
[15] N. V. Ganapathi Raju, K. Prasanna Lakshmi, K. G. Praharshitha and C. Likhitha, "Prediction of chronic kidney disease (CKD) using Data Science," International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, pp. 642-647, 2019.
[16] S. Vashisth, I. Dhall and S. Saraswat, "Chronic Kidney Disease (CKD) Diagnosis using Multi-Layer Perceptron Classifier," 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, pp. 346-350, 2020.
[17] S. -M. Chiu, F. -J. Yang, Y. -C. Chen and C. Lee, "Deep learning for Etiology of Chronic Kidney Disease in Taiwan," IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE), Yunlin, Taiwan, pp. 322-325, 2020.
[18] M. Desai, "Early Detection and Prevention of Chronic Kidney Disease," 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, pp. 1-5, 2019.
[19] H. A. Wibawa, I. Malik and N. Bahtiar, "Evaluation of Kernel-Based Extreme Learning Machine Performance for Prediction of Chronic Kidney Disease," 2nd International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, pp. 1-4, 2018.
[20] E. Avci, S. Karakus, O. Ozmen and D. Avci, "Performance comparison of some classifiers on Chronic Kidney Disease data," 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, Turkey, pp. 1-4, 2018.
[21] O. Zaim, T. Saidi, N. El Bari and B. Bouchikhi, "Assessment Of "Breath Print" In Patients With Chronic Kidney Disease During Dialysis By Non-Invasive Breath Screening Of Exhaled Volatile Compounds Using An Electronic Nose," IEEE International Symposium on Olfaction and Electronic Nose (ISOEN), Fukuoka, Japan, pp. 1-4, 2019.
[22] C. Suresh, B. C. Pani, C. Swatisri, R. Priya and R. Rohith, "A Neural Network based Model for Predicting Chronic Kidney Diseases," Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 157-162, 2020.
[23] A. A. Johari, M. H. Abd Wahab and A. Mustapha, "Two-Class Classification: Comparative Experiments for Chronic Kidney Disease," 4th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, pp. 789-792, 2019.
[24] B. V. Ravindra, N. Sriraam and M. Geetha, "Chronic Kidney Disease Detection Using Back Propagation Neural Network Classifier," International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, pp. 65-68, 2018.
[25] H. Zhang, C. Hung, W. C. Chu, P. Chiu and C. Y. Tang, "Chronic Kidney Disease Survival Prediction with Artificial Neural Networks," IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, pp. 1351-1356, 2018.