JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CHRONIC DISEASE DIAGNOSIS USING MACHINE LEARNING ALGORITHM

[1] Dr. Ayesha Banu, [2] Zareena Begum

[1]Associate Professor,[2]Assistant Professor

[1] Department of Computer Science and Engineering (Data Science)

[1]Vaagdevi college of Engineerimg, Bollikunta, Warangal, Telangana, India.

Abstract : Chronic kidney disease (CKD), which is also akey risk factor for other diseases, kills and disables people all over the world. Becausethere are no evident signs in the early stages ofCKD, it might go unnoticed. Medicine thatdecreases the progression of renal disease can be used to prevent it from progressing in patients who are diagnosed early. Clinicians can achieve their objectives more quickly by using machine learning models. This study suggests a CKD diagnosis approach based on machine learning. In the UCI machine learningrepository, missing values in the CKD data set were discovered. The most similar measures from a large number of full samples were used to fill in the missing data in each partial sample using KNN imputation. Patients may forget to take measures in the real world for a variety of reasons, resulting in missing data. Six machine learning approaches were employed to construct models when the missing data set was completed: logistic regression, k-nearest neighbor, naive Bayes classifier, and feed forward neural network.With a diagnosis accuracy of 99.75 percent, Random Forest is the most accurate of these machine learning models. After ten simulations based on the errors generated by the constructed models utilizing the integrated model, an average accuracy of 99.83 percent can be achieved. As a result, we came to the conclusion that this method may be used to diagnose more complex clinical disorders.

Keywords: Machine Learning, Logistic Regression, KNN, Neural Network, Random Forest.

I. INTRODUCTION

Chronic kidney disease (CKD) affects roughly 10% of the world's population, making it a significant public health issue. [3] Chronic kidney disease (CKD) affects 10.8% of Chinese individuals, compared to 10–15% in the US. Unemployment in Mexico has risen to 14.7 percent, according to a new survey. Renal function declines over time, and the kidneys eventually cease to function. When kidney disease is in its early stages, there are no visible indicators of it. It's likely that the illness won't be diagnosed until the kidney has lost around 25% of its function. Chronic kidney disease has a severe impact on the human body, with high rates of morbidity and mortality. It can lead to cardiovascular disease. Once CKD has started, it is impossible to stop it. Patients who are recognized earlier in the disease's progression can receive treatment to halt or stop it, which is why early identification and diagnosis are so critical. A computer programme that combines data and deductive reasoning to learn the features of a particular pattern is an example of machine learning. This technology may be a viable tool for diagnosing CKD in patients due to its capacity to reliably and economically diagnose illnesses. Electronic health records have evolved into a new type of medical equipment with a wide range of uses as a result of their rapid expansion. Machine learning has previously been used in the medical field to identify human body status, analyze disease factors, and diagnose a variety of diseases.

Heart disease, diabetes, retinopathy, acuterenal injury, cancer, and other disorders are just a few examples which can be diagnosed using machine learning algorithms. Regression tree, neural network, probability and decision surface approaches were used in these models. Hodneland et al used image registration to discover kidney morphological alterations in order to diagnose CKD. Vasquez-Morales et al. used large CKD datasets to construct classifiers that had a 95% accuracy rate on their test data.

Chemo metrics can be used to investigate the relationships between various items and factors using a multivariate method. Theapplication of chemo metric-based multivariateclassifiers in the diagnosis of chronic kidney disease (CKD) may be beneficial. F u z z y logic and fuzzy mathematics diagnostic methods can be used on patients data with chronic kidney disease to better understand and diagnose their illness.

L.A. Zadeh released his seminal work on fuzzy set theory in 1965. "Fuzzy set theory," an infinite-valued logic, allows for less-than-perfect reasoning. Crisp ingredients may or may not be included in the kit. Components in fuzzy sets are only marginally significant. The borders of the set get blurry as a result of this. Fuzziness is a metric of representational ambiguity, not likelihood, rather than assessing the frequency with which something will occur. Fuzzy logic has already been used to handle a variety of challenges in bioinformatics and system

biology. C.T. Zhang et al. used fuzzy clustering to predict protein structure classes on the basis of amino acid content. H.B. Shen et al. combined supervisory fuzzy clustering and fuzzy K- nearest neighbor to predict the protein structural classes and membrane protein types. A fuzzy support vector machine was used for predicting protein structure based on pseudo amino acid content.

A diverse number of biological challenges were dealt using the fuzzy K-nearest neighbor technique which also include the nuclear receptor subfamilies and G-protein receptors discovery by X. Xiao et al. It was also used by Xiao et al. to predict the channel-drug interaction. Classification problems that can be solved with fuzzy approaches include endometrial cancer and cervical cancer diagnosis, herbal drug identification, and diabetes mellitus diagnosis. One of the benefits of using FOAM for single-class classification is that it exploits feature similarities inside the class. FuRES, to be used as a supplementary classification method, all classes must be defined differently from FOAM. For this discussion, we assume that the response matrix Y considers labels that characterize the sample categories corresponding to prediction matrix X. This is a rare example of PLS PLS-DA (partial least squares discriminant analysis) in action.

The ultimate goal of machine learning is to construct models that can swiftly generalize and categorize from instances observed before they were created (ML). ML creates these models by either constructing or learning functional correlations between the user- selected input and output feature domains. Patients with CKD can be identified by sorting the information gathered from their symptoms (features) into helpful categories (groups of healthy individuals, CDK individuals, or individuals with the some other type of ailment).

CKD is a long-term health issue that affects nearly 10% of the world's population. In real- life cases of CKD, however, cardiovascular disease and renal function decline are frequently linked to an increased risk of hospitalization, morbidity, and even death (s). Chronic kidney disease (CKD) patients are more likely than the general population to develop atherosclerosis and other symptoms.

These conditions have a significant negative impact on a person's quality of life. A diagnosis of CKD indicates that the kidneys have been damaged. A wide range of symptoms and risk factors increase an individual's likelihood of developing chronic kidney disease (CKD). The creation of computer models based on machine learning principles that can predict the course of Chronic Kidney Disease (CKD) can substantially improve our understanding of the disease. For the diagnosis of CKD, many researchers have suggested fuzzy logic models. The main goal of this study is to create a basic classifier that is more accurate at classification. Many well-known machine learning techniques, such as the ANN, SVM, k-nearest neighbor (k-NN), C4.5 decision tree, and random forest, were used in our study to build the highly accurate CKD diagnosis model (RF). We used real data (the CKD dataset) from the UCI machine learning database to demonstrate the utility and effectiveness of our suggested methodologies.

2. RELATED WORK:

To help patients with chronic diseases, Gut Predictive models are built using temporal data from electronic health records (EHRs), Anima Singh et.al. These data have a number of technical flaws, such as irregular sampling and a wide range of patient histories. This study uses a patient's temporal EHR data to describe and evaluate three different machine learning methods for generating prediction models. The values of the patient's medical history predictors are gathered for the first procedure. It is customary to use this non-temporal technique. The data's temporal dynamics are used in the other two processes. Both systems handle temporal information and missing data in different ways. Models that are frequently used to forecast the reduction in estimated glomerular filtration rate (GFR) were learned and evaluated using data from Mount Sinai Medical Center's electronic health record (EHR) (eGFR). According to our findings, combining temporal information with a patient's medical history can help predict renal failure. They also emphasize the importance of how the information is delivered. According to our findings, multi-task learning is a great strategy for capturing the temporal dynamics of EHR data since the relative importance of distinct predictors alters with time. We demonstrate how the model may be used to identify people at high risk of short-term renal function deterioration using an example from the literature.

Luxia Zhang et.al. Performed a cross-sectional study with a sample of Chinese citizens from around the country to find out more information when Haiyan Wang came up with the idea. Chronic renal disease was defined as the presence of albuminuria or an eGFR of less than 60 mL/min per 73 m2. The blood pressure values and urine samples were taken along with filling out a questionnaire about their lifestyle and medical history. The rate of glomerular filtration was calculated using the amount of creatinine in the bloodstream. Urine albumin and creatinine levels were measured to determine albuminuria. The crude prevalence of kidney damage markers was obtained, and factors associated with the presence of chronic renal disease were investigated using logistic regression.

This paper includes the findings of Zewei Chen et.al. FuRES and FOAM, two fuzzy rule-building expert systems, were put to the test on people with chronic renal disease to determine how effectively they could diagnose them. PLS- DA, a linear classifier, was employed as a benchmark in the evaluation. The UCI Machine Learning Repository provided all of the CKD data utilized in this work. To examine how well the two fuzzy techniques stood up to the increasing stress, a variety of proportional noise was added into the datasets. This approach is a 11-step process where 11 levels of proportional noise is added to each of the training and prediction sets numerical attributes. 121 pairs of simulation data is taken and the categorization rates were compared using a grid of 121 groups.

For this datasets 11 levels of random noise is distributed to every numeric attribute. Using 200 bootstrap Latin partitions, FOAM and FuRES averaged 97.2, 1.2 and 98.1, 0.5 percent respectively. The PLS-DA also produced approximately the same results 94.3 percent. The classification performance of FuRES, FOAM, and PLS- DA models were investigated using confluent datasets including both the original and updated datasets. From 200 bootstrapped evaluations, FuRES and FOAM achieved average prediction rates of 99.2 percent and 99.0 percent, respectively. The accuracy of PLS-DA , 95.9 is just 0.6 percent better than PLS. Both FuRES and FOAM are effective at identifying CKD patients, but FuRES is more robust. These two fuzzy classifiers are robust enough to diagnose CKD patients, but they can also be used to diagnose other patients.

A. PRELIMINARIES:

This section includes the data set and operating environment, missing value imputation, and feature vector extraction before developing the models. The dataset is considered from [32] and has 400 samples. This CKD data collection includes one categorical response variable as

well as 24 predictive factors (11 numerical and 13 categorical) (class). Each class has two options: CKD (sample with CKD) or nonckd (sample without CKD) (sample without CKD). 250 of the 400 samples are from people with CKD, while 150 are from those who don't have CKD. As a result, it's critical to point out that the data contains a substantial number of missing values.

Table 1: Sample Data Set

Variables	Explain	Class	Scale	Missing Rate
age	Age	Numerical	age in years	2.25%
bp	Blood Pressure	Numerical	in mm/Hg	3%
sg	Specific Gravity	Nominal	(1.005,1.010,1.015,1.020,1.025)	11.75%
al	Albumin	Nominal	(0,1,2,3,4,5)	11.5%
su	Sugar	Nominal	(0,1,2,3,4,5)	12.25%
rbc	Red Blood Cells	Nominal	(normal,abnormal)	38%
pc	Pus Cell	Nominal	(normal,abnormal)	16.25%
pcc	Pus Cell clumps	Nominal	(present,notpresent)	1%
ba	Bacteria	Nominal	(present,notpresent)	1%
bgr	Blood Glucose Random	Numerical	in mgs/dl	11%
bu	Blood Urea	Numerical	in mgs/dl	4.75%
sc	Serum Creatinine	Numerical	in mgs/dl	4.25%
sod	Sodium	Numerical	in mEq/L	21.75%
pot	Potassium	Numerical	in mEq/L	22%
hemo	Hemoglobin	Numerical	in gms	13%
pcv	Packed Cell Volume	Numerical	-	17.75%
wbcc	White Blood Cell Count	Numerical	in cells/cumm	26.5%
rbcc	Red Blood Cell Count	Numerical	in millions/cmm	32.75%
htn	Hypertension	Nominal	(yes,no)	0.5%
dm	Diabetes Mellitus	Nominal	(yes,no)	0.5%
cad	Coronary Artery Disease	Nominal	(yes,no)	0.5%
appet	appet	Nominal	(good,poor)	0.25%
pe	Pedal Edema	Nominal	(yes,no)	0.25%
ane	Anemia	Nominal	(yes,no)	0.25%
class	Class	Nominal	(ckd,notckd)	0%

B. DATA PROCESSING:

Each nominal (categorical) variable was coded to make it easier for a computer to process. The codes for normal and abnormal in rbc and pc were 1 and 0, respectively. The presence and absence of pcc and ba were assigned numerical values of 1 and 0 respectively. Htn, Dm, Cad, Pe, and Ane were all coded as 1 and 0, and each question was responded with a "yes" or "no" response. Appet received a 1 for excellent performance and a 0 for poor performance. Despite the fact that three variables were categorical in the original data description: sg (status), al (alphabetical), and su (status), the data was interpreted numerically (numerical). Category variables have been replaced by factor variables. Each sample was given a unique identifier, which may vary from 1 to 400. Although there are numerous missing values in the data set, there are 158 complete occurrences. For a variety of reasons, patients may fail to take measurements that are essential for disease diagnosis. If the diagnostic categories of samples are unclear, missing values will show in the data, necessitating an imputation strategy.

The original CKD data set's categorical variables were first analysed and filled in. For each missing value in the dataset, KNN imputation was employed, which selects the K complete samples with the least Euclidean distance. All missing values are filled using data from K complete samples, and if there are more than K complete samples, data is used for missing values in a given variable category. Because people in similar physical settings should have similar physiological data, KNN is utilised to fill in the missing physiological measurements. The constancy of physiological measurements in healthy people is an example of this. Physiological measurements should be the same for people with the same level of sickness. Differences in physical data between people in similar circumstances should not be substantial. This method, which has already been used to analyse hyperuricemia diagnostic data, should be applied to other diseases.

C. EXTRACTING FEATURE VECTORS OR PREDICTORS:

Extraction of features or predictors could remove factors that are not predictive nor connected to the response variables, and so prevent these unrelated variables from being included in the prediction process. By changing the model's architecture, which leads to more accurate predictions [34]. We used RF and optimum subset regression to find the most essential predictor variables in our investigation. Optimal subset regression is a technique that examines all possible combinations of predictors to identify the one that produces the best results. RF calculates the contribution of each variable to the decline in the Gini index. As the Gini index rises, categorization uncertainty increases. The variables with a 0 value are regarded redundant due to their small impact. Each dataset was submitted to the same approach to ensure thorough feature extraction.

Figures 1 and 2 demonstrate the results of KNN imputation on a single complete data set, and this data set was obtained when K equals 9.

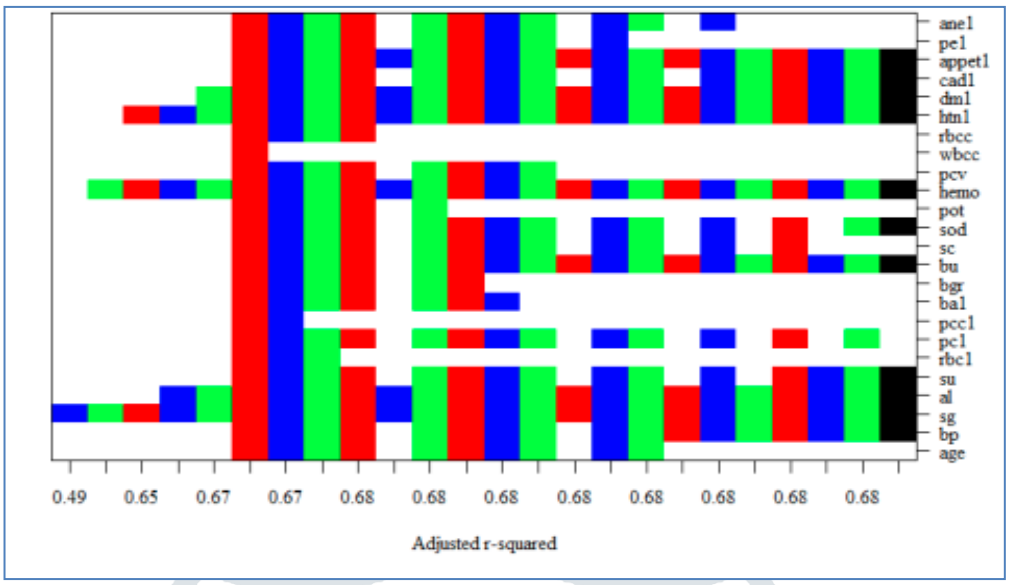

FIGURE 1. The results of utilizing optimal subset regression at K = 9 to extract key variables.

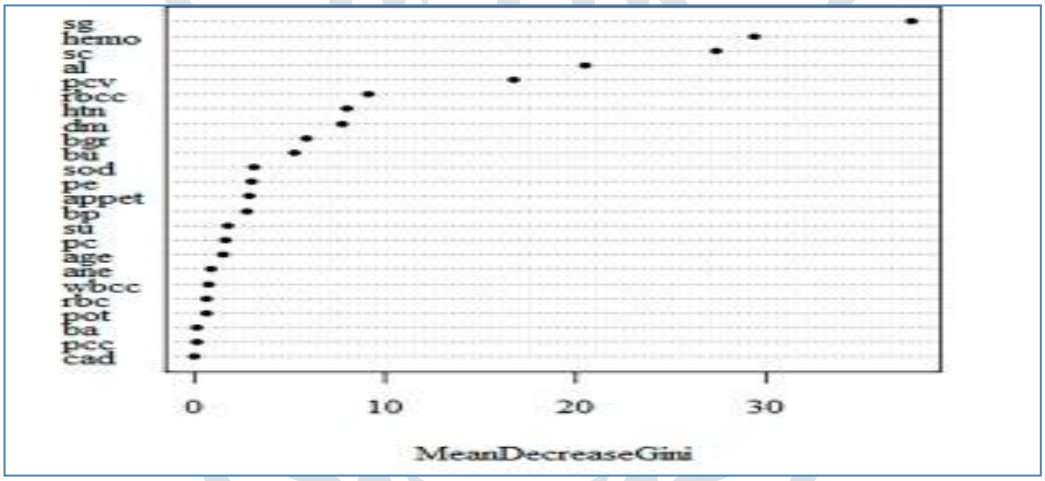

FIGURE 2. Important variables were extracted using RF at K = 9 and shown here.

3.PROPOSED MODEL:

Before creating classifiers, different machine learning approaches were applied to diagnose the data samples in this section. From this group, some of the top models were picked for consideration as potential components. By examining their incorrect assumptions, they were able to figure out the component models. An integrated model was then created in order to increase performance.

A. ESTABLISHING AND EVALUATING INDIVIDUAL MODELS:

The following machine learning models for diagnosing CKD were created using a portion of the complete CKD data sets. Models based on regression include: RF is a tree-based model in this circumstance. SVMs (Simultaneous Vector Machines) are a type of parallel vector machine (SVM) The fourth model is the KNN distance model: The NB model is based on probability. 6) A neural network is referred to as a FNN. Samples for disease diagnosis are typically dispersed across multiple dimensions. This section contains predictors that are used to classify data (ckd or notckd). Because of their diverse classifications, data samples are grouped in different sections of the space. As a result of the decrease in distances between samples in each group, there is a border between the two categories.

We use the aforementioned approaches for disease diagnosis because of their classification capacities. To calculate the weights of each predictor, LOG uses linear regression. Based on the total of their effects, all predictors will be used to determine whether the sample is classed as ckd or not ckd. Random sampling of training samples and predictors can produce a large number of decision trees. The goal of each decision tree is to find a boundary that maximizes the difference between the ckd and notckd values. To achieve a conclusive diagnosis of the disease, all trees' forecasts are considered. The samples' predictors are organized into a multidimensional decision surface using SVM. Based on the distances between the test sample and the training samples, KNN votes on which diagnostic category the test sample should be assigned to. The number of ckd and notckd samples in the measurement interval is used to compute the conditional probability of a sample. Hidden and output layers of the FNN utilized the sigmoid activation function to analyze non-linear relationships in the data.

Training stage
Input: Training data
Output: Integrated model (LOG, RF and perception).
Procedure
1. Use training data to train the model of LOG.
2. Use training data and default parameters to train the model of RF.
3. Input training data into LOG and RF to record the probabilities that the samples are judged as notckd by them.

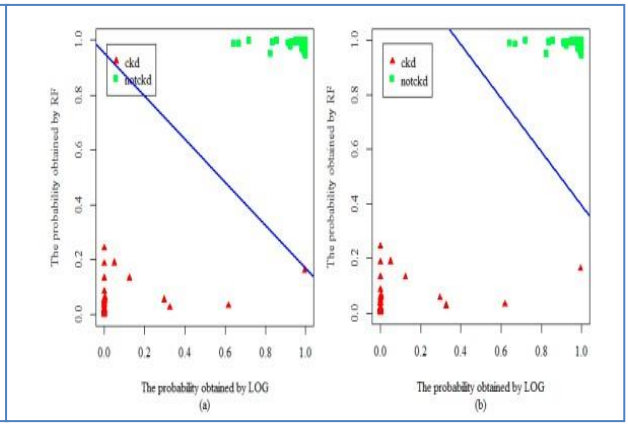

Figure 3: Steps for Training and values

When the perceptron is trained on a dataset, it generates a decision line (blue line). Probabilities of samples being classified as notckd by models are represented by the axes. The response variable is the label of training data, and the predictors are the probability of being recorded. The perceptron is initialized, W is created at random, and b is set to zero. Using the new training data set, go over all of the samples. If (9) is not met, use (13) to update W and b. (15). When all of the sample is satisfied, repeat step 6. (9). Return LOG, RF, and perceptual information**.**

Testing stage
Input: Test data
Output: Sample category
Procedure
1. Input the data into LOG and RF to record the probabilities that the samples are judged as notckd by them.
2. Input the probabilities into the perceptron to obtain the result.

Figure 4: Steps for Testing

4. EXPERIMENTS AND EVALUATIONS

Based on this, we attempted to build and test an integrated model for each data set using random number seed 1234, as shown in Table 2. Table 2 shows the results of the confusion matrices. When K = 3 and 11, an integrated model outperforms component models in terms of performance, as illustrated in Tables 9 and 5. With K equal to 5, 7, and 9, integration improves Log performance while maintaining Rf model accuracy. Seed 1234, which assisted in partitioning the data into four subsets and determining the RF, had been deleted to allow for a more thorough examination. After that, the entire datasets were run through the algorithm ten times. Table 2 indicates that the integrated model outperforms the component models in detecting the two distinct groups when using data from both models. The matrix of confusion that the integrated models produce.

Table 2: Performance of Integrated Model

Models at different values of K	Actual	Prediction	
		ckd	notckd
Integrated model at K = 3	ckd	250	0
	notckd	0	150
Integrated model at K = 5	ckd	250	0
	notckd	1	149
Integrated model at K = 7	ckd	248	2
	notckd	1	149
Integrated model at K = 9	ckd	249	1
	notckd	0	150
Integrated model at K = 11	ckd	250	0
	notckd	0	150

It has the best accuracy and F1 scores in almost every situation. When compared to component models, integrated models have varied degrees of improvement in accuracy and F1 scores, and component sensitivity is also improved. As a result, we can conclude that an integrated model is capable of producing the best possible result. The approaches employed in this study (LOG, RF, and the integrated model) were compared to those used in previous research with the same data (contrast models), with the results shown in Table 3.

Table 3. The proposed model's performance on the same data as the other models.

K	LOG				RF				Integrated model			
	Acc	Sen	Spec	F1	Acc	Sen	Spec	F1	Acc	Sen	Spec	F1
3	98.50	97.84	99.60	98.79	99.58	99.32	100	99.66	99.78	99.80	99.73	99.82
5	98.50	97.88	99.53	98.79	99.75	99.60	100	99.80	99.78	99.88	99.60	99.82
7	98.58	97.92	99.67	98.85	99.68	99.48	100	99.74	99.63	99.52	99.80	99.70
9	98.95	98.44	99.80	99.15	99.65	99.44	100	99.72	99.83	99.84	99.80	99.86
11	98.85	98.16	100	99.07	99.73	99.56	100	99.78	99.83	99.76	99.93	99.86

Acc, Sen, Spec and F1 represent the accuracy, sensitivity, specificity and F1 score, respectively. Their unit is %.

Having stated that, despite the fact that this model's performance is lower than some prior models, it still outperforms roughly half of the contrast models. Although there are various models that are equivalent to the RF, the majority of the models produced in previous works outperform the RF. Individual models perform better, and almost all contrast models are inferior, with the maximum accuracy and F1 score hitting 100% in Table 4. This method appears to be a viable option, according to our findings. For LOG, RF, SVM, and FNN, KNN imputation allowed them to exceed the competition. In this table, we compare our suggested integrated model to the LOG, RF, and RF models.

Table 4: Performance Evaluation

Model	Accuracy (%)	Sensitively (%)	Specificity (%)	F1 score (%)
Fuzzy rule-building expert system [1]	99.60	99.30	100	-
KNN [2]	95.75	93.20	100	96.48
RF [2]	100	100	100	100
SVM without feature selection [6]	97.75	96.40	100	98.17
SVM with Filter SubsetEval with Best first [6]	98.50	97.60	100	98.79
KNN [27]	99.70	100	99.30	-
SVM [27]	99.70	100	99.30	-
LOG [28]	98.10	98.97	96.77	98.40
MLP [28]	98.10	98.97	96.77	98.40
C4.5 decision tree [29]	99.00	99.60	98.00	99.20
MLP [29]	99.75	99.60	100	99.80
KNN [29]	95.75	93.20	100	96.48
Neural network [30]	99.75	99.60	100	99.80
SVM [30]	97.75	96.40	100	98.17
decision tree [31]	99.10	-	-	-
The best average result in this study				
LOG	98.95	98.44	99.80	99.15
RF	99.75	99.60	100	99.80
Integrated model	99.83	99.84	99.80	99.86

When mean/mode imputation and random imputation were both utilized When the diagnostic categories are uncertain, KNN imputation can fill in the missing values in a data set, bringing the data set closer to the real medical condition. The component models for this study were derived from a study on judgment errors. Most samples in the data set are linearly separable, according to the LOG's 98.75 percent accuracy. The RF outperformed the LOG by roughly 99.75 percent in terms of accuracy. Both LOG and RF make different mistakes in almost every situation, as shown in Tables 7 and 8, and the relevant computation times are also fast in most cases.

The component models' performance was improved by using an integrated model that included both LOG and RF data. According to the simulation results, combining multiple classifiers is viable and effective. Perhaps in the future, this method could be used to more challenging challenges. To build models for processing more complicated data, several alternative methods are first tested. After misjudgment analysis, component models are developed from better algorithms that produce different types of errors. The performance of the classifier is then improved by layering an integrated model on top of the previous one.

Tables 3 and 4 illustrate a comparison of the proposed methodology to previous studies by comparing the performance of the independent models to the models proposed in previous studies. We employed Euclidean distance to assess the distance between samples to further examine sample similarity, and KNN was able to reach a high accuracy of 99.25 percent based on Euclidean distance in this investigation. Because the CKD data set comprises both quantitative and categorical characteristics, mixed data approaches can be utilized to assess sample similarity. This is why we didn't examine the similarity of samples in any other way.

5. RESULTS

The required data can be uploaded clicking on the "upload dataset" button." Shown in figure 5.

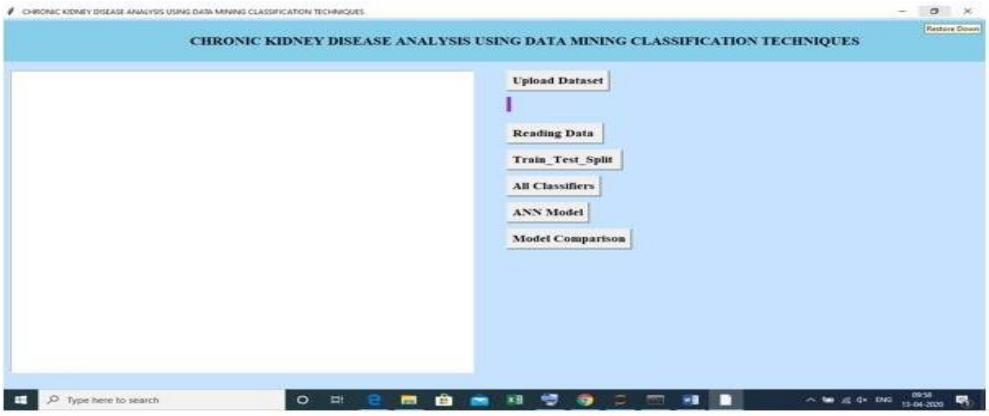

Figure 5. Screen for data upload

We can divide the data into training and testing by clicking on the "Train Test split" button. The results are shown in figure 6.

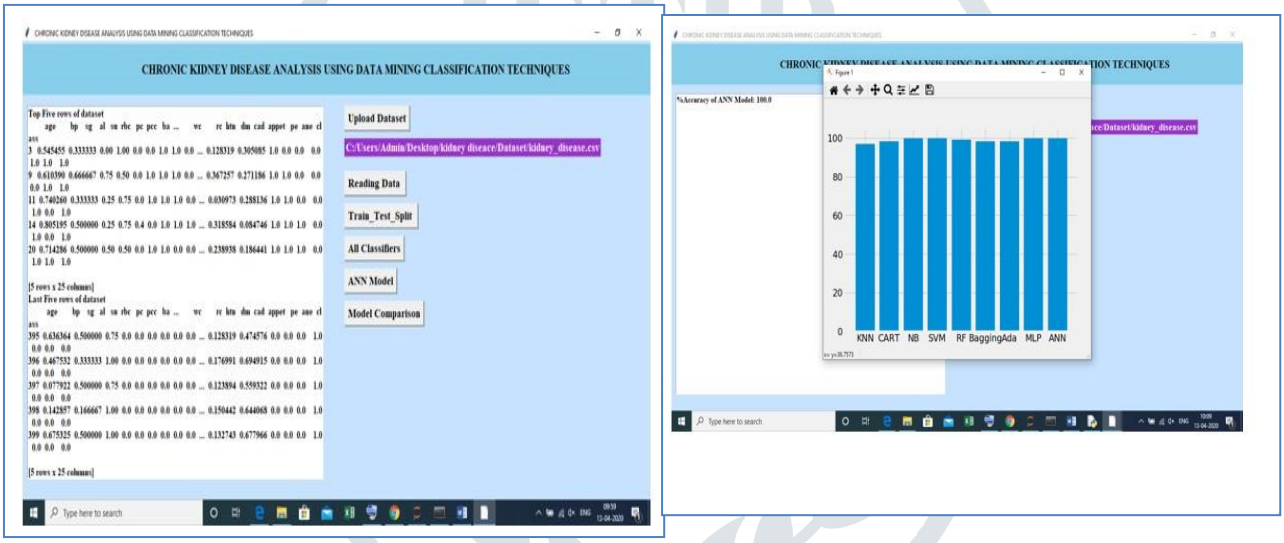

Figure 6: Results

6. CONCLUSION:

The proposed CKD diagnostic methodology is feasible in terms of data imputation and sample diagnosis. KNN imputation allowed the integrated model to achieve a decent degree of accuracy for missing data. The use of this approach to reliably diagnose CKD would therefore be useful in our opinion. The clinical data of various disorders may also be used in actual medical diagnosis using this methodology. Accordingly, the number of data sets accessible for model development is limited, with just 400 samples available for use in this procedure. It is possible that this will limit the model's ability to generalize. Furthermore, the model is unable to determine the severity of CKD because the data set contains just two types of samples (ckd and notckd). To increase the model's ability to generalize and detect disease severity, a vast amount of more sophisticated and representative data will be gathered over time. As data sets grow in quantity and quality, we expect this model to improve even further.

REFERENCES:

[1] Z. Chen, Z. Zhang, R. Zhu, Y. Xiang, and P. B. Harrington, ‘‘Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers,’’ Chemometrics Intell. Lab. Syst., vol. 153, pp. 140–145, Apr. 2016.
[2] A. Subasi, E. Alickovic, and J. Kevric, ‘‘Diagnosis of chronic kidney disease by using random forest,’’ in Proc. Int. Conf. Med. Biol. Eng., Mar. 2017, pp. 589–594.
[3] L. Zhang, ‘‘Prevalence of chronic kidney disease in China: A crosssectional survey,’’ Lancet, vol. 379, pp. 815–822, Mar. 2012.
[4] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, ‘‘Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration,’’ J. Biomed. Informat., vol. 53, pp. 220–228, Feb. 2015.
[5] A. M. Cueto-Manzano, L. Cortés-Sanabria, H. R. Martínez-Ramírez, E. Rojas-Campos, B. Gómez-Navarro, and M. Castillero-Manzano, ‘‘Prevalence of chronic kidney disease in an adult population,’’ Arch. Med. Res., vol. 45, no. 6, pp. 507–513, Aug. 2014.
[6] H. Polat, H. D. Mehr, and A. Cetin, ‘‘Diagnosis of chronic kidney disease based on support vector machine by feature selection methods,’’ J. Med. Syst., vol. 41, no. 4, p. 55, Apr. 2017.
[7] C. Barbieri, F. Mari, A. Stopper, E. Gatti, P. Escandell-Montero, J. M. Martínez-Martínez, and J. D. Martín-Guerrero, ‘‘A new machine learning approach for predicting the response to anemia treatment in a large cohort of end stage renal disease patients undergoing dialysis,’’ Comput. Biol. Med., vol. 61, pp. 56–61, Jun. 2015.
[8] V. Papademetriou, E. S. Nylen, M. Doumas, J. Probstfield, J. F. Mann, R. E. Gilbert, and H. C. Gerstein, ‘‘Chronic kidney disease, basal insulin glargine, and health outcomes in people with dysglycemia: The ORIGIN Study,’’ Amer. J. Med., vol. 130, no. 12, pp. 1465.e27–1465.e39, Dec. 2017.
[9] N. R. Hill, ‘‘Global prevalence of chronic kidney disease—A systematic review and meta-analysis,’’ PLoS ONE, vol. 11, no. 7, Jul. 2016, Art. no. e0158765.
[10] M. M. Hossain, R. K. Detwiler, E. H. Chang, M. C. Caughey, M. W. Fisher, T. C. Nichols, E. P. Merricks, R. A. Raymer, M. Whitford, D. A. Bellinger, L. E. Wimsey, and C. M. Gallippi, ‘‘Mechanical anisotropy assessment in kidney cortex using ARFI peak displacement: Preclinical validation and pilot in vivo clinical results in kidney allografts,’’ IEEE Trans. Ultrason., Ferroelectr., Freq. Control, vol. 66, no. 3, pp. 551–562, Mar. 2019.