International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Hybrid Machine Learning Based Approach to Reduce the Features for Prediction of Long-Term Renal Ailment

Jayashree M[1,*], Dr. Anitha N[2]

Submitted: 10/09/2022 **Accepted:** 20/12/2022

Abstract: The most intensifying communal health problem is chronic kidney disease (CKD) caused by various conditions that reduce the efficiency of the kidneys which results in other health complications, that ultimately lead to the demise of the affected individual. This paper describes an experiment that uses a novel approach to fill missing values and combines related attributes based on domain knowledge and it is evaluated by 4 different Machine Learning (ML) classification algorithms – Logistic Regression, an Extended Gradient Boosting trees, Artificial Neural Networks (ANN), and Support Vector Machine (SVM). Among these four algorithms, ANN and XG-Boost provide the better result of 0.97 F1- Score. The standard related data is taken from a repository called UCI machine learning, which has 400 individual data, 250 were reported to have CKD.

Keywords: *Dimensionality Reduction, Logistic Regression, Random Forest, Support Vector Machines, Extended Gradient Boosting Trees, Machine Learning, Chronic Kidney Disease.*

1. Introduction

Kidneys are the major excretory and most important of the vital organs that purge unwanted substances and extra liquid from the human body. Also, they regulate the PH levels in the body to maintain balance in the fluids. They control blood pressure, synthesize a chemical 'erythropoietin' used to produce red blood cells, and produce Vitamin D in its active form to promote the strong and healthy structure of the jaw bone. Conditions like diabetes mellitus, Hypertension, cardiovascular diseases, ethnicity, and kidney issues in the family history, old age, and recurrent use of few suppositories affect the functions of the Renal adversely.

Kidney diseases can be classified into acute and chronic kidney diseases (AKD and CKD). AKD occurs when kidneys stop working suddenly due to reduced blood flow (reduced BP), direct damage to the kidneys, or retention of urine in the kidneys. Loss of blood due to traumatic injury, severe infection, autoimmune diseases, enlarged prostate blocking the flow of urine, complications due to pregnancy such as 'Eclampsia', certain medications that affect the kidneys directly, a sudden increase in protein levels in the blood due to dehydration or muscle tissue break down, liver failure, and cardiovascular diseases lead to Acute disease. Gradual attenuation of kidneys and progressive degradation in renal function for a duration of three months and to several years cause chronic kidney disease. The patient may be asymptomatic in its early stages. Generally, High blood pressure and high glucose levels in the blood will affect renal disease. It may also be caused by Polycystic kidney disease (a genetic condition also called PKD), inflammations in kidneys caused by scars due to repeated infections, autoimmune diseases, and prolonged use of certain medications like as 'NSAIDs' (nonsteroidal anti-inflammatory drugs).

CKD leads to loss of weight, muscle weakness, fatigue, pedal edema, gastrointestinal symptoms, etc. National Kidney Foundation classifies kidney disease into five stages based on the amount of kidney damage and function. It is determined by the glomerular filtration rate (GFR), a numerical expression that uses a person's gender, age, serum creatinine level and ethnicity. Stage 1 damage is mild with GFR > 90mL/min, and probably has no symptoms, whereas in stage 5 with GFR <15 mL/min kidneys have stopped working already.

[1] Research Scholar, EPCET and Sr. Assistant Professor, New Horizon College of Engineering Computer Science, VTU, Bangalore, India.
[2] BNM Institute of Technology, Professor, Computer Science, VTU, Bangalore, India, anitha.mhp@gmail.com.
* *Corresponding Author Email:* jayashreevtu123@gmail.com

CKD is a major health condition affecting individuals worldwide with increasing incidents and prevalence, and high management costs. According to [1], around 2.5% – 11.2% of grown men and women across Northern America, Asia, Australia, and Europe are comparatively highly affected by CKD. The USA is home to at least 27 million affected individuals [2]. The Kidney Foundation for Kidneys at the National level states that about 59% of the population has a high probability of developing chronic kidney disease in their lifetime in the USA [3].

Initial detection and medication of renal disease can prevent some of these adverse conditions [4]. Even though there is a rise in overall knowledge of chronic renal disease, it is still far too low. Less than 5% of patients were identified with stage 1 or stage 2, 10% were diagnosed with stage 3, and 45% were diagnosed with stage 4, according to a survey of National Health and Nutrition Examination reports [5]. Primary care providers (PCPs) are left with the burden of management of chronic kidney disease because there are very few practicing nephrologists, and they are unable to cope with diagnosing patients with the disease. [6] Shows that kidney disease awareness among PCPs is unacceptably low. Family practitioners, even with substantial practice, have very poor knowledge of disease management. Hence an automated and accurate Kidney detection system will make it convenient for involved physicians to manage CKD efficiently.

Many classification and regression ML algorithms can be used to classify CKD. The main goal of this experiment is to build a pipeline consisting of three steps. Step 1 uses a combination of RF algorithm and domain knowledge to fill missing values; step 2 involves the combined use of unsupervised and supervised ML algorithms to combine relevant features; step 3 evaluates 4 ML classification algorithms, Logistic Regression, ANN, SVM, and XGBoost to classify individuals into persons with and without CKD.

The structure of the paper is ordered under subsequent sections:

Section 2 provides some information about related work in the detection of CKD; section 3 provides information about the dataset considered and the methodology used; section 4 lists the results, and section 6 is on the take away from the paper.

2. Related Work

Asif Salekin Et.al considered 24 predictive parameters to evaluate three Machine Learning Algorithms, KNN, RF, and the MLP-based Neural Network approach to predict CKD. The IBK algorithm (KNN) and C4.5 trees (RF) were able to handle the missing values from the dataset, while the MLP ignored the missing values. They concluded that the RF classifier yields an accuracy of detection is 0.993 F1-measure and 0.1084 RMSC on all 24 attributes [14, 15].

M.Doğruyol Başar, Et.al used the WEKA tool and 10-fold-cross-validation to measure 7 different classification ML algorithms (NB, Hoeffding Tree, Random Tree, REPTree Adaboost, and IBk) on [14,17] dataset, for the diagnosis of the CKD. They concluded that individual random tree and IBK classifiers yield better results than the considered 7, over 6 reduced predictive parameters.

Guneet Kaur, Et.al evaluated the SVM, KNN, and combined SVM/KNN using the MATLAB tool on Hadoop based dataset [14, 18] to predict CKD. They used a multi-layer classification approach and observed that the combination of SVM and KNN attained better accuracy, precision, and recall.

S. Revathy, Et.al considered the dataset for their experiments. They evaluated the Decision Tree, RF, and SVM ML algorithms to predict CKD in 400 patients. [14, 21] They filled up the lost values in the dataset using mean for mathematical values and mode for categorical values. They observed that the RF yielded maximum accuracy of 99.16%.

S. Ramya and et.al [23] evaluated ANN, Neural Network with Radial Basis Function, and RF on a dataset consisting of 1000 instances of CKD, with 15 attributes. They collected this data from various laboratories in Coimbatore. They used the R programming Language for their work. They concluded that the neural network algorithm with Radial Basis Function yields the best results of 85.3% accuracy.

L. Jerlin Rubini et.al used the Weka tool and 10-fold cross-validation technique to measure RBFN, MLP, and logistic regression classifiers on the dataset [14, 24] to predict CKD. They concluded that MLP yields a prediction accuracy value of 99.75, specificity value of 100, the sensitivity of 99.33, and F-measure of 99.66.

N. Pavithra et.al employs the FCM clustering algorithm using the MATLAB tool, to understand the risk of CKD with the available information from the dataset [14, 25]. Their experiment initializes the fuzziness between 0 and 1 and outputs 3 clusters, the first corresponding to individuals with risk of CKD, the second corresponding to normal individuals, and the third corresponding to the mispredictions. Consists of 14 attributes and in the 14^{th} attribute, the output corresponds to the presence or absence of risk. Their experiment yielded a prediction accuracy of 92%.

Milan deep Arora and Et. Al employed the WEKA tool to compare the performance of NB, J48 RF, and SMO ML algorithms with cross validation-10 fold technique on the dataset [14,27] to predict CKD. They considered all 25 attributes and concluded that the RF classifier yielded the best results of 99% class accuracy, and 0.0225 mean absolute error.

Misir R Et.al implemented the feature selection and dimensionality reduction stage using correlation, discretization, Random Search, Greedy Stepwise Linear Forward Selection, Genetic Search, Exhaustive Search, and Scatter Search on the dataset [14, 28]. They compared the performance of Incremental Back Propagation Learning Networks (logistic activation function and 1 hidden layer) and Levenberg–Marquardt Classifiers to predict CKD from the dataset.

Siddheshwar Tekale [30] used the WEKA tool to evaluate the SVM and Decision Tree ML algorithms on the dataset [14] to predict CKD. They reduced the number of features to 13 and applied arithmetic mean / mode to fill the NAS. They concluded that SVM performed better with 0.1803 Root mean squared Error, 6.9308% Relative absolute error, 37.2379% Root relative squared error, 0.9579 F1 Measure, 0.9308 Precision, and 0.9866 Recall values.

3. Dataset Used in this Experiment

The Experiment described in this paper was performed using the data set [14] which consists of twenty-five attributes, the twenty-fifth being the target class which corresponds to the absence or presence of CKD. The attributes present in the data set are:

Table 1. The Features from the Data set (24 attributes)

ATTRIBUTE	SHORT FORM	TYPE	UNITS
Age	Age	Numeral	Years
Systolic blood pressure	BP	Numeral	Mm/Hg
Specific Gravity (urine)	SG	Numeral	None
Albumin (in urine)	Al	Categorical / Nominal	1 – 5
Sugar (in urine)	Su	Categorical / Nominal	0 – 4
Red blood cells (in urine)	RBC	Categorical/nominal	0 – 1 (normal / abnormal)
Pus Cells (in urine)	PC	Categorical / Nominal	0 – 1 (normal / abnormal)
Pus Cell clumps (in urine)	PCC	Categorical / Nominal	0 – 1 (present / not present)
Bacteria (in urine)	BA	Categorical/nominal	0 – 1 (present / not present)
Blood glucose random	BGR	Numeral	Mg/dl
Blood Urea	BU	Numeral	Mg/dl
Serum Creatinine	SC	Numeral	Mg/dl
Sodium	Sod	Numeral	mEq/L
Potassium	Pot	Numeral	mEq/L
Hemoglobin	Hemo	Numeral	Gms
Packed Cell Volume	PCV	Numeral	
White Blood Cell Count	WC	Numeral	cells/cumm
Red Blood Cell Count	RC	Numeral	millions/cmm
Hypertension	HTN	Categorical/nominal	0 – 1 (yes / no)
Diabetes Mellitus	DM	Categorical/nominal	0 – 1 (yes / no)
Coronary Artery Disease	CAD	Categorical/nominal	0 – 1 (yes/no)
Appetite	appet	Categorical/nominal	0 – 1 (good/poor)
Pedal Edema	PE	Categorical/nominal	0 – 1 (yes/no)
Anemia	Ane	Categorical/nominal	0 – 1 (yes/no)
Classification	Classification	Categorical/nominal	0 – 1 (CKD/NOCKD)

The dataset used in this experiment [14] includes data collected from 400 individuals of whom 250 were reported to have CKD. The following paragraphs explain briefly some of the important attributes listed in Table 1 which are associated with kidney diseases in general.

Elevated blood pressure and Hypertension: an increase in BP constricts, narrows, and blood vessels get weak all over the body including the kidneys. The flow of blood is reduced. Damaged blood vessels in the kidneys prevent them from functioning properly and purge the human body from toxins and extra fluids [7].

Coronary artery disease (CAD): Recent research shows that the significant risk for kidney diseases is the failure of the Heart. when pumping of the blood becomes congested, it causes the pressure to build up in the main vein that gets connected with the kidneys which intern leads to congestion in the kidneys.

Blood glucose random and Diabetes Mellitus: Diabetes Mellitus [DM) is one of the major risk factor for CKD. High blood glucose from DM causes damage to the kidneys by narrowing and clogging the blood vessels in the kidneys thereby limiting the blood supply to the kidneys.

Blood Urea: Urea in the blood is measured using the Blood Urea Nitrogen test and determines the functioning of the kidneys. The build-up of urea in the blood, 'Uremia', occurs when the kidneys stop filtering toxins from the blood. Uremia signals the end stage of CKD.

Serum creatinine: creatinine is generated from muscles as a waste product. The Kidneys eliminate the creatinine from the blood through the urine. Creatinine level in the blood determines the wellness of the kidneys and is determined by a lab test. The level of Serum creatinine gauges the glomerular filtration rate and is an oblivious marker of CKD. A raise of Serum creatinine

from 0.8 mg/dl to 1.2 mg signals a decrease of GFR by 33%.

The number of omitted values in the dataset for each attribute is shown in the graph below shows. The hybrid approach is considered in filling these values and updated data set with no missing value is taken to an eliminated set of attributes for the prediction of CKD. Fig 1 shows the number of missing values in each attribute.

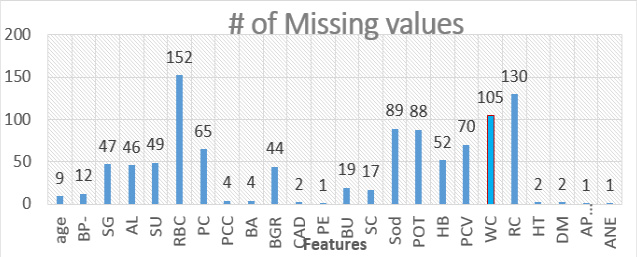

Fig. 1. The Missed Values in the UCI Repository Dataset

4. Methodology Used

The objective of the study is to detect whether the person is affected by kidney disease. The two main objectives associated with this experiment are to reduce the attributes set and to evaluate the ANN, Logistic Regression, SVM, and XGBoost ML algorithms for maximum accuracy to predict the CKD or NON-CKD individuals.

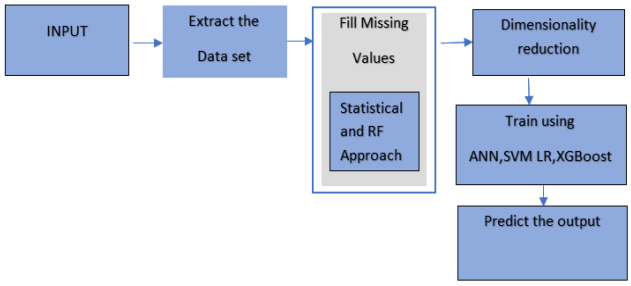

Fig. 2. General Block diagram of the process

4.1. Addressing the Missing Values

From the description of the dataset above, it can be seen that the attributes can be grouped into 5 groups:

Table 2. The Attributes Group

Attribute	Attribute Group
Anemia	Hemoglobin, packed cell volume, RBC count, and Anemia
Hypertension	Age, Systolic blood pressure, Sodium level, potassium level, Hypertension, and coronary artery disease.
Diabetes	Age, blood glucose random, and Diabetes Mellitus
Infection	bacteria, pus cells, pus cell count, red blood cells in the urine, white blood cell count
Critical	Specific Gravity, albumin in urine, sugar in the urine, Serum Creatinine, blood urea, Appetite, and Pedal Edema

The proposed methodology addresses missing values in each group using different strategies. Considering the group of attributes R = {p1, p_2, p_3, …, p_n}, Where p_1, p_2, p_3, …, p_{n-1} are predictors and p_n is the dependent value. Let S denote the set of R containing no missing values and M be the group with dependent value to be determined. The below three algorithms addresses the missing values in each group.

SelectFirstMatchingValue: Identify the group from S that best matches with the attributes from M and fill the dependent value from the matching group.

SelectAvg: Get all groups of attributes from S that match best with M, compute the mean if the dependent variable is numerical or the arithmetic mode, and fill the missing dependent variable with the value.

RF_Regression: To predict the dependent variable in M, utilise S to train a Random Forest model.
The following paragraph explains how missing values are addressed in each of the groups.

Addressing the missing values in the Hypertension and Diabetes Meletus groups:

To address omitted values which are available in the Hypertension and Diabetes Meletus groups apply the above algorithms to handle the missing values.

1. Apply **SelectFirstMatchingValue** on the set containing the age, BP, and Hypertension attribute to fill the lone missing value in the Hypertension attribute in the hypertension group.
2. Apply **SelectAvg** in the set from the previous step to fill the missing BP values.
3. Apply **RFRegression** to the set from the previous step consisting of age, BP, sodium, potassium, and Hypertension attributes successively to fill missing sodium and potassium attribute values.
4. Initialize the missing coronary artery disease values to ‘no’.
5. Follow steps 1 and 2 on the group consisting of age, blood glucose, and Diabetes Meletus attributes to fill missing Diabetes Meletus and blood glucose values respectively.

Addressing the missing values in the Anaemia group of attributes:

Anemia is a significant factor that determines the problem if a person has kidney illness, and it is predicted in such cases. The deciding attribute ‘Anemia’ has only one missing value that can be filled using the hemoglobin

value manually. Since this attribute has its values filled for all the cases studied, missing values from other attributes may be filled using the following strategy:

1. Apply **SelectAvg** on the group containing Haemoglobin, RBC count, and Anemia attributes successively to fill missing Haemoglobin and RBC count values
2. The packed cell volume does not impact the group, this attribute can be dropped off.

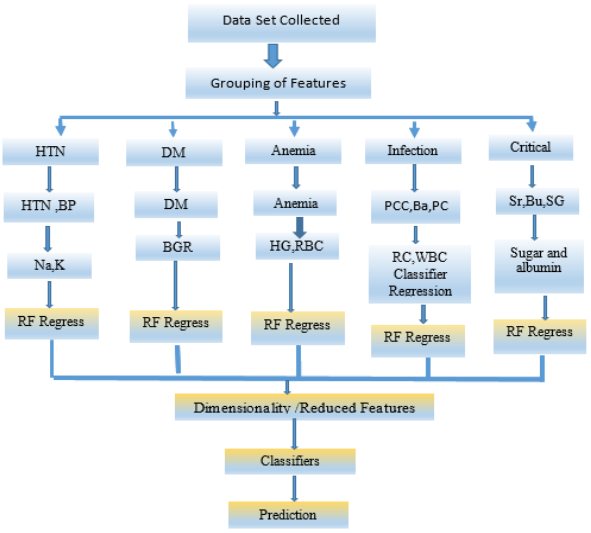

Fig 3: Flow of filling the missing values in the data set

Addressing the missing values in the Critical group:

The critical group of attributes forms the basic group of attributes that decides if the person has a kidney disease at all. Furthermore, all attributes in the critical group depend on creatinine and blood urea for addressing the missing values. Pedal edema and appetite attributes do not have a missing value. The proposed technique addresses the missing values as follows:

1. Apply **SelectAvg** on the group containing specific gravity, creatinine, appetite, pedal edema, and anemia (from the anemia dataset above) to fill the missing creatinine value.
2. Apply **SelectAvg** on the group containing specific gravity, blood urea, creatinine, appetite, pedal edema, and anemia from step 1 to fill the missing blood urea values
3. Apply **RFRegression** on the group containing specific gravity, albumin, sugar, blood urea, creatinine, appetite, pedal edema, and anemia from step 2 successively to predict missing specific gravity, albumin, and sugar values

4.2. Reduction of features

This section discusses the dimensionality reduction process in some detail. The proposed methodology reduces the dataset [14] from 24 attributes to 12 as follows:

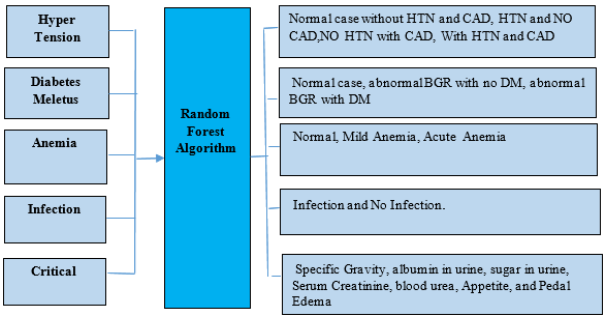

Fig. 4. The Reduction of attributes using the RF Algorithm

The Classification of attributes in each group using the RF algorithm is explained in the following paragraphs.

Classifying the Hypertension group of attributes using the RF algorithm into four classes:

Normal cases without Probable HTN and CAD, Cases with Probable HTN and no CAD, Cases with no Probable HTN and with CAD, Cases with Probable HTN and CAD.

Classifies the attributes from the Diabetes Meletus group into three classes. Normal case, abnormal BGR with no DM, and abnormal BGR with DM.

Classifies the attributes from the Anemia group into three classes. Normal, Mild and acute anemia.

classifies the attribute from the infection group into two classes, Infection, and No Infection.

We need to consider the Age attribute and attributes from the critical group are considered individually.

Table 3: Reduced Features

ATTRIBUTE	SHORT FORM	TYPE	UNITS
Age	Age	Numeral	Years
Specific Gravity (urine)	SG	Numeral	None
Albumin (in urine)	Al	Categorical / Nominal	1 – 5
Sugar (in urine)	Su	Categorical / Nominal	0 – 4
Blood Urea	BU	Numeral	Mg/dl
Serum Creatinine	SC	Numeral	Mg/dl
BP State	BP	Categorical	0 – 1 (yes / no)
Glucose State	BGR	Categorical	0 – 1 (yes / no)

Appetite	appet	Categorical/nominal	0 – 1 (good/poor)
Pedal Edema	PE	Categorical/nominal	0 – 1 (yes/no)
Anemia	Ane	Categorical	0 – 1 (yes/no)
Infection	Inf	Categorical	0 – 1 (yes/no)
Classification	Classification	Categorical/Nominal	0 – 1 (CKD/No CKD)

4.3. The CKD Detection

The proposed methodology applies 4 ML algorithms, **Logistic Regression**, **SVM**, **ANN**, and **XGBoost** to the reduced dataset. The following are the parameters that are used in the below 4 algorithms.

The Logistic Regression approach: The following are the parameters that are used to build the logistic regression for the given data.**Max_iter:**10**, multiclass:** "auto", **penalty:** "l2". It results in an accuracy of **90%.**

The SVM approach: The following are the parameters that are used to build the logistic regression for the given data.**Kernel:** "RBF kernel", **Gamma value:** "1.0". It results in an accuracy of **56%.**

The ANN approach: The following are the parameters that are used to build the logistic regression for the given data.**Layer:** "Dense layer", **Activation function:** "Relu", **Loss function:** "binary_crossentropy", **Optimizer:** "Adam", **Output layer:** "sigmoid" activation, **Epochs:** "250". It results in an accuracy of **97%.**

The XG Boost approach: The following are the parameters that are used to build the logistic regression for the given data.**Base_score:** "0.5", **Booster:** "gbtree", **Learning_rate:** "0.30", **Max_depth:** 6, **n_estimators**: 100, **Predictor:** "auto". It results in an accuracy of **97%.**

4.4. Result in Analysis

4.4.1. Support Vector Machine (SVM)

There are training data and testing data in the data set. The data set has 396 rows and 13 columns, with 80% of the data being used for training and 20% being used for model testing. The network is trained using the SVM model. It results in an accuracy of 56%. The classification report for the SVM model says that for class 0 the precision is 56% and the f1 score for class 0 is 72%. For class 1 precision is around 99%, recall is 3% and f1 score is around 5%. The average weight for precision, Recall, and F1Score is 0.76, 0.56, and 0.42.

4.4.2. XG Boost

There are training data and testing data in the data set. The data set has 396 rows and 13 columns, with 80% of the data being used for training and 20% being used for model testing. The network is trained using the XG Boost model. It results in an accuracy of 96.95%. The classification report for the XG Boost model says that precision is around 94%, recall is 95% and f1 score is 97%. The average weight for precision, Recall, and F1Score is 0.97. It explains that out of 121 samples it classifies 58 samples as true positive. It classifies 69 samples as true negative. It classifies 04 samples as false positives.

4.4.3. Logistic regression

Training data and testing data are separated from the data set. The data set has 396 rows and 13 columns, and 80% of the data is utilized to train the model and 20% to test it. The network is trained using the Logistic Regression model. It results in an accuracy of 90%. The classification report for the Logistic regression says that precision is around 82%, recall is 99% and f1 score is 90%. It explains that out of 278 samples it classifies 125 samples as true positive. The average weight for precision, Recall, and F1Score is 0.82, 0.9, and 0.9. It classifies 125 samples as true negative. It classifies 27 samples as false positive and it classifies 1 sample as a false negative.

4.4.4 Artificial Neural Network (ANN)

Training data and testing data are separated from the data set. The data set has 396 rows and 13 columns, and 80% of the data is utilized to train the model and 20% to test it. Artificial neural networks are used to train the network. The Relu activation function is used as the activation function in the model, which has been trained for 250 epochs. It results in an accuracy of 97.50%. The classification report for the ANN model says that precision is around 97%, recall is 98% and f1 score is 97%. The average weight for precision, Recall, and F1Score is 0.97. It explains that out of 80 samples it classifies 32 samples as true positive. It classifies 46 samples as true negative. It classifies 01 samples as a false positive and it classifies 01 samples as a false negative. The comparison of Performance measures with the graph is shown below.

Table 4. Comparison of Performance Measures

Classification/ performance Measure	**SVM**	**ANN**	**LR**	**XGBoost**
Precision	0.76	0.97	0.82	0.97
Recall	0.56	0.97	0.9	0.97
f1 score	0.42	0.97	0.9	0.97
Accuracy	0.97	0.56	0.9	0.97

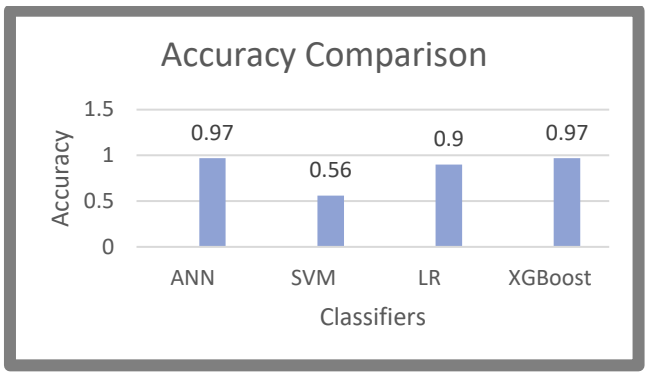

Fig. 5. Accuracy Comparison

Fig. 6. Performance Measure Comparison

5. Conclusion

The dataset studied in this paper comprised noisy and many missing values**.** The proposed approach is considered a hybrid approach combining statistical averaging and random forest algorithms to fill the missing values. It reduced the 24 attributes of the dataset to 12 using a novel grouping and prediction technique. It grouped the attributes into 5 groups based on their purpose namely, BP, blood glucose level, infection state, anemia state, and critical attributes, and used the random forest to predict the intermediate attributes required for the final classification algorithm. The number of classes depended on the group. This approach considers the 7 attributes from the critical group namely serum creatinine, albumin in urine, Sugar in the urine, Specific Gravity, blood urea, Appetite, and Pedal Edema directly in the final classification algorithm without grouping because they are critical for determining if the patient had any kidney disease. The approach proposed in this paper evaluated 4 supervised machine learning algorithms Support Vector Machine, Logistic Regression, Artificial Neural Network, and XG Boosted Trees algorithms to predict cases with CKD. It considered a dataset of 400 individuals, of whom 250 had CKD. Application of these 4 algorithms on the dataset with reduced features yielded better results, and ANN and XG Boosted Trees yielded better accuracies.

6. Future Enhancement

The proposed solution can be extended further by adding components that capture the relevant information from unstructured records using natural language processing. Also, this component can extract relevant information from generic data of the patient to get critical information. The extracted information can be merged with the implementation details proposed in this paper to prognosticate any form of kidney disease before the reduction in the functioning of the kidney, saving the patient's life.

7. References

[1]. Zhang Q.L., Rothenbacher D., *Prevalence of chronic kidney disease in population-based studies: a systematic review*, BMC public health, 2008. **8**(1): p. 117.

[2]. Baumgarten M., Gehr T., *Chronic kidney disease: detection and evaluation*, American family physician, 2011. **84**(10): p. 1138.

[3]. Moyer V. A., US Preventive Services Task Force*, *screening for chronic kidney disease: US Preventive Services Task Force recommendation statement*. Annals of internal medicine, 2012. **157**(8): p.567-570

[4]. Remuzzi G., Ruggenenti P., Perico N., *Chronic renal diseases: renoprotective benefits of renin-angiotensin system inhibition*, Annals of internal medicine, 2002. **136**(8): p.604-615.

[5]. Patil, P. ., D. D. . Waghole, D. V. . Deshpande, and D. M. . Karykarte. "Sectoring Method for Improving Various QoS Parameters of Wireless Sensor Networks to Improve Lifespan of the Network". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 6, June 2022, pp. 37-43, doi:10.17762/ijritcc.v10i6.5622.

[6]. Plantinga Laura C., Boulware L.E., Coresh J., Stevens L.A., Miller E.R., Saran R., Messer K.L., Levey S.A., and Powe N.R., *Patient awareness of chronic kidney disease: trends and predictors*, Archives of internal medicine , 2008. **168**(20): p.2268-2275.

[7]. Plantinga L. C., Tuot D. S., Powe N. R., *Awareness of chronic kidney disease among patients and providers*", Advances in chronic kidney disease, 2010. **17**(3): p. 225–236.

[8]. Chaudhary, D. S. . (2022). Analysis of Concept of Big Data Process, Strategies, Adoption and Implementation. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 05–08. https://doi.org/10.17762/ijfrcsce.v8i1.2065

[9]. Peralta C.A., Whooley M.A., Ix J.H., Michael G., *Kidney Function and Systolic Blood Pressure New Insights From Cystatin C: Data from the Heart and Soul Study*, American Journal of Hypertension, 2006. **19**(9): p. 939–946.

[10]. Bigotte Vieira M., Neves J.S., Leitão L., Baptista R.B., Magriço R., Dias C.V., Oliveira A., Carvalho D., Mc Causland F.R., *Impaired Fasting Glucose and Chronic Kidney Disease, Albuminuria, or Worsening Kidney Function: A Secondary Analysis of SPRINT*, The Journal

of Clinical Endocrinology & Metabolism, 2019. **104**(9): p. 4024–4032.

[11]. Jha V., Garcia-Garcia G., Iseki K., Li Z., Naicker S., Plattner B., Saran R., Wang A.Y.M., Yang C.W., *Chronic kidney disease: global dimension and perspectives* Lancet. , 2013: p. 260-272.

[12]. Afsar B., Turkmen K., Covic A., Kanbay M., *An Update on Coronary Artery Disease and Chronic Kidney Disease*, International Journal of Nephrology, 2014.

[13]. Shankar, M., Narasimhappa S., Madhura N.S., *Urinary tract infection in chronic kidney disease population: a clinical observational study*. *Cureus*, 2021. **13**(1).

[14]. *UCI Machine Learning Repository Pima Indians Diabetes Data Set- online*, https://archive.ics.uci.edu › support › diabetes.

[15]. Dubey A., *A classification of ckd cases using multivariate k-means clustering*, *International Journal of Scientific and Research Publications*, 2015. **5**(8): p.1-5.

[16]. *Chronic Kidney Disease Dataset* (URL 'http://archive.ics.uci.edu/ml/datasets/Chronic Kidney Disease'

[17]. Salekin A., Stankovic J., *Detection of chronic kidney disease and selecting important predictive attributes*, In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, 2016: p. 262-270.

[18]. *The New York Times health guide for swelling*, accessed: 2016-3-11 http://www.nytimes.com/health/guides/symptoms/swelling/overview.html.

[19]. Doğruyol Başar M., Aydin A.K.A.N., *Chronic Kidney Disease Prediction with Reduced Individual Classifiers*, Istanbul University, Istanbul, Turkey. Electrica, 2018, **18**(2): p. 249-255,

[20]. Ahmed Cherif Megri, Sameer Hamoush, Ismail Zayd Megri, Yao Yu. (2021). Advanced Manufacturing Online STEM Education Pipeline for Early-College and High School Students. Journal of Online Engineering Education, 12(2), 01–06. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/47

[21]. Kaur G., Sharma A., *Predict Chronic Kidney Disease Using Data Mining Algorithms In Hadoop*, Amritsar College of Engineering and Technology, Amritsar, India. International Journal of Advances in Electronics and Computer Science, Apr.-2018. **5**(4), ISSN: 2393-2835.

[22]. Pitchumani Angayarkanni S., *Predictive Analytics of Chronic Kidney Disease using Machine Learning Algorithm*, Lady Doak College, Madurai, Tamil Nadu, India. International Journal of Recent Technology and Engineering, July 2019. **8** (2), ISSN: 2277-3878, DOI: 10.35940/ijrte.B1727.078219

[23]. Kumar C.S., Thangaraju P., *Improving Classifier Accuracy for diagnosing Chronic Kidney Disease Using Support Vector Machines*, Al, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India. International Journal of Engineering and Advanced Technology, 2019. **8**(6), ISSN: 2249 – 8958, DOI: 10.35940/ijeat.F9377.088619.

[24]. Revathy S., Bharathi B., Jeyanthi P., Ramesh M., *Chronic Kidney Disease Prediction using Machine Learning Models*, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India. International Journal of Engineering and Advanced Technology, October 2019. **9** (1), ISSN: 2249 – 8958, DOI: 10.35940/ijeat.A2213.109119

[25]. Almasoud M., Ward T.E., *Detection of Chronic Kidney Disease using Machine Learning Algorithms with Least Number of Predictors*, College of Computer and Information Science, Princess Nourah bint Abdul Rahman University, Riyadh, Saudi Arabia. International Journal of Advanced Computer Science and Applications, 2019. **10**(8).

[26]. Hermina, J. ., Karpagam, N. S. ., Deepika, P. ., Jeslet, D. S. ., & Komarasamy, D. (2022). A Novel Approach to Detect Social Distancing Among People in College Campus. International Journal of Intelligent Systems and Applications in Engineering, 10(2), 153–158. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/1823

[27]. Ramya S., Radha N., *Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms*, PSGR Krishnammal College for Women, Coimbatore, India. International Journal of Innovative Research in Computer and Communication Engineering, 2016. **4**(1).

[28]. Rubini L., Eswaran P., *Generating comparative analysis of early-stage prediction of Chronic Kidney Disease*, Alagappa University, Karaikudi, Tamil Nadu, India International OPEN ACCESS Journal of Modern Engineering Research, 2015. **5**(7), ISSN: 2249–6645.

[29]. Pavithra N., Shanmugavadivu R., *Efficient Early Risk Factor Analysis of Kidney Disorder Using Data mining Technique*, PSG College of Arts & Science, Coimbatore, India International Journal of Innovative Research in Computer and Communication Engineering, 2017. **5**(2), ISSN (Print): 2320-9798, DOI: 10.15680/IJIRCCE.2017. 0502084 1690.

[30]. Kumar M., *Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm*, Banaras Hindu University, Varanasi, India. International Journal of Computer Science and Mobile Computing, 2016. **5** (2): p. 24-33, ISSN 2320–088X.

[31]. Arora M., Sharma E.A., *Chronic Kidney Disease Detection by Analyzing Medical Datasets in Weka*, ACET, Amritsar, India. International Journal of Computer Application, 2016. **6**(4).

[32]. Misir R., Mitra M., Samanta R.K.A., *Reduced Set of Features for Chronic Kidney Disease Prediction*, Vidyasagar University, Medinipur, West Bengal, India. Journal of Pathology Informatics, 2017. DOI: 10.4103/jpi.jpi_88_16.

[33]. Sharma S., Sharma V., Sharma A., *Classification Techniques for Chronic Kidney Disease Diagnosis*, University Of Jammu, Jammu, India.

[34]. Tekale S., Shingavi P., Wandhekar S., Chatorikar A., *Prediction of Chronic Kidney Disease Using Machine Learning Algorithm*, Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering & Technology, Baramati, Pune, Maharashtra, India. International Journal of

Advanced Research in Computer and Communication Engineering, October 2018. **7**(10), ISSN (Print) 2319-5940, DOI: 10.17148/IJARCCE.2018.71021 92.