International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Robust Chronic Kidney Impact Identification System Using Prab Algorithm

S. Mohammed Imran[1], N. Prakash[2], K. Hazeena[3], S. Sivasubramanian[4]

Submitted: 02/10/2023 **Revised**: 21/11/2023 **Accepted**: 02/12/2023

Abstract: Kidneys act as one of the important organs of the body. Kidney disease needs to be treated in the early stages to protect Human life. Chronic kidney disease slowly impacts the functionality of kidneys that make severe damage to kidneys if untreated. Chronic kidney disease affects other organs of the body and creates life threatening problems. Early prediction of kidney disease can save the life of the person and reduce the financial cost taken for treatment purposes. The proposed approach is focused on developing a robust algorithm named probability weighted AdaBoost to predict chronic kidney disease. The adaptive weight calculation of Adaboost algorithm (PRAB) is modified here to provide the reduction of incorrectly classified values on kidney disease detection. The existence of false positive data is one of the major drawbacks produces that induces similarity issues in classification. To avoid the weak classification criteria PRAB algorithm is utilized here. The adaptive selections of classifier with removal of weak data analyzer enhance the accuracy of prediction.

Keywords: *Chronic kidney disease, adaptive boosting, data analytics, machine learning, and exploratory data analysis*

1. Introduction

Kidney is one of the important organs of the body. Chronic kidney prediction has significant research scope in healthcare industry [1]. A chronic disease needs continuous medical support and treatments throughout their life. People cannot survive without kidneys. It is highly demandable to have a robust prediction system with high accuracy. The current internet of things (IoT) world provides the feasibility of many medical records collected from patient and stored in the cloud [2]. Big Data Analytics plays a major role in making predictive analysis and systematicdecisions. Duringtreatments procedures various patient related decisions are made systematically. The computerized patient information based on physiological data, health records whichare helpful for analyzing various chronic diseases. Big Data Analytics is highly demandable technology for sorting out these data into a variety of attributes. Prediction of chronic kidney disease is highly important in current scenario to safeguard the people from lifelong threatening health problems [3]. The problems such as joint pain, cold, ortho, diabetics are some of the frequently occurring problems come under the category of chronic disease. Persistence of medical sickness leads to long lasting chronic disease [4]. People often leave the basic problems without having any diagnostic treatment.

[1]*Department of Computer Applications, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India*
[2]*Department of Information Technology, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India*
[3]*Department of Computer Applications, MEASI Institute of Information Technology, Chennai, India*
[4]*Department of Information Technology, New Prince Shri Bhavani College of Engineering and Technology, Anna University, Chennai, India*
Corresponding author E-mail id: prakash@crescent.education

Predictive data analysis [5]is considered as an cost effective way of detecting the future outcomes based on the present data collected from the healthcare department. Algorithms such as EDA (exploratory data analysis) [6]Feature selection is important to reduce the number of features in a dataset to prevent over fitting and improve accuracy. [7] are helpful for supporting the decision making on healthcare devices. Continuous monitoring and treatment validation of patient are required in healthcare departments to make the predictive analysis of future outcomes. HealthCare Centers often focus on creating more optimized applications to ensure the patient health and upcoming problems immediately also these intelligent systems helpful for the patient to provide immediate medical support [7][8]. A significant amount of data is used to identify successful analysis, treatment, and decision-making. To avoid death, people must receive good treatment and heal [9][10].

The existing knowledge base in the field of medicine is embedded with a huge repository of information exploited to make important decisions about diagnosis and treatment regimens.Machine learning approaches have recently gained prominence in the medical field because of their high predicted accuracy for chronic disease and low computational effort and cost. Early prediction of a chronic kidney disease is possible with the use of machine learning algorithms, which lowers the risk of permanent repercussions that might otherwise arise owing to a late or inaccurate diagnosis.[11][12]. A number of different machine learning algorithms were presented [13][14]. A comprehensive strategy utilizing association rule mining with classification algorithms can greatly increase classification accuracy and be more beneficial for CKD prediction. Likewise, this study has several theoretical and practical ramifications for the healthcare and medical industries [15][16].

- The proposed approach is focused on creating a robust methodology to detect the chronic kidney disease based on recorded attributes of patient data collected from various Healthcare centers, as UCI CHKD dataset is considered here.
- The dataset is preprocessed in which removal of junk data, and values that are not relevant to the processing type are removed completely. These data are large accumulation of recorded patient information such as clinical records.
- Based on relativity extraction between each physiological data, such as blood glucose, blood pressure, habitual data and previous health records, the features are mapped. Each feature maps have specific covariate points.
- The proposed Probability reweighted adaboost (PRAB) algorithm is utilized here. The proposed method is compared with Random Forest (RF), Gradient boosted decision tree (GBDT), J48 algorithm (J48A), XGBoost model (XGB) respectively in terms of accuracy.

2. Related Works

Chronic kidney disease is one of the frequently occurring problems with Diabetics and without Diabetics patients. Utilizing enormous information available on cloud platforms to anticipate the chance of infection in future is highly inevitable. Numerous calculations and statistical measures are required to focus on detection of chronic kidney disease. The proposed approach considers various existing systems and their existing methodology to formulate the proposed idea.

Ogunleye and coauthors [17] provided a survey that established a soft computing technique for diabetic prediction. The framework considered physiological date of gathered from the patients which are examined utilizing brain organizations to make a forecast of presence of Diabetics in the beginning phases. Using XGBoost model the system achieved the static accuracy of 100% with pre-extracted features. The early detection of disease can save many lives.

Derevitskii and coauthors [18] The creator present a search model based diabetic forecast system in which different dataset associated from the patients are dissected as far as the physiological boundaries, for example, blood glucose treatment under taken and different other medical problems are thought of Utilizing this information and with the assistance of information mining procedure the yearly forecast of Diabetics is dissected. The system compares 9 different machines learning technique and concluded with XGboost model that performs with better accuracy.

Ilyas and coauthors [19]discuss the presented system explored about the early prediction strategies with unexpected consequences towards chronic kidney disease. The specified process considers random forest algorithm with J48 algorithm to make clear detection of disease with various health record considered. The comparative analysis offers J48 algorithm on chronic kidney disease detection achieved 85.5% accuracy.

Chaudhuri and coauthors [20]discuss the system presented here provides a feature elimination-based comparison study of chronic kidney disease and no-chronic kidney disease detection system. The presented systems consistently analyze the disease parameters impacted through enhanced decision tree (EDT) model. The systems have high consistency in continuous learning process on chronic kidney disease parameters.

Z. Ye and coauthors [21] discussed chronic kidney disease detection as well as coronary artery disease detection method using machine learning. The dataset is collected from high mortality critical care patients. The comparative analysis of various machine learning techniques is utilized in which gradient boosted decision tree achieved highest predictive AUC (Area under curve) around 94.6%. The data is collected from MIMIC-IV publicly available online access data with 50000 records collected from ICU patients.

Hassan, M.M. and coauthors [22] explored the comparative study of various impacts in chronic kidney disease detection with the help of patient clinical records. Intelligent analysis of impacted parameters in chronic kidney disease is discussed with comparison of support vector machine (SVM), random tree forest (RT), Bagging tree model (BT), Neural network (NN) etc. feature selection is modeled using XGBoost model. The dataset utilized for analysis is collected from UCI repository.

3. Methodology

The proposed system is focused on analyzing the chronic kidney disease related information. The goal of the system is to provide a robust procedure to detect the chronic kidney disease using recorded historical data of patients and deep analysis through probability reweighted AdaBoost algorithm. The processing stages consider the raw dataset and remove the unwanted noise present in the dataset. Presence of junk value, not a number (NAN) value and repeated duplicate datasets are removed. Figure 1 Shows the System architecture of proposed method using Probability reweighted adaptive boosting algorithm.

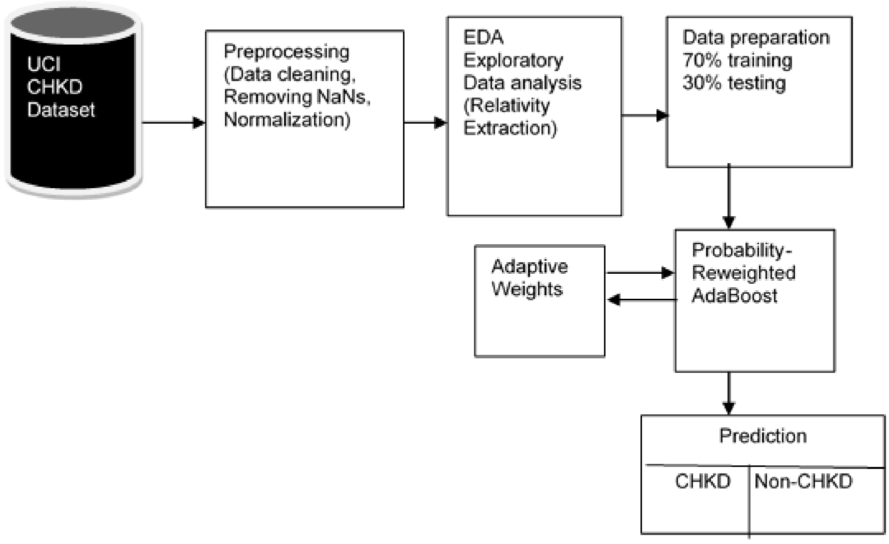

Fig 1. System architecture of proposed method

Exploratory data analysis:The exploratory data is required in the initial investigations on non-linear data. It is required to discover the pattern of the dataset to identify the relevant parameters that repeatedly occur at every analysis. Exploratory data analysis is focused on making graphical representations of the input data provided with summary of statistical measures. It is required to structure the non-linear data into a normalized data for prediction process. The proposed system is developed using Python Tool incorporated with Google Collaborator for analyzing the non-linear input with various statistical scores. Google Collab is integrated with Python libraries; hence it acts as an open-source tool for running the Python code in online. The proposed design utilized some of the libraries in python such as NumPy, scikit-learn, SciPy,matplotlib. These libraries are helpful to access the basic system functionality such as Input outputs access and other core modules.

Proposed PRAB algorithm pseudo code:

Input: features of dataset
Output: class value {CKD, Non-CKD}
Initialize w[i] = 1/N,
N= Training Samples
Ta= Number of classifiers
Loop: For t = 1 to Ta:
Train classifier using w[i]
H_t(x) = {1,-1}
Err_rate=∑w[i] when h_t(x[i])! = y[i]
Else Err_rate=0
Weight_of_Classifier=log((1 - error_rate) / error_rate)
- w[i] = w[i] * exp(alpha_t)
Normalize the weights
- w[i] = w[i] / sum(w[i] for i in range(N))
End Loop;
To predict new data, initialize y_hat=0;
-y_hat+=alpha_t*h_t(x)
Return absolute (-y_hat)

The proposed PRAB algorithm pseudo code is explored above. The inputs are features extracted from the dataset. These features have the unique statistical ranges to determine the presence of CKD class, non-CKD class etc. each class have the unique statistical pattern as evaluating the mean values. The weights are initially declared as zero. As each pattern updated in every varying sample of the dataset, the weights are getting updated and normalized. PRAB model considers the logistic regression and binary regression as base classifier. The evaluation of weak classifier is adaptively selected through the PRAB model through AdaBoost classifier.

The weight update and normalization of values are continuing until all the samples in the dataset considered for validation. The final absolute value compared with the normal range equivalent to 1, and abnormal range assigned to -1 determines the class.

4. Material and Method

AdaBoost techniques have the advantage of incorporating various week classifiers into strong classifier. Adaptive boosting algorithms revote the existing classifier at each iteration and widen the analysis system to improve the detection accuracy. The proposed approach is focused on creating probability reweighted adaboost algorithm in which the performance is compared with the standard boost algorithm. The probabilities of the weight are evaluated every iterationthe adjusted iterations are added up to the standard weights and for these rates; it can change the complete analysis of the proposed values. Probability reweighted Adaboost incorporate the problem of overhead issue in normal AdaBoost algorithm. The proposed paper considersprobability-based strategy to get improved classification accuracy and further selecting the strong classifier using the time taken to make the complete analysis is being reduced. The proposed technique adaptively changes the weight of the boosting algorithm based on positive classification as well as negative classification. The weight updating of the probability booster adaptive algorithm based on false positive and false negative rate is discussed below.

5. Results and Discussions

Dataset Selection: The chronic kidney disease dataset was collected from the UCI website. The dataset collected from Apollo health Service India in 2015 assumed control with a time limit of 2 months. It consists of 400 participants of patients with the knowledge record of 250 having chronic kidney disease patients, and under the record of 400, almost 150 patients are without chronic kidney disease. The basic factors recognized to detect are available in the blood glucose level, albium level, sugar and blood pressure are helpful to determine the chronic kidney disease. Chronic kidney disease is based on various factors such as lifestyle habits, diabetics, accidental kidney damage etc.

CKD analysis:

Blood glucose is highly important to differentiate the chronic kidney patient and normal patient. Both the patient has relatively equal number of sicknesses in blood level red platelet and hemoglobin has high connections. From the given dataset the chronic kidney disease are impacted by potassium level of 39 and 49 should be taken care. The sodium level which is 14.5 is also considered. The missing values in the dataset are round off to 20. These data have many junk values that are not relevant to the present analysis. The missing values are analyzed and normalized using K-nearest neighbor technique. In data mining techniques, the data reduction is the process of scaling the dimension of the date before making into analysis. To make the connection between the data and calculate the relativity score, these data need to be cleaned. Data transformation is used to switch the data from the given format into zero score normalization in which the converted data is used for the analysis. In the proposed approach, the existing frameworks such as Decision tree, Naïve Bayes, Support vector Machines, AdaBoost are compared with the proposed method.

Figure 2 shows the obtained result for the count of each class distributed labels. Figure 3 shows how the Explorer data analysis displays parameters and principal attributes. Figure 4 It is shown that the proposed algorithm outperforms other algorithms. Figure 5 represents the comparison of existing and proposed precision scores. The proposed model achieves a high value. Figure 6 shows the performance of classification algorithms on the dataset based on recall. Figure 7: In comparison to other approaches, the system's F1 Score is relatively high due to its better sensitivity.

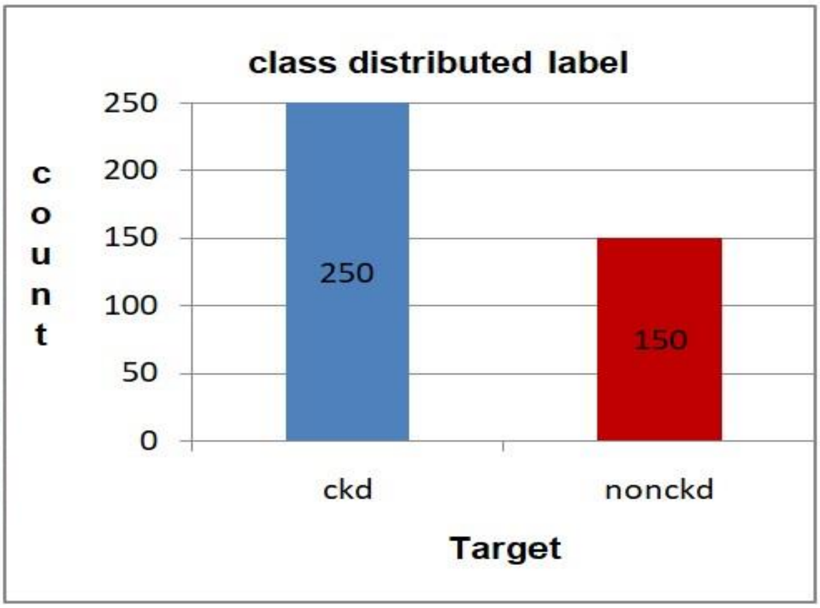

Fig2. Class Distributed Labels

Figure 2shows the class's distributed labels. 400 samples of data are collected from the patients, of which 150 are non-CKD and normal people, and 250 contain CKD data.

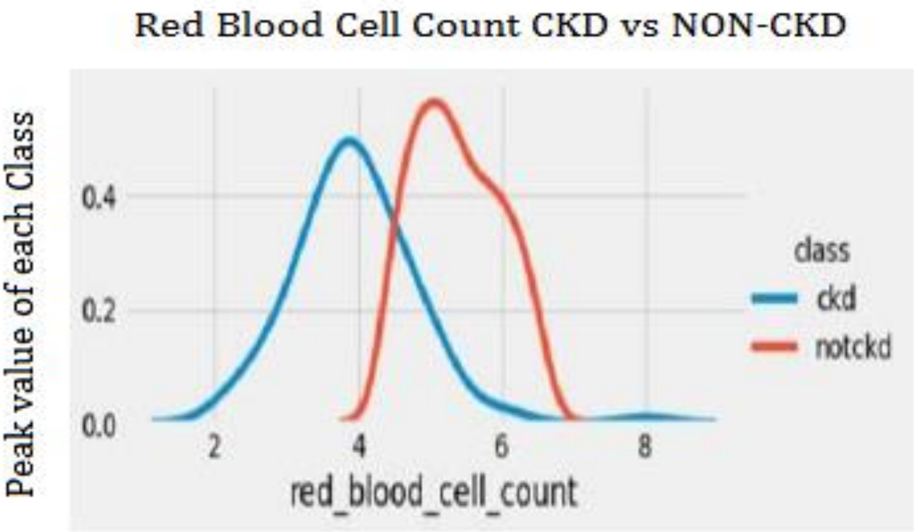

Fig 3. Red Blood Cell Count CKD vs. NON-CKD

Figure 3 Shows the Red Blood cell count of CKD vs. Non-CKD as per the exploratory data analysis report. The proposed approach considers Red Blood cell count as an important physiological parameter to make the classification process. The representation shows the Explorer data analysis of collected dataset in which the various parameters and principal attributes are shown.Infected individuals having under 5 red platelets and 400 stuffed cell volumes. Red blood cell versus volume also important to analyses the difference. Red platelet and hemoglobin have high connections chronic kidney disease infected individual are having under 5 platelets provided with 12.5 hemoglobin level.

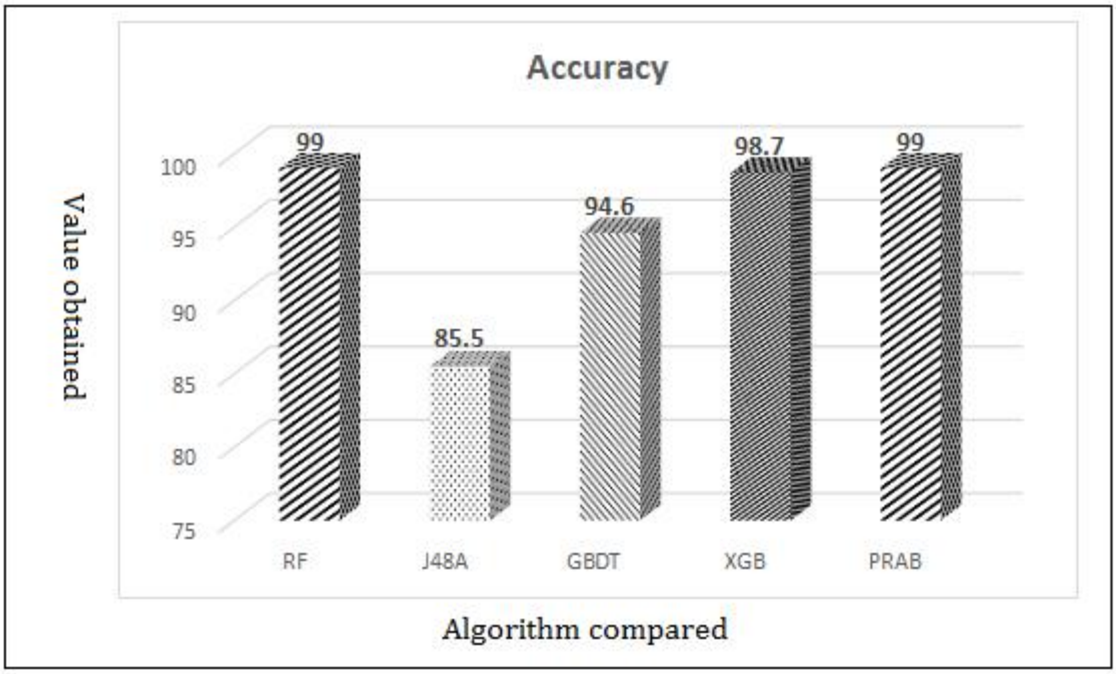

Fig 4. Comparison of Accuracy with Existing and proposed Models

Figure 4 shows the comparative graph showing the accuracy of existing and proposed model. The proposed approach PRAB model achieved the accuracy of 99%, where random forest (RF) 99%, J48A with 85.5%, gradient boosted decision tree (GBDT) with 94.6%, and XGBoost with 98.7% is achieved respectively.

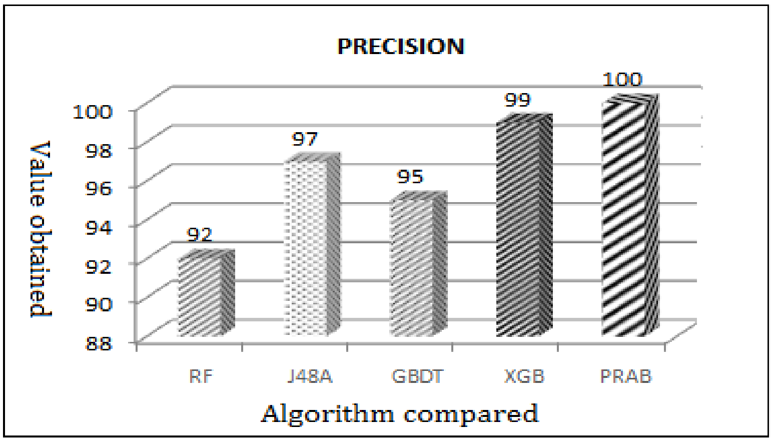

Fig 5. Comparison of Precision of Existing and proposed Models

Figure 5shows the comparison of existing and proposed precision score achieved. Further achieving 100% precision is rare in consideration of various cases. The proposed model achieved such a high value here

.

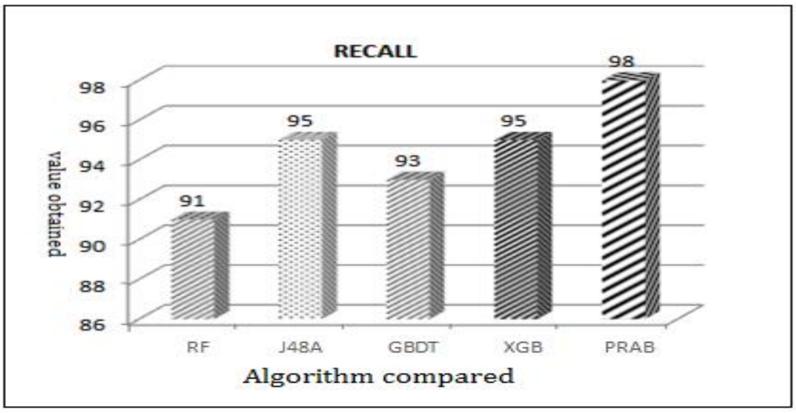

Fig 6. Comparison of Recall value of existing and proposed model

Figure 6 shows the Recall value of the existing and proposed methodology, in which the Recall is used to analyze the prediction capability of the proposed approach.

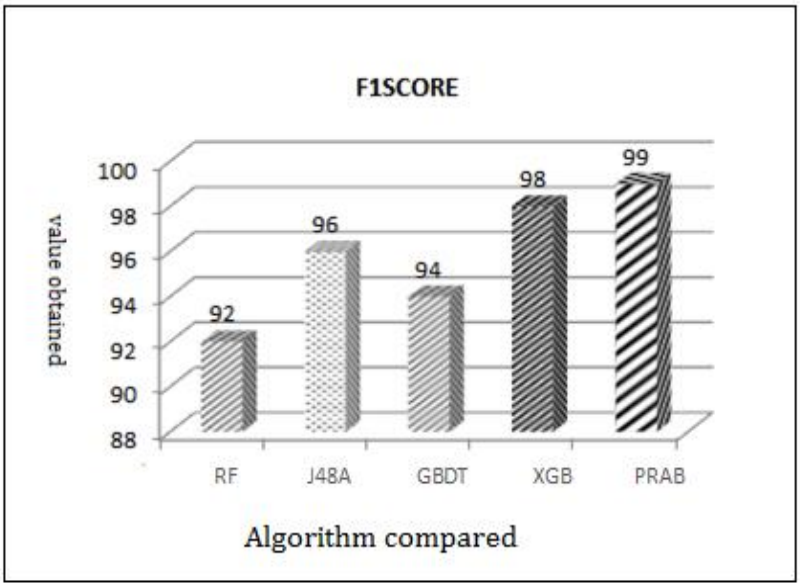

Fig 7. F1Score comparison

Figure 7 shows the comparison of F1Score achieved with existing and proposed model. The system achieved with higher sensitivity hence the F1Score is relatively high comparing other methods.

Table1. Comparison of existing and proposed method

References	Method	Accuracy	Precision	Recall	F1score	Dataset
Rashed Al-Mahfuz M et al.,[23]	Random forest model(RF)	99%	92%	91%	92%	'DB-I', 'DB-II', 'DB-III', 'DB-IV', 'DB-V', and 'DB-IV'
Ilyas et al.,[24]	J48 algorithm (J48A)	85.50%	97%	95%	96%	Clinical data
Z. Ye et al.,[25]	Gradient boosted decision tree (GBDT)	94.60%	95%	93%	94%	MIMIC-IV
Ogunleye et al., [26]	XGBoost model (XGB)	98.70%	99%	95%	98%	UCI
Proposed system	Probability reweighted adaboost(PRAB)	99%	100%	98%	99%	CKD - UCI

Table 1 shows the comparison of existing state of art approaches, with proposed model. One of the major challenges in the proposed work is the non-linear data handling and time taken to normalize the unbalanced data. Further the system can be improved by adopting deep exploratory data analysis and considering automated dimensionality reduction techniques such as principal component analysis and Self organized mapping models. Further the existing challenges in uploading and handling the health care [23] data in cloud are discussed in detail, the big data stream [24] almost the complex process and in future enhancements need to be developed with lightweight architecture that perform the disease interpretations [25] Secure transmission of medical data ensures timely access to quality care. [26], analysis and prediction done with the cloud environment.

6. Conclusion

Chronic kidney disease is one of the highly impacted diseases that affect the people life in a resilient way. Leaving the chronic kidney diseasewithout any treatment leads to life threatening problems the proposed research work is focus on analyzing the chronic kidney diseasewith the help of various physiological parameters and using machine learning algorithms. In terms of analyzing the chronic kidney disease using the dataset connected from UCI website the results are evaluated here. Based on blood glucose level and the relative parameters incorporated by mining the relativity extraction analysis the proposed approach considers various attributes from the UCIdataset. Between these dataset analyses, considering various clinical records, Diagnostic procedures, changes that help the medical practitioners and patients to survive in the early stages. The proposed PRABmodel achieved the accuracy of 99%. Further the system can be improved by formulating Ensemble algorithms and techniques with multiple machine learning models.

References

[1] Jackins V, Vimal S, Kaliappan M, Lee MY. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. The Journal of Supercomputing. 2021 May; 77:5198-219.

[2] Aswathy RH, Suresh P, Sikkandar MY, Abdel-Khalek S, Alhumyani H, Saeed RA, Mansour RF. Optimized tuned deep learning model for chronic kidney disease classification. Comput. Mater. Contin. 2022 Jan 1; 70:2097-111.

[3] Bhaskar N, Suchetha M, Philip NY. Time series classification-based correlational neural network with bidirectional LSTM for automated detection of kidney disease. IEEE Sensors Journal. 2020 Oct 5; 21(4):48118.

[4] Elkholy SM, Rezk A, Saleh AA. Early prediction of chronic kidney disease using deep belief network. IEEE Access. 2021 Sep 21; 9:135542-9.

[5] Hosseinzadeh M, Koohpayehzadeh J, Bali AO, Asghari P, Souri A, Mazaherinezhad A, Bohlouli M, Rawassizadeh R. A diagnostic prediction model for chronic kidney disease in internet of things platform. Multimedia Tools and Applications. 2021 May; 80:16933-50.

[6] Ahmed I, Ahmad M, Jeon G, Piccialli F. A framework for pandemic prediction using big data

analytics. Big Data Research. 2021 Jul 15; 25:100190.
[7] Janani J, Sathyaraj R. Diagnosing Chronic Kidney Disease Using Hybrid Machine Learning Techniques. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 2021 Jul 11; 12(13):6383-90.
[8] Rajarajeswari S, Tamilarasi T. Chronic kidney disease (CKD) prediction using supervised data mining techniques. International Journal of Advanced Networking and Applications. 2021 May 1; 12(6):4776-80.
[9] Shankar S, Verma S, Elavarthy S, Kiran T, Ghuli P. Analysis and prediction of chronic kidney disease. Int. Res. J. Eng. Technol. 2020 May; 7(5):4536-41.
[10] Gupta M, Gupta P. Predicting chronic kidney disease using machine learning. Emerging Technologies for Healthcare: Internet of Things and Deep Learning Models. 2021 Aug 30:251-77.
[11] Balakrishnan S. Feature selection using improved teaching learning based algorithm on chronic kidney disease dataset. Procedia Computer Science. 2020 Jan 1; 171:1660-9.
[12] Revathy S, Bharathi B, Jeyanthi P, Ramesh M. Chronic kidney disease prediction using machine learning models. International Journal of Engineering and Advanced Technology. 2019 Oct; 9(1):6364-7.
[13] Almustafa KM. Prediction of chronic kidney disease using different classification algorithms. Informatics in Medicine Unlocked. 2021 Jan 1; 24:100631.
[14] Debal DA, Sitote TM. Chronic kidney disease prediction using machine learning techniques. Journal of Big Data. 2022 Dec; 9(1):1-9.
[15] Alaiad A, Najadat H, Mohsen B, Balhaf K. Classification and association rule mining technique for predicting chronic kidney disease. Journal of Information & Knowledge Management. 2020 Mar 11; 19(01):2040015.
[16] Pinto A, Ferreira D, Neto C, Abelha A, Machado J. Data mining to predict early stage chronic kidney disease. Procedia Computer Science. 2020 Jan 1; 177:562-7.
[17] Ogunleye A, Wang QG. XGBoost model for chronic kidney disease diagnosis. IEEE/ACM transactions on computational biology and bioinformatics. 2019 Apr 17; 17(6):2131-40.
[18] Derevitskii IV, Kovalchuk SV. Machine Learning-Based Predictive Modeling of Complications of Chronic Diabetes. Procedia Computer Science. 2020 Jan 1; 178:274-83.
[19] Ilyas H, Ali S, Ponum M, Hasan O, Mahmood MT, Iftikhar M, Malik MH. Chronic kidney disease diagnosis using decision tree algorithms. BMC nephrology. 2021 Dec; 22(1):1-1.
[20] Chaudhuri AK, Sinha D, Banerjee DK, Das A. A novel enhanced decision tree model for detecting chronic kidney disease. Network Modeling Analysis in Health Informatics and Bioinformatics. 2021 Dec; 10:1-22.
[21] Ye Z, An S, Gao Y, Xie E, Zhao X, Guo Z, Li Y, Shen N, Ren J, Zheng J. The prediction of in-hospital mortality in chronic kidney disease patients with coronary artery disease using machine learning models. European Journal of Medical Research. 2023 Dec; 28(1):1-3.
[22] Hassan MM, Hassan MM, Mollick S, Khan MA, Yasmin F, Bairagi AK, Raihan M, Arif SA, Rahman A. A Comparative Study, Prediction and Development of Chronic Kidney Disease Using Machine Learning on Patients Clinical Records. Human-Centric Intelligent Systems. 2023 Feb 22:1-3.
[23] Rashed-Al-Mahfuz M, Haque A, Azad A, Alyami SA, Quinn JM, Moni MA. Clinically applicable machine learning approaches to identify attributes of Chronic Kidney Disease (CKD) for use in low-cost diagnostic screening. IEEE Journal of Translational Engineering in Health and Medicine. 2021 Apr 15; 9:1-1.
[24] Yang Y, Li Y, Chen R, Zheng J, Cai Y, Fortino G. Risk prediction of renal failure for chronic disease population based on electronic health record big data. Big Data Research. 2021 Jul 15; 25:100234.
[25] Sarosh P, Parah SA, Bhat GM, Muhammad K. A security management framework for big data in smart healthcare. Big Data Research. 2021 Jul 15; 25:100225.
[26] Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. Kidney International Supplements. 2022 Apr 1; 12(1):7-11.
[27] Kothandaraman, D., Praveena, N., Varadarajkumar, K., Madhav Rao, B., Dhabliya, D., Satla, S., Abera, W. Intelligent Forecasting of Air Quality and Pollution Prediction Using Machine Learning (2022) Adsorption Science and Technology, 2022, art. no. 5086622,
[28] Mehraj, H., Jayadevappa, D., Haleem, S.L.A., Parveen, R., Madduri, A., Ayyagari, M.R., Dhabliya, D. Protection motivation theory using multi-factor authentication for providing security over social networking sites (2021) Pattern Recognition Letters, 152, pp. 218-224.