Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

Classifying the status of CKD using Randomized Weighted Optimization Model

Ms. A. Vijayalakshmi,
Research Scholar, Computer Science, Vels Institute of Science, Technology & Advance studies, Chennai,India.
E-mail: nitmanviji@gmail.com

Dr. V. Sumalatha,
Research Supervisor, Computer Science, Vels Institute of Science, Technology & Advance studies, Chennai, India.
E-mail: sumalatha.scs@velsuniv.ac.in

Abstract

chronic kidney disease (CKD) is a substantial health-care burden owing to its growing occurrence, enhanced danger of development to end-stage renal disease, and reduced mortality and morbidity prognosis. It is quickly accelerating to a worldwide well-being catastrophe. The main causes of CKD include high blood pressure and diabetes. The major function of the kidney is filtering waste from a human body. When kidneys fail, waste builds up in bodies and eventually leads to death. Researchers all around the globe utilize the Glomerular Filtration Rate (GFR) and kidney damage indicators to define CKD as a disorder that causes decreasing renal function over a period of time. A person having CKD is more likely to die early. Doctors have a tough time recognizing the several disorders associated with CKD early enough to avoid the condition. Data mining approaches have recently been exposed to considerable research in CKD diagnosis, with a focus on accuracy, either through the simplicity of illness by conducting feature selection in addition to pre-processing or not before classification. This research work determines creatinine level is a kind of blood metabolite that has a significant relationship with GFR and Blood Urea Nitrogen (BUN) BUN Creatinine Ratio (BCR). Since measuring GFR and BCR is challenging, this work focused in analysing creatinine value is used to determine Estimated GFR (EGFR) and BCR from serum creatinine indirectly utilizing the attributes available in the dataset whereas the 28 attributes are considered inclusive of "Gender" with 523 records. The EGFR and BCR computing results are used in this prediction analysis for providing an accurately categorizing the status of CKD which has been utilized by Randomized Weighted Optimization Model (RWOM). Furthermore, the proposed RWOM with Neural Network (NN) is compared to the RWOM with Logistic Regression (LR), NN, LR in terms of analysing the better categorization status of CKD from patient records.

Keywords: CKD, eGFR, RWOM, BUN, creatinine, non-CKD, classification.

1. Introduction

Because of its high mortality rate, CKD has gotten a lot of attention. As per World Health Organization (WHO), chronic illnesses have emerged as a chief threat to developing nations [1]. CKD can be termed as an illness in kidney that could be diagnosed in its primary phases; nevertheless, it ultimately causes renal failure. In 2016, 753 million people died from CKD worldwide, with men accounting for 336 million fatalities and females accounting for 417 million [2]. The kidney illness is named "chronic" because it develops slowly and persists a long period, impairing urine system's function. The build-up of waste materials in the circulation causes numerous health issues, including diabetes, high and low Blood Pressure (BP), bone difficulties and nerve damage, all of those contribute to cardiovascular disease. Diabetes, high BP and cardiovascular disease (CVD) are all risk issues for patients suffering from CKD [3]. Patients with CKD experience adverse effects, particularly in the late stages that harm the neurological and immunological systems. Patients in impoverished nations may approach the end of their lives and require dialysis or kidney transplants. GFR, which describes renal function, is used by doctors to diagnose kidney disease. By means of using patient data such as the blood test results, gender, age and various additional parameters, GFR is computed [4]. Doctors can divide CKD into five phases based on the GFR number. Table 1 depicts the progression of renal disease as measured by GFR.

TABLE.1 GROWTH OF CKD STAGES

Stage	Description	GFR (mL/min/ 1.73 m2)	Treatment stage
1	Functioning of Kidney is regular	≥90	Control of BP based on the observation
2	Minor damage of Kidney	60–89	Control of BP and various hazard issues based on the observation
3	Reasonable damage to Kidney	30–59	Control of BP and various hazard issues based on the observation
4	Severe damage of Kidney	15–29	Renal failure comes to end stage
5	Recognized failure of kidney	≤ 15	Treatment choices

The preliminary discovery and CKD treatment can help in averting renal failure. The finest method to diagnose CKD is to recognize it at the preliminary stages itself; however, waiting until it is advanced may result in renal failure, requiring kidney transplantation or dialysis to live a normal life. A blood test used for examining glomerular filtrate or a urine test to assess albumin are utilized in medical diagnosis of CKD. Due to growing quantity of chronic renal patients, the paucity of specialized doctors and the expensive costs of treatment and diagnosis, particularly in emerging nations, computer-assisted diagnostics are needed for aiding radiologists and physicians in making diagnostic judgments. CKD symptoms aren't illness specific. Symptoms appear gradually, and some people may have none at all. As a result, early detection of the disease becomes extremely difficult.

The role of Machine Learning (ML) in the recent times has been very vital in illness diagnosis by simply examining existing patient records and building a model to predict future patient behaviour. ML is an Artificial Intelligence area in which the computer system learns on its own and improves its predictions as a result of training. Supervised learning is a type of ML that can predict or categorize data sets. ML has been utilized quite well in a variety of disciplines, particularly in the biomedical industry for illness identification and classification. To anticipate sicknesses, many ML methods may be applied, each with its own set of advantages and downsides. Decision-tree delivers better precise kidney-related illness classification reports amongst all these. As a result, it appears to be a good candidate for use in developing a forecast scheme for diagnosing kidney illnesses at a preliminary stage. CKD is acknowledged as a major public health issue. Every year, several people expire as a result of insufficient healthcare, deficiency of health education and the exorbitant expense of treating CKD. As per worldwide information regarding kidney illnesses, CKD affects 13.4% of the global population. Many research has been carried out to prediction of CKD using various classification algorithms, with the predicted results of their suggested model being generated. This research work describes an calculating eGFR and BCR to identify CKD status, as well as categorization of CKD using various data mining approaches. The CKD-EPI equation is based on a combined data set from 10 studies of individuals with and without CKD, each of which evaluated GFR using blood creatinine levels and urine iothalamate clearance in line with international reference standards. BUN and creatinine levels are both affected by kidney function, but they are linked to provide information on the location of damage, such as whether it is an intrinsic renal or prerenal disease. The EGFR and BUN-CR tests are used to determine the precise Sc concentration in patients. As a result, the focus of this study is on obtaining EGFR and BCR values to accurately categorize CKD status using a RWOM technique that is not computationally intensive for quick and accurate diagnosis.

The breakdown of the paper's structure is as follows: Section 2 discusses the literature support for classifying and detecting CKD using various ML approaches, and Section 3 discusses the suggested RWOM methodology for improving CKD classification, Section 4 compares RWOM with NN and RWOM with LR in terms of performance measures, and Section 5 concludes that RWOM with NN has superior accuracy for CKD status classification than RWOM with LR.

2. Literature Review

Different researchers have conducted experiments on CKD and obtained the accuracy of several ML algorithms using various tools.

Baidya et.al [5] presents a method by using eight different ML algorithms that can promptly detect the infection of CKD considering the health condition dataset information of the patient. A dataset is used of nearly two months of that period delivered by the hospital to identify the plausibility of CKD. According to the analysis results, K-Nearest Neighbours (KNN) and Extra Tree Classifier have performed better than other algorithms for achieving an accuracy of 99% preceding the Gradient Boost, which stands at 98%. Sathyaraj et al. [6] proposed a hybrid model that can perform better to diagnose CKD by collecting data from the UCI ML repository, and in order to get a clean data set, they followed pre-processing steps such as label encoding, as well as min-max normalizing. Eventually, they found that the AdaBoost and Random Forest (RF) algorithms maintained a better accuracy rate. Ifraz et al. [7] propose a technique for forecasting CKD status from clinical information that comprises data pre-processing, data aggregation, and feature extraction. The LR and decision tree technique, according to the study's findings, may be used to better correctly forecast chronic renal disease. As per the study, their precision and their accuracy are 96.25 percent and 97 percent respectively.

Poonia et al. [8] have created a fantastic feature-based prediction algorithm for identifying renal illness. A LR classifier model attains the accuracy of 98.75 chosen using the Chi-Square approach. Vasquez et al. [9] created a NN classifier based on huge amounts of CKD data, and the model predicted the development of diabetic kidney disease 95 percent of the time. Makino et al. [10] used textual data to collect patient treatment information in order to analyse the course of diabetic kidney disease. The XGBoost approach, developed by Z. Segal and colleagues [11], employs an ensemble tree-based ML algorithm to detect kidney disorder in its initial stages. The comparison models such as regression, CatBoost and RF are used. The proposed model improves all metrics with accuracy of 0.93, specificity of 0.958 and sensitivity of 0.715 percent. Khamparia et al. [12] presents an early detection of CKD using a stacked autoencoder model to extract features from hypermedia data. This work used a SoftMax classifier to anticipate the last class, which they originate to be exact.

Polat et al. [13] utilized the SVM approach to forecast renal illness. Their accuracy rate is 97.5 percent. They used a range of feature selection approaches to improve the accuracy. Using feature selection, they are able to enhance accuracy by 1%. KNN, Naïve Bayes (NB), and decision tree classifiers are proposed by Panwong et al. [14]. They also used the wrapper approach to decrease the number of characteristics. They achieved an accuracy of 85 percent using the decision tree approach. Dulhare et al. [15] proposed utilizing the NB classifier in conjunction with the R trait selector to diagnose renal disease. They correctly diagnosed kidney disease 97.5 percent of the time. The work by Senan et al. [16] supports specialists in investigating CKD prevention approaches through early identification utilizing ML techniques. The data for this study came from 400 patients and included 24 characteristics. The misplaced numerical and nominal data is replaced using the mean and mode statistical examination methods. The RF technique outstripped all other models obtaining 100 percent of accuracy.

Kim et al. [17] recommended a NN-based GA, in which the GA improved the weight vectors used to train the NN. For CKD diagnosis, the system outperforms standard NN. Ilyas et al. [18] provide a study in which ML classification algorithms is applied to a dataset built from CKD patients' medical records to predict the various stages of the disease. The RF and J48 algorithms are used to develop a model that is both sustainable and practical for diagnosing numerous phases of CKD with high medical precision. As per the comparison of the data, J48 predicted CKD better than RF at all phases, with an accuracy of 85.5 percent. Ren et al. [19] used an Electronic

Health Records (EHR) data set to construct a predictive model for CKD detection. This suggested approach uses NN architecture to encode and decode textual and numerical data from EHR. To identify chronic renal illness, Ma F. et al. [20] design a deep NN model. In comparison to ANN and SVM, the given model has the greatest accuracy.

3. Research Methodology

The study is done with the help of the CKD dataset [21]. This dataset has 523 rows and 27 columns. The value of "class" in the output column is either "1" or "0." The patient has no CKD, as shown by the value "0.", whereas the value "1" specifies that the patient has. Before pre-processing, the total quantity of CKD and non-CKD items in the output column is shown in Figure 1. There are 380 CKD data points in total, with 143 non-CKD data points.

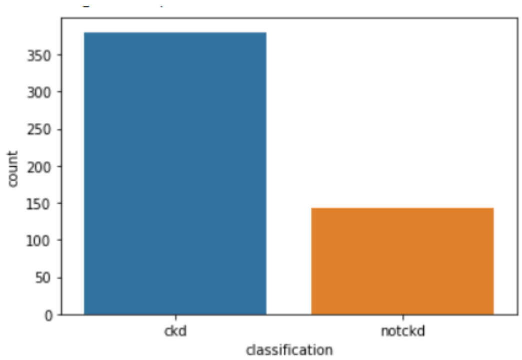
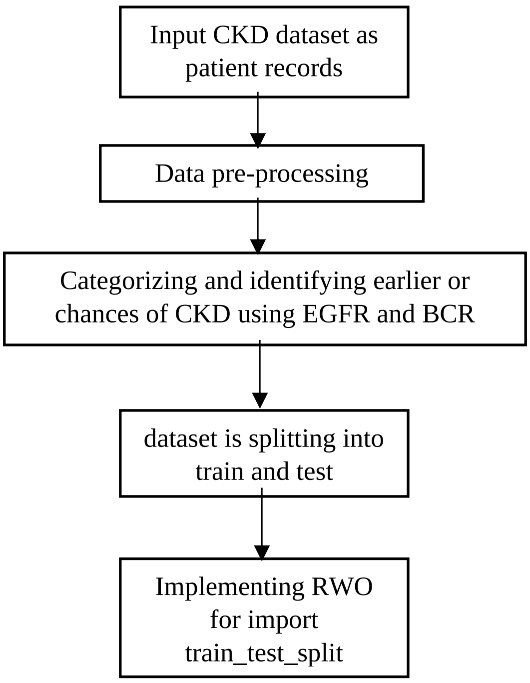


Figure 1 Over-all number of CKD and non-CKD data

Data pre-processing is necessary before constructing a model to eliminate undesired noise and outliers from the dataset that may cause the model to deviate from the intended training set. This step focuses on removing any obstacles to the model's efficiency. Data must be cleansed and readied for model creation once it has been collected. After that, the dataset is checked for null values. The output values "False" and "0" indicate that there are no null values. Once the data has been obtained and saved, it must be kept hidden, as must missing data imputing. As a result, records with multiple missing data are deleted, and data cleansing entails clearing the records from the analysis of 523 records. To improve task accuracy and efficiency, the data is divided into training and testing segments using an 75% of training and 25% of testing. Following division, the model is trained using a variety of classification approaches. In this work, the RWOM classification methods are used. Figure 2 depicts the workflow of the proposed RWOM architecture, which demonstrates the purpose and advantage of determining the ideal weights for ML models like NN and LR models to identify the greater accuracy in diagnosing the stage of CKD using EGFR and BCR.



Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

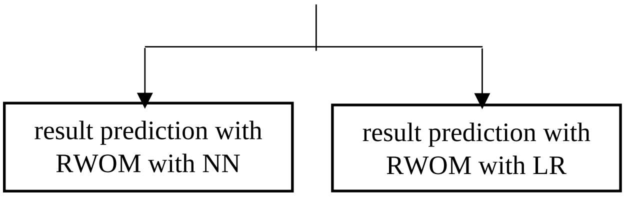


Figure.2 Proposed flow diagram of RWOM

3.1 Categorise the risk of CKD using EGFR and BUN Creatinine Ratio (BUN-CR)

EGFR: A patient's normal GFR may not necessarily correlate to the normal serum creatinine reference interval. Based on the clinical laboratories to compute GFR and report the result when serum creatinine is detected in patients 18 and older, if appropriate and possible, because blood creatinine alone can indicate mild to severe renal impairment. An EGFR derived from blood creatinine using an Isotope Dilution Mass Spectrometry (IDMS) noticeable equation is an unpretentious and operative technique for laboratories to assist health care practitioners in detecting CKD in patients with hazard factors such as diabetes, hypertension, cardiovascular disease, or a family history of kidney disease. Once albuminuria has been discovered, it is crucial to use EGFR to evaluate kidney function. Providers may also utilize EGFR to follow patients who have already been diagnosed with CKD. CKD-EPI equations consider gender, age, and race, which may help clinicians diagnose CKD even when blood creatinine levels appear to be within or just over the usual reference interval. The recommendations computing the GFR equation for adults and identifying whether or not individuals have CKD.

The CKD-EPI equation is used to describe the association between GFR and serum creatinine using a 2-slope "spline" (Sc). Furthermore, the equation contains numerous factors such as age, gender, and race. As a result, the equation has been converted to Sc in mg/dL in the table below. Thus, equations (1) and (2) with distinct parameters involved in calculating EGFR are shown below.

$$EGFR = 141 \times (Sc/S,1)^{\alpha} \times (0.993)^{Age} \qquad (1)$$

$$EGFR = 144 \times (Sc/S,1)^{\alpha} \times (0.993)^{Age} \qquad (2)$$

Where,
S = Sex kappa factor as 0.7 for female and 0.9 for male
α = Sex alpha factor as -0.329 and -1.209 for female with ≤0.7 and >0.7 respectively. Similarly, -0.411 and -1.209 for male with ≤0.9 and >0.9 respectively.

The influence of creatinine assay imprecision on the uncertainty of an EGFR result, on the other hand, is greater with higher EGFR levels and should be considered when determining the highest EGFR value to report.

BUN-CR: It is useful in identifying whether you have a renal or nutrition issue. However, to further the diagnosis and determine the specific cause of the renal impairment, it employs two serum-measured values: creatinine and blood urea nitrogen. BUN and creatinine levels are both affected by kidney function, but they work together to reveal the site of the damage, whether it is intrinsic renal or prerenal. Equation (3) represents the CR formula.

BUN-CR = BUN (mg/dL) / serum creatinine (mg/dL) ……………………(3)

If the BUN-CR is high, it signifies that BUN-CR >20 indicates a prerenal source, whereas BUN-CR 10 indicates an intrinsic renal cause. Serum creatinine levels should be between 0.7 and 1.3 mg/dL (62 and 115 mol/L). However, the EGFR and BUN-CR are used to determine the precise Sc concentration in the patients, as shown in figure 3.

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

	ID	age	Gender	bp	sg	al	su	rbc	pc	pcc	...	dn	cad	appet	pe	ane	classification	EGFR	BCR	Age_group	EGFR_status
0	1	48	Male	80	1.020	1.0	0.0	normal	normal	notpresent	...	yes	no	good	no	no	ckd	71.077494	30.000000	Adult	No CKD
2	3	62	Male	80	1.010	2.0	3.0	normal	normal	notpresent	...	yes	no	poor	no	yes	ckd	39.457290	29.444444	Adult	High risk
3	4	48	Female	70	1.005	4.0	0.0	normal	abnormal	present	...	no	no	poor	yes	yes	ckd	13.295099	14.736842	Adult	Kidney failure
4	5	51	Female	80	1.010	2.0	0.0	normal	normal	notpresent	...	no	no	good	no	no	ckd	43.534033	18.571429	Adult	High risk
5	6	60	Female	90	1.015	3.0	0.0	normal	normal	notpresent	...	yes	no	good	yes	no	ckd	54.701241	22.727273	Adult	Moderate risk
6	7	68	Male	70	1.010	0.0	0.0	abnormal	normal	notpresent	...	no	no	good	no	no	ckd	1.651094	2.250000	Adult	Kidney failure
7	8	24	Male	80	1.015	2.0	4.0	normal	abnormal	notpresent	...	yes	no	good	yes	no	ckd	93.462311	28.181818	Adult	No CKD
8	9	52	Female	100	1.015	3.0	0.0	normal	abnormal	present	...	yes	no	good	no	yes	ckd	29.883664	31.578947	Adult	Very High risk
9	10	53	Male	90	1.020	2.0	0.0	abnormal	abnormal	present	...	yes	no	poor	no	yes	ckd	7.864905	14.861111	Adult	Kidney failure
10	11	50	Female	60	1.010	2.0	4.0	abnormal	abnormal	present	...	yes	no	good	no	yes	ckd	12.321339	13.750000	Adult	Kidney failure

10 rows × 31 columns

Figure 3 Computation of EGFR and BUN-CR(BCR)

Figure 4 depicts the analysis of precise accuracy of CKD and determining the earlier or likelihood of CKD using EGFR and BCR.

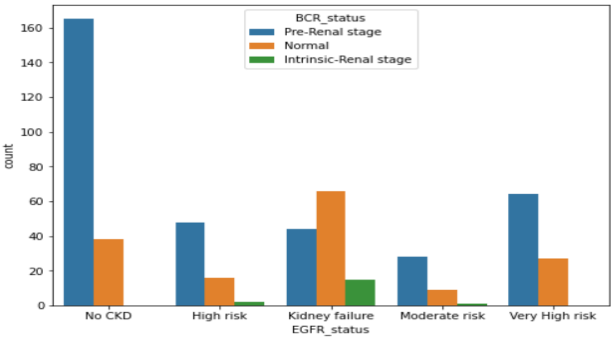


Figure.4 EGFR and BCR status to identify CKD stages

Figure 5 depicts EGFR status studies utilizing age group and BUN Cretinine Ratio for better evaluation of CKD.

	ID	age	Gender	bp	sg	al	su	rbc	pc	pcc	...	cad	appet	pe	ane	classification	EGFR	BCR	Age_group	EGFR_status	BCR_status
0	1	48	1	80	1.020	1.0	0.0	1	1	0	...	0	1	0	0	1	71.077494	30.000000	1	No CKD	Pre-Renal stage
2	3	62	1	80	1.010	2.0	3.0	1	1	0	...	0	0	0	1	1	39.457290	29.444444	1	High risk	Pre-Renal stage
3	4	48	0	70	1.005	4.0	0.0	1	0	1	...	0	0	1	1	1	13.295099	14.736842	1	Kidney failure	Normal
4	5	51	0	80	1.010	2.0	0.0	1	1	0	...	0	1	0	0	1	43.534033	18.571429	1	High risk	Normal
5	6	60	0	90	1.015	3.0	0.0	1	1	0	...	0	1	1	0	1	54.701241	22.727273	1	Moderate risk	Pre-Renal stage
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
605	606	55	1	80	1.020	0.0	0.0	1	1	0	...	0	1	0	0	0	121.995772	98.000000	1	No CKD	Pre-Renal stage
606	607	42	1	70	1.025	0.0	0.0	1	1	0	...	0	1	0	0	0	74.137275	25.833333	1	No CKD	Pre-Renal stage
607	608	12	1	80	1.020	0.0	0.0	1	1	0	...	0	1	0	0	0	153.102965	43.333333	0	No CKD	Pre-Renal stage
608	609	17	1	60	1.025	0.0	0.0	1	1	0	...	0	1	0	0	0	110.162943	50.000000	0	No CKD	Pre-Renal stage
609	610	58	1	80	1.025	0.0	0.0	1	1	0	...	0	1	0	0	0	73.605618	16.363636	1	No CKD	Normal

523 rows × 32 columns

Figure.5 computation of age group for CKD status using EGFR and BCR

3.2 Proposed method of Randomized Weighted Optimization Model (RWOM)

Once the dummies are formed, the dataset is relatively simple to study the variable in an easy manner, and the dataset has been separated into train dataset and test dataset, with the train dataset using 75% of the total dataset as a sample. This suggested RWOM may be used to discover the best weights for ML models like NN and regression models like LR. The process of classifying the parameter values that minimize a pre-specified loss function for a certain training set is known as fitting the model parameters for a variety of different ML models. NN, linear regression, and LR models are examples of such models, and the best model weights for such models are typically discovered via gradient descent. In contrast, fitting the parameters (or weights) of a ML model may be considered as a continuous-state optimization problem in which the loss function replaces the fitness function and the aim is to minimize this function. By framing the problem in this way, the model parameters may be fitted using any of the randomized optimization approaches established for continuous-state optimization problems.

Solving ML Weight Optimization Problems with mlrose: The Python mlrose module allows you to design and apply some of the most often used randomization and search algorithms to a variety of optimization issues. mlrose offers built-in support for solving the weight optimization problem for three types of ML models: (traditional) NN, linear regression models, and LR models. The Neural Network (), Linear Regression (), and Logistic Regression () classes are used for this. A fit method is included in each of these classes, and it implements the three phases for solving an optimization issue for a given training set. That is,

1. Create an object that represents a fitness function.
2. Create a problem object for optimization.
3. Pick a randomized optimization algorithm and run it.

Finding the best model weights is just a means to an end when fitting a ML model. Not because we want to know the ideal weight values, but because we need to identify the optimal model weights so that our study may utilize our fitted model to accurately forecast the labels of future observations. As a result, the aforementioned classes feature a predict method, which utilizes the fitted model to anticipate the labels for a particular test set when called after the fit function.

The following are the steps involved in using mlrose to solve an ML weight optimization problem:

1. Create an object for the ML weight optimization issue.

2. In order to obtain the optimum model weights for a certain training set, use the fit method of the object you constructed in Step 1.
3. Predict the labels for a test set using the predict method of the object generated in Step 1.

To fit the model weights, the user can utilize randomized hill climbing, the genetic technique, simulated annealing or gradient descent. The gradient descent strategy is only available in mlrose for solving the ML weight optimization problem, and it is primarily provided for benchmarking purposes, as it is one of the most often used strategies in fitting NN and LR models.

Algorithm for RWOM:

Step 1: Import the Python's scikit-learn library.

```
Step 2: Set random seed
    if inst1 (rand_state, int) and rand_state > 0:
    sp.rand.seed (rand_state)
Step 3: Set problem, period and efforts counter
       if in_state1 is Nothing:
            {
             prob1.reset()
             }
      Else
             {
             prob1.set_state(in_state1)
             }
Step:4 if bend:
             fitne_bend = []
  efforts = 0
  iters = 0
  optimum_fitne = prob1.get_max () *prob1.get_fitne ()
  optimum_state = prob1.get_state ()
  while (efforts < maximum efforts) and (iters < maximum iters):
```

```
    iters += 1
Step:5 Update weights

    update = flat_weig(prob1.cal_update ()
Step:6 if next_fitne > prob1.get_fitne ():
          efforts = 0
      else:
          efforts += 1
Step:7 if next_fitne > prob1.get_maximum () *optimum_fitne:
        optimum_fitne = prob1.get_maximum () *next_fitne
        optimum_state = next_state
    if curve:
        fitne_bend.append(prob1.get_fitne())
    prob1.set_state(next_state)
   if bend:
    return optimum_state, optimum_fitne, np.asarray(fitne_bend)
  return optimum_state, optimum_fitne
```

Step:8 To identify the optimal weights, initialize a fit object based on mlrose as a Randomized Hill Climbing technique, with a maximum of 2000 iterations of the algorithm and 100 trials at each step, learning_rate = 0.001, clip_max=5 and random_state=3.

Step:9 Once fitted, the model to predict the labels for our training and test sets can be utilized, and we can utilize these forecasts to quantify the model's training and accuracy of the test.

4. RESULT AND DISCUSSION

After data purification, the dataset of kidney disease patient records has 523 occurrences, with 75 percent of the dataset considered for the training set and the remaining 25% as a random sample contained in the dataset considered for the testing set. The proposed RWOM with NN, RWOM with LR, NN and LR classifier models are used to find the best classification based on CKD status such as 'No CKD', 'High risk', 'Kidney failure', 'Moderate risk', 'very high risk'. The implementing of experiment is executed by the python programming language. The Confusion Matrix (CM) parameter for proposed RWOM model and individual classifier testing status is identified and calculated below.

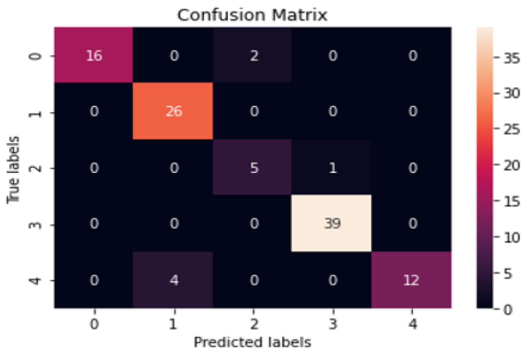


Figure 6 CM for multi-class classification of RWOM with NN

Figure 6 illustrates that five type of classification issues can be determined through this kind of categorization. A similar as binary classification, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) plays the major role but it not as typical as binary classification. For example, TP, TN, FP and FN value for RWOM are computed and array '0' as No CKD, '1' as High risk, '2' as Kidney failure, '3'as Moderate risk and '4' as very high risk.

- **For No CKD status (0)**

TP = 16

TN = (26+0+5+0) = 31

FP = 0+0+0+0 = 0

FN = 0+2+0+0 = 2

Similarly, the as 1, 2, 3 and 4 status is calculated but multi-label binary problems need to be manipulated for generating Micro F1-score which can be utilized in assessing multi-label binary quality. Micro F1-score is the aggregated contribution of all classes namely No CKD, High risk, Kidney failure, Moderate risk and very high risk. As per micro F1-score metrics calculated globally are measured to be equal is shown in equation 4.

$$Precision = Recall = Micro\ F1 - Score = Accuracy \tag{4}$$

The micro precision and micro recall can be calculated using total TP, total FP and total FN whereas the five classes are accumulated. The same instance of proposed RWOM with NN is shown in Table 2.

Table 2 Multi-class classification of five with binary classification

Classifier Model Description	**TP**	**TN**	**FP**	**FN**
RWOM.NN (0)	16	31	0	2
RWOM.NN (1)	26	45	4	0
RWOM.NN (2)	5	51	2	1
RWOM.NN (3)	39	28	1	0
RWOM.NN (4)	12	42	0	4
	98		**7**	**7**

- **Micro Precision** formulae is as like as general precision formula of binary classification but instead of TP and FP, total TP and total FP is used.

$$Micro\ precision = \frac{Total\ TP}{(Total\ TP + Total\ FP)}$$

$$Micro\ precision = \frac{98}{(98+7)} = 0.9333$$

- **Micro Recall** formulae is as like as general recall formula of binary classification but instead of TP and FN, total TP and total FN is used.

$$Micro\ Recall = \frac{Total\ TP}{(Total\ TP + Total\ FN)}$$

$$Micro\ Recall = \frac{98}{(98+7)} = 0.9333$$

Hence, as per equation 4 Micro F1-score is 0.9333 which is said to be accuracy of the RWOM with NN model. Thus, the accuracy of RWOM with NN model is 0.9333 (93.33%). Similarly, there are various classifier methods are calculated and shown in Table 3 and figure 7.

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

Table 3 Accuracy of various classifier model with proposed RWOM

Classifier Model	Accuracy (%)
NN	23.80
LR	19.047
RWOM with NN	**93.33**
RWOM with LR	22.85

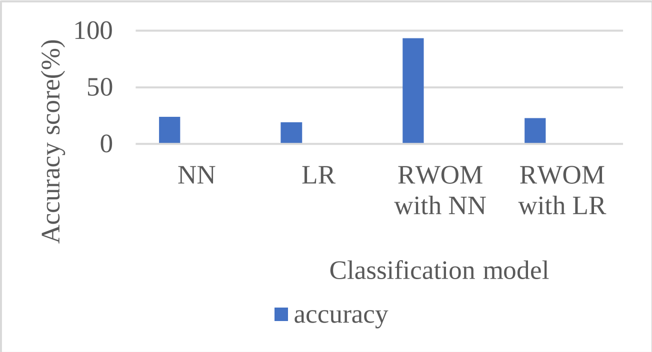


Figure 7 Accuracy score for CKD status using classification algorithm

The CM parameters like precision, recall and F1-Score for each category in the proposed RWOM have been evaluated to the respective ML techniques are calculated and considered as macro average F1-score. Similarly, the weighted score is obtained by considering weight for each category are made to sum up the total number of samples. The macro average and weighted average of precision, recall and F1-Score with respect to ML techniques are shown in Table 4.

Table 4 Macro Avg. and weighted Avg. of CM parameters for various classifier

ML Technique	Macro Average precision score	Macro Average Recall score	Macro Average F1-score value	Weighted Average Precision Score	Weighted Average Recall Score	Weighted Average F1-Score value
NN	0.2209	0.1947	0.1899	0.3319	0.2366	0.2762
RWOM.NN	0.9112	0.8944	0.7428	0.9414	0.9434	0.9424
LR	0.2090	0.2365	0.1572	0.1650	0.2930	0.2111
RWOM.LR	0.2615	0.2624	0.1924	0.2615	0.3064	0.2822

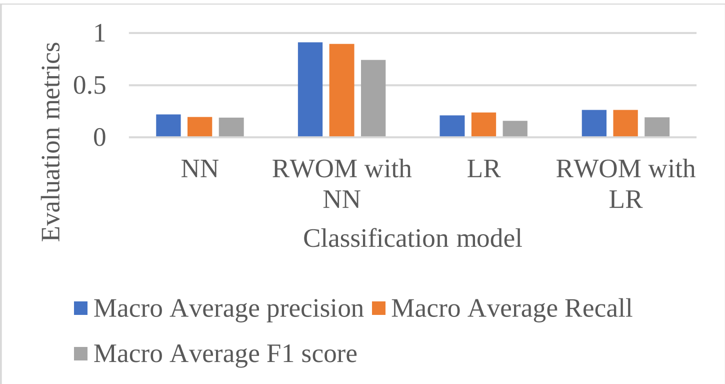


Figure.8 Macro Avg Precision, Recall and F1-Score value for classification of CKD status using various ML models

Figure 8 illustrates the macro avg precision, recall and F1-Score for various ML techniques that shows RWOM with NN performance of macro avg precision, recall and F1-Score is higher than other ML technique. The macro avg precision of RWOM with NN is 0.911, macro avg recall for RWOM with NN is 0.8944 and macro avg F1-Score is 0.7428 whereas RWOM with LR is 0.26, 0.26 and 0.19 respectively. Simultaneously, the RWOM with NN method has higher CM parameter Macro score while compared to other ML techniques.

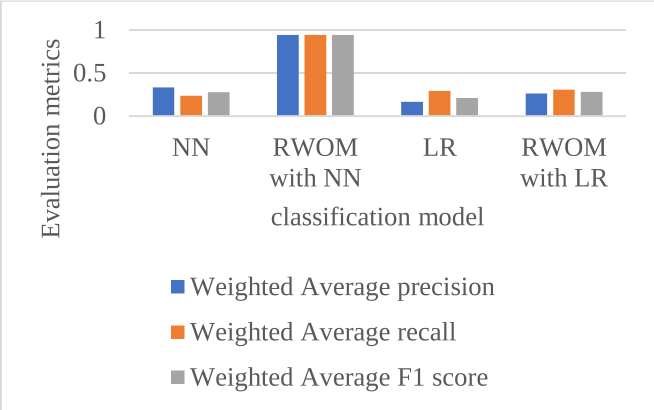


Figure.9 Weighted Avg Precision, Recall and F1-Score value for classification of CKD status using various ML models

Figure 9 illustrates the weighted avg precision, recall and F1-Score for various ML techniques that shows RWOM with NN performance of weighted avg precision, recall and F1-Score is higher than other ML technique. The weighted avg precision of RWOM with NN is 0.94, weighted avg recall is 0.9434 and weighted avg F1-Score is 0.9424 whereas RWOM with LR is 0.26, 0.306, 0.2822 respectively. Simultaneously, the RWOM with NN method has higher CM parameter weighted score while compared to other ML techniques.

5. Conclusion

The main focus of this research is to analyse the exact precision of CKD and identifying the earlier or chances of CKD may be justified using EGFR and BCR to detect and determine severity for implementing kidney disease treatment. In order to defining the stage of CKD for understanding the severity as well as identify the earlier stage CKD patients using eGFR and BCR to determine the risk of CKD with normal or pre-Renal stage or Intrinsic Renal stage. This paper has presented RWOM for improving the accuracy of classification to provide effective CKD diagnosis. Therefore, the benefit of proposed RWOM model allows in improving the output by involving randomized weights of these three classifiers such as NN and LR are considered. The research has utilized kidney disease dataset to classify and detect accuracy of these RWOM with NN mechanism for enhancing accuracy in classifying CKD status. The proposed RWOM with NN method performance is evaluated and compared with existing RWOM with LR, NN and LR methods have illustrated that better accuracy. The obtained accuracy in RWOM with NN is 93.33% in classifying CKD status which is higher than RWOM with LR, NN and LR method. These assist the expert to diagnosis the exact kidney disease and prefer the correct treatment to save the patients from the kidney disease as earlier as possible. In future work, the accuracy scores will be increase by recommending certain more rules as well as increasing the CKD data. Moreover, recommended rule-based procedure is evaluated through deep learning methods to provide high accuracy in predicting the research studies present in the kidney disease of the patient.

REFERENCES

[1] World Health Organization, Preventing Chronic Disease: A Vital Investment, WHO, Geneva, Switzerland, 2005.

[2] B. Bikbov, N. Perico, and G. Remuzzi, "Disparities in chronic kidney disease prevalence amongmales and females in 195 countries: analysis of the global burden of disease 2016 study," Nephron, vol. 139, no. 4, pp. 313–318, 2018.

[3] Z. Chen, X. Zhang, and Z. Zhang, “Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models,” International Urology and Nephrology, vol. 48, no. 12, pp. 2069–2075, 2016.
[4] Glomerular Filtration Rate (GFR), National Kidney Foundation, New York, NY, USA, 2020, https://www.kidney.org/ atoz/content/gfr.
[5] Deepanita Baidya, Umme Umaima, Md Nazrul Islam, F. M. Javed Mehedi Shamrat, Anik Pramanik and Md. Sadekur Rahman, “A Deep Prediction of Chronic Kidney Disease by Employing Machine Learning Method”, International Conference on Trends in Electronics and Informatics, 2022.
[6] J. Janani and R. Sathyaraj, “Diagnosing Chronic Kidney Disease Using Hybrid Machine Learning Techniques”, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(13), 6383-6390,2021.
[7] Gazi Mohammed Ifraz, Muhammad Hasnath Rashid, Tahia Tazin, Sami Bourouis and Mohammad Monirujjaman Khan, “Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods”, Hindawi Computational and Mathematical Methods in Medicine Volume 2021, Article ID 6141470, 10 pages https://doi.org/10.1155/2021/6141470.
[8] Ramesh Chandra Poonia, Mukesh Kumar Gupta, Ibrahim Abunadi , Amani Abdulrahman Albraikan, Fahd N. Al-Wesabi, Manar Ahmed Hamza and Tulasi B, “Intelligent Diagnostic Prediction and Classification Models for Detection of Kidney Disease”, Healthcare 2022, 10, 371. https://doi.org/10.3390/ healthcare10020371.
[9] Vasquez-Morales, G.R, Martinez-Monterrubio, S.M, Moreno-Ger, P, Recio-Garcia, J.A, “Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning” IEEE Access 2019, 7, 152900–152910.
[10] M.Makino, R. Yoshimoto, M. Ono, T. Itoko, T. Katsuki, A.Koseki, M. Kudo, K. Haida, J. Kuroda, R.Yanagiya,et.al, “Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning”, Sci. Rep. 2019, 9, 11862.
[11] Z. Segal, D. Kalifa, K. Radinsky, B. Ehrenberg, G. Elad, G. Maor, M. Lewis, M.Tibi, L.Korn and G.Koren, “Machine learning algorithm for early detection of end-stage renal disease”, BMC Nephrol. 2020, 21, 518.
[12] A. Khamparia, G. Saini, B. Pandey, S. Tiwari, D. Gupta and A. Khanna, “KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network”, Multimed. Tools Appl. 2020, 79, 35425–35440.
[13] H. Polat, H.D. Mehr, A. Cetin, “Diagnosis of chronic kidney disease based on support vector machine by feature selection methods” J. Med. Syst. 2017, 41, 55.
[14] P.Panwong, N. Iam-On, “Predicting transitional interval of kidney disease stages 3 to 5 using data mining method”, In Proceedings of the 2016 Second Asian Conference on Defence Technology (ACDT), Chiang Mai, Thailand, 21–23 January 2016; pp. 145–150.
[15] U.N.Dulhare and M.Ayesha, “Extraction of action rules for chronic kidney disease using Naïve bayes classifier”, In Proceedings of the 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, India, 15–17 December 2016; pp. 1–5.
[16] Ebrahime Mohammed Senan , Mosleh Hmoud Al-Adhaileh , Fawaz Waselallah Alsaade , Theyazn H. H. Aldhyani , Ahmed Abdullah Alqarni , Nizar Alsharif, M. Irfan Uddin , Ahmed H. Alahmadi , Mukti E Jadhav, and Mohammed Y. Alzahrani, “Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques”, Hindawi Journal of Healthcare Engineering Volume 2021, Article ID 1004767, 10 pages https://doi.org/10.1155/2021/1004767.
[17] D. -H. Kim and S. -Y. Ye, “Classification of chronic kidney disease in sonography using the GLCM and artificial neural network,” Diagnostics, vol. 11, no. 5, p. 864, 2021.
[18] Hamida Ilyas, Sajid Ali, Mahvish Ponum, Osman Hasan, Muhammad Tahir Mahmood, Mehwish Iftikhar and Mubasher Hussain Malik, “chronic kidney disease diagnosis using decision tree algorithms”, BMC Nephrology ,2021, 22:273 https://doi.org/10.1186/s12882-021-02474-z.
[19] Y. Ren, H. Fei, X. Liang, D. Ji and M. Cheng, “A hybrid neural network model for predicting kidney disease in hypertension patients based on electronic health records”, BMC Med. Inf. Decis. Mak. 2019, 19, 131–138.
[20] F. Ma, T. Sun, L. Liu and H. Jing, H, “Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network” Future Gener. Comput. Syst. 2020, 111, 17–26.
[21] Kaggle,“Chronic Kidney Disease Dataset,” https://www.kaggle .com/abhia1999/chronic-kidney-disease