Enhanced Optimized Classification Model of Chronic Kidney Disease

Shahinda Elkholy[1], Amira Rezk[2], Ahmed Abo El Fetoh Saleh[3]
Information Systems Dept.-Faculty of Computers and Information, Mansoura University,
Mansoura, Egypt

***Abstract*—Chronic kidney disease (CKD) is one of the leading causes of death across the globe, affecting about 10% of the world's adult population. Kidney disease affects the proper function of the kidneys. As the number of people with chronic kidney disease (CKD) rises, it is becoming increasingly important to have accurate methods for detecting CKD at an early stage. Developing a mechanism for detecting chronic kidney disease is the study's main contribution to knowledge. In this study, preventive interventions for CKD can be explored using machine learning techniques (ML). The Optimized deep belief network (DBN) based on Grasshopper's Optimization Algorithm (GOA) classifier with prior Density-based Feature Selection (DFS) algorithm for chronic kidney disease is described in this study, which is called "DFS-ODBN." Prior to the DBN classifier, whose parameters are optimized using GOA, the proposed method eliminates redundant or irrelevant dimensions using DFS. The proposed DFS-ODBN framework consists of three phases, preprocessing, feature selection, and classification phases. Using CKD datasets, the suggested approach is also tested, and the performance is evaluated using several assessment metrics. Optimized-DBN achieves its maximum performance in terms of sensitivity, accuracy, and specificity, the proposed DFS-ODBN demonstrated accuracy of 99.75 percent using fewer features comparing with other techniques.**

Keywords*—Machine learning (ML); feature selection (FS); chronic kidney disease (CKD); deep belief network (DBN); grasshopper's optimization algorithm (GOA)

I. Introduction

Chronic renal disease, or chronic kidney disease (CKD), worsens over time, eventually causing kidney failure. Most of the time, it can go undetected for up to 25% of its usefulness before being discovered. For people who are unaware of kidney failure's symptoms, this might make it difficult to diagnose and treat the condition. Treatment for kidney failure aims to manage the causes and slow the progression of renal failure. Patients in the last stages of renal failure will require dialysis or a kidney transplant if treatment fails [1]. Renal failure affects four out of every 1000 people in the UK, while dialysis keeps more than 300,000 Americans with kidney failure alive [2]. More people in South Asia, Africa, and the rest of the world have renal illness, according to the National Health Service (NHS). Because chronic renal failure cannot be detected until it has progressed to an advanced state, recognizing kidney failure at an early stage is critical. By detecting renal disease at an early stage, the likelihood of permanent damage to the kidneys is reduced. As a result, patients should have regular checkups and early diagnosis to avoid serious risks of renal failure and related disorders [1]. Doctors can make therapy decisions that slow the rate of progression. Measuring parameters allow for differentiation, and patients' medical records can be used to classify and predict disease using data mining techniques [3].

Data mining techniques allow for the extraction of meaningful information from large and hidden databases. As a method of gaining knowledge from unstructured information, data mining techniques can be applied even when the information is not directly related to medicine [4]. Data mining has three stages: data processing, data modeling, and data post processing. Data mining jobs in data modeling include classification/predictive algorithms and regression algorithms that are learned through a supervised learning process. As a result of missing and unneeded data being stored in the hospital database, it is difficult to mine the patient data. As a result, prior to implementing data mining techniques, it is necessary to improve data processing and data reduction methodologies [5]. Accurate and reliable data makes the identification of CKD easier and faster. Data classification can be used to identify CKD from a patient's medical records but an important part of the classification process is the establishment of a link between the feature values and the class labels for the data being processed. Hence, classification is a supervised procedure. An algorithm for classification and prediction uses training data to build a model, which is then used to predict test data [6]. Using artificial intelligence (AI) approaches, categorization models have recently been improved. Multiple issues arose as a result of the high-dimensional nature of the medical data, including high processing complexity, overftting, and low finishing model interoperability. Feature selection (FS) is the quickest and most effective way to address the issue. This method aims to reduce the number of features to a manageable subset by removing redundant or irrelevant ones. In order to save calculation time, it makes use of a small number of characteristics to extract the maximum amount of data from a dataset [7]. The selected feature subset is useful in modeling these functions. To improve prediction results, FS approaches are used in a wide range of applications such as machine learning (ML), data mining, and pattern recognition [8]. Wrapper, embedded, and filter-based methods are all examples of FS validation methods. To validate the feature subset in a filter technique, one uses fixed measurements rather than learners and a predetermined set of features. However, the wrapper method uses the learning strategy as a sub-process of evaluation to determine whether or not a feature set is better than it was previously. This method, which is widely used, has certain drawbacks, such as a high computational cost, difficulty in recognizing user-defined parameters of the learner, and

built-in constraints on the learners [9]. Embedded techniques are simpler than wrapper approaches in terms of complexity. But the selection of features is based on the learning technique [10]. The filter and wrapper methods are both integrated into the embedding method, which removes their respective drawbacks. Despite the cheap computational complexity of the filter approaches, the feature subset used for classification was shown to be unreliable. The wrapper approaches, on the other hand, achieve better classification results while using a lot more time. All three techniques have improved the classification of the features. In addition, the FS procedure has improved features rather than the classifier. High classification performance was achieved with high computing complexity using wrapper and hybrid approaches.

This paper aims to enhance the accuracy of CKD classification using the DFS-ODBN method which an innovative wrapper strategy for CKD detection that incorporates density-based feature selection (DFS) with optimized DBN based on GOA to tackle these difficulties. Used techniques are capable of solving the problem of class imbalance in the dataset which may affect classification performance in addition to this heuristic way that determine whether or not a feature is worthwhile. To improve the final classifier outputs of the DFS-ODBN algorithm, the addition of optimization algorithm called GOA as parameter adaptation to increase the performance of the classifier helps to enhance the accuracy of the optimized DBN classifier. From the UCI repository, a benchmark CKD dataset is used to test the efficiency of the DFS-ODBN method. The model was evaluated using metrics like accuracy, sensitivity, and F-measure. The results showed that the provided DFS-ODBN strategy outperforms the evaluated approaches in terms of classification performance.

The following is the article's flow: Other diagnostic approaches are reviewed in Section II; Section III presents a review of related methods and technologies; Section IV provides a discussion of applied methodology, and Section V includes simulation test results for the proposed method, comparing the results with other related studies. Section VI provides the future work with the concluded presented work.

II. Related Work

Much research has used data mining algorithms to accurately predict CKD in patients based on their medical records. From the original set of features, there has been a focus on a subset of relevant features that play a significant role in the medical diagnosis sectors because of the high dimension of required multidimensional medical multimedia data for CKD prediction. All of the studies used performance indicators including specificity, accuracy, and sensitivity to support their CKD prediction approach. We'll go through some current research on predicting kidney disease later on.

The UBFST (Union Based Feature Selection Technique) was developed for the rapid and accurate classification and diagnosis of chronic kidney disease (CKD). This method uses SVM, regression tree, and random forest to classify CKD [11]. Using a Las Vegas Wrapper Feature Selection approach (LVW-FS) and an ensemble learning-based model, hemodialysis treatment time can be predicted with acceptable accuracy. Using the LVW selection methodology, they have developed a new way of extracting crucial vital indicators. As a method of classification, a group of learners was used in this study to give numerous classifiers. Through a variety of trials with different learners, the suggested model based on LVW and the ensemble learning method was shown to have the greatest influence in decreasing hyperthyroidism characteristics and excluding noise [12]. Their technology can forecast the health of the kidney based on factors such as age, albumin and glucose levels, and more.

An evolutionary algorithm (GA) based on neural networks optimize weight vectors to train a neural network. For CKD diagnosis, the system outperforms existing neural networks in terms of accuracy [13]. For the dataset of CKD, using multilayer perceptrons (MLPs), probabilistic neural networks (PNNs), radial basis functions (RBFs), and SVM. The PNN algorithm surpassed the SVM, MLP, and RBF algorithms in terms of performance [14]. In Colombian population neural networks were used to predict the likelihood of CKD in people [15].

The study [16] proposed a method for diagnosing chronic renal illness based on grey wolf optimization (GWO) and hybrid kernel support vector machines (HKSVM). The UCI ML repository's chronic kidney dataset yielded a 97.26% accuracy rate. CKD can be identified using two fuzzy classifiers known as FuRES and FOAM, which are both fuzzy rule-building expert systems (FuRES). FuRES provides a minimum NN-based classification tree. The weight vector with the least fuzzy entropy is determined by the categorization criteria. The 386 CKD patients were identified using two fuzzy classifiers. FuRES, on the other hand, performs better than FOAM in cases where the training and prediction processes are both noisy. In the detection of CKD, both FOAM and FuRES performed better, although FuRES outperformed FOAM.

PCA and SVM were used to diagnose cervical cancer. A total of 32 potential risk factors as well as four specific outcomes were examined in this study: Hinselmann, Schiller, cytology, and biopsies. SVM with recursive feature removal and SVM with PCA were used to classify the target objects (PCA-SVM). PCA-SVM came out on top over the other two methods [17].

Using three classification algorithms Olex-GA, Ant Colony Optimization ACO and PSO compared them to developed system for diagnosing CKD. The ACO method with Density Based Feature Selection to select the important features showed superior performance and the highest accuracy [18].

Seven ML techniques are comprised to predict the CKD, J48, SVM, NBTree, LR, MLP, Naïve Bayes and Composite Hypercube on Iterated Random Projection (CHIRP) are utilized. The results of experiments show better performance for CHIRP [19].

C5.0, Artificial neural network, CHAID, logistic regression, random tree, K-Nearest neighbors and linear support vector machine were the seven classifier algorithms used to predict CKD. Results were calculated using all features from the classifier, features chosen by CFS, features chosen by Wrapper, LASSO regression, SMOTE, and chosen features

SMOTE with all the features, LASSO with selected features. The LSVM was found to have the highest accuracy in SMOTE with all features [20].

The research [21], has an additional strategy based on Recursive Feature Elimination (RFE). The most strongly representative features of CKD were chosen using the RFE algorithm. SVM, KNN, decision tree, and random forest were used to classify the features. All classifier parameters were fine-tuned to provide the best classification results, and all methods produced promising results. For all measures, the random forest approach surpassed all other algorithms. Multiclass statistical analysis was used to investigate and evaluate the system, and the empirical results of SVM, KNN, and decision tree algorithms revealed significant values of 96.67 percent, 98.33 percent, and 99.17 percent in terms of accuracy metric.

To properly pick the features subset, several feature optimization techniques were described to see the effect of them on the performance of the ML model which was tested on five influential classification models Logistic regression , Random forest (RF), SVM , K-nearest neighbors and Xtreme gradient boosting (XGB), experiments have shown that the accuracy of the model can be enhanced by using Linear discriminant analysis (LDA) feature optimizer that performs the highest outstanding result [22].

III. Related Methodologies

A. Feature Selection

Pattern recognition, knowledge discovery, and statistical research all rely heavily on feature selection. It is the goal of feature selection to eliminate unneeded inputs. No predictive class information is required to determine which features are relevant. Reducing the dimensionality of features and omitting features that are not relevant to classification can result in a comprehensive model. As the name implies, the fundamental issue in feature reduction lies in identifying the optimal collection of features to maximize classification performance [23]. Simplifying the data collection, reducing the problem of over fitting, and reducing the amount of data stored are all benefits of feature selection [24].

Three types of feature selection approaches exist. A variety of methods, including filtering, embedding, and wrapping methods. The filter method chooses the highest-ranking features, and the resulting subset can then be used in any classification algorithm that is needed to be used. Various classification methods could be tested after feature selection using the filter approach as shown in Fig. 1 [25]. The classifier's performance can be improved by reducing processing time and making better use of the dataset's optimized data while making a suitable feature selection [26]. Fastness and scalability are two more advantages of the filter approach in feature selection [27].

The classifier algorithm is used as a black box to determine scores for feature sets based on estimated power [28]. Testing and training on a given dataset are used to evaluate a subset. All features of subsets can be searched for in a wrapper approach that uses the wrapper algorithm around the classifier shown in Fig. 2 [25]. In spite of the advantages of working with correlated data and identifying the relevant correlations, over-fitting difficulties may arise. Feature selection is included in the construction of the classifier in the embedded method illustrated in Fig. 3.

When using an embedded technique, you don't have to deal with costly computations and can instead interact directly with your classification model. Table I displays the benefits and drawbacks of various feature selection approaches.

TABLE I. Benefits and Drawbacks of Various Feature Selection Approaches

approach	Benefits	Drawbacks
Filter	Scalable and quick without being dependent on any classifier,	Interaction with classifiers is not considered.
Wrapper	- A higher degree of computation efficiency, Simple, Classificati accuracy models with dependenci are the best kind. - The classifier interacts with it. - Wrapper approaches have a high computational complexity. Mode include a hierarchy of dependencies	Over fitting is a possibility, and the process is computationally intensive.

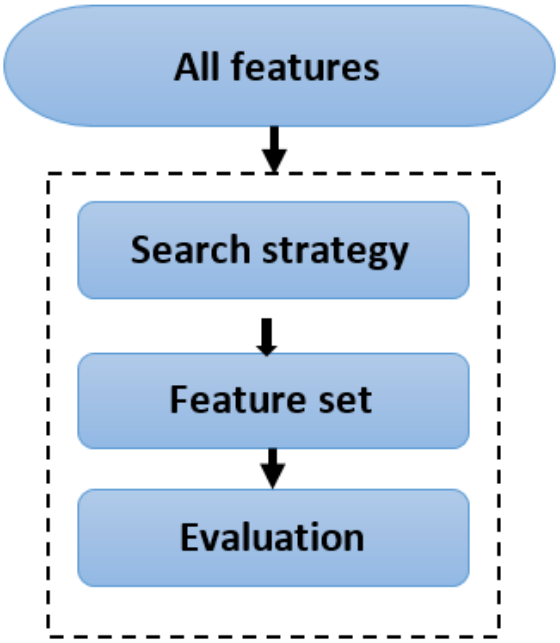

Fig. 1. Filter approach of feature selection

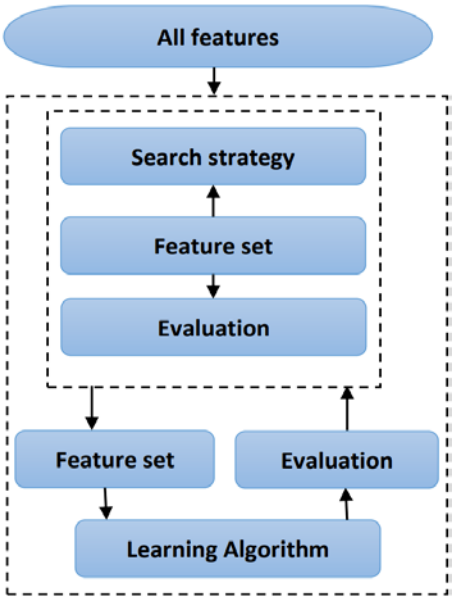

Fig. 2. Wrapper approach of feature selection

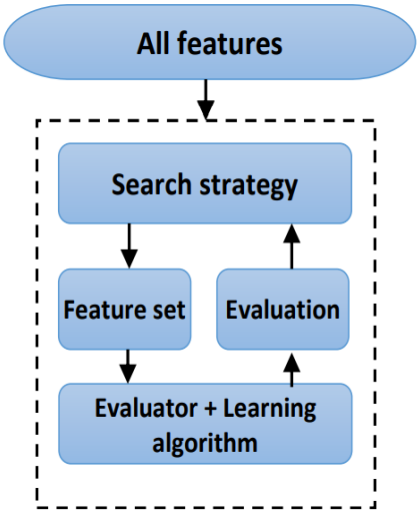

Fig. 3. Embedded approach for feature selection

B. DBN

Using an unsupervised machine learning paradigm, deep belief networks are a more advanced kind of generative neural network. A DBN can be created by stacking and training individual Restricted Boltzmann Machines (RBM) in a layered configuration. The pre-training stage is a step of unsupervised learning. Stacking sub networks, each with two processing levels, divides the network into groups. These weights are provided to the network as a way to avoid the issues that can arise from using random numbers to initialize the connection weights. Unsupervised learning is used to steer the learning phase of an energy-based stochastic neural network (RBM), which includes two layers of neurons, hidden and visible nodes. Unsupervised training may be used for this function by using the Greedy Layer Wise unsupervised training algorithm. RBM has a hidden layer including nodes and a visible layer including nodes, respectively.

Using an unsupervised learning technique, one can learn an RBM, a generative stochastic neural network. There are two processing tiers in the RBM's network, as depicted in Fig. 4. Construction and reconstruction operations can be carried out independently of one another because these layers are linked together [29].

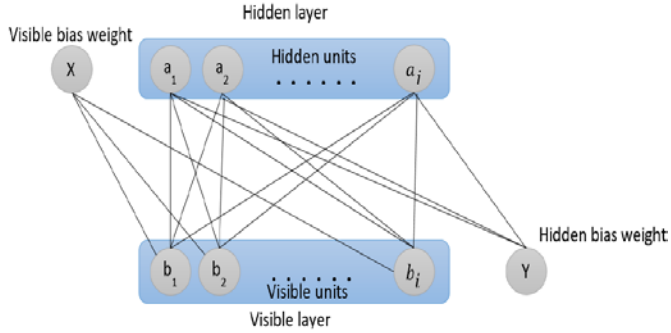

Fig. 4. The RBM architecture

In the visible layer (b), there are visible units $(b_1, b_2, b_3, \dots, b_i)$ that represent the features of the pattern, while in the hidden layer (a), there are hidden units $(a_1, a_2, a_3, \dots, a_j)$ that accept their data from the visible units and are able to reconstruct the pattern's characteristics from them. The visible node B_i and the hidden node H_j have a weight of U_{ij} in the $k \times l$ matrix U, which represents the weights between visible and hidden levels. Let's assume that the binary and hidden units are B and A. In this case, $A \in \{0,1\}^l$ and $\in \{0,1\}^k$. As shown below in (1), the RBM energy function is a quadratic function of the square of:

$$\mathrm{Ene(B,A)} = -\sum_{i=1}^{k} \mathrm{R_i B_i} - \sum_{j=1}^{l} \mathrm{Z_j A_j} - \sum_{i=1}^{k} \sum_{j=1}^{j} \mathrm{B_i A_j U_{ij}} \quad (1)$$

where Z and R are the basis vectors of the hidden and visible layers, respectively. Also, the likelihood of the (B, A) configuration is shown in the following equation.

$$\mathrm{pr(B,A)} = \frac{\mathrm{e^{-Ene(B,A)}}}{\sum_{B,A} \mathrm{e^{-Ene(B,A)}}} \quad (2)$$

Normalization is reflected in the denominator of the aforesaid equation. Stochastic gradient descent (SGD) is used to optimize RBM parameters Z, R, and U based on the training data's log likelihood. In order to calculate the probability of a given sample over all potential hidden vectors, one can use the following formula:

$$\mathrm{pr(B)} = \frac{\sum_{B} \mathrm{e^{-Ene(B,A)}}}{\sum_{B,A} \mathrm{e^{-Ene(B,A)}}} \quad (3)$$

Z, R, and U are used to generate stochastic gradient ascent derivatives of $pr(B)$, which lead to the following equations:

$$\mathrm{U^{n+1}} = \mathrm{U^{n}} + \xi\left(\mathrm{pr(A/B)B^{T}} - \mathrm{pr(\tilde{A}/\tilde{B})\tilde{B}^{T}}\right) - \eta \mathrm{U^{T}} + \alpha \Delta \mathrm{U^{n-1}} \quad (4)$$

$$\mathrm{Z^{n+1}} = \mathrm{Z^{n}} + \xi(\mathrm{B} - \tilde{\mathrm{B}}) + \alpha \Delta \mathrm{Z^{n-1}} \quad (5)$$

$$\mathrm{R^{n+1}} = \mathrm{R^{n}} + \xi(\mathrm{prob(A/B)} - \mathrm{prob}(\tilde{\mathrm{A}}/\tilde{\mathrm{B}})) + \alpha \Delta \mathrm{R^{n-1}} \quad (6)$$

With

$$\mathrm{pr}(\mathrm{A_j} = 1/\mathrm{B}) = \mathrm{sigma}\left(\sum_{i=1}^{k} \mathrm{U_{ij}B_i} + \mathrm{R_j}\right) \quad (7)$$

And,

$$\mathrm{pr}(\mathrm{B_j} = 1/\mathrm{A}) = \mathrm{sigma}\left(\sum_{i=1}^{l} \mathrm{U_{ij}A_i} + \mathrm{Z_j}\right) \quad (8)$$

where η, n, ξ, and α denote the weight decays, number of hidden nodes, learning rate, and momentum weights, in that order. The Softmax function is symbolised by the symbol sigma in the logistic model . For the visible and hidden nodes, two different learning methods are used to figure out their weights and biases. Both CD and persistent contrastive divergence (PCD) fall under this category (PCD).

DBN uses an error propagation method and fine-tuned optimal performance to identify the starting weights, which are obtained by undertaking unsupervised pre-training. RBM pre-training is still lacking in the optimal number of layers and nodes. The outcomes are affected by the number of layers and also the number of nodes, but the ideal value relies on the type of dataset and the attributes to be learned. Therefore obtaining the global optimum value has certain downsides. GOA is used to solve the problem of DBN in our job. GOA is a technique for determining the DBN's ideal value and, as a result, reducing error.

IV. Proposed Model

The proposed DFS-ODBN model consists of three phases namely: preprocessing phase, feature selection phase and classification phase. The first phase, preprocessing, involve dealing with the dataset nature which include many missing data due to the archive of data in hospitals. Preprocessing also include normalization of scale of data to make all attribute in a specified range. The second phase involves selecting a subset of attributes to reduce the complexity and time of classification phase. Feature selection phase involve selecting the most appropriate features among all available features using DFS feature selection algorithm in wrapper approach which repeatedly apply DFS method. In last phase, classification, the DBN classifier is used to predict the case of data which ckd or NOT ckd. The classifier parameters are estimated and tuned using an optimization algorithm called GOA. The use of optimization algorithm helps in parameter adaptation to increase the performance of the classifier. The last step is model evaluation which assesses the performance of the proposed system according to many metrics such as: accuracy, sensitivity, f-measure, and precision. The evaluation proved that the proposed system is better than many other related methods. The proposed system is shown in Fig. 5.

A. Description of Dataset

UCI's Machine Learning Repository now has the CKD dataset used in this study which was collected and uploaded by the Apollo Hospital in India in 2015. There are 400 data points total, with 25 different properties, 11 of which are numeric and 14 of which are nominal [30]. 250 of the dataset's 400 instances have been assigned to the ckd class, while the remaining 150 have been assigned to the NOTckd class. Table II shows the CSD dataset's attribute breakdown.

B. Preprocessing

The quality of the data used in data mining operations must be high in order to achieve a high level of performance at a cheap cost. Anomaly type characteristics will be converted to numeric in the preprocessing step. A total of 14 nominal attributes will be transformed into numerical attributes.

TABLE II. Chronic Kidney Disease Dataset Information

Features	Type
Red blood cell count (Rc)	Numeric
Sugar (Su)	Nominal
Hypertension (Htn)	Nominal
Sodium (Sod)	Numeric
Red blood cells (Rbc)	Nominal
Packed cell volume (Pcv)	Numeric
Pus cell (Pc)	Nominal
Age (Age)	Numeric
Appetite (appet)	Nominal
White blood cell (Wbcc)	Numeric
Bacteria (Ba)	Nominal
Diabetes mellitus (Dm)	Nominal
Specific gravity (Sg)	Nominal
Serum creatinine (Sc)	Numeric
Anemia (Ane)	Nominal
Pus cell clumps (Pcc)	Nominal
Blood glucoses (Bgr)	Numeric
Coronary artery disease (Cad)	Nominal
Pedal edema (Pe)	Nominal
Blood pressure (Bp)	Numeric
Blood urea (Bu)	Numeric
Albumin (Al)	Nominal
Haemoglobin (Hemo)	Numeric
Potassium (Pot)	Numeric
Class (class)	Nominal

1) Missing values: More than half of the variables in the CKD dataset are missing, necessitating the handling of missing values in order to improve accuracy. The mode method is used to replace the empty value with the attribute's maximum frequency when a value is lacking. Attributes can be univariate, monotonous in their missing values, or arbitrary. Only one characteristic has all of the missing values in univariate analysis (feature). If at least three attributes are missing values, the model is said to be monotonous. If the missing values are of random characteristics, then it is arbitrary [31].

2) Data normalization: There are numerous approaches to data normalization. Keep the data in a range for each input feature in order to reduce the neural network's preference for one feature over another. Training time can be reduced by normalizing data such that all features are trained at once. It is particularly beneficial for modelling applications when the inputs are often on a wide range of scales. The features or outputs are rescaled using the Min-Max normalizing method from one range of values to another. Most of the time, the features are rescaled to fall between 0 and 1 or -1 and 1. It is common to perform the rescaling by applying a linear interpretation formula like:

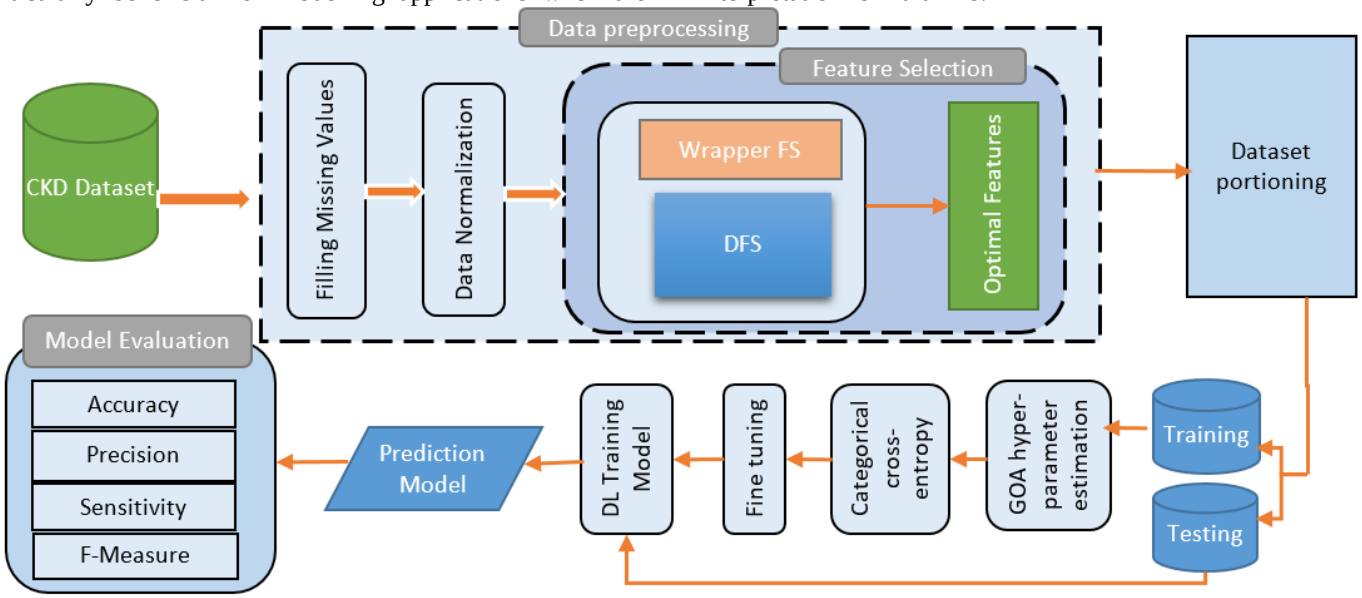

Fig. 5. Proposed framework

$$x_i' = ((max_{target} - min_{target})x\frac{(x_i - min_{value})}{(max_{value} - min_{value})} + min_{targe}) \quad (9)$$

where $(max_{value} - min_{value}) = 0$. when $(max_{value} - min_{value}) = 0$ for a feature, it shows that that feature in the data has a constant value. Feature values having a constant value should be deleted from the data set because they do not contribute any useful information to the neural network. Min-max normalization maintains the same range of values for each feature when it is applied. The advantage of using min-max normalization is that it keeps all of the data's relationships intact.

C. DFS for Feature Selection

In order to select the best features, the proposed algorithm employs the following procedures. It is necessary to organize the CKD dataset into groups once the preprocessing is complete. Selecting a collection of features in each iteration is done using DFS. The most important feature in the classification process is a subset of the best features in the raw dataset.

Features can be evaluated using the DFS method, a heuristic approach. A feature is considered good if it has less overlap with other classes than other features. The DFS method takes into account the distribution of features across all classes and their associations for determining rankings. As a first stage in DFS, every feature in every class is given a probability density function (PDF). The next step is to rank the features depending on the overlap area, and this is done next. Parametric and non-parametric methods of computing PDF are the most frequent approaches [32]. The first method assumes that data has a Gaussian distribution, so the work of estimating density simply entails choosing appropriate values for the distribution's mean and variance. Instead of making assumptions about the shape of the density function, non-parametric techniques simply calculate the density from the observed data. Many pattern recognition applications lack a standard format for estimating the density of raw data. When using random distributions and non-parametric approaches, it is unnecessary to know the fundamental density forms before using these methods [32]. Because of this, the proposed solution is described as using the following parametric approach:

$$p(x) \cong \frac{k}{MVo}$$

Here, the value of the derived PDF for instance x is represented by the expression $p(x)$, whereas M the total number of examples, Vo is the volume surrounding x, and k the number of instances within V are all given. To get a more accurate PDF, try increasing M and decreasing Vo. After estimating PDFs for each class, the next step is to compare the value of each feature based on the calculated PDFs for each class. As previously stated, a feature is considered good if it has less class overlap than the rest. The estimation of PDF for each class label and feature is used to estimate the amount of overlap between occurrences of a certain feature class. The significance of a feature for class label prediction decreases as the overlapping region grows larger, and this leads to a decline in classification performance. The overlapping value of a feature r in class ω can be calculated using the formula 2.

$$\mathrm{Ov}\ (\mathrm{r}, \omega) = \int \mathrm{Minimum}\left(\mathrm{Maximum}\left(\mathrm{PDF}(\omega_i)\right), \mathrm{PDF}(\omega)\right)$$

such that i is not equal to ω and $1 \leq i \leq$ # of classes

D. Optimized DBN based on GOA

"Reference [33] proposed the GOA algorithm. It was inspired by the swarming activity of grasshoppers in nature. A grasshopper's flight path in a swarm is affected by the following three factors: Those three factors are: human social connection, gravitational pull, and wind advective forces. The GOA algorithm is used for minimizing the value of error in order to get the ideal DBN value. The GOA has a series of steps, as outlined below:

1) Step 1: initial step: It is necessary to initialize the RBM parameters in GOA, as well as the number of candidate solutions, MIN_{CA}, MAX_{CA} , and maximum number of evaluations, before performing any analysis. A decrease in the repulsion area, comfort area, and attraction area is represented here by the parameter CA. The starting population is then generated at random, and using the objective function, each solution in the population is assessed.

2) Step 2: assessment of fitness: For each search agent, the fitness function is calculated and computed after initialization. For the sake of this study, we define fitness as the minimization of DBN's mean square error:

$$\text{Fitness } = \mathrm{Min}_{\text{error}} = \frac{1}{\mathrm{E}} \sum_{i=1}^{E} \left(\frac{P(x,y) - T_y}{T_y} \right)^2 \quad (10)$$

The mean square error denotes the average distance between the predicted and observed values. T_y and $P(x, y)$ represent the target value for the appropriate data y and the estimated probability of the appropriate data x, respectively.

The RBM parameter is tuned using this algorithm to reduce the amount of error that occurs. Using a DBN for unsupervised pertaining and supervised fine tuning was the basis for the proposed method. GOA is used throughout the entire process of relating and fine tuning. The range of hyper-parameters in our model was identified by conducting random search experiments and then selecting their values at random until we achieved the best performance. It takes less time and effort to train a network using random search. Because not all hyper-parameters are equally important for tuning, random experimentation based on hyper-parameter values is more efficient.

3) Step 3: updating: Update the best target's position in each evaluation and factor C in this phase. In actuality, there is no goal because we don't know what the global optimum is or what the precise goal is. The target must be identified at each stage of the optimization process. During optimization in GOA, it is assumed that the fittest grasshopper is the target. Grasshopper must advance toward the good target in order for GOA to save it in search space during this iteration. Position and C updates are calculated as follows:

$$x_m^d = C\left(\sum_{n=1, n\neq m}^{Z} C \frac{d_{ub} - d_{lb}}{S} S\left(\left|X_n^d - X_m^d\right|\right) \frac{X_n - X_m}{d_{mn}}\right) + \hat{d}_T \quad (11)$$

In the d-dimension, d_{lb} and d_{ub} denote the lower and upper bounds, respectively. The d-dimensional goal is d_T . The following formula can be used to determine the strength of social forces represented by the function S:

$$S(f) = ge^{-\frac{f}{r}} - e^{-f} \quad (12)$$

where g and r stand for the intensity of attraction and the length scale of attraction, respectively. C represents a decreasing coefficient that reduces the repulsion zone, comfort zone, and attraction zone. The GOA algorithm's primary governing parameter is C, which is kept current using the eq. 13 below.

$$C = \mathrm{Max}_C - y \frac{\mathrm{Max}_C - \mathrm{Min}_C}{Y} \quad (13)$$

where Y denotes the most iterations possible and the current iteration is referred by y, $\mathrm{Max}_C = 1$ and $\mathrm{Min}_C = 0.00001$.

4) Step 4: termination: When the maximum number of iterations is achieved, the position is updated iteratively. Finally, the best objective was returned, with the global optimum being a combination of position and fitness. The GOA algorithm is used to figure out what the ideal value of DBN should be. Once DBN has classified the ckd or NOTckd, the process is complete. This means that the result from the classifier is either positive or negative. Below, the GOA algorithm 1 is illustrated and the steps are shown in "Fig. 6".

Line #	**Algorithm 1:** DBN optimized based on GOA algorithm
1	**Input**: candidate solution, RBM parameters, MAX_C, maximum # of iters, MIN_C,
2	**Output**: Optimal parameter combination
3	Set initialization of the candidate solution and random population of parameters of RBM
4	Set initialization of MAX_c, maximum # of iters, MIN_c
5	Compute the fitness value fit (i) for each individual agent using mean square error as fitness function
6	Set best individual search agent $= W$
7	***While loop*** ($Q <$ max # of iters)
8	modify C using eq. 3
9	***foreach*** individual:
10	The current individual position is modified using eq. 2
11	reset the individuals above and below the bounds
12	**for end**
13	Updating W to a better solution
14	$Q = Q + 1$
15	**while end**
16	Return W as optimal

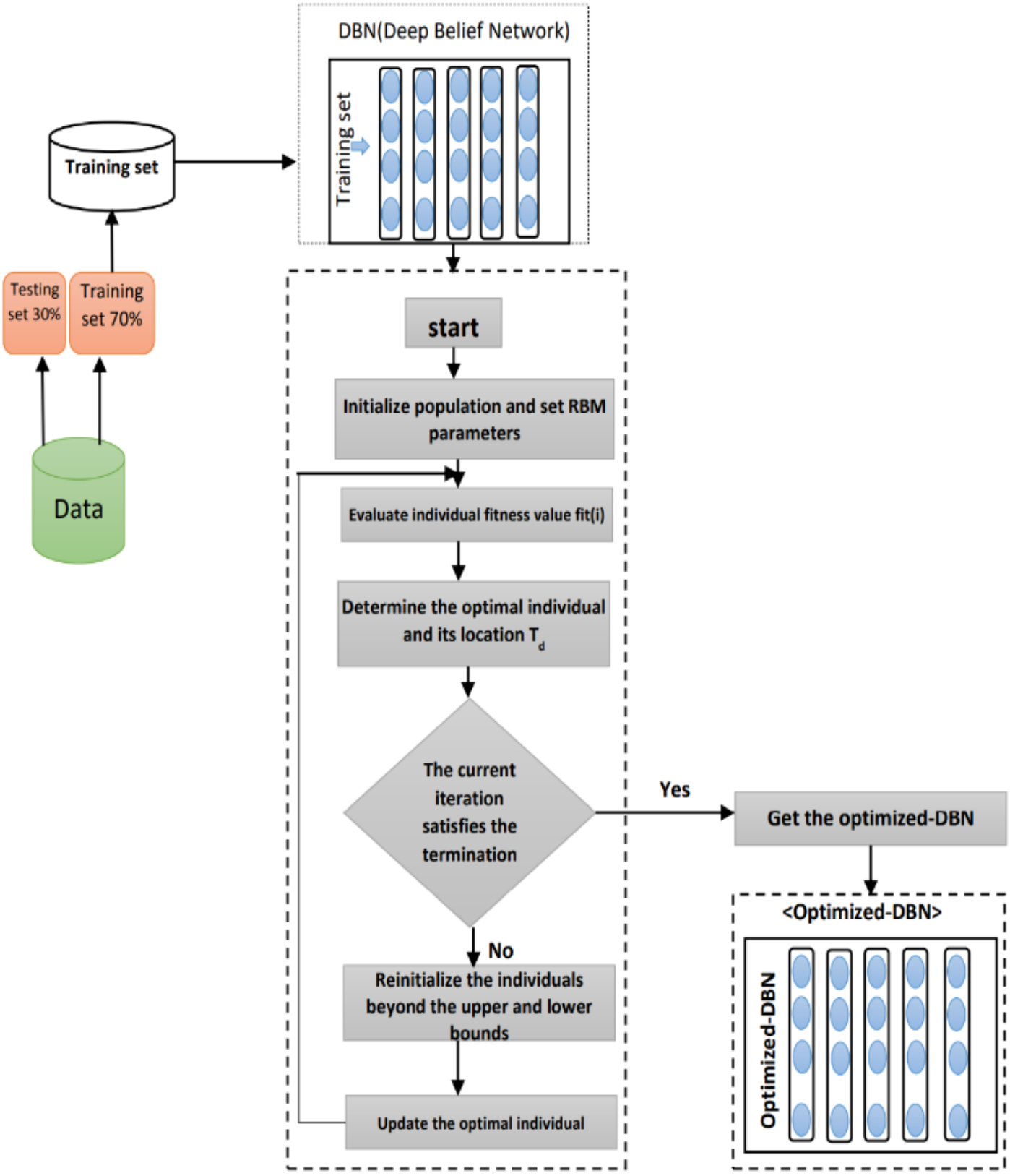

Fig. 6. DBN optimized based on GOA algorithm

V. EXPERIMENTAL RESULTS

A. Environmental Setup

The proposed model was developed in MATLAB 2016a, with some preprocessing done in Weka. MATLAB is a powerful data mining tool. Models and applications are built based on data analysis. A Core i7 machine with an NVIDIA graphics card is utilized for testing and assessments.

B. Performance Metrics

In order to train the model, the data set is randomly divided into two parts, with the first portion containing 70% of all the total collection of data. The data from the second section is used for testing (30% of the time).The suggested model is evaluated and validated using six performance measures. Accuracy, precision, sensitivity, and F-Measure [34] are some examples of these parameters. By measuring performance measures, the confusion matrix describes the performance of categorization algorithms. The following measures were utilized in this study to evaluate the performance of the methodologies in use.

1) True positive (TP): This indicates occurrences of positive outputs that have been appropriately classified.

2) False negatives (FN): These are false negatives that are not actually false negatives.

When unfavorable outcomes are mistakenly labeled as positive ones, the term "False Positive" (FP) is used.

3) False negative (FN): good events that were mistakenly labeled as negative in the report.

- **Accuracy:** An image's accuracy in a database is determined by how closely the image's coordinates match the database's real value. Accuracy measures are quite close to the genuine value, and they are processed as a true proportion of the outcomes:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

- **Sensitivity (Recall):** Sensitivity is a term used to describe the state of being sensitive. A test's sensitivity or recall refers to its capacity to correctly identify people who suffer from a certain illness (True Positive Rate). In this way, it can be stated:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (15)$$

- **Specificity:** When a test has high specificity, it can accurately identify people who do not have the condition (True Negative Rate). According to the definition, it is:

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (16)$$

- **Precision:** Predictive Value (PPV) or Precision is a measure of the accuracy of a categorization result. The following is the formula used to arrive at this result:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (17)$$

- **F-measure:** According to this method, the harmonic mean of precision and recall is calculated as follow:

$$\text{F} - \text{Measure} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (18)$$

C. Performance Comparison

As a result of using a variety of feature selection methods, including filtering and wrapping, CKD diagnosis was improved. In all of the methods used, the original dataset can be reduced in dimension to produce a new dataset. Twenty-five different variables were included in the dataset that was used. The dataset was condensed down to 11 attributes using the DFS wrapper approach. On a smaller dataset, the optimized-DBN classifier was able to identify 2 FPs and 3 FNs. The value of FN in the Optimized-DBN classifier method is lower than the values of FN in the other methods. When compared to the current system, the proposed combination of Optimized-DBN and unsupervised training has better accuracy than other research work. Layers (visible layer, hidden layer, and output layer), nodes, and weights and biases for the layers were used to develop the final architecture of Optimized-DBN. Our competitors' quantitative measures are compared to ours in this section. Table III represents the different performance evaluation metrics, i.e., accuracy, f-measure, precision, and sensitivity for the proposed system DFS-ODBN and other different related methods.

A sensitivity and accuracy analysis of the proposed DFS-ODBN technique is shown in Fig. 7. As shown in the figure, the FDS-ODBN technique has improved in both sensitivity and precision.

TABLE III. PERFORMANCE METRICS FOR PROPOSED SYSTEM AND OTHER SYSTEMS

Methods	Results(%)			
	Sensitivity	Precision	F-measure	Accuracy
Shrivas, Sahu and Hota [11]	94.00	95.14	94.57	93.25
Rady and Anwar [14]	95.61	95.98	95.79	94.75
Rubini and Perumal [16]	91.61	92.33	91.97	95.00
Elhoseny et al [18]	96.00	96.46	96.00	95.00
Chittora et al [20]	98.00	96.67	98.31	98.86
M.M. Hossain [22]	99.43	99.82	99.60	99.50
Proposed model	99.63	98.81	99.63	99.70

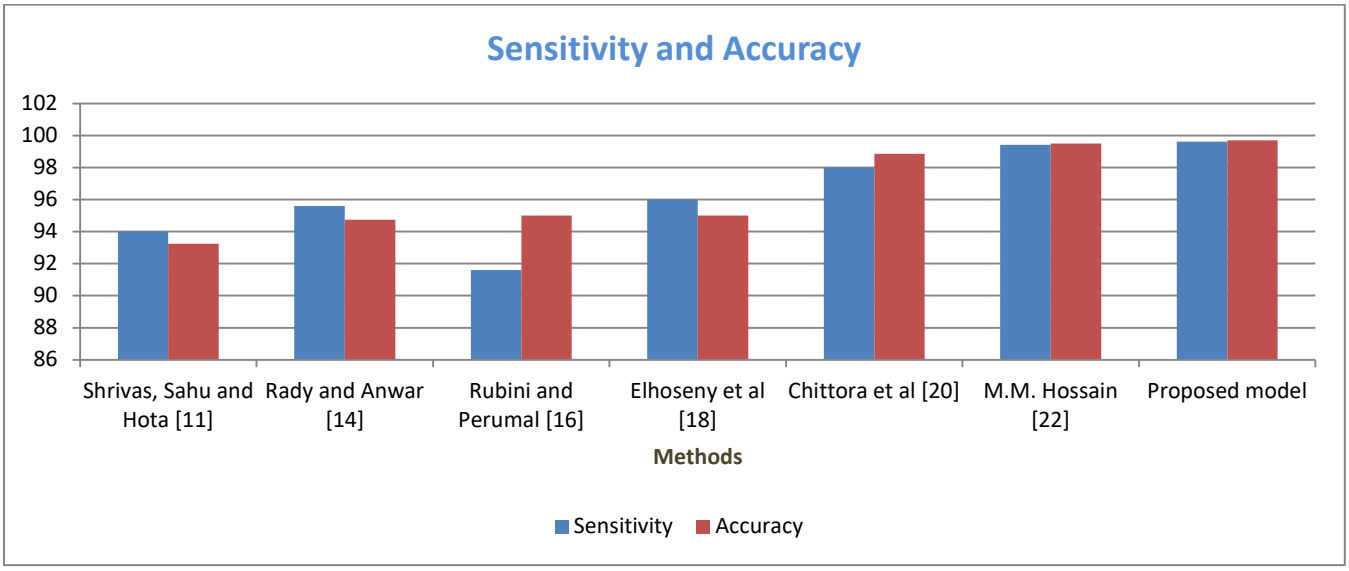

Fig. 7. Evaluation metrics (sensitivity and accuracy) Vs. classifiers

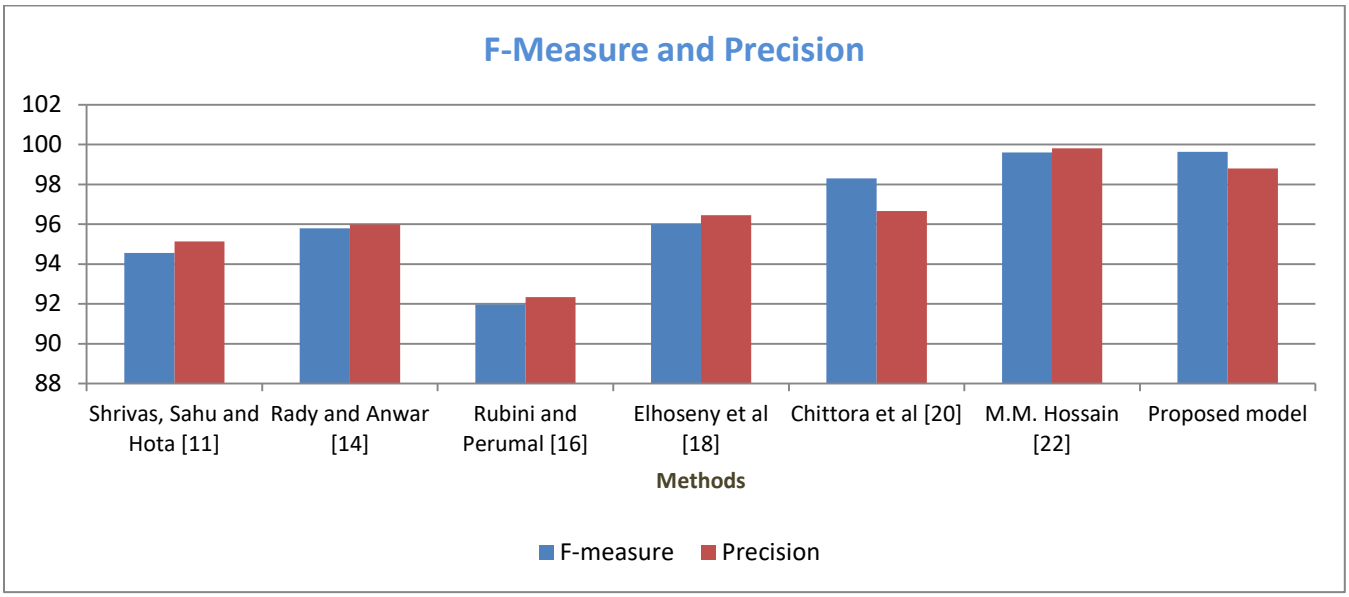

Fig. 8. Evaluation metrics (F-measure and precision) Vs. classifiers

This method is, 4.7 percent more accurate than [16]. 6.52 percent more accurate than [11].percent more accurate than [14]. 4.7 percent higher [18] and 1.1 percent higher [20].

F-measure and precision analysis are shown in Fig. 8 in terms of the feature selected for the proposed technique DFS-ODBN. According to the graph, the FDS-ODBN technique now has higher precision and precision values. In terms of precision, this method is 3.66308 percent higher than [11], and 2.82084 percent higher than [14]. The proposed method is 6.46607 percent higher [16] and 3.364 percent higher [18] in terms of the F-measure.

VI. Conclusion and Future Work

The study looked at the classification of medical data in order to identify the patient's disease at an early stage of development. The most difficult part of classifying medical data is selecting the best subset of attributes from the dataset under consideration. A DFS feature selection algorithm was used to select the best features from a preprocessing stage in which the missing values were eliminated. Based on the presence or absence of CKD, the dataset was divided into two classes: the ckd class and the NOT ckd class. The Deep Belief Network algorithm was used for this classification because it is the best method for data classification. It was necessary to use the GOA algorithm in order to obtain the DBN network hyper-parameters. GOA has strong capabilities to explore the search space and it benefits from high exploration and exploitation.

Furthermore, the classification issues can be overcome with an average computational cost. Using CKD datasets, Optimized-DBN was able to achieve its maximum performance in terms of sensitivity, accuracy, and specificity. When compared to other techniques, the proposed DFS-ODBN

demonstrated accuracy of 99.75 percent. In the future, with algorithms designed or prediction techniques, the data classification can be enlarged and missing values can be removed by new imputation approaches, classification, and prediction. The development of hybrid and novel optimization algorithms for the classification of medical data and feature selection is recommended as a focus for future contributions

REFERENCES

[1] H. Polat, H. D. Mehr, and A. Cetin, "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods," J. Med. Syst., vol. 41, no. 4, p. 55, 2017.

[2] A. S. Go, G. M. Chertow, D. Fan, C. E. McCulloch, and C. Hsu, "Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization," N. Engl. J. Med., vol. 351, no. 13, pp. 1296–1305, 2004.

[3] P. Kathuria and B. Wedro, "Chronic kidney disease quick overview." IOP Publishing emedicinehealth, 2016.

[4] M. H. A. Elhebir and A. Abraham, "A novel ensemble approach to enhance the performance of web server logs classification," Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 7, no. 1, pp. 189–195, 2015.

[5] G. A. Afzali and S. Mohammadi, "Privacy preserving big data mining: association rule hiding using fuzzy logic approach," IET Inf. Secur., vol. 12, no. 1, pp. 15–24, 2018.

[6] M. Shardlow, "An analysis of feature selection techniques," Univ. Manchester, vol. 1, no. 2016, pp. 1–7, 2016.

[7] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Feature extraction: foundations and applications, vol. 207. Springer, 2008.

[8] R. Jensen, "Combining rough and fuzzy sets for feature selection," Citeseer, 2005.

[9] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," IEEE Trans. Evol. Comput., vol. 20, no. 4, pp. 606–626, 2015.

[10] D. Mladenić, "Feature selection for dimensionality reduction," in International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection", 2005, pp. 84–102.

[11] A. K. Shrivas, S. K. Sahu, and H. S. Hota, "Classification of chronic kidney disease with proposed union based feature selection technique," in Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT), 2018, pp. 26–27.

[12] C. Xiong, M. Su, Z. Jiang, and W. Jiang, "Prediction of hemodialysis timing based on LVW feature selection and ensemble learning," J. Med. Syst., vol. 43, no. 1, pp. 1–8, 2019.

[13] S. Ravizza et al., "Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data," Nat. Med., vol. 25, no. 1, pp. 57–59, 2019.

[14] E.-H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms," Informatics Med. Unlocked, vol. 15, p. 100178, 2019.

[15] G. R. Vásquez-Morales, S. M. Martinez-Monterrubio, P. Moreno-Ger, and J. A. Recio-Garcia, "Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning," IEEE Access, vol. 7, pp. 152900–152910, 2019.

[16] L. J. Rubini and E. Perumal, "Hybrid kernel support vector machine classifier and grey wolf optimization algorithm based intelligent classification algorithm for chronic kidney disease," J. Med. Imaging Heal. Informatics, vol. 10, no. 10, pp. 2297–2307, 2020.

[17] L. Jerlin Rubini and E. Perumal, "Efficient classification of chronic kidney disease by using multi-kernel support vector machine and fruit fly optimization algorithm," Int. J. Imaging Syst. Technol., vol. 30, no. 3, pp. 660–673, 2020.

[18] M. Elhoseny, K. Shankar, and J. Uthayakumar, "Intelligent diagnostic prediction and classification system for chronic kidney disease," Sci. Rep., vol. 9, no. 1, pp. 1–14, 2019.

[19] B. Khan, "Empirical Evaluation of ML Techniques for CKD Prophecy," IEEE Access, vol.8, March 2020, DOI:10.1109/ACCESS.2020.2981689.

[20] P. Chittora et al, "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," IEEE Access, vol.9 January 2021 DOI 10.1109/ACCESS.2021.3053763.

[21] E. M. Senan, M. H. Al-Adhaileh, and F. W. Alsaade, "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques," Journal of Healthcare Engineering ., vol. 2021, https://doi.org/10.1155/2021/1004767.

[22] M. M, Hossain, Reshma. A. Swarna, "Analysis of the performance of feature optimization techniques for the diagnosis of machine learning-based chronic kidney disease," Machine Learning with Applications, https://doi.org/10.1016/j.mlwa.2022.100330.

[23] R. K. Singh and M. Sivabalakrishnan, "Feature selection of gene expression data for cancer classification: a review," Procedia Comput. Sci., vol. 50, pp. 52–57, 2015.

[24] C.-T. Su and C.-H. Yang, "Feature selection for the SVM: An application to hypertension diagnosis," Expert Syst. Appl., vol. 34, no. 1, pp. 754–763, 2008.

[25] B. Kumari and T. Swarnkar, "Filter versus wrapper feature subset selection in large dimensionality micro array: A review," 2011.

[26] O. Villacampa, "(Weka - Thesis) Feature Selection and Classification Methods for Decision Making: A Comparative Analysis," ProQuest Diss. Theses, no. 63, p. 188, 2015, [Online]. Available: https://login.pallas2.tcl.sc.edu/login?url=https://search.proquest.com/docview/1721469983?accountid=13965%0Ahttp://resolver.ebscohost.com/openurl?ctx_ver=Z39.882004&ctx_enc=info:ofi/enc:UTF8&rfr_id=info:sid/ProQuest+Dissertations+%26+Theses+Global&rft_v.

[27] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, "Feature Subset Selection Problem using Wrapper Approach in Supervised Learning," Int. J. Comput. Appl., vol. 1, no. 7, pp. 13–17, 2010, doi: 10.5120/169-295.

[28] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," Proc. 2014 Sci. Inf. Conf. SAI 2014, no. October, pp. 372–378, 2014, doi: 10.1109/SAI.2014.6918213.

[29] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," in Neural networks: Tricks of the trade, Springer, 2012, pp. 599–619.

[30] J. Li et al., "Imputation of missing values for electronic health record laboratory data," NPJ Digit. Med., vol. 4, no. 1, pp. 1–14, 2021.

[31] T. Jo and N. Japkowicz, "Class imbalances versus small disjuncts," ACM Sigkdd Explor. Newsl., vol. 6, no. 1, pp. 40–49, 2004.

[32] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification. 605 Third Avenue." New York, USA: John Wiley \& Sons, 2012.

[33] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: theory and application," Adv. Eng. Softw., vol. 105, pp. 30–47, 2017.

[34] M. Liu, C. Xu, Y. Luo, C. Xu, Y. Wen, and D. Tao, "Cost-sensitive feature selection via f-measure optimization reduction," 2017.