Distribution Based Chicken Swarm Optimization And Enhanced Support Vector Machine Algorithm For Chronic Kidney Disease Classification

M. Lincy Jacquline[1], N. Sudha[2]*

[1]Assistant Professor, Department of Computer Science, Nirmala College for Women,
Email: jacquline1990@gmail.com
[2]Associate professor, Department of Computer Science, Bishop Appasamy College of Arts and Science,
Email: sudhamuruganathan@yahoo.co.in
*Corresponding Author

Received: 13.04.2024 Revised : 18.05.2024 Accepted: 26.05.2024

ABSTRACT

One of the major issues with death rates in healthcare today is Chronic Kidney Disease (CKD), that is slow to detect and often goes undiagnosed. Because of this important problem, millions of men and women suffer every year as a result of inadequate early screening programs and treatment. Nonetheless, prompt identification of the illness at its initial stages can save people's lives. Furthermore, given a trustworthy dataset, the machine learning (ML) method's assessment procedure can identify the stage of this fatal illness much more quickly. In this work, Distribution based Chicken Swarm Optimization (DCSO) algorithm and Enhanced Support Vector Machine (ESVM) algorithm is suggested. Initially, the dataset is collected which is pre-processed utilizing K-Means Clustering (KMC) system. It is utilized to handle the missing values and error rates efficiently. Formerly, the feature selection is completed by DCSO process to select the more relevant and important disease features from the pre-processed dataset. It is done based on features including age, sugar level, haemoglobin through best fitness values. After that Weighted Fuzzy C Means clustering (WFCM) process is applied to predict the data sample'sclass label to reduce the misclassification results. Finally, the CKD medical dataset classification is performed using ESVM algorithm. It performs training and testing process via weight based support vectors which improves the CKD classification accuracy significantly. According to the study findings, the suggested DCSO-ESVM method outperforms the current methods according to of greater accuracy, precision, recall, and f-measure.

Key words: Chronic Kidney Disease (CKD), Distribution based Chicken Swarm Optimization (DCSO) algorithm and Enhanced Support Vector Machine (ESVM) algorithm

1. INTRODUCTION

The kidneys in humans serve as the body's filter for blood, eliminating toxins. Toxins travel from the kidneys to the bladder and are subsequently eliminated from the body by urination [1] [2]. The body becomes overburdened with toxins when waste products are not removed from the circulation. Because kidney disease results in problems eliminating urine and causes blood loss to the kidneys, kidney failure can be fatal. There are two types of kidney problems: acute and chronic. Acute kidney disease is characterized by a high degree of waste product accumulation in the bloodstream, which leads to an abrupt loss of kidney function. CKD is a progressive form of kidney disease.

CKD is the result of long-term kidney damage or dysfunction. The body waste system can malfunction as a result of CKD. Complications include elevated anemia, brittle bones, inadequate nutrition might occur [3]. It is also well recognized that kidney disease raises a person's risk of blood vessel and heart problems. Hypertension, diabetes, and other conditions that progressed gradually can cause CKD. Because renal failure eventually results in dialysis or a kidney transplant, kidney disease has the potential to be lethal.

The medical sector needs tools in this day and age to help with better resource management, hospital infection control, and smarter treatment approaches. The field of medical data mining has extensively investigated several data mining jobs, with association rule mining, clustering, and categorization. As anoutcome, when deciding operational or strategic decisions, the healthcare sector must conduct in-depth analysis of vast amounts of medical data [4]. Diabetes, Parkinson's illness, liver cancer, and breast cancer are a few instances of medical or clinical data. Other data mining approaches, such as the multilayer perceptron, decision forest, radial basis functions network and logistic regression, naive Bayes, time-series analysis, and artificial neural networks, were used specifically to the CKD prediction and classification [5-7]

This research's primary goal is to classify CKD utilizing the DCSO-ESVM method. Although several studies and approaches have been presented, there is still a substantial lack of assurance regarding the accuracy of the CKD classifier. The current methods suffer from high mistake rates and imprecise categorization outcomes. The DCSO-ESVM technique is suggested to solve the issues and improve general quality of classification. Preprocessing, feature selection, clustering, and illness classification are the main accomplishments of this research. The recommended methodology yields improved outcomes for the provided CKD dataset with the implementation of efficient techniques.

The remaining content is arranged as follows: Section 2 provides a brief overview of a portion of the literature on feature selection, preprocessing, clustering, and illness classification techniques. Section 3 explains the optional DCSO-ESVM scheme. Section 4 contains the experimental data and a description of the performance assessment. Section 5 explains the conclusions.

2. Related work

In [8], Pujitha et al (2023) discussed that CKD is an extremely dangerous medical illness that, if left untreated or not identified early on, can cause a host of issues. It created a system to identify the illness by employing feature selection, data preprocessing, and machine learning methods, including LR, RF, SVM, DT, and naive bayes. To accurately anticipate the disease, the accuracy of different methods is examined and contrasted. The disease prediction system selects the one that has produced the best results. By eliminating irrelevant characteristics from the dataset and only collecting the most useful ones, it has improved its efficacy and performance.

In [9], Polat et al (2017) suggested SVM classification process. The wrapper and filter techniques, were designated to minimize the CKD dataset's dimensions to identify the condition. In the wrapper strategy, two search engines were employed: The classifier's subset evaluator utilizes the greedy stepwise search engine, whereas the wrapper subset evaluator employs the ideal First search engine. The filter approach employing an association feature selection subset evaluator with a greedy stepwise search engine and an examined subset evaluator with the Ideal First search engine. In comparison to other chosen approaches, findings demonstrated that the SVM classifier employing the filtered subset evaluator with the Greatest First search engine feature selection system had a greater accuracy rate in the identification of CKD.

In [10], Lamprea-Montealegre et al (2021) discussedthe way CKD staging and detection might assist in prioritizing individuals who would profit most from cardiovascular preventive interventions. Furthermore, the crucial role that CKD staging and diagnosis play in the start of cardiorenal preventive treatments such sodium-glucose cotransporter 2 inhibitors, that have a wealth of data supporting their ability to protect the heart. In order to address historical obstacles to CKD identification and medical care, options were covered in the conclusion.

In [11], Jerlin Rubini, L., and Eswaran Perumal (2020) suggested MKSVM and FFOA.FFOA is implemented to choose the finest features from the variety of features that are presented. The selected features from the health dataset are examined and supplied into the MKSVM for classification the medical data. MATLAB provides a simulation of the CKD classification algorithm. The unique benchmark CKD dataset utilized to test the dataset. Accuracy, specificity, positive predictive value, negative predictive value, and false negative rate are employed to estimate the efficacy of the suggested CKD classification system.

In [12], Singh et al (2022) presented a deep learning framework. Creating a deep neural network and evaluating its performance against other ML techniques. During testing, the mean of the relevant features was employed to fill in all missing values in the database. The optimal parameters of the neural network were then established by configuring it and carrying out multiple trials. RFE determined which traits were most essential. Key characteristics in the RFE included hemoglobin, specific gravity, red blood cell count, albumin,and hypertension. ML system received a subset of the features to be classified.

In [13], Suresh et al (2020) mentioned CKD is becoming more and more of an issue for developing nations. The most important known health issues, early detection is becoming more and more important for successful and long-lasting treatment. gradually, a variety of variables contribute to the ongoing deterioration in kidney function. In terms of how serious the condition is, CKD has multiple phases. Therefore, to stop further progression of CKD, a more accurate diagnosis is needed. In addition to providing an analysis of the bio-clinical phenotype of cases with high-risk CKD, this effort will continuously evaluate patients over an extended period of time, assist in the classification of CKD stages, and subsequently offer improved suggestions to stop the progression of CKD.

3. PROPOSED METHODOLOGY

Here, DCSO algorithm and ESVM process is proposed for earlier stages of CKD disease prediction. The keyrole is pre-processing, feature selection, clustering and CKD prediction. The general architecture of the suggested DCSO-ESVM method is depicted in Fig 1.

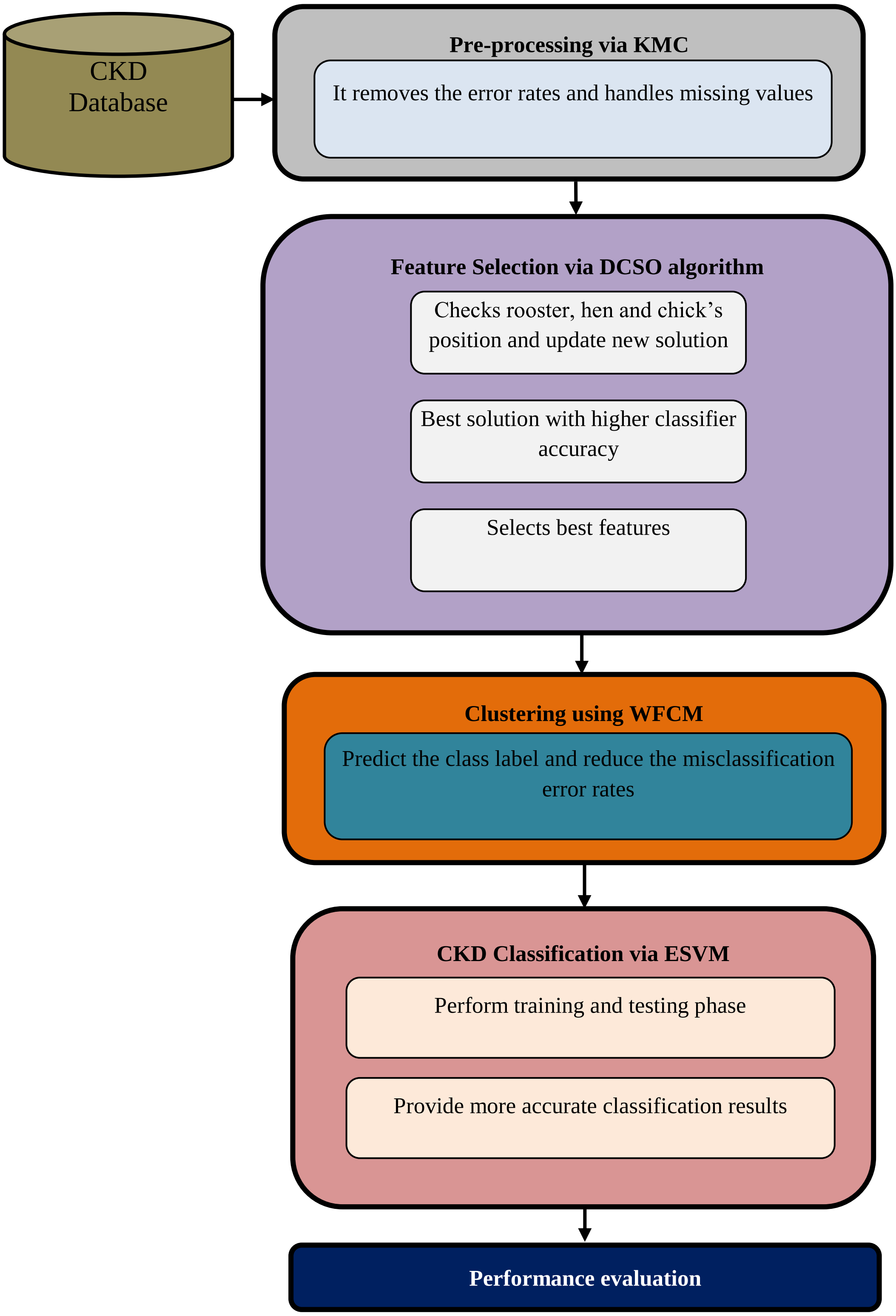

Fig 1. Overall block diagram of the proposed work

3.1 Pre-processing using K-Means Clustering (KMC) algorithm

The KMC system is employed for pre-processing to optimize the accuracy of the CKD healthcare dataset. KMC is a powerful clustering method that splits comparable data into clusters according to their original centroids [14]. The clusters'centroids are calculated utilizing the Euclidean distance concept. Assuming a random partitioning as a starting point, the algorithm iteratively determines the present cluster centers and redirects each data thing to the cluster whose center is closest to it. When there are no further reassignments, it ends. This minimizes the intra-cluster variance locally, which is the sum of squares of the variances across the cluster centers that correspond to the data characteristics.

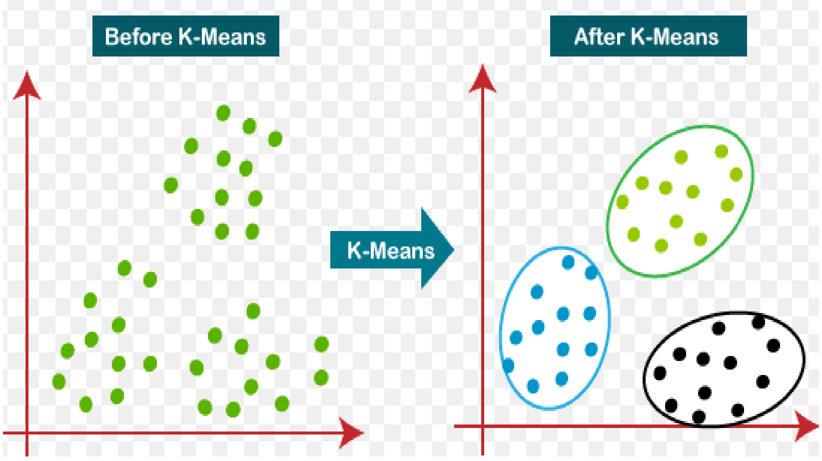

Fig 2. Example of KMC algorithm

The simplicity of implementation and linear runtime of K-means generate an effective device for data analysis. The cluster number and the quantity of classes are maintained equal in this process. Utilizing the formula provided below, establish the Euclidean distance to locate the cluster centroids.

$$d(i,j) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

Where x_i and y_i are two points in Euclidean n-space

Algorithm 1: KMC algorithm

1. Select a quantity of clusters k from CKD Dataset
2. Set cluster centers µ1,... µk
3. Assign cluster centers to k selected data points.
4. Using clusters as a mechanism, allocate points to them at random.
5. Establish cluster center, each data point is closest to, then use (1) to determine the distance measure needed to locate the missing values.
6. Insert this cluster's data point in its place.
7. Re-compute cluster centers
8. Discover and eliminatemissing value and error values
9. Once there are no further reassignments, stop.

The dataset is separated into objects from the original dataset that lack certain properties. Two sets are created from the dataset: one set has complete instances with no missing values, and the other set has incomplete cases with some missing values. To generate clusters of entire instances, KMC is utilized to the entire instances set. As a result, each instance is examined individually, and any missing variables are filled in with potential values. After applying KMC to the dataset from the resulting clusters, it is verified whether or not the newly added instance has been clustered in the right class. The process moves on to the next instance if it is in the right cluster and the assigned value becomes permanent. In the event that it is in the incorrect cluster, the instance will be moved to the correct cluster by assigning and comparing the next possible value. thereby, by applying the KMC method, the preprocessing strategy successfully improves the CKD classification accuracy.

3.2 Feature selection using Distribution based Chicken Swarm Optimization (DCSO) algorithm

Feature selection is completed by DCSO processto improve the classification accuracy. A successful classification task minimizes the dimensionality of the features space through the method of feature selection, which involves choosing features from a larger set of features. A technique for determining the key features and eliminating superfluous or unnecessary information from a dataset is feature selection.

It improves prediction efficiency, lessens the need for computation, lessens the impact of the "curse of dimensionality," and assists in interpreting data [15].

With the assistance of a novel bio-inspired method called CSO, which replicates the behaviors and hierarchical structure of a swarm of chickens, withhens, roosters, and chicks problems optimally solved by effectively extracting the intellect of the chickens. As a result, CSO is employed in feature selection to minimize the quantity of selected features while searching the feature space for the best feature combination to maximize the efficacy of classification. The method emulates the actions of each individual chicken in the swarm as well as the classifiedframework of a swarm. A chicken swarm's hierarchical order is separated into multiple groups, each has a rooster and numerous chicks and hens. The laws of motion that apply to various kinds of chickens. Chickens' social lives are significantly influenced by a hierarchical order. A group of chickens will be dominated by its stronger members. The more subservient hens and roosters who position at the outside of the group coexist with the more dominating hens who stay close to the head roosters. Fig. 3 illustrates the CSO algorithm's nature.

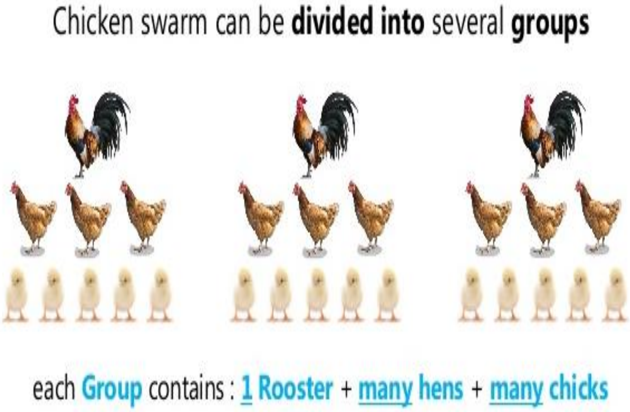

Fig 3. Nature of CSO algorithm

Chickens Movements

Rooster Movement: Greater levels of fitness enable roosters to look for food in more locations.

Hen movement: A bunch of hens look for food by trailing behind roosters. In addition, while being suppressed by the other chickens, they might additionally haphazardly steal the tasty food that they discovered. The more assertive hens would possess an edge over the more timid ones while vying for sustenance.

Chick movement: The chicks follow their mother around in an attempt to get food.

The following guidelines, which encapsulate the actions of the chickens, are the foundation of the mathematical explanation of CSO presented in [16]:

1) There are various groupings within the swarm of chickens. Several hens and chicks obey the head rooster in every group.
2) The social order is determined by the chicken’s fitness values; the roosters, who possess the highest fitness values, are the group leaders, whereas the chicks, who possess the lowest fitness values, are classified into distinct birds. The hens will replace the others.
3) Within a set, the mother-child bond, swarm hierarchy, and domination dynamic will not alter. Only many (G) time steps separate updates to these states.
4) The N virtual chickens that constitute the swarm are separated into the following groups: RN, CN,HN, and MN, which stand for the quantity of roosters, chicks, hens, and mother hens, accordingly. Positions of each person in a D-dimensional space are expressed as

$$x_{i,j}(i \in [1,\dots,N], j \in [1,\dots,D]). \quad (2)$$

In this work, chicken swarm is considered as number of autism features, classification accuracy is main objective function which is taken into ESVM classification phase. Update the solution using best objective values to increase the relevant CKD data for the given dataset.

When it comes to acquiring food, better fitness scores are preferred above lower fitness scores among the roosters. The scenario wherein higher fitness roosters are capable to look for food in a greater variation of locations than those with lower fitness values might be utilized to replicate this instance, for sheer simplicity. The normal CSO have problem with error rate hence Gaussian distribution is introduced to avoid the error rates significantly. Then the DCSO can be formulated below

$$x_{i,j}^{t+1} = x_{i,j}^{t} * (1 + \mathrm{Randn}(0, \sigma^2)) \quad (3)$$

$$\sigma^2 = \begin{cases} 1, & \text{if } f_i \le f_k \quad k \epsilon [1, N], k \ne i \\ \exp\left(\frac{(f_k - f_i)}{|f_i| + \varepsilon}\right), & \text{otherwise} \end{cases} \quad (4)$$

Where Randn $(0, \sigma^2)$ is a Gaussian distribution by a standard deviation of $\sigma^2$and mean of 0.ε is utilized to evade zero-division-error, is the least constant in the computer. k, a rooster's index is assigned at random from the set of roosters, f is the fitness value of the equivalentx.

When it comes to finding food, the hens might follow the roosters in their community. In addition, while being suppressed by the other chickens, they might additionally haphazardly steal the tasty food that they discovered. In a competition for food, the more assertive chickens might have a gain over the more timid hens. These occurrences have the following mathematical explanation.

$$x_{i,j}^{t+1} = x_{i,j}^{t} + S_1 * \mathrm{rand} * \left(x_{r1,j}^{t} - x_{i,j}^{t}\right) + S_2 * \mathrm{rand} * (x_{r2,j}^{t} - x_{i,j}^{t}) \quad (5)$$

$$S_1 = \exp((f_i - f_{r1}) / (\mathrm{abs}(f_i) + \varepsilon) \quad (6)$$

$$S_2 = \exp((f_{r2} - f_i)) \quad (7)$$

Here, a uniform random number across [0, 1] is called rand. $r2\epsilon[1, .. N]$is a ranking of the chicken which is selected at random from the swarm $r1 \ne r2$, whereas $r1\epsilon[1, .. N]$is rooster index, that is the ith hen's group-mate.

Apparently,$f_i > f_{r1}, f_i > f_{r2}$, so S2 <1< S1. Accept S1=0, Subsequently, the hen search for sustenance, trailed by the other chickens. The superior the disparity across the fitness values of the two hens, the lower the S2 value and the greater the distance across their locations. Consequently, there could be a smaller chance of the hens stealing food that other fowl found. There are competitions, which is reason the mathematical form for S1 is different from that of S2. The competitions amongst hens in a group are utilized to simulate the chicken's fitness values according to the rooster's fitness value for purposes of simplicity. In the event where S2=0, the ith hen might appear for food within its own domain. The rooster has a distinct fitness value for that particular group. Therefore, the closer S1 approximates to 1 and the narrower the distance separating the positions of the ith hen and its group-mate rooster, the lower the fitness value of the ith hen. Therefore, there's a greater chance that the dominant hens will consume the food than the obedient ones.

With the purpose of finding nourishment, the chicks travel about their mother. This is expressed as follows:

$$x_{i,j}^{t+1} = x_{i,j}^{t} + FL * \left(x_{m,j}^{t} - x_{i,j}^{t}\right) \quad (8)$$

Where $x_{m,j}^{t}$is the place of the ith chick's mother ($m\epsilon[1, N]$). FL(FLϵ(0,2)) is a variable, which the chick would go food scavenging with its mother. With each chick's FL choosing a random number among 0 and 2, consideration of their distinct variations.

Algorithm 2: DCSO

Input: a population of N chickens (CKD database)

Objective function: Best solution x_{best} (higher classifier accuracy)

Output: Optimal feature selection (age, blood pressure, sugar level, hemoglobin and so on)

```
Initialize the parameters such as $R_n$, $H_n C_n$, and $M_n$
Find the objective function f(x),   $x = (x_1, \dots, x_n)^T$
Determine the values of the N chickens' objective functions (accuracy)
While t < Maximum iteration do
    If t%G==0 then
        Create a level of hierarchy in the swarm by order the goal function values of the chickens.
        Separate the swarm into many groups, then ascertain how each group's chicks and mother hens interact
    End
    For i=1 to N features do (higher classifier accuracy)
        If i==rooster then
            Find $r_i^t$
            Update the solution using (3)
        End
        If i==hen then
            Update the solution using (5)
        End
        If i==chick then
            Update the solution utilizing (8)
```

19. End
20. Estimate the novel solution
21. Update it if the newly created solution outperforms the previous one.
22. End
23. End
24. Return x_{best} (better accuracy)

Utilizing the CKD feature data's random position for chick assignment and determination. To enhance the relevance feature utilizing CKD data, the chicken’s highest fitness value is improved. Utilizing equation (8), calculate the best values for all chickens and identify the best hens with the global best searching food technique fitness value. Utilizing a small number of carefully chosen features, the suggested approach optimizes classification accuracy by combining features utilizing the DCSO method. Finding the ideal point in the search space that optimizes the specified fitness function demands intelligence because the feature space, where each feature is characterized by a separate dimension with a span of 0 to 1, is extremely large. Maximizing accuracy in classification across the validation set with the available training data is the DCSO's fitness function.

3.3 Clustering using Weighted Fuzzy C means clustering (WFCM)

After feature selection it might need to predict the class labels to reduce the misclassification errors. This work uses WFCM clustering for class label prediction. a kind of fuzzy clustering where each data point might be a part of several clusters. Clustering is a method of allocating data points to groups so that things within the same cluster are as similar as feasible and those within separate clusters are as diverse as is practical. Similarity metrics, such as intensity, connection, and distance, are employed to identify clusters. The selection of various comparable metric depends on the application [17] [18].

Disadvantages of FCM

Some characteristics in a high dimensional signal setting should be unimportant, while others may have different clustering relevance. These features must be incorporated into the clustering process for better clustering. To mitigate these problems, WFCMclustering approach is recommended.

Weighted Fuzzy c-Means Clustering

FCMassist in data set X = $\{x_1,\ldots, x_i,\ldots x_n\}$ ($1\le i\le n$)allocating into c clusters rendering to membership degree matrix $U=(u_{ti})c*n$ when the minimal value is reached by objective function J. X's x_i is p-dimensional; u_{ti} stands for membership degree indicators that indicate which sample x_i is a member of cluster center v_t. Cluster centers $V=\{v_1,\ldots v_t,\ldots v_c\}$ $1\le t\le c$, which are initialized randomly, are employed to indicate c clusters in this instance. The calculation of membership degree u_{ij}:

$$u_{ij} = \frac{1}{\sum_{Z=1}^{c}(d_{ti}/d_{zi})^{2/(m-1)}} \tag{9}$$

here d_{ti} signifies Euclidean distance amongst sample x_i to cluster center v_t, m signifies power exponent. cluster centres in iteration:

$$v_t=\frac{\sum_{i=1}^{n} u_{ti}^{m}\, d_{ti}}{\sum_{i=1}^{n} u_{ti}^{m}} \tag{10}$$

The objective function J:

$$J=\sum_{i=1}^{n}\sum_{t=1}^{c} u_{ti}^{m}\, d_{ti}^{sy\,2} \tag{11}$$

Utilizing weighted Euclidean distance, feature-weight learning is employed. The weighted Euclidean distance is illustrated by d_{ij}^{w}, and d_{ij} is utilized Euclidean distance.

$$d_{ij}^{w} = \sqrt{\sum_{k=1}^{s} w_k\left(x_{jk}-v_{tk}\right)^2} \tag{12}$$

Algorithm 3: FCM algorithm

Step 1: Establish a threshold value and a maximum quantity of clusters (c). Let m be an appropriate constant.

Step 2: Set up centers and memberships according to FCM.

Step 3: Utilizing (6), calculate u_{ij} . Update the v_t utilizing the novel calculated u_{ij} in accordance with (10).

Step 4: Employ (11) to determine the objective function J

Step5:It stops if it converges or if the variance between two successive calculated values of the goal function J is smaller than the designated threshold. If not, proceed to step 3.

Based on the above procedure normal subjects and Kidney Disease subjects are separately by different cluster

3.4 CKD Classification using Enhanced Support Vector Machine (ESVM) algorithm

CKD classification is completed by ESVM process. Searching for a linear separation hyperplane with the maximum margin for dividing data in a higher dimensional space is the goal of this potent ML method for data categorization. The normal SVM algorithm has long training time. To overcome the issue, weight based SVM is introduced called ESVM.

In separating the various classes, the ESVM achieves a separation level which is near optimum. With the assistance of geometry and linear algebra, data that must be divided utilizing nonlinear rules might be implicitly embedded by ESVM into a high dimensional feature space. It increases the distance between either class and the hyperplane while utilizing the hyperpalne to split the greatest proportion of training data on the same class [16]. The ESVM, in general, separates the high-dimensional space by building a hyperplane. The ESVM classifier utilizes the labeled vector of expulsions of each provided sample feature as input, producing an estimate of the training data that will be utilized to dynamically categorize unseen feature displacements. The training dataset structure is employed in the testing procedure. ESVMs utilize more informative features to achieve excellent accuracy in classification for training sets.

The fundamental principle behind ESVM is to give each data point a unique weight based on the relative significance of the data points, so that different data points influence the decision surface's learning in different ways. Assuming that the weights are assigned, the training dataset turns into

$$\{(x_i, y_i, W_i)\}_{i=1}^{l} x_i \epsilon R^N, \quad y_i \epsilon \{-1, 1), \quad W_i \epsilon R \qquad (13)$$

here the scalar $0 \leq Wi \leq 1$ is the weight allotted to data point x_i

When building a cost function, the ESVM attempts to obtain strong generalization ability by minimizing the classification error and maximizing the margin of difference. Throughout the training process, all training data points are given the same weight and C is fixed. The effects of the less important data points is reduced by the ESVM balancing the penalty term. The problem of restricted optimization expressed as

$$\text{Minimize } \Phi(w) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} W_i \xi_i \qquad (14)$$

Assuming

$$y_i(\langle w, \phi(x_i) \rangle + b \geq 1 - \xi_i, \quad i = 1, \dots l \qquad (15)$$

$$\xi_i \geq 0 \ i = 1, \dots l$$

It assigns the weight W_i to the data point x_i in the above formula. The dual formula converts

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j z_i z_j K(x_i x_j) \qquad (16)$$

Assuming

The upper bounds of α_i in SVM are constrained by a constant C, whereas in ESVM, they are constrained by dynamical boundaries represented by weight values CWi.This algorithm is focused to classify the given input is CKD data or not

Algorithm 4: ESVM

1. Input: CKD database
2. Do pre-processing utilizing KMCprocess
3. Do feature selection utilizing DCSO features
4. Select more relevant and important features
5. Perform ESVM classification process
6. Do training and testing process
7. Classify the input samples into CKD or Non CKD
8. Use (15) and (16) to get accurate classification results

4. Experimental result

The studies conducted on the suggested framework are analyzed in this section. MATLAB is employed to execute the implementation. In this section evaluation of the existing NN, MKSVM, EANFIS algorithms and the proposed EBO-IANN algorithm are completed according to precision, specificity, accuracy and F-measure for the CKD Data Set (https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease). It contains 25 Attributes, 400 number of Instances.

Accuracy

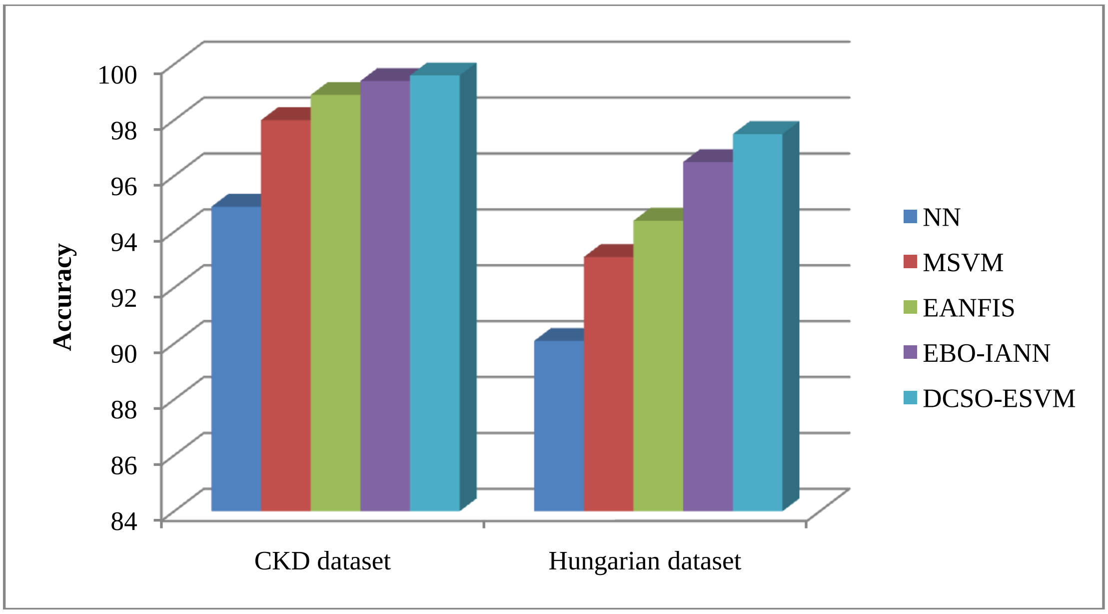

Fig 4. Accuracy

Figure 4 indicates that the accuracy of a comparable measure is assessed utilizing the current and optional approaches. The methods are represented on the x-axis, and the accuracy value is displayed on the y-axis. Whereas the DCSO-ESVM method provides improved accuracy for the supplied CKD and Hungarian medical datasets, current techniques, including naïve bayes and NN, MKSVM, EANFIS, and EBO-IANN strategies, produce lower accuracy. By choosing more pertinent data, the suggested approach increases accuracy. So, the outcome indicates that by using the best features, the suggested DCSO-ESVM method improves the CKD classification performance.

Precision

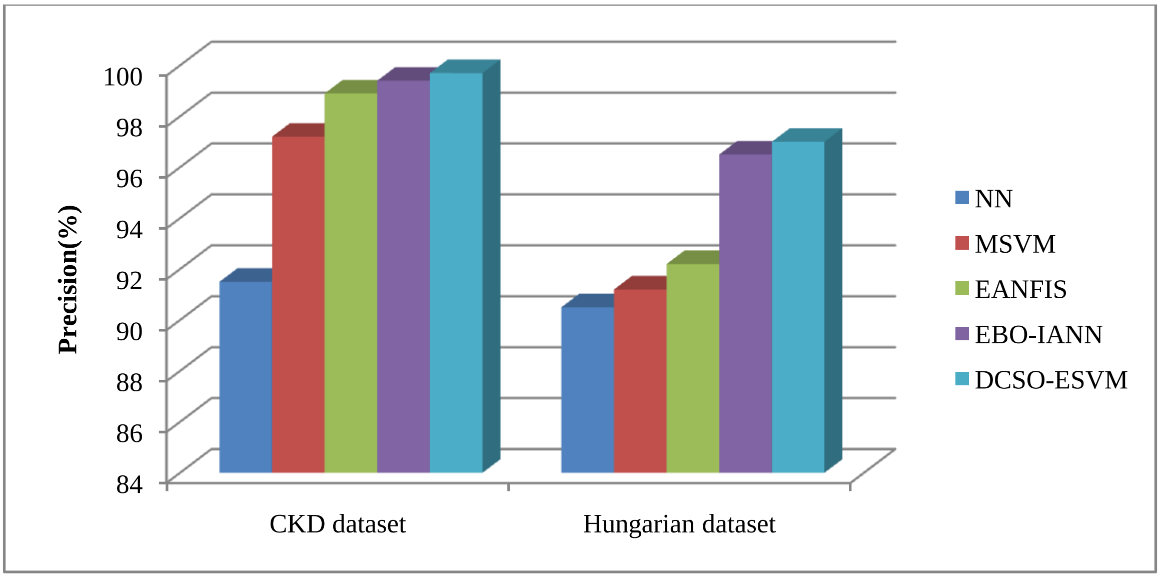

Fig 5. Precision

Figure 5 shows that the precision of an examination metric is assessed utilizing the current and optionalapproaches. The methods are represented on the x-axis, and the precision value is displayed on the y-axis. For the given CKD and Hungarian medical datasets, the DCSO-ESVM method gives more precision than the current techniques, including naïve bayes and NN, MKSVM, EANFIS, and EBO-IANN techniques. By choosing more pertinent data, the suggested approach increases precision. So, the outcome indicates that by utilizing the best features, the suggested DCSO-ESVM method improves the CKD classification performance.

Recall

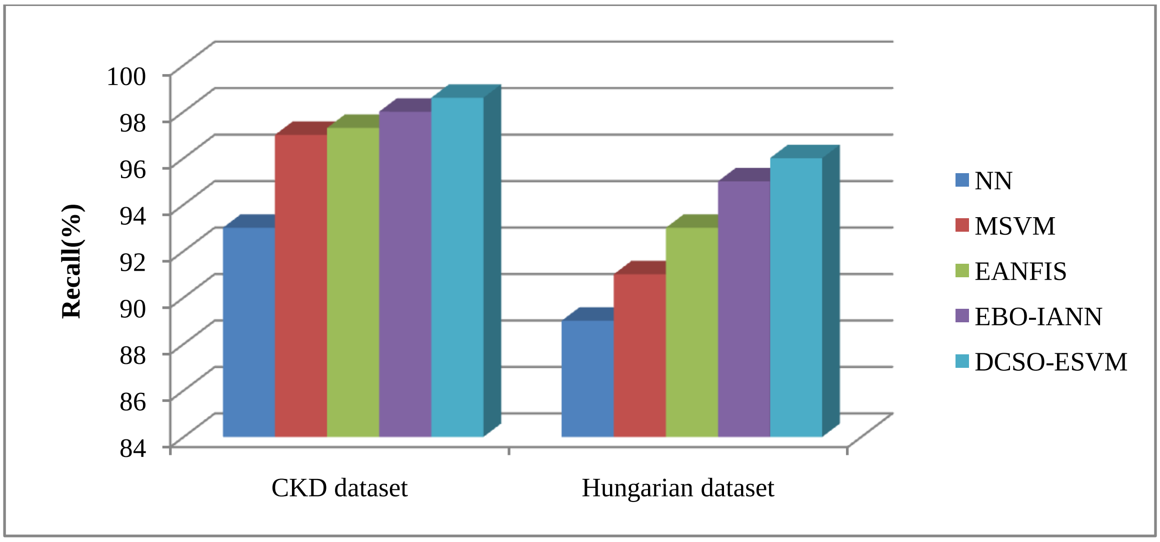

Fig 6. Recall

Figure 6 illustrates how the comparison measure is assessed according to the recollection applying the present and optionalapproaches. The techniques are represented on the x-axis, while the recall value is displayed on the y-axis. For the given CKD and Hungarian medical datasets, the DCSO-ESVM technique gives higher recall than current techniques including naïve bayes and NN, MKSVM, EANFIS, and EBO-IANN techniques. By choosing more pertinent data, the suggested approach increases precision. So, the outcome indicates that by utilizing the best features, the suggested DCSO-ESVM method improves the CKD classification performance.

F-measure

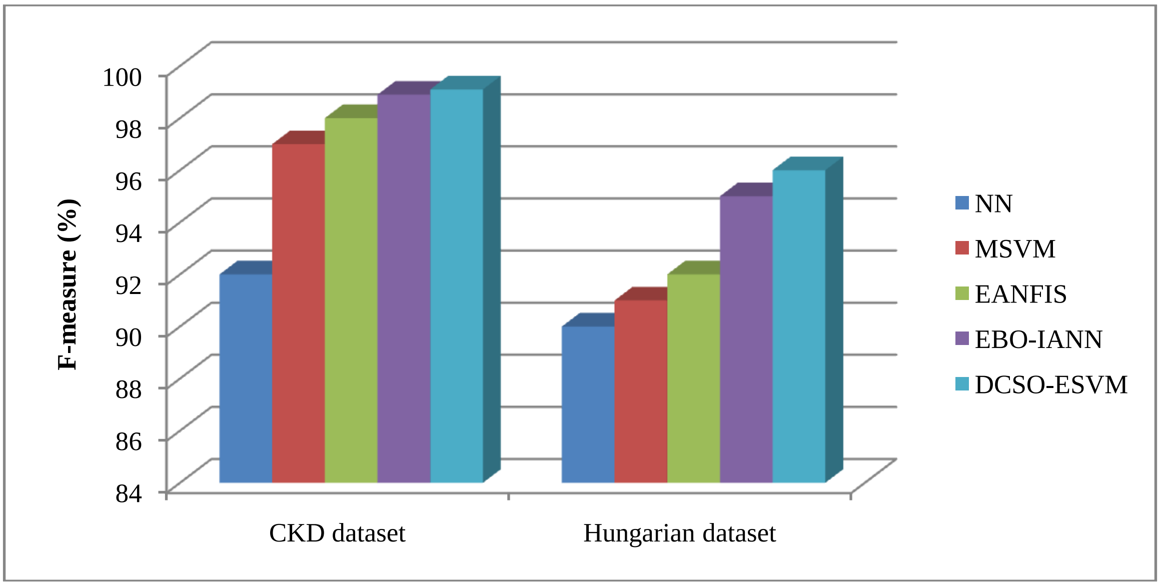

Fig 7. F-measure

The comparative values for the F-measure metric utilizing the existing and suggested algorithms are assessed from Fig 7. For the provided CKD and Hungarian medical datasets, the proposedDCSO-ESVMalgorithm yields higher F-measure compared to the existing NN, MKSVM, EANFIS and EBO-IANN approaches. The proposed classifier demonstrated that F1 score of 99.1 and 96% for CKD and Hungarian medical datasets respectively, without any incorrectly identified features. Also, the findings depict that the existing NN, MSVM, EANFIS and EBO-IANN techniques produce lower f-measure for both datasets. DCSO-ESVMprocess utilized to serve optimal features. Therefore, for the providing CKD and Hungarian medical datasets, the optional approach suggestions improved classification accuracy and greater efficiency

Error rate

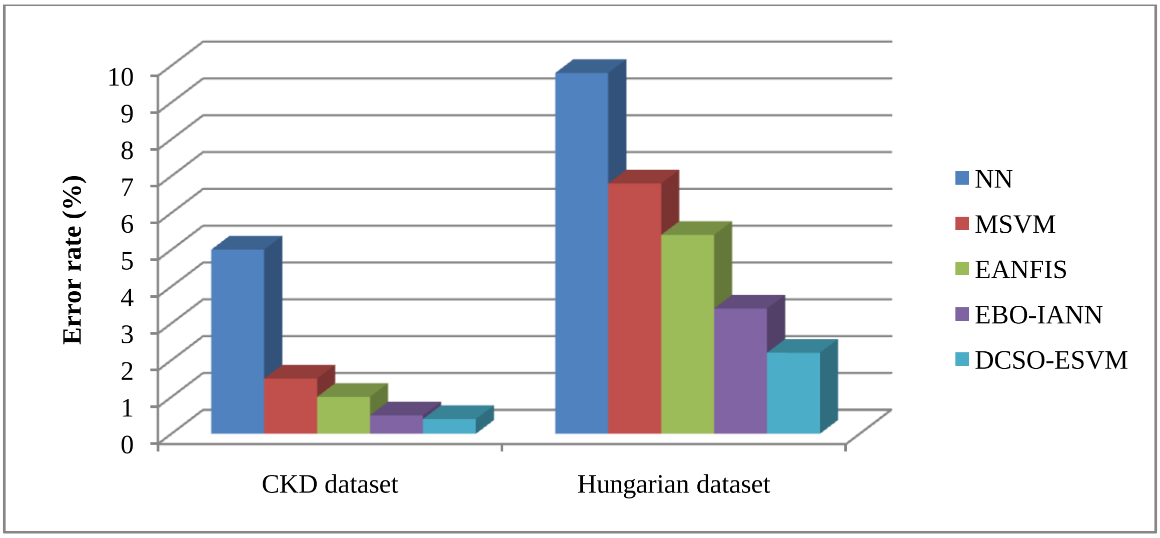

Fig 8. Error rate

The comparative values for the error rate metric utilizing the existing and suggested algorithms are assessed from Fig 8. For the provided CKD and Hungarian medical datasets, the proposed DCSO-ESVM algorithm yields lower error rate compared to the existing NN, MKSVM, EANFIS and EBO-IANN approaches. The proposed classifier demonstrated that error rate of 0.4% and 2.2% for CKD and Hungarian medical datasets respectively. Also, the findings depict that the existing NN, MSVM, EANFIS and EBO-IANN techniques produce higher error rate for both datasets. Therefore, for the provided CKD and Hungarian medical datasets, the proposed DCSO-ESVMapproach offers improved classification accuracy and greater efficiency

5. CONCLUSION

Here, DCSO-ESVMprocess is suggested to optimise the CKD medical dataset classification performance. Four primary elements comprise this research: pre-processing, feature selection, clustering, and classification. The KMC method is employed to reduce noise and fill in missing variables to improve classification accuracy. The important and beneficial features are then chosen utilizing DCSO, a feature selection method. Ultimately, the ESVM method is employed for classification, providing a more precise classification performance for the CKD dataset. From the findings, it concluded that the DCSO-ESVM system gives greater accuracy, recall, precision, f-measure and error ratein stead of the current algorithms. To improve the CKD efficiency, a neural network technique utilizing hybrid optimization may be created in the future.

REFERENCES

[1] Ahmad, Mubarik, et al. "Diagnostic decision support system of chronic kidney disease using support vector machine." 2017 second international conference on informatics and computing (ICIC). IEEE, 2017.
[2] Ghosh, Pronab, et al. "Optimization of prediction method of chronic kidney disease using machine learning algorithm." 2020 15th international joint symposium on artificial intelligence and natural language processing (iSAI-NLP). IEEE, 2020.
[3] Elhoseny, Mohamed, K. Shankar, and J. Uthayakumar. "Intelligent diagnostic prediction and classification system for chronic kidney disease." Scientific reports 9.1 (2019): 9583.
[4] Wang, Weilun, Goutam Chakraborty, and Basabi Chakraborty. "Predicting the risk of chronic kidney disease (ckd) using machine learning algorithm." Applied Sciences 11.1 (2020): 202.

[5] Nelson, Robert G., et al. "Development of risk prediction equations for incident chronic kidney disease." Jama 322.21 (2019): 2104-2114.

[6] Snegha, J., et al. "Chronic kidney disease prediction using data mining." 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). IEEE, 2020.

[7] Turiac, Andreea S., and Małgorzata Zdrodowska. "Data mining approach in diagnosis and treatment of chronic kidney disease." acta mechanica et automatica 16.3 (2022): 180-188.

[8] Pujitha, K., et al. "Chronic Kidney Disease Detection Using Machine Learning Approach." 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN). IEEE, 2023

[9] Polat, Huseyin, Homay Danaei Mehr, and Aydin Cetin. "Diagnosis of chronic kidney disease based on support vector machine by feature selection methods." Journal of medical systems 41 (2017): 1-11.

[10] Lamprea-Montealegre, Julio Alejandro, Michael G. Shlipak, and Michelle M. Estrella. "Chronic kidney disease detection, staging and treatment in cardiovascular disease prevention." Heart (2021).

[11] Jerlin Rubini, L., and Eswaran Perumal. "Efficient classification of chronic kidney disease by using multi-kernel support vector machine and fruit fly optimization algorithm." International Journal of Imaging Systems and Technology 30.3 (2020): 660-673.

[12] Singh, Vijendra, Vijayan K. Asari, and Rajkumar Rajasekaran. "A deep neural network for early detection and prediction of chronic kidney disease." Diagnostics 12.1 (2022): 116.

[13] Suresh, Chalumuru, et al. "A neural network based model for predicting chronic kidney diseases." 2020 second international conference on inventive research in computing applications (ICIRCA). IEEE, 2020.

[14] Mohamad, Ismail Bin, and Dauda Usman. "Standardization and its effects on K-means clustering algorithm." Research Journal of Applied Sciences, Engineering and Technology6.17 (2013): 3299-3303

[15] Hamdi, Monia, et al. "Chicken Swarm-Based Feature Subset Selection with Optimal Machine Learning Enabled Data Mining Approach." Applied Sciences 12.13 (2022): 6787.

[16] Wu, Dinghui, Shipeng Xu, and Fei Kong. "Convergence analysis and improvement of the chicken swarm optimization algorithm." IEEE Access 4 (2016): 9400-9412

[17] Nayak, J., Naik, B. and Behera, H., 2015. Fuzzy C-means (FCM) clustering algorithm: a decade review from 2000 to 2014. Computational intelligence in data mining-volume 2, pp.133-149.

[18] Qin, J., Fu, W., Gao, H. and Zheng, W.X., 2016. Distributed $ k $-means algorithm and fuzzy $ c $-means algorithm for sensor networks based on multiagent consensus theory. IEEE transactions on cybernetics, 47(3), pp.772-78

[19] Razzaghi, T., Roderick, O., Safro, I., & Marko, N. (2016). Multilevel weighted support vector machine for classification on healthcare data with missing values. PloS one, 11(5), e0155119.