Coronary Artery Disease Classification Using Deep Neural Network and Ensemble Models Optimized by Particle Swarm Optimization

Pratibha Verma, Dr. C.V. Raman University, Bilaspur, India
Vineet Kumar Awasthi, Dr. C.V. Raman University, Bilaspur, India
Sanat Kumar Sahu, Govt. K.P.G. College, Jagdalpur, India*
Akhilesh Kumar Shrivas, Guru Ghasidas Vishwavidyalaya, Bilaspur, India

ABSTRACT

Many people are suffering from several health-related issues of which coronary artery disease (CAD) is an important one. Identification, prevention, and diagnosis of disease is a very challenging task in the field of medical science. This paper proposes a new feature optimization technique known as PSO-Ensemble1 to reduce the number of features from CAD datasets. The proposed model is based on particle swarm optimization (PSO) with Ensemble1 classifier as the objective function and is compared with other optimization techniques like PSO-CFSE and PSO-J48 with two benchmark CAD datasets. The main objective of this research work is to classify CAD with the proposed PSO-Ensemble1 model using the ensemble technique.

KEYWORDS

Coronary Artery Disease (CAD), Deep Neural Network (DNN), Ensemble Model, Naive Bayes (NB), Naive Bayes and Decision Tree (NBTree), Particle Swarm Optimization (PSO), Random Forest (RF)

INTRODUCTION

Coronary Artery Disease (CAD) is developed by the formation of plaques inside the walls of coronary arteries, resulting in the narrowing of lumens of coronary arteries (Pal & Chakraborty, 2011). CAD is a serious health issue in human life. Today, a variety of medical technology available in healthcare industries provide improved diagnosis of heart and coronary diseases. But, in some cases, the diagnosis is beyond the scope of common man. The healthcare industry nowadays produces a huge amount of unreadable and complex data related to patients, hospital resources, disease diagnosis, electronic patient records, medical devices, etc. This data is the main resource to be processed and analysed for knowledge extraction and used as guidance for decision-making and cost-savings (El-bialy, Salamay, Karam, & Khalifa, 2015).
In this study, we have created a model based on the Data Mining and Machine Learning (ML) techniques that can handle the problem of CAD classification. Data mining is the method of finding previously unknown patterns and developments in databases and making analytical models with that information. In healthcare, data mining is an area of high significance and has become more effective and essential (Han, Kamber, & Pei, 2012; Pujari, 2013). A lot of techniques have been introduced in the field of data mining and ML classification task; but it is more complicated to enhance the performance of an individual classifier significantly. Nowadays researchers are taking interest in combining many classifiers to achieve improved performance. One such method is the Ensemble, also called the amalgamation method. In earlier researches, ensembles have been verified theoretically and empirically for performing precisely than any individual classifier (Chen, Wong, & Li, 2014). To make a perfect ensemble to reach the expected results, two necessary elements have to be considered carefully. The first is to initiate sufficient variety into the elements of an ensemble. The second is to decide an appropriate combining technique to combine the various outputs to a single output (Polikar, 2006). Variety is the foundation of an ensemble. In this study, we have used a voting based ensemble technique with two combination rules (Average of Probabilities and Majority Voting). The classifiers namely NB, RF and NBTree were used to create eight ensemble models (Ensemble1 to Ensemble8) and one extra DNN has been used for comparing with other classifiers. We have proposed a new feature optimization technique PSO-Ensemble1 model to optimize the features from CAD dataset. The main objective of this research work is to develop a computationally efficient and robust model using ensemble classifier with the proposed new PSO-Ensemble1 model which gives a better accuracy compared with other existing classifiers. The novelty of this research work is to optimize the feature subset using PSO with base ensemble classifier that is PSO-Ensemble1 model. This paper contributes in the following three ways:

1) Examines the performance of the individuals and ensemble classifiers in terms of accuracy, sensitivity, specificity and F1-score.
2) Use the PSO in different configurations and propose PSO-Ensemble1 model for feature optimization.
3) Compare and analyse the performance of individuals and ensemble classifiers with optimized features in terms of accuracy, sensitivity, specificity and F1-Score.

Many authors have used several feature selection techniques and ensemble classifiers to achieve better classification accuracy. Kumari et al. (2017) proposed a Particle Swarm Optimization (PSO) based Ensemble classification model with high dimensional microarray datasets. The experimental outcomes showed that the proposed model performed better compared to other conventional feature selection based classification models. Aburomman et al. (2016) proposed the Weighted Majority Voting (WMV) to merge results from different classifiers. They combined the classifiers’ opinions using the PSO, meta-optimized PSO and the WMA approaches and their finding showed that the proposed model outperformed WMA in terms of classification accuracy. Hedeshi et al. (2014) proposed a PSO algorithm with a boosting approach to extract rules for recognizing the presence or absence of CAD in a patient. The outcomes showed that the model could identify CAD with acceptable accuracy. Various authors have worked on other feature selection and optimization techniques to reduce the irrelevant features from datasets and improve the performance of models. Nilashia et al. (2018) used dimensionality reduction method and Self-Organizing Map method for clustering tasks and ensembles of Neuro-Fuzzy Inference System for predicting the hepatitis disease. They also used the decision trees for selecting the significant features from dataset. The proposed method performed better than the intelligent learning system for hepatitis disease diagnosis. Ledezma et al. (2010) proposed the GA-Stacking model using excellent stacking configuration and genetic search. They also used GA parameters and GA-stacking which have meta-level and base classifiers. The GA-Stacking was comparable to the best results obtained whereas Ordóñez et al. (2008) proposed GA-Ensemble as an
extension to their earlier techniques. The GA-Ensemble was related to a Genetic Algorithm (GA) in searching the configuration based on various datasets exclusive of apriori hypothesis. GA-Ensemble inherited from GA-Stacking is its flexibility and extensibility. Wang et al. (2004) proposed an enhanced Ant Colony Optimization (ACO)-Miner. The findings revealed that ACO-Miner had a considerably better accuracy and a much reduced set of rules than Ant-Miner. Chen et al. (2014) proposed an ensemble structure technique that applies ACO to the stacking ensemble structure procedure to create domain-specific configurations. They conducted a lot of experiments to compare the proposed ACO-Stacking approach with some well-known ensemble methods on 18 benchmark datasets. The results showed that the new approach could generate better stacking ensembles. The above literature shows that our proposed feature optimization PSO-Ensemble gives a better accuracy compared with other PSO based feature optimization techniques. The limitation of this model is that the computational time is high because of taking ensemble classifier as the base classifier with PSO. This research work has used ensemble techniques to achieve better classification accuracy with CAD. Many authors have also worked on various ensemble models to achieve better classification accuracy. Alizadehsani et al. (2016) proposed the SMO, Naive Bayes and an ensemble algorithm to conduct the analyses for CAD. The results of the different algorithms were calculated on the dataset using tenfold cross-validation. The suggested ensemble model gave up to 88.5% accuracy. Kabari et al. (2019) used two kinds of ensemble learning namely bagging and voting, and combined many classifiers to achieve enhanced classification accuracy. They compared the obtained results and showed that bagging is an effective ensemble learning method. Hamze-Ziabari et al. (2018) proposed four different predictive techniques namely CART, Ensemble CART, Ensemble M5 and Ensemble of M5 and CART to evaluate the Peak Ground Velocity, Peak Ground Displacement and Peak Ground Acceleration. The result indicates that Ensemble M5, and Ensemble of M5 and CART appreciably outperformed the CART and Ensemble CART models. Pandey et al. (2014) worked on five ensemble techniques based on four delegated learning methods namely Bagging, Adaboost, Rotation Forest and Random Forest. The outcome revealed that the Rotation Forest algorithm performed better compared with the Random Forest algorithm for student performance prediction. Gandhi et al. (2015) used a hybrid ensemble classifier that combined the Instance-based learner Decision Tree and Naive Bayes Tree algorithms with voting scheme configuration. They used the ensemble model on the 28 benchmark dataset and compared the ensemble model with the Naive Bayes, Rule Learner, Decision Tree, Bagging and Boosting Algorithms. The finding reveals that the proposed method performed better than the single classifiers. In this research work, we have used individual classifiers to compare the performance of ensemble classifier with our proposed PSO-Ensemble feature optimization technique. Many authors have worked on different individual classifiers with different healthcare datasets and compared the performance of classifiers. Sharma et al. (2020) proposed the Deep Neural Network and compared it with KNN, SVM, Naive Bayes and Random Forest for classification of heart disease. The results showed that Hyper-parameter optimization (Talos) gave the highest accuracy of 90.78% as compared to other models. Amutha et al. (2018) worked on the data mining approach forecast of the results of TMT. They used the ensemble method and an adjusted variant of K-Nearest Neighbor (KNN) method, K-Sorting and Searching (KSS). The outcome revealed that the TMT predicted with 84% accuracy. Kim et al. (2017) worked on the prediction of Coronary Heart Disease (CHD) using Neural Network (NN) and NN-based Feature Correlation Analysis (NN-FCA). They found that NN-FCA gave a better accuracy for predicting the risk of CHD. Bektas et al. (2017) proposed different models like Logistic regression RNA, NN, FST like Relief-F and Independent t-test analysis with Cardiovascular dataset. The outcomes showed 84.1% accuracy with NN and that the Relief-f attribute selection technique performed better than other techniques. Lohita et al. (2015) have suggested classification techniques like J48, Random Forest, Bagging, REP Tree, Naive Bayes, CART and Decision Stump for classification of CHD. The outcomes showed that the Bagging algorithm achieved the maximum accuracy compared to other algorithms. Alizadehsani et al. (2013) proposed the Naive Bayes (NB), K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Artificial Neural Network
(ANN) techniques. The results showed that SVM and ANN performed better for the prediction of CAD. Alizadehsani et al. (2013) suggested various classification techniques like C 4.5, Naive Bayes, Sequential Minimal Optimization (SMO) and K-Nearest Neighbor (KNN) for the classification of CAD. The outcome showed that the SMO model provided incredibly high sensitivity (97.22%) and accuracy (92.09%).

PROPOSED FRAMEWORK

The proposed framework is used for feature selection and classification of CAD dataset.

Figure 1. Shows the Proposed Framework
DESCRIPTION OF THE DATASETS

The experimental dataset is available in the UC Irvine Machine Learning Repository. The datasets Z-Alizadeh Sani (ZADS) and Extension of Z-Alizadeh Sani (EZADS) hold the records of 303 patients, each of which has 56 features (55 features with 1 class level) and 59 features (58 features with 1 class level) (Sani, Alizadehsani, & Roshanzami, 2017; “Z-Alizadeh Sani dataset” 2016) respectively.

Table 1. CAD Datasets Features Details

| Dataset Name | Features Details | Feature Sets |
|--------------|------------------|-------------|
| ZADS         | $f_1=\text{Age}, f_2=\text{Weight}, f_3=\text{Length}, f_4=\text{Sex}, f_5=\text{BMI}, f_6=\text{DM}, f_7=\text{HTN}, f_8=\text{Current Smoker}, f_9=\text{EX-Smoker}, f_{10}=\text{FH}, f_{11}=\text{Obesity}, f_{12}=\text{CRF}, f_{13}=\text{CVA}, f_{14}=\text{Airway disease}, f_{15}=\text{Thyroid Disease}, f_{16}=\text{CHF}, f_{17}=\text{DLP}, f_{18}=\text{BP}, f_{19}=\text{PR}, f_{20}=\text{Edema}, f_{21}=\text{Weak Peripheral Pulse}, f_{22}=\text{Systolic Murmur}, f_{23}=\text{Diastolic Murmur}, f_{24}=\text{Typical Chest Pain}, f_{25}=\text{Dyspnea}, f_{26}=\text{Function Class}, f_{27}=\text{Atypical}, f_{28}=\text{Nonanaginal}, f_{29}=\text{Exertional CP}, f_{30}=\text{Low TH Ang}, f_{31}=\text{ST Elevation}, f_{32}=\text{Depression}, f_{33}=\text{Tension}, f_{34}=\text{LVH}, f_{35}=\text{Poor R Progression}, f_{36}=\text{BBB}, f_{37}=\text{FBS}, f_{38}=\text{CR}, f_{39}=\text{TG}, f_{40}=\text{LDL}, f_{41}=\text{HDL}, f_{42}=\text{BUN}, f_{43}=\text{ESR}, f_{44}=\text{K}, f_{45}=\text{Na}, f_{46}=\text{WBC}, f_{47}=\text{Lymph}, f_{48}=\text{Neut}, f_{49}=\text{PLT}, f_{50}=\text{EF-TTE}, f_{51}=\text{Region RWMA}, f_{52}=\text{VHD}, \text{(class=Cath/Normal Details)}$ | Fs-55 |
| EZADS        | $f_1=\text{Age}, f_2=\text{Weight}, f_3=\text{Length}, f_4=\text{Sex}, f_5=\text{BMI}, f_6=\text{DM}, f_7=\text{HTN}, f_8=\text{Current Smoker}, f_9=\text{EX-Smoker}, f_{10}=\text{FH}, f_{11}=\text{Obesity}, f_{12}=\text{CRF}, f_{13}=\text{CVA}, f_{14}=\text{Airway disease}, f_{15}=\text{Thyroid Disease}, f_{16}=\text{CHF}, f_{17}=\text{DLP}, f_{18}=\text{BP}, f_{19}=\text{PR}, f_{20}=\text{Edema}, f_{21}=\text{Weak Peripheral Pulse}, f_{22}=\text{Systolic Murmur}, f_{23}=\text{Diastolic Murmur}, f_{24}=\text{Typical Chest Pain}, f_{25}=\text{Dyspnea}, f_{26}=\text{Function Class}, f_{27}=\text{Atypical}, f_{28}=\text{Nonanaginal}, f_{29}=\text{Exertional CP}, f_{30}=\text{Low TH Ang}, f_{31}=\text{ST Elevation}, f_{32}=\text{Depression}, f_{33}=\text{Tension}, f_{34}=\text{LVH}, f_{35}=\text{Poor R Progression}, f_{36}=\text{BBB}, f_{37}=\text{FBS}, f_{38}=\text{CR}, f_{39}=\text{TG}, f_{40}=\text{LDL}, f_{41}=\text{HDL}, f_{42}=\text{BUN}, f_{43}=\text{ESR}, f_{44}=\text{K}, f_{45}=\text{Na}, f_{46}=\text{WBC}, f_{47}=\text{Lymph}, f_{48}=\text{Neut}, f_{49}=\text{PLT}, f_{50}=\text{EF-TTE}, f_{51}=\text{Region RWMA}, f_{52}=\text{VHD}, f_{53}=\text{LAD}, f_{54}=\text{LCX}, f_{55}=\text{RCA}, \text{(class=Cath/Normal Details)}$ | Fs-58 |

Features Optimization Technique

The feature optimization (selection) technique (FOT/FST) is an important data pre-processing tool in data mining. It has been a dynamic area of research and development in the previous years. As datasets are becoming larger both in terms of instance and feature count in the area of healthcare medical research, this causes scalability and performance issues in learning a lot of algorithms. Feature selection solves the scalability problem and enhances the performance of classification techniques by removing unrelated, redundant or noisy features from high dimensional datasets (Karthika & Divakar, 2014; Karthika & Prakash, 2014). Feature selection is a method of choosing a subset of significant features by applying certain estimation criteria (Tran, Xue, & Zhang, 2014). The feature selection procedure consists of three steps. It begins with choosing a subset of the original attributes and evaluating the value of each feature in the subset. Secondly, using this estimation, several features in the subset can be eliminated or limited to existing subsets. Third, it examines whether the final subset is good enough using some evaluation criterion (Hall, 1999). The procedure of feature (attribute) selection can be supervised, unsupervised and semi-supervised foundation on class labels. In the selection of a supervised feature, the estimation of the features is decided to utilize their similarity to the class similarity. For semi-supervised, we use limited details on the label to enhance the collection of unsupervised feature selection. They are filters, wrappers and hybrids. Filters evaluate the value of a feature without any learning algorithms. Wrappers have a predefined learning algorithm for determining a feature’s merit within a subset. Hybrid is a mixture of filters and
wrappers. We have used the meta-heuristic search method namely the Particle Swarm Optimization technique for feature optimization. It is a wrapper based FST which evaluates and selects attributes based on accuracy estimates by the target learning classifiers or induction algorithms (Dag, Sayin, Yenidogan, Albayrak, & Acar, 2012; Yildirim, 2015). Another form of the filter based FST is a filter with CFSE (Correlation-based Feature Selection) as a fitness function.

**Proposed Particle Swarm Optimization (PSO)**

In data mining and machine learning, a meta-heuristic is a high-level approach designed to fix, generate or select a heuristic, specifically providing an appropriate solution to an optimization problem; but it has limited computing capability (Bianchi, Dorigo, Maria, & Gutjahr, 2009). PSO is meta-heuristic because it searches for very large locations of candidate solutions (Blum & Roli, 2001). PSO is a population-based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995 and inspired by bird flocking or fish schooling social behaviour (Kennedy & Eberhart, 1995; Shi & Eberhart, 1998). These swarms conform to a cooperative way to find food, and each member in the swarm continues shifting the search pattern according to the learning experiences of its own and other members (D. Wang, Tan, & Liu, 2018).

In this paper, we have used three different ways for feature/attribute selection namely proposed PSO-Ensemble1, PSO-J48 and PSO-CFSE.

A. Proposed PSO-Ensemble1

The proposed Wrapper based FST is the PSO with objective function or learning classifiers as (Ensemble of NB and RF or NB+RF) or Ensemble1 to select the relevant features from CAD datasets. In this study, the wrapper based FST estimates the feature sets by using an objective function as ensemble learning classifier. (Ian H. Witten & Frank, 2004). The objective function of our proposed ensemble learning classifier combines two classifiers namely NB and RF. The voting (Bach, Werner, Zywiec& Pluskiewicz, 2017) ensemble technique with the average of probability combining rule is used. The voting is a cumulative method that merges the results of multiple classifiers. The 5-fold cross-validation is used to estimate the accuracy of the learning scheme for a set of features. We call this FST as PSO-Ensemble1 or PSO-Ensemble (NB+RF). In PSO, a candidate solution is represented by a particle. A population of particles move ("fly") together in the search space to find the optimal solutions. During the movement, each particle has a position shown by \( x_i = (x_{i1}, x_{i2}, \ldots x_{iD}) \) and a velocity shown by \( v_i = (v_{i1}, v_{i2}, \ldots, v_{iD}) \), where D is the dimensionality of the search space. During the search process, each particle can remember its best position visited so far called personal best (denoted by \( p\text{best} \)), and the best previous position visited so far by the whole swarm called global best (denoted by \( g\text{best} \)). Based on \( p\text{best} \) and \( g\text{best} \), PSO iteratively updates \( x_i \) and \( v_i \) of each particle to equation (1) and (2).

\[
\begin{align*}
  v_{id}^{t+1} & = w * v_{id}^t + c1 * r1(p_{id} - x_{id}') + c2 * r2(p_{gd} - x_{id}'), \\
  x_{id}^{t+1} & = x_{id}^t + v_{id}^{t+1}
\end{align*}
\]  

Where \( v_{id}^{t+1} \) shows the velocity of particle \( i \) in the \( d \)th dimension at the \((t + 1)\)th iteration, \( x_{id}^{t+1} \) shows the position value of particle \( I \), \( w \) is the inertia weight reflecting the influence of the previous velocity, \( c_1 \) and \( c_2 \) are acceleration constants, \( r_{i1}, r_{i2}, \) and \( rand() \) are random values, which are uniformly distributed in \([0, 1]\) and \( p_{id} \) and \( p_{gd} \) show the values of \( p\text{best} \) and \( g\text{best} \) in the \( d \)th dimension respectively. The pseudo code of the proposed PSO-Ensemble1 algorithm is given below:
B. PSO-J48: In this wrapper based FST namely PSO with objective function as J48 (C4.5), classifiers are used to select the relevant features from the original feature space of CAD datasets. The Wrapper based FST estimates the feature sets by using an objective function as learning classifier. (Ian H. Witten & Frank, 2004). The 5-fold cross-validation is used to estimate the accuracy of the learning scheme for a set of features. We call this FST as PSO-J48.

C. PSO-CFSE: The filter based FST namely PSO, with CFSE (Correlation-based Feature Selection) as a fitness function is used to select relevant features from original feature space of CAD datasets. The worth of a subset of attributes is assessed by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with the class while having low inter correlation are preferred(Ian H. Witten & Frank, 2004). We call this FST as PSO-CFSE.

Proposed Technique for Classification

Classification techniques follow the principle of supervised learning techniques. They assist in predicting the various classes of labels (Han et al., 2012; Haykin, 2008). The data classification process is carried out in two stages. In the first stage, the classifier is constructed. The ethics of the training algorithm helps to create classifiers using classification algorithms. The second stage is the processing of classifiers created in the first stage. The classifier is used to estimate the value of unseen data during the stages of the use of the classifier. We use algorithms NB, RF, NBTree, Ensemble

Algorithm 1: Proposed PSO-Ensmble1 Algorithm

| 1. Start                                                                 |
| 2. Initialize the position and velocity of the particles               |
| 3. Determine stopping criteria                                         |
| 4. PSO Search: Performs a search                                       |
| PSO Search ()                                                         |
| {}                                                                     |
| 5. Assess each particle's fitness using the objective function        |
| Objective function ()                                                  |
| {}                                                                     |
| 6. Update the individual/global best positions                        |
| 7. Update position/velocity of each particle based on equation (1) and (2) |
| 8. End                                                                 |

Algorithm 2: Proposed Vote Ensemble Model

Set Seed, Set Number Cross folds
Set Classifier as Ensemble (NB+RF) Call
Set IR Class Value, Set Do not Check Capabilities, Set Evaluation Measure
Set Threshold

6. Update the individual/global best positions
7. Update position/velocity of each particle based on equation (1) and (2)
models (Ensemble1, Ensemble2, Ensemble3, Ensemble4, Ensemble5, Ensemble6, Ensemble7 and Ensemble8) and DNN to classify the CAD with original and reduced feature sets.

**Naive Bayes**

In data mining and machine learning, Naive Bayes (NB) classifiers are a family of simple “probabilistic classifiers” based on applying Bayes’ theorem with strong independence assumptions between the features. They are among the simplest Bayesian network models (Haykin, 2008; Lantz, 2013). The Naive Bayes methods have been named so because they create a pair of “naive” perception about the data. Naive Bayes classifier is predominantly flexible; it has needed various parameters linear in the quantity of feature (character/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, instead of by costly iterative estimation as done for some different sorts of classifiers (Chakraborty, Ghosh & Saha, 2020; Solankar, Jagatap, Rupnawar, Shitole & Kumbhar, 2015). Given a set of \( N \) samples \( D = \{ D_1, D_2, \ldots, D_n \} \) (training dataset) where every sample \( D \) is represented as an \( n \)-dimensional vector \( \{ X_1, X_2, \ldots, X_n \} \) value correspond to the attribute \( A_1, A_2, \ldots, A_n \) respectively, provided there are \( K \) classes \( C_1, C_2, \ldots, C_K \) where every sample belong to any of these classes. If there is an additional sample \( D \) (unseen sample), the class for \( D \) can be predicted using the class with the highest posterior probability. The basic relation of Bayes Theorem is given as

\[
P(C/A) = \frac{P(A/C) \cdot P(C)}{P(A)}
\]  

(3)

As \( P(A) \) is a constant of all class, only the product \( P(A/C) \cdot P(C) \) will be maximized. Using the assumption of conditional independence, the estimate class is given as:

\[
\text{Estimate} = \arg \max_{C} P(C) \cdot \prod_{i=1}^{n} P(A_i)
\]  

(4)

Where \( A \) = attribute value of the dataset sample, \( C \) = class

**Random Forest**

The decision tree is a good classifier because of its high execution speed. Random Forest (RF) or Random Decision Forest (RDF) is an ensemble learning method for regression and classification tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees (HO, 1998; Ho, 1995).

**Naive Bayes Tree (NBTree)**

The NBTree or Decision Tree with Naive Bayes is a hybrid algorithm. It is the combination of both Decision Tree and Naïve Bayes. NBTree which deploys an NB classifier on each leaf node of the built decision tree has demonstrated remarkable classification performance in terms of classification accuracy (Kohavi, 1996; S. Wang, Jiang, & Li, 2015). NBTree is also ideal to work on large datasets. Naive Bayes Tree is superior to both the decision tree and the Naive Bayes classifier.

**Deep Neural Network (DNN)**

Deep learning is a sub-area of Machine Learning (ML) that produces a multi-layered representation of data, usually makes use of ANN and has enhanced the robustness in a variety of ML tasks.
It permits computational models that can be composed of several processing layers to learn representations of data with more than one level of abstraction. It discovers complex configurations in massive information sets by the usage of the backpropagation algorithm to specify how a system should change its interior parameters which are used to compute the representation in each layer from the illustration inside the preceding layer (Bengio, Courville, & Vincent, 2013; Lecun, Bengio, & Hinton, 2015; Schmidhuber, 2015). DNN compose computations performed by many layers. Denoting the output of hidden layers by $h^{(l)}(x)$, the computation for a network with $L$ hidden layers is:

$$f(x)=f[a^{(l+1)}(h^{(l)}(a^{(l)}(\ldots (h^{(2)}(a^{(2)}(h^{(1)}(a^{(1)}(x)))))))]].$$  \hfill (5)

Each pre-activation function $a^{(l)}(x)$ is typically a linear operation with matrix $W^{(l)}$ and bias $b^{(l)}$, which can be combined into a parameter $\theta$:

$$a^{(l)}(x)=W^{(l)}x+b \quad (6)$$
$$a^{(l)}(\hat{x})=\theta^{(l)} \hat{x}, \quad l=1 \quad (7)$$
$$a^{(l)}(\hat{h}^{(l-1)})=\theta^{(l)} \hat{h}^{(l-1)}, \quad l>1 \quad (8)$$

The “hat” notation $\hat{x}$ indicates that 1 has been appended to the vector $x$. Hidden-layer activation functions $h^{(l)}(x)$ often have the same form at each level, but this is not a requirement.

Activation functions generally operate on the reactivation vectors in an element-wise fashion. We use softmax activation instead of sigmoid with the cross-entropy loss because softmax activation distributes the probability throughout each output node.

$$\sigma(z)^i = \frac{e^{Z_i}}{\sum_{j=1}^{k} e^{Z_j}} \quad (9)$$

**Ensemble Methods**

An ensemble learning technique combines multiple classifiers and gives the final prediction (classification results). The ensemble model combines a set of classifiers to produce a single compound model that offers higher accuracy (Gandhi & Pandey, 2015; Pandey & Taruna, 2014). Ensemble methods can be termed as a committee, classifier fusion, combination or aggregation, etc. Previous studies show that classification from a composite model gives better results compared to single classification performance. Ensemble methods can be classified as Homogeneous Ensemble Methods (Breiman, 1996) and Heterogeneous Ensemble Methods (Dietterich, 2000). Homogenous ensemble techniques use a single learning algorithm on different training datasets to construct multiple classifiers such as Bagging, Boosting, Random Subspaces and Random Forest, while Heterogenous ensemble Methods use a variety of learning algorithms and manipulate training datasets to make multiple models. Some of the Heterogeneous methods are voting, stacking, etc. The voting ensemble learning consists of two steps namely construction of base classifier and combination rule. Many theoretical frameworks for the combination have been proposed by Kittler, but, for this study we used only one rule. (Kittler, Hatef, Duin, & Matas, 1998).
Construction of Base Classifier

In this study, three classifiers namely NB, RF and NBTree have been used as base classifiers. The prediction results of the ensemble classifiers need to be combined to get the final results.

Combination Rule

We used two combination rules namely Average of Probabilities and Majority Voting.

**Average of Probabilities**: This combination rule of voting calculates the mean value of each class among the independent classifiers and the class having the maximum value among the means is assigned the input pattern (I H Witten, Frank, & Hall, 2011).

**Majority Voting**: Voting is a cumulative method that merges the results of the multiple classifiers. Three editions of voting namely unanimous voting, plurality voting and majority voting are given in (Wolpert, 1992). In unanimous voting, all classifiers agree to the final decision, majority voting considers more than 50% vote for final decision and in plurality voting, most of the votes decide the final prediction.

Algorithm 2: Proposed Vote Ensemble (VE) Model

| Input: CAD datasets: DS = DS_{train} U DS_{test} |
| Output: O: Best classification output |
| 1. Begin Calculate the importance of each feature using a FST; |
| 2. Select the number of base classifiers in the VE model and select the Combination Rule in the VE model |
| Combination rule using **Average of Probabilities** or **Majority Voting** |
| 3. for i=1…, I do |
| Train the base machine learning algorithms, |
| Model using DS_{train}; |
| End |
| 4. Classify unseen records from DS_{test} using the VE model; |
| 5. Return the best classification outcome using the best establish model; |
| 6. End |

Compare Performance of classifiers

We calculated the performance of classifiers based on the confusion matrix as given in Table 2.

| Table 2. Confusion matrix |
| Actual Vs. Predicted | Positive | Negative |
| Positive | True Positive (TP) | False Negative (FN) |
| Negative | False Positive (FP) | True Negative (TN) |

The performance of the classification models was measured using four performance methods namely accuracy, sensitivity, specificity and f-measure. Accuracy is the percentage of perfectly
classified instances among all the instances. Sensitivity analysis techniques measure the rate of change at the output of a sample due to adjustments in the input of a variable. Specificity analysis is the proportion of actual negatives that are perfectly identified. F-measure is the harmonic mean of Precision and Recall, and provides a measure of the imperfectly classified cases.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}
\]

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \tag{11}
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} \tag{12}
\]

\[
\text{F1-Score} = \frac{2TP}{2TP + FP + FN} \tag{13}
\]

The accuracy based performance was used to recognize the important features. The accuracy measure provides a better acceptance of the overall behaviour of the specific combination of features to detect important features whereas accuracy, sensitivity, specificity and F-measure were used to identify classification techniques to produce the best performing models. On the other hand, classification techniques analysis focuses on the best performing models that can achieve a high accuracy in CAD datasets because accuracy and F1-Score are the most intuitive performance estimation matrices. Performance of each classifier has been calculated and all the findings are perfectly reported for further study.

Finally, we compared the performance of classifiers in both the datasets of CAD in two ways.

§ Compared the classifier’s performance of NB, RF, NBTree, Ensemble Models (Ensemble1, Ensemble2, Ensemble3, Ensemble4, Ensemble5, Ensemble6, Ensemble7 and Ensemble8) and DNN with all the features of CAD datasets.

§ Compared the classifier’s performance of NB, RF, NBTree, Ensemble Models (Ensemble1, Ensemble2, Ensemble3, Ensemble4, Ensemble5, Ensemble6, Ensemble7 and Ensemble8) and DNN with selected features of CAD datasets.

RESULTS AND DISCUSSION

In the preliminary experiments we have used many algorithms in the Waikato Environment for Knowledge Analysis - WEKA - (Ian H. Witten & Frank, 2004). In this section, we discuss the following: results of the analysis of feature selection technique, results of the analysis of classification and comparison of the proposed model with the previous survey.
Analysis of Feature Selection Technique

In this section, we describe the design of the experiments performed to evaluate the various PSO models. The FSTs or ASMs have two sections in WEKA namely Attribute Evaluators and Search Methods (Ian H Witten, Frank, & Hall, 2011).

Attribute Evaluators: Evaluate attribute sets in two ways by using a learning scheme. We have used Correlation-based Feature Subset Evaluation considering the individual predictive ability of each feature along with the degree of redundancy between them. In attribute evaluators, we have used the wrapper Subset Evaluators with learning classifiers J48 and Ensemble1. Table 3 shows the parameters used in the attribute evaluators.

Table 3. Parameters used in Attribute Evaluators

| Attribute Evaluators                        | Parameters                                      |
|---------------------------------------------|-------------------------------------------------|
| Correlation-based Feature Subset Evaluation | seed – 1
folds -- 5
classifier – NULL
IRClassValue -- Null
doNotCheckCapabilities -- False
evaluationMeasure – Accuracy
threshold – 0.01 |
| WrapperSubsetEval                          | seed – 1
folds -- 5
classifier – J48
IRClassValue -- NULL
doNotCheckCapabilities -- False
evaluationMeasure – Accuracy
threshold – 0.01 |
| WrapperSubsetEval                          | seed – 1
folds -- 5
classifier – Ensemble (NB+RF)
IRClassValue -- Null
doNotCheckCapabilities -- False
evaluationMeasure – Accuracy
threshold – 0.01 |

Table 3 shows the PSO search parameters with common settings value. The Correlation-based Feature Subset Evaluation evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them (Hall, 1999). We have used the following ways: The number of threads to use and the size of the thread pool is 1. The output debugging info is off. The Treat missing as a separate value is False. The locally predictive attributes is true value. The Wrapper Subset Evaluation evaluates attribute sets by using a learning scheme. Cross validation is used to estimate the accuracy of the learning scheme for a set of attributes (Ron Kohavi & John, 1997). We have used seed for randomly generating splits 1. The fitness function or objective with PSO were used as classifiers. Firstly, J48 classifiers and secondly Ensemble1 (NB+RF) were used as the base classifier. The learning classifiers of PSO are used with 5-fold cross validation technique to estimate subset accuracy. The class label or 1-based index of the class label was used when evaluating subsets with an IR metric (such as f-measure or AUC). We have used IRClass Value as Null. If set, evaluator capabilities are not checked before evaluator is built (Use with caution to reduce runtime). We set the doNotCheckCapabilities as False. The evaluation
measure used to evaluate the performance of attribute combinations. Finally we set the threshold as – 0.01. The evaluation measure of classifiers were calculated in terms of accuracy.

**Search Methods:** In this study, we have used the meta-heuristic search method namely Particle Swarm Optimization (PSO) for feature optimization. Table 4 shows the parameters used in PSO.

| Search Methods | Parameters |
|----------------|------------|
| Particle Swarm Optimization (PSO) | seed – 1, populationSize – 20, reportFrequency – 20, iterations – 20, logFile-Weka-3.8, socialWeight -- 0.33, individualWeight = 0.34, mutationType – bit-flip, inertiaWeight – 0.33, startSet – NULL, mutationProb -- 0.01 |

Table 4 shows the different parameters of PSO search method with common value settings. We have fixed the value of random seed as 1. The PSO used the particle in the swarm and used a population of size 20. It set the report frequency to 20. The number of iterations to perform search was set as 20. A log file was used to keep track of the fitness and the best fitness through all the iterations was found to be 3.8. In this work, we set the social weight as 0.33 and individual weight as 0.34. The other parameters like bit-flip as mutation operators, the inertia weight was set as 0.33. We set the start point for searching the start set with NULL. The start set became one of the population members of the initial population. We used the probability of mutation with 0.01.

With the help of Table 3 and Table 4, we have found three different feature selection techniques:

| Name of CAD datasets | Search Methods | Features | Name of Features | Feature set |
|----------------------|----------------|----------|-----------------|-------------|
| ZADS                 | PSO-J48        | 21       | f1,f5,f7,f8,f9,f11,f15,f18,f22,f23,f24,f25,f27,f29,f35,f38,f41,f42,f46,f51,f54 | Fs-21       |
|                      | PSO-CFSE       | 13       | f1,f6,f7,f18,f25,f28,f32,f34,f35,f45,f47,f53,f54 | Fs-13       |
|                      | PSO-Ensemble1  | 22       | f1,f4,f5,f7,f8,f9,f12,f14,f15,f22,f24,f25,f29,f33,f35,f38,f46,f47,f48,f49,f51,f54 | Fs-22       |
| EZADS                | PSO-J48        | 14       | f10,f18,f22,f26,f29,f41,f42,f44,f46,f50,f52,f56,f57,f58 | Fs-14       |
|                      | PSO-CFSE       | 16       | f1,f6,f9,f13,f16,f25,f27,f29,f30,f40,f47,f50,f52,f55,f56,f57,f58 | Fs-16       |
|                      | PSO-Ensemble1  | 17       | f10,f18,f19,f22,f26,f29,f50,f52,f53,f56,f57,f58 | Fs-17       |
• PSO-CFSE: Correlation-based Feature Subset Evaluation attributes evaluator with PSO search method.
• PSO-J48: Wrapper Subset Evaluation attribute evaluator with learning classifier C4.5 (J48) and PSO search method.
• PSO-Ensemble1: Wrapper Subset Evaluation attribute evaluator with learning classifier Ensemble1 and PSO search method.

Table 5 shows the attributes (features) selected by proposed feature selection technique. The ZADS dataset has a total of 55 features denoted by \( (f_1, f_2, f_3, \ldots, f_{55}) \). Similarly, EZADS has a total of 58 features denoted by \( (f_1, f_2, f_3, \ldots, f_{58}) \). Table 5 shows the features selected by proposed PSO and existing PSO techniques.

Figure 2 displays the number of features of both ZADS and EZADS datasets with the proposed and existing PSO based feature selection techniques. The PSO algorithm has used three separate base classifiers namely C4.5, CFSE and Ensemble1 and applied them to two CAD (ZADS and EZADS) datasets. The PSO-J48 selects 21 important features out of the 55 features from ZADS dataset while 14 important features out of the 58 features from EZADS dataset. The PSO-CFSE selects 13 important features out of the 55 features in the ZADS dataset while 16 important features out of the 58 features from EZADS dataset. The proposed PSO-Ensemble1 selects 22 important features out of the 55 features from ZADS dataset while 17 important features out of the 58 features in the EZADS dataset.

Waikato Environment for Knowledge Analysis (WEKA) contains NB, RF and NBTree and an ensemble technique. The voting ensemble techniques have been chosen using the average of probabilities and majority voting. A total of eight ensembles were chosen from the pool of three classifiers (NB, RF
Table 6. Accuracy of classifiers in percentage(%) 

| Name of algorithms | DSN: ZADS | DSN: EZADS |
|--------------------|----------|-----------|
|                    | Feature Sets | Feature Sets |
|                    | Fs-55 | Fs-21 | Fs-13 | Fs-22 | Fs-58 | Fs-14 | Fs-16 | Fs-17 |
| NB                 | 79.87 | 86.14 | 85.48 | 86.14 | 93.40 | 97.69 | 97.03 | 99.01 |
| NBTree             | 78.22 | 82.84 | 82.84 | 84.16 | 99.01 | 99.34 | 97.03 | 99.34 |
| RF                 | 83.17 | 83.50 | 85.15 | 87.46 | 94.39 | 99.34 | 99.00 | 99.34 |
| Ensemble1          | 84.16 | 88.78 | 86.14 | 89.11 | 93.74 | 99.34 | 97.36 | 99.67 |
| Ensemble2          | 83.83 | 86.47 | 85.15 | 87.46 | 94.72 | 98.35 | 97.35 | 99.34 |
| Ensemble3          | 81.85 | 88.12 | 84.49 | 85.48 | 97.03 | 98.35 | 97.03 | 99.34 |
| Ensemble4          | 78.54 | 83.83 | 84.49 | 85.15 | 93.74 | 99.34 | 97.36 | 99.67 |
| Ensemble5          | 81.52 | 84.82 | 84.82 | 85.81 | 99.34 | 99.67 | 99.01 | 99.01 |
| Ensemble6          | 80.20 | 84.49 | 85.15 | 85.81 | 97.03 | 99.34 | 97.35 | 99.67 |
| Ensemble7          | 85.15 | 87.79 | 86.80 | 88.45 | 97.36 | 99.67 | 99.01 | 99.67 |
| Ensemble8          | 84.81 | 86.14 | 85.81 | 87.79 | 96.69 | 99.67 | 98.35 | 99.67 |
| DNN                | 85.48 | 86.47 | 85.15 | 86.14 | 88.78 | 91.75 | 92.41 | 91.09 |

Table 7. Accuracy difference (in %) between the selected features and all features of ZADS and EZADS CAD datasets 

| Algorithm | DSN: ZADS | DSN: EZADS |
|-----------|-----------|-----------|
|           | Feature Selection Techniques | Feature Selection Techniques |
|           | Fs-55 | Fs-21 | Fs-13 | Fs-22 | Fs-58 | Fs-14 | Fs-16 | Fs-17 |
| NB        | 79.87 | 6.27 | 5.61 | 6.27 | 93.40 | 4.29 | 3.63 | 5.61 |
| NBTree    | 78.22 | 4.62 | 4.62 | 5.94 | 99.01 | 0.33 | -1.98 | 0.33 |
| RF        | 83.17 | 0.33 | 1.98 | 4.29 | 94.39 | 4.95 | 4.61 | 4.95 |
| Ensemble1 | 84.16 | 4.62 | 1.98 | 4.95 | 93.74 | 5.60 | 3.62 | 5.93 |
| Ensemble2 | 83.83 | 2.64 | 1.32 | 3.63 | 94.72 | 3.63 | 2.63 | 4.62 |
| Ensemble3 | 81.85 | 6.27 | 2.97 | 4.95 | 98.35 | 0.33 | -0.33 | 0.99 |
| Ensemble4 | 78.54 | 5.29 | 5.95 | 6.94 | 97.03 | 1.32 | 0.00 | 2.31 |
| Ensemble5 | 81.52 | 3.30 | 3.30 | 4.29 | 99.34 | 0.33 | -0.33 | -0.33 |
| Ensemble6 | 80.20 | 4.29 | 4.95 | 5.61 | 97.03 | 2.64 | 0.32 | 2.64 |
| Ensemble7 | 85.15 | 2.64 | 1.65 | 3.30 | 97.36 | 2.31 | 1.65 | 2.31 |
| Ensemble8 | 84.81 | 1.33 | 1.00 | 2.98 | 96.69 | 2.98 | 1.66 | 2.98 |
| DNN       | 85.48 | 0.99 | -0.33 | 0.66 | 88.78 | 2.97 | 3.63 | 2.31 |
and NBTree). Further, we have used a 10-fold cross-validation method for data partitions in training and testing.

**Analysis of Classification**

Accuracy is the percentage of correctly classified instances among all instances. The accuracy percentage of the proposed classification algorithms used in these experiments is reflected in Table 6.

Table 6 shows the accuracy in percentage (%) of the classifiers with ZADS and EZADS CAD datasets. In this research work, we have compared the accuracy of the classifiers with all features (Fs-55) of the ZADS CAD datasets and feature subsets selected using the proposed FSTs (Fs-21, Fs-13 and Fs-22). The Ensemble3 and DNN models obtained higher accuracy with Fs-21 features. The proposed classifiers NB, RF, NBTree and ensemble models (Ensemble1 to Ensemble8) obtained higher accuracy with all the cases of ZADS dataset with proposed PSO-Ensemble1 optimization technique using Fs-22 features compared with other FSTs. Now we have compared the accuracy of the classifiers obtained by all features (Fs-58) of the EZADS CAD datasets and feature subsets chosen from the proposed FSTs (Fs-14, Fs-16 and Fs-17). The Ensemble5 model obtained higher accuracy with Fs-14 features and also DNN obtained higher accuracy with Fs-16 features. The proposed classifiers NB, RF and NBTree and ensemble models (Ensemble1 to Ensemble8) obtained higher accuracy with all the cases of EZADS dataset with the proposed PSO-Ensemble1 optimization technique using Fs-17 features compared with other FSTs. It is also important to note the variations in accuracy, as one of the purposes of this study was to demonstrate the performances of PSO-Ensemble1 and other FSTs. Table 7 shows the accuracy difference between the selected features and all features of ZADS and EZADS CAD datasets.

Tables 7 show the accuracy difference of classifiers between the selected features and all features of ZADS and EZADS CAD datasets. This table shows that the Ensemble4 model achieves the highest accuracy difference (6.94%) with the selected Fs-22 features with ZADS dataset. The highest accuracy difference is obtained by the Ensemble1 model (5.93%) with the selected Fs-17 features with EZADS dataset. It is also significant to mention that the differences in accuracy in the context of this research are illustrated with respect to the selected features of PSO-Ensemble1 and all features in CAD datasets. Figure 3 displays the accuracy chart of the proposed classifiers in the case of Fs-55 and reduced Fs-22 features.

![Figure 3. Comparison chart of classifiers with accuracy in Fs-55 and Fs-22 features](image-url)
Figure 3 shows the comparative chart of all the proposed models of NB, NBTree, RF, Ensemble Models (Ensemble1, Ensemble2, Ensemble3, Ensemble4, Ensemble5, Ensemble6, Ensemble7, Ensemble8) and DNN with the proposed PSO-Ensemble1 FST with ZADS dataset. If we compare this graph between Fs-55 features and Fs-22 features, it clearly shows that the proposed FST achieves the highest accuracy in all the cases by all classifiers with the selected Fs-22 features.

Figure 4 displays the accuracy chart of the proposed classifiers in the case of 58 and 17 features. The figure shows the comparative charts of all proposed models with the proposed PSO-Ensemble1 FST with EZADS dataset. If we compare this graph with Fs-58 features and Fs-17 features, then this chart clearly depicts that the proposed FST obtains the highest accuracy in all the cases (except Ensemble5) by all classifiers with Fs-17 selected features.

Table 8. Sensitivity of Classifiers in percentage (%)
The sensitivity analysis techniques measure the rate of change at the output of a sample due to adjustments in the input of a variable. They are primarily used to decide which input parameter is the most relevant or sensible to arrive at the correct data. The sensitivity percentage of the proposed classification algorithms used in these experiments is reflected in Table 8.

The results of the sensitivity parameters of individual models NB, RF, NBTree and DNN are given for both ZADS and EZADS datasets and also compared with the Ensemble models. The proposed FST PSO-Ensemble1 significantly outperforms the other FSTs with respect to the sensitivity parameters mentioned in this study. Thus, the proposed Ensemble1 to Ensemble8 and DNN models exhibit improved sensitivity in the case of Fs-55 as well as the selected Fs-22 features in the ZADS dataset. But in the case of the EZADS dataset, the sensitivity of Ensemble3, Ensemble5 and DNN models fluctuate with the selected 17 features.

Specificity analysis is the proportion of actual negatives that are correctly identified. The specificity percentage of the proposed classification algorithms used in these experiments is reflected in Table 9.

Specificity analysis is the proportion of actual negatives that are correctly identified. The specificity percentage of the proposed classification algorithms used in these experiments is reflected in Table 9.

Table 9. Specificity of classifiers in percentage (%)

| Algorithm | DSN: ZADS | DSN: EZADS |
|-----------|------------|------------|
|           | Feature Sets | Feature Sets | |
|           | Fs-55 | Fs-21 | Fs-13 | Fs-22 | Fs-58 | Fs-14 | Fs-16 | Fs-17 |
| **NB**    | 75.86  | 85.06  | 83.91  | **86.21** | 88.51  | 93.10  | 95.40  | **96.55** |
| **RF**    | 52.87  | 58.62  | **70.11** | 68.97  | 85.06  | 97.70  | 97.70  | **97.70** |
| **NBTree** | 58.62  | 65.52  | 64.37  | **67.82** | 98.85  | 97.70  | 94.25  | **98.85** |
| **Ensemble1** | 75.86  | **82.76** | 80.46  | 79.31  | 87.36  | 97.70  | 95.40  | **98.85** |
| **Ensemble2** | 68.96  | 79.31  | 78.16  | **80.46** | 88.50  | 95.40  | 95.40  | **97.70** |
| **Ensemble3** | 66.67  | **80.46** | 72.41  | 72.41  | 95.40  | 95.40  | 96.55  | **98.85** |
| **Ensemble4** | 60.92  | 74.71  | 75.86  | **75.86** | 94.25  | 95.40  | 93.10  | **97.70** |
| **Ensemble5** | 58.62  | 68.97  | 70.11  | **70.11** | 98.85  | **100** | 98.85  | 98.85  |
| **Ensemble6** | 54.02  | 67.82  | **71.26** | 67.82  | 93.10  | 98.85  | 95.40  | **98.85** |
| **Ensemble7** | 68.96  | **77.01** | 77.01  | 75.86  | 93.10  | 98.85  | 96.55  | **98.85** |
| **Ensemble8** | 67.82  | 73.56  | **79.31** | 78.16  | 93.10  | 98.85  | 97.70  | **98.85** |
| **DNN** | 87.36  | **88.51** | 87.36  | 86.21  | 98.85  | **100** | 98.85  | 98.85  |

Table 9 shows the specificity parameters of the individual models NB, RF, NBTree, Ensemble models and DNN for both ZADS and EZADS datasets. The proposed PSO-Ensemble1 FST significantly outperforms other FSTs in terms of specificity. The proposed Ensemble2, Ensemble4 and Ensemble5 models achieved enhanced specificity for the selected Fs-22 features in the case of the ZADS dataset. However, in the case of the EZADS, the specificity of the models Ensemble1, Ensemble2, Ensemble3, Ensemble4, Ensemble6, Ensemble7 and Ensemble8 have enhanced in the selected Fs-17 features.

F-measure (F-score or F1-score) has been introduced to balance between sensitivity and specificity. The F1-Score percentage of the proposed classification algorithms used in these experiments is reflected in Table 10.
The proposed classifiers achieved an excellent F1-score with Fs-22 features of the ZADS dataset. The Ensemble2 model achieved the highest F1-Score of 95.82% with the 22 features selected by PSO-Ensemble1 FST. Similarly, better F1-Scores were achieved by the proposed Ensemble models with Fs-17 features in the EZADS dataset. This F1-Score was compared with all features of CAD datasets and two other feature selection techniques namely PSO-J48 and PSO-CFSE.

### Comparative Analysis of Proposed Model with Existing Work

The accuracy of proposed model using Fs-22 relevant features of ZADS dataset and Fs-17 relevant features of the EZADS dataset were compared with other existing models which have used the CAD datasets in the recent years, where the proposed model has achieved acceptable accuracy compared with others existing models.

The above table shows the highest accuracy of the proposed models and the existing models. The proposed ensemble models namely Ensemble1, Ensemble6, Ensemble7 and Ensemble8 obtained the highest accuracy of 99.67% with Fs-17 and DNN obtained the highest accuracy of 92.41% with Fs-16 features of the EZADS dataset. The proposed Ensemble1 model obtained the highest accuracy of 89.11% with Fs-22 and DNN obtained the highest accuracy of 86.47% with Fs-21 features of the ZADS dataset as shown in the above table. Based on this comparison, it is apparent that this research has achieved higher accuracy using the proposed ensemble models with the proposed PSO-Ensemble1 FST. Hence, it is concluded that the proposed classification model has given better accuracy compared with other existing models.

### CONCLUSION

CAD is developed by the formation of plaques inside the walls of coronary arteries, resulting in the narrowing of lumens. Machine learning techniques can be useful in detecting the disease. In this paper, we have proposed a new feature optimization technique namely PSO with learning classifier as ensemble based on voting technique and combination rule (average probability). The PSO has used the ensemble model as the learning classifier. We have further described about their parameters, attribute evaluators and including the learning classifier. We have followed an approach to ensemble construction, applying PSO to the method of voting ensemble structure to generate domain-specific
configurations. The proposed PSO-Ensemble1 FST selected more relevant features compared with other FSTs like PSO-J48 and PSO-CFSE. These performances were consolidated by comparing all the features of CAD datasets and two other FSTs (PSO-J48 and PSO-CFSE). After this, we proposed ensemble models of classifiers based on voting technique with two combination rules (average probability and majority voting). An extensive study was conducted and the results from the proposed techniques were compared with the frequently used classification methods such as NB, RF, NBTree, their ensemble models and DNN. The results from the combination of different classifiers depicted better accuracy compared with individual classifiers. In this study, it has been observed that the proposed Ensemble models performed better than the different individual classifiers such as NB, RF, NBTree and DNN. The results show that the Ensemble1 model assured an accuracy of 89.11% with Fs-22 selected features (features selected by FST PSO-Ensemble1) in the ZADS dataset. The results also show that the various ensemble models Ensemble1, Ensemble2, Ensemble3, Ensemble4, Ensemble5, Ensemble6, Ensemble7 and Ensemble8 performed better with the selected features of both the CAD datasets. Overall, the findings reveal that the proposed FST PSO-Ensemble1 model is an efficient and encouraging approach towards selecting features for the CAD datasets.

Table 11. Comparison of the proposed model and existing models

| Name of author and year | Dataset | Algorithms | Accuracy (%) |
|-------------------------|---------|------------|--------------|
| The proposed models     | EZADS   | NB, RF, NBTree, Ensemble1 to Ensemble8 and DNN FSTs PSO-CFSE, PSO-J48, PSO-Ensemble1 | Ensemble1-Fs-17: 99.67 Ensemble6-Fs-17: 99.67 Ensemble7-Fs-17: 99.67 Ensemble8-Fs-17: 99.67 DNN-Fs-16: 92.41 |
| The proposed models     | ZADS    | Classifiers NB, RF, NBTree, Ensemble1 to Ensemble8 and DNN FSTs PSO-CFSE, PSO-J48, PSO-Ensemble1 | Ensemble1-Fs-22: 89.11 DNN-Fs-21: 86.47 |
| Verma & Mathur, (2019)  | CHD     | RE-DNN     | RE-DNN: 85.48 |
| Miao & Miao (2018)      | CHD     | EDNN-RD    | EDNN-RDM: 83.67 |
| Bektas et al. (2017)    | CVD     | Logistic regression RNA, NN, FST like Relief-F, independent t-test analysis, | NN-Relief-F: 84.10 |
| Paul, Shill, Rabin, & Akhand (2016) | CAD | NN with Fuzzy | NN-Fuzzy: 80.00 |
| Verma, Srivastava, & Negi (2016) | CAD | DT | DT: 80.68 |
| Elbialy et al. (2015)   | CAD     | DT         | DT: 78.54 |
| Hedeshi & Abadeh (2014) | CAD     | fuzzy boosting PSO (En-PSO2) | En-PSO2: 85.76 |
| Alizadehsani et al. (2013) | ZADS | Bagging and C4.5 classification | C4.5: 79.54 |
| Alizadehsani et al. (2012) | ZADS | SMO, Naive Bayes, C4.5 and AdaBoost | SMO: 79.86 |
The performance of PSO-Ensemble1 FST model can be further improved by generating new domain-specific configurations. The number of classifiers and other ensemble techniques will play an important role in further classification of CAD.

Table 12. List of abbreviations

| Abbreviations | Description |
|---------------|-------------|
| NB            | Naive Bayes |
| NBTree        | Decision Tree with Naive Bayes |
| RF            | Random Forest |
| Ensemble1     | Ensemble model –Combination of Naive Bayes and Random Forest Combination Rule – Average of Probabilities |
| Ensemble2     | Ensemble model –Combination of Naive Bayes and Random Forest Combination rule -Majority Voting |
| Ensemble3     | Ensemble model –Combination of Naive Bayes and NBTree Combination Rule – Average of Probabilities |
| Ensemble4     | Ensemble model –Combination of Naive Bayes and NBTree Combination Rule – Majority Voting |
| Ensemble5     | Ensemble model –Combination of Random Forest and NBTree Combination Rule – Average of Probabilities |
| Ensemble6     | Ensemble model –Combination of Random Forest and NBTree Combination Rule – Majority Voting |
| Ensemble7     | Ensemble model –Combination of Naive Bayes, Random Forest and NBTree Combination Rule – Average of Probabilities |
| Ensemble8     | Ensemble model –Combination of Naive Bayes, Random Forest and NBTree Combination Rule – Majority Voting |
| DNN           | Deep Neural Network |
| PSO           | Particle Swarm Optimization |
| DSN           | Dataset Name |
| CFSE          | CFsubsetEval |
| FST           | Feature Selection Technique |
| ASM           | Attribute Selection Method |
| CAD           | Coronary Artery Disease |
| Fs-55         | Z-Alizadeh Sani with all features |
| Fs-21         | FST: PSO-J48 and DSN: Z-Alizadeh Sani |
| Fs-13         | FST: PSO-CFSE and DSN: Z-Alizadeh Sani |
| Fs-22         | FST: PSO- Ensemble1 and DSN: Z-Alizadeh Sani |
| Fs-58         | extension of Z-Alizadeh Sani with all features |
| Fs-14         | FST: PSO-J48 and DSN: extension of Z-Alizadeh Sani |
| Fs-16         | FST: PSO-CFSE and DSN: extension of Z-Alizadeh Sani |
| Fs-17         | FST: PSO- Ensemble1 and DSN: extension of Z-Alizadeh Sani |
| ZADS          | Z-Alizadeh Sani |
| EZADS         | extension of Z-Alizadeh Sani |
REFERENCES

Aburomman, A. A., & Reaz, M. B. I. (2016). A novel SVM-knn-PSO ensemble method for intrusion detection system. Academic Press.

Alizadehsani, R., Zangoeei, M. H., Hosseini, M. J., Habibi, J., Khosravi, A., Roshanzamir, M., … Nahavandi, S. (2016). Coronary Artery Disease Detection Using Computational Intelligence Methods. Knowledge-Based Systems, 109. https://Doi.Org/10.1016/J.Knosys.2016.07.004

Alizadehsani, R., Habibi, J., Sani, Z. A., Mashayekhi, H., Boghrati, B., Ghandeharioun, A., … Alizadeh-Sani, F. (2013). Diagnosing Coronary Artery Disease via Data Mining Algorithms by Considering Laboratory and Echocardiography Features. Journal of Rajaie Cardiovascular Medical and Research Center. 10.5812/cardiovascmed.10888

Alizadehsani, R., Hosseini, M. J., Boghrati, R., Ghandeharioun, A., Khozeimeh, F., & Sani, Z. A. (2012). Exerting Cost-Sensitive and Feature Creation Algorithms for Coronary Artery Disease Diagnosis. Exerting Cost-Sensitive and Feature Creation Algorithms for Coronary Artery Disease Diagnosis, 3(March), 59–79. doi:10.4018/jkdb.2012010104

Amutha, A. J., Padmajavalli, R., & Prabhakar, D. (2018). A Novel Approach For The Prediction Of Treadmill Test In Cardiology Using Data Mining Algorithms Implemented As A Mobile Application. Indian Heart Journal, 70(4), 511–518. doi:10.1016/j.ihj.2018.01.011 PMID:30170646

Bektas, J., Ibrikci, T., & Ozcan, I. T. (2017). Classification Of Real Imbalanced Cardiovascular Data Using Feature Selection And Sampling Methods: A Case Study With Neural Networks And Logistic Regression. International Journal of Artificial Intelligence Tools, 26(06), 1750019. doi:10.1142/S0218213017500191

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives. Academic Press.

Bianchi, L., Dorigo, M., Maria, L., & Gutjahr, W. J. (2009). A Survey On Metaheuristics For Stochastic Combinatorial Optimization. https://Doi.Org/10.1007/S11047-008-9098-4

Blum, C., & Roli, A. (2001). Metaheuristics In Combinatorial Optimization : Overview And Conceptual Comparison Metaheuristics In Combinatorial Optimization : Overview And Conceptual Comparison. ACM Computing Surveys, 35(3), 268–308. doi:10.1145/937503.937505

Breiman, L. (1996). Bagging Predictors. Machine Learning, Springer, 140(2), 123–140. doi:10.1007/BF00058655

Chen, Y., Wong, M., & Li, H. (2014). Applying Ant Colony Optimization To Configuring Stacking Ensembles For Data Mining. Expert Systems with Applications, 41(6), 2688–2702. doi:10.1016/j.eswa.2013.10.063

Dag, H., Sayin, K. E., Yenidogan, I., Albayrak, S., & Acar, C. (2012). Comparison Of Feature Selection Algorithms For Medical Data. International Symposium On Innovations In Intelligent Systems And Applications. doi:10.1109/INISTA.2012.6247011

Dietterich, T. G. (2000). Ensemble Methods. Machine Learning, 1–15.

El-Bialy, R., Salamay, M. A., Karam, O. H., & Khalifa, M. E. (2015). Feature Analysis Of Coronary Artery Heart Disease Data Sets. Procedia Computer Science, 65, 459–468. https://Doi.Org/10.1016/J.Procs.2015.09.132

Gandhi, I., & Pandey, M. (2015). Hybrid Ensemble Of Classifiers Using Voting. International Conference On Green Computing And Internet Of Things (1 Cgclot). 399–404. doi:10.1109/ICGClot.2015.7380496

Hall, M. (1999). Correlation-Based Feature Selection For Machine Learning. Methodology, 21, i195-120. https://Doi.Org/10.1.1.149.3848

Hamze-Ziabari, S. M., & Bakhshpoori, T. (2018). Improving The Prediction Of Ground Motion Parameters Based On An Efficient Bagging Ensemble Model Of M5 And CART Algorithms. Applied Soft Computing, 68, 147–161. doi:10.1016/J.asoc.2018.03.052

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts And Techniques (3rd ed.). Elsevier.

Haykin, S. (2008). Neural Networks And Learning Machines. Pearson Prentice Hall. https://Doi.Org/978-0131471399
Hedeshi, N. G., & Abadeh, M. S. (2014). A Fuzzy-Boosting PSO Approach. Computational Intelligence and Neuroscience, 2014, 1–12. doi:10.1155/2014/783734 PMID:24817883

Ho, T. (1998). The Random Subspace Method For Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(20), 832–844. https://Doi.Org/10.1109/34.709601

Ho, T. K. (1995). Random Decision Forests. In 3rd International Conference On Document Analysis And Recognition (Vol. 47, pp. 4–8). Academic Press.

Kabari, L., & Onwuka, U. C. (2019). Comparison Of Bagging And Voting Ensemble Machine Learning Algorithm As A Classifier. Academic Press.

Kam, H. J., & Kim, H. Y. (2017). Learning Representations For The Early Detection Of Sepsis With Deep Neural Networks. Computers In Biology And Medicine. https://Doi.Org/10.1016/J.Comphiomed.2017.08.015

Karthika, D., & Divakar, S. (2014). Improving The Efficiency Of Fast Using Semantic Similarity Algorithm. International Journal Of Scientific And Research Publications, 4(1), 1–5.

Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. Academic Press.

Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On Combining Classifiers. https://Doi.Org/10.1109/ICPR.1996.547205

Kohavi, R. (1996). Scaling Up the Accuracy Of Naive-Bayes Classifiers: A Decision-Tree Hybrid. In Proceedings Of The Second International Conference On Knowledge Discovery And Data Mining. AAAI Press.

Kumari, L., & Rao, N. M. (2017). An Efficient PSO Based Ensemble Classification Model On High Dimensional. https://Doi.Org/10.5121/Ijsc.2017.8401

Lang, S., Bravo-marquez, F., Beckham, C., & Hall, M. (2019). WekaDeeplearning4j : a Deep Learning Package for Weka based on DeepLearning4j. 10.1016/j.knosys.2019.04.013

Lantz, B. (2013). Machine Learning With R (First). Packt.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. 10.1038/nature14539

Ledezma, A., Aler, R., Sanchis, A., & Borrajo, D. (2010). GA-Stacking : Evolutionary Stacked Generalization. https://Doi.Org/10.3233/IDA-2010-0410

Lohita, K., Sree, A. A., Poojitha, D., Renuga Devi, T., & Umamakeswari, A. (2015). Performance Analysis Of Various Data Mining Techniques In The Prediction Of Heart Disease. Indian Journal of Science and Technology, 8(35). Advance online publication. doi:10.17485/ijst/2015/v8i35/87458

Miao, K. H., & Miao, J. H. (2018). Coronary Heart Disease Diagnosis Using Deep Neural Networks. International Journal of Advanced Computer Science and Applications, 9(10), 1–8. doi:10.14569/IJACSA.2018.091001

Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O., & Akbari, E. (2018). A Predictive Method For Hepatitis Disease Diagnosis Using Ensemblesof Neuro-Fuzzy Technique. Journal Of Infection And Public Health. https://Doi.Org/10.1016/J.Jiph.2018.09.009

Ordóñez, J., Ledezma, A., & Sanchis, A. (2008). Genetic Approach For Optimizing Ensembles Of Classifiers. Academic Press.

Pal, D., & Chakraborty, C. (2011). Data Mining Approach For Coronary Artery Disease Screening. ICIIP. doi:10.1109/ICIIP.2011.6108972

Pandey, M., & Taruna, S. (2014). A Comparative Study Of Ensemble Methods For Students' Performance Modeling. International Journal Of Computer Applications. https://Doi.Org/10.5120/18095-9151

Paul, A. K., Shill, P. C., Rabin, R. I., & Akhand, M. A. H. (2016). Genetic Algorithm Based Fuzzy Decision Support System For The Diagnosis Of Heart Disease. Academic Press.

Polikar, R. (2006). Ensembles Based Systems. IEEE Circuits and Systems Magazine, 6(3), 21–45. doi:10.1109/MCAS.2006.1688199

Pujari, A. (2013). Data Mining Techniques (3rd ed.). University Press.
Sani, Z. A., Alizadehsani, R., & Roshanzami, M. (2017). Z-Alizadeh Sani Data Set. UC Irvine Machine Learning Repository. Retrieved From https://archive.ics.uci.edu/ml/datasets/extension+of+z-alizadeh+sani+dataset

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. doi:10.1016/j.neunet.2014.09.003 PMID:25462637

Sharma, S., & Parmar, M. (2020). Heart Diseases Prediction Using Deep Learning Neural Network Model. *International Journal of Innovative Technology and Exploring Engineering, 9*(3), 2244–2248. https://doi.org/10.35940/ijitee.C9009.019320. doi:10.35940/ijitee.C9009.019320

Shi, Y., & Eberhart, R. (1998). *A Modified Particle Swarm Optimizer*. Academic Press.

Solankar, P. A., Jagatap, T. B., Rupnawar, S. H., Shitole, V. J., & Kumbhar, S. L. (2015). Student Performance Evaluation In Education Sector Using Prediction And Clustering Algorithms. Academic Press.

Tran, B., Xue, B., & Zhang, M. (2014). Improved PSO For Feature Selection On High-Dimensional Datasets. In Asia-Pacific Conference On Simulated Evolution And Learning (pp. 503–515). Springer. doi:10.1007/978-3-319-13563-2_43

Verma, L., & Mathur, M. K. (2019). *Deep Learning Based Model For Decision Support With Case Based Reasoning*. Academic Press.

Verma, L., Srivastava, S., & Negi, P. C. (2016). A Hybrid Data Mining Model To Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data. *Journal Of Medical Systems*. https://doi.org/10.1007/s10916-016-0536-Z

Wang, D., Tan, D., & Liu, L. (2018). *Particle Swarm Optimization Algorithm: An Overview*. https://doi.org/10.1007/S00500-016-2474-6

Wang, S., Jiang, L., & Li, C. (2015). Adapting Naive Bayes Tree For Text Classification. *Knowledge and Information Systems, 44*(1), 77–89. doi:10.1007/s10115-014-0746-y

Yildirim, P. (2015). Filter Based Feature Selection Methods For Prediction Of Risks In Hepatitis Disease. *International Journal of Machine Learning and Computing, 5*(4), 258–263. doi:10.7763/IJMLC.2015.V5.517

Z-Alizadeh Sani Dataset. (2016). Retrieved from https://archive.ics.uci.edu/ml/datasets/extension+of+z-alizadeh+sani+dataset
Pratibha Verma is currently a Ph.D. candidate in Department of Information Technology at Dr. C.V. Raman University, Bilaspur (Chhattisgarh), India. She received her master's Degree in Computer Application from Chhattisgarh Swami Vivekananda Technical University Bhilai (Chhattisgarh), India in 2012. Her research interests include machine learning and deep Learning.

Vineet Kumar Awasthi is the assistant professor at Dr CV Raman University, Bilaspur (CG). He has completed PhD in Computer Science at 2019. He has published different research papers in Scopus indexed, UGC listed journals. He has also presented different papers in national and international conferences. His research area is in optimization techniques.

Sanat Kumar Sahu is working as Assistant Professor in Department of Computer Science, Govt. Kaktiya PG College, Jagdalpur (Bastar) Chhattisgarh. He has received his Master’s Degree in Computer Application from Guru Ghasidas Vishwavidyalaya, Bilaspur in 2008 and M. Phil in Computer Science from Dr. C.V. Raman University, Kota, Bilaspur in 2009. He has more than 7 years teaching and 02 years research experience. He has published more than 12 research paper in reputed journals and attended workshop and conference at national and international level. His area of interest includes soft computing, Machine learning, and data mining.

Akhilesh Kumar Shrivas is an Assistant Professor in the Department of Computer Science and Information Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur, India. He has six years teaching experience. His research interests are data mining-based classification, soft computing and information security. He has published approximate 75 research papers in National and International Journals/Proceedings.