INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

Check for updates

Corresponding author.
 S Omprakash

Research Scholar, Department of Computer Science, SRMV College of Arts and Science, Coimbatore, Tamil Nadu, India. Tel.: 9894959107
subramaniamomprakash@gmail.com

Coronary artery disease prediction using hidden Markov model based support vector machine

S Omprakash[1]*, **M Ravichandran**[2]

1 Research Scholar, Department of Computer Science, SRMV College of Arts and Science, Coimbatore, Tamil Nadu, India. Tel.: 9894959107
2 Associate Professor, Department of Computer Science, SRMV College of Arts and Science, Coimbatore, Tamil Nadu, India

Abstract

Background: Medical data classification has become a hot research domain in data mining, but still it faces the increased classification accuracy issues. **Methods/Statistical Analysis:** Novel Hidden Markov Model based Support Vector Machine (HMM-SVM) is proposed to classify and predict Coronary Artery Disease (CAD). The features are extracted using HMM, and normalized using SVM. Feature Extraction assist the classification algorithm to get better results. HMM-SVM performs classification by extracting the features of Z-AlizadehSani dataset and finally selects the appropriate feature to perform classification. **Findings:** Z-AlizadehSani dataset holds 303 records with 4 different types of features, which are demographic, symptom and examination, ECG, and laboratory cum echo. For extracting these features and finding hidden information there exists no common algorithm. In HMM-SVM, HMM is applied to extract features by finding the hidden and previous stage values, and SVM is applied to perform classification on extracted features. To analyze the performance of HMM-SVM benchmark performance metrics are utilized. Discriminative performance results of internal validations are high in the task of binary classification (i.e., sensitivity- 98.2%; specificity-97.96%). False Positive Rate of HMM-SVM is entirely low (i.e.,1.87%) when comparing with previous algorithms. HMM-SVM holds the classification accuracy as 98.02% and which is the better cum expected results towards the prediction of CAD. **Novelty:** Detailed analysis indicates HMM-SVM have better effects towards classifying and predicting CAD. Furthermore, care needs to be placed in adhering to ethical principles while utilizing the models that are automated. Future studies should make use of bio-inspired concepts to get even better results.

Keywords: CAD; Classification; SVM; HMM; Markov

1 Introduction

Coronary artery disease (CAD) is one among the many heart diseases that lead to sudden death without any symptoms, and it is increasing in South Asian countries like India. First and most common sign of CAD is heart attack. Root cause for CAD is plaque that gets build in arteries wall. Knowledge discovery is the efficient process to analyze and understand the enormous amount of available data. It involves the process of identifying the valid, potential, novel, and useful patterns in the data. Data mining is the method to apply machine learning in datasets to extract the hidden, unknown, and better patterns in the datasets. Data mining is seen as a better step in discovering the knowledge in vast data.

Angiography is the best method to detect CAD, but it involves more extensive cost and knowledge. Because of limitations like this, researchers started applying data mining techniques in medical field to make decisions. Predictive data mining is better computing method to build a model using the features in dataset. It ranges from easy level to complex level, and persuasive ones involve hidden markov model (HMM). The main objective of this paper is to make use of HMM for identifying the CAD and compare its classification accuracy with other classification techniques.

In order to enhance the prediction of heart disease from Parkinson disease, optimized crow search algorithm[1] proposed, where it attempted to predict the heart disease more accuracy to provide on-time treatment. The results showed that the proposed algorithm is not fit for dataset related to heart disease, where the classification accuracy becomes very low. Two-Class Classification[2] proposed with the framework of machine learning by utilizing artificial neural network classification concept. The classifier works by selecting the spectral features of sub-band. The result shows that classifier could not perform well when there exist noisy data more than the remarkable range, where the false positive gets increases. Robust algorithm[3] was proposed to localize the heart beats and classify by utilizing the variables and threshold values. During the process of classification, dataset labels were utilized to find the variations in heart beat. Localization errors in the results shows that the algorithm is not suitable large datasets. Multistage classification[4] was proposed to classify and provide diagnose to the patient of congestive heart failure. It performs the analysis based on variation in heart rate, and computes the features related to heart rate. For computing the features, it uses the domains of time and frequency. The result provided a ineffective result regarding the true negative which cannot be used to provide treatments or medications to the patients.

Automated classifier based on support vector machine[5] proposed to classify the electrocardiograms towards predicting the heart disease. It depends on the time of electrocardiograms, to train the support vector machine to select the feature. The results proved that the classification accuracy went down due to feature selection concept, where the classifier omitted the important feature for classification. Modified version of ant colony optimization[6] proposed to increase the classification accuracy towards predicting the coronary artery disease, where it uses the least square model of regression. The correlation coefficients calculated for checking the fitness level between the selected features. The result came with low classification accuracy. Deep learning strategy[7] for the classification of ECG towards heart disease prediction was proposed. This strategy was proposed with the target of classifying in a automatic manner. The result showed that the results were not efficient when comparing with the existing algorithms in the term of sensitivity. Treadmill advanced prediction method[8] was proposed to predict the coronary artery disease. Instead of following the traditional statistics approach, data mining strategy was used. In this decision tree, k-nearest neighbor, k-sorting and searching algorithms were ensembled. Features selection were also performed to increase the accuracy, but f-measure value got decreased.

Imbalanced data classifier[9] was proposed to predict the heart disease even there exist missing values in the dataset. Symmetric Minority Over Sampling Technique was applied to increase the performance of classifier, resulting that it is not suitable for all dataset by giving the poor f-measure. Hybrid classifier[10], which was an ensemble of neural network and genetic algorithm, was proposed for the classification of coronary artery disease. Initially neural network performed and then the genetic algorithm was used. The result showed that the specific hybrid classifier was not fit for the prediction of coronary artery disease, where the results came with very low classification accuracy. Mobile health service platform[11] was proposed to analyze and classify the sounds of heart in order to predict the heart disease. It aims to monitor the patients from remote location by using the wireless technology. The service platform is built by integrating the Hidden Markov Model and Mel Frequency Cepstral Coefficient. The result provides inaccurate results varying with different mobile devices.

Disease gene classification method[12] proposed with the utilization of metagraph representation method, where it has integrated the terms, which describe the protein. The result showed the method is not suitable to predict the heart disease only by the gene, where the result came with increased false positive and false negative. Deep learning based convolutional neural network[13] proposed to classify the heartbeat to predict the level of heart disease. It utilized the function of batch based weight loss to measure the loss and overcome the problem of imbalance, which occur between classes. The accuracy has become low due to the dynamic change of class and batches. Computer aided diagnosis system[14] proposed to predict the valvular heart disease by utilizing the signals of impedance cardiography. It uses the concepts of selecting features by using the support vector machine and k-nearest neighbors algorithms. The result came with very low true positive which affecting the accuracy. Extreme Gradi-

ent Boosting based Classifier [15] proposed to detect the heart disease by analyzing the electrocardiogram signals. It extracts the features from six broad categories and finds the best feature by utilizing the recursive feature elimination concept. During the feature extraction phase, important features were discarded leading to misclassification.

2 Proposed Methodology

With the increased development in medical field, the classification is becoming a major idea to handle and organize medical data. The classification in medical field is possible to formulate as the following. Consider the training set of labeled texts $S = \{(s_1, b_1), \ldots, (s_m, c_m)\}$, where s_j indicates every text that are contained in set S, and b_j that belongs to s_j is expected to be available in $B = (b_1, \ldots, b_n\}$. The main intention of classification is to develop a better learning algorithm that takes the input as training set S and result in generating a classifier $g : E \rightarrow B$, which have the ability to perform classification in a more accurate manner from S. Even though adequate methods are proposed for classification, support vector machine (SVM) is addressed as identified as better classifier, which falls under binary classification family. This paper attempts to combine HMM with SVM for classifying the medical data towards prediction of CAD. Initially, HMM is applied to extract the features to generate output that involves discrimination information. Secondly, Outputs of HMM are normalized into new feature vectors. Finally, new feature vectors are fed as input to SVM for classification.

2.1 HMM based Feature Extraction

Selecting the features and better technique for classification are discussed in this part. In medical data classification, term values involved are used for classifying the features, but there exist no guarantee for better classification accuracy when in involves all values and features. Hence, there exist a need for eliminating the information that are not useful for classification.

Consider $P = (P_1, \ldots, P_s)$ as sequence of medical data from the dataset, where P_s indicates the corresponding tokens to medical data. Precisely, every token act as a token for extracted vectors. Linguistic tags Z_j are added to selected tokens of P_s. The feature extraction maps $P_1, \ldots, P_s$ to a unique sequence tags $(\tau_1, \ldots \tau_s)$ where $\tau_s \in (Z_1, \ldots, Z_m, \Lambda\}$.

An HMM $\lambda = (\pi, C, D)$ consist restricted number of states $(T_1, \ldots, T_m\}$. The probability at initial state T_j will be $\pi_j = Q(f_1 = T_j)$ and the probability at transition state (i.e., T_j to T_k) will be $c_{jk} = Q(f_{s+1} = T_k | f_s = T_j)$. Over different number of observations every state in HMM will get characterize by $d_j(P_s) = Q(P_s | f_s = T_j)$ probability distribution. $P = P_1, \ldots, P_s$ is considered as a observational sequence. According to the theorem of Bayes, for every P_s observation its necessary to return z_j tag which increases probability $Q(\tau_s = Z_j | P)$. It indicates the necessity of identifying the states $f_1, \ldots, f_s$ to increase $Q(f_s = T_j | P, \lambda)$ which will return the Z_j tag corresponding to T_j state for P_s token.

2.2 HMM's forward-backward ethod

Forward-Backward (*FoBa*) is a method which interferes HMM to calculates post values of overall variables in hidden state $\alpha_s(j) = Q(f_s = T_j, P_1, .., P_s | \lambda)$ which is considered as the forward variable that quantifies the reaching state T_j at time t that observes the sequence $P_1, \ldots, P_s$. $\beta_s(j) = Q(P_{s+1}, \ldots P_s | f_s = T_j, \lambda)$ is considered as the backward variable that quantifies the possibility to observe $P_{t+1}, \ldots, P_s$ rest sequence state T_j at time t. The task of interference is called as smoothing. *FoBa* utilize dynamic programming principle to calculate post values distribution in 2 different passes. First pass proceeds forward in unique time and second pass proceeds in different unique time. *FoBa* is utilized to denote any algorithm that belongs to classes of algorithm and operates in a simultaneous manner. It is necessary to calculate $\alpha_s(j)$ and $\beta_s(j)$ to express the probability in state T_j at time t.

Finally, the features are extracted using HMM as below:

Input text $S = (V_1, \ldots, V_s)$; HMM λ; set of tags $Z_1, \ldots, Z_m$ correspond to objective HMM states $T_1, \ldots, T_m$.

Generate the sequence $P = (P_1, \ldots P_s)$, where P_s denotes the vector that contains the detection V_s.

Call Forward-Backward algorithm Calculate $f_s^* = max_t{}_j \gamma_s(j)$, $1 \leq t \leq S$.

If $f_s = T_j \in (T_1, \ldots, T_m\}$

then Output $\langle Z_j \rangle$ V_s $\langle Z_j \rangle$;else output V_s.

When the feature extraction process gets over, the outputs are normalized to form a new feature vector, and then SVM classifier is applied for classification.

FoBa algorithm starts with some initial estimate of the HMM parameters $\lambda = (A, B)$. We then iteratively run two steps. Like other cases of the EM (expectation-maximization) algorithm, the forward-backward algorithm has two steps: the expectation step, or E-step, and the maximization step, or M-step.

The Expectation-Maximization (EM) algorithm is the method to discover maximum-likelihood estimations for model parameters when dataset faces incompleteness problem, missing data points, or unobserved latent variables. It is an iteration-based method to approximate the function of maximum likelihood. When maximum likelihood estimation gets the ability to find "best fit" model for few data in dataset, then it will not work well.

In the E-step, we calculate the expected state occupancy count γ and the expected state transition count ξ from the earlier A and B probabilities. In the M-step, we use γ and ξ to extract the features by recalculating new A and B probabilities.

$function\ FoBa\ (observations\ of\ len\ T,\ output\ vocabulary\ V,\ hidden\ state\ set\ Q)$

$returns\ HMM = (A, B)$

$initialize\ A\ and\ B$

$iterate\ until\ convergence$

E-Step

$\gamma_t(j) = \frac{\alpha_t(j)\beta_t(j)}{\alpha_T(q_F)}\ for\ all\ t\ and\ j$

$\xi_t(j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\alpha_T(q_F)}\ for\ all\ t,\ i\ and\ j$

M-Step

$a_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\sum_{k=1}^{N}\xi_t(i,k)}$

$b_j(v_k) = \frac{\sum_{t=1 s.t.\ O_t=v_k}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)}$

return A,B

2.3 SVM Classifier

SVM is one of the great classification algorithm coming under supervised learning category. SVM working mechanism works based on risk minimization concept derived from the theory of computational learning. Increased generalizing capability in SVM makes it best suitable for dataset having high number of features. In multiple r[16,17]searches$ it has been shown that SVM can perform better than different classification algorithms.

Consider a collection of samples $((z_1, x_1), (z_2, x_2), \ldots, (z_l, x_l))$ where $z_l \in R^m$, and $x_l \in (-1, +1)$. Let the decision functions be $sgn((W \bullet \chi) + b)$, where $(W \bullet \chi)$ indicates the product that is calculated using W and χ. Hence, the decision function $g_{w,b}$ have the properties as shown in Eq.(1).

$$\xi_t(j) = \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\alpha_T(q_F)}\ for\ all\ t,\ i\ and\ j \tag{1}$$

In most cases, the hyperplane used for separation does not exist. In order to make the possibilities to violate Eq.(1), slack variables are necessarily to be introduced which are $\xi_j \geq 0, j = 1, \ldots, l$. Hence, the issues in SVM can be illustrated as Eq.(2)

$$minimize \phi(W, \xi) = (W \bullet W) + B\sum_{j=1}^{l}\xi_j$$
$$Subject\ to\ x_j((W \bullet \chi) + b) \geq 1 - \xi_j, 1 \leq j \leq l \tag{2}$$

The minimization issue mentioned in Eq.(2) is constrained-quadratic-programming issue, that can be formulated as convex constrained-quadratic-programming issue as shown in Eq.(3)

$$maximize \sum_{j=1}^{l}c_j - \frac{1}{2}\sum_{j=1}^{l}\sum_{k=1}^{l}c_j c_k x_j x_k (z_j \bullet z_k)$$
$$Subject\ to\ \sum_{j=1}^{l}x_j\alpha_j = 0, 0 \leq \alpha_j \leq B\ (j = 1, \ldots, l) \tag{3}$$

where Lagrange multipliers are indicated as α_j, the parameter B is used to assign a penalty for the misclassification. As a result of solving the Eq.(3), it provides a option to form a decision function and it is shown in Eq. (4)

$$g(z) = sgn\left(\sum_{j=1}^{l} \alpha_j x_j (z_j \bullet z) + d\right\} \tag{4}$$

where bias term is indicated as d. A tiny fraction portion of coefficients α_j is considered as nonzero. The equivalent pairs of entries are treated as support vectors and it defines the function used for taking decision.

Eq.(4) provides a way to generalize the issue in nonlinear case and it is possible to achieve by mapping problem data into increased space of dimensional feature (H) which is done by transforming $(z_j \bullet z_k)$ as $\phi(z_j)\phi(z_k)$. Mapping function is completely defined from definite kernel function $(z_j, z_k) = \phi(z_j)\phi(z_k)$. Hence, the function used for taking decision can be redrafted as

$$g(z) = sgn\left(\sum_{j=1}^{l} \alpha_j x_j L(z_j, z) + d\right\} \tag{5}$$

2.4 Combined HMM with SVM

The proposed new approach for classifying the medical data that ensemble HMM and SVM is discussed in this current section. Feature vectors are made to extract by HMM in the same dimension and it is made to embed in appropriate vector space. Feature vectors dimensions are necessarily to be reduced to solve the speed and size complexity in SVM. Generation of feature vectors is one of the important step in HMM.

Let L classes $h = (h_1, \ldots, h_L\}$ with its appropriate HMM set $\lambda' = (\lambda_1, \ldots, \lambda_L\}$, where $\lambda^j = \left(\lambda_1^j, \ldots, \lambda_{M_j}^j\right\}$, so the total count of HMM will be $M = \sum_{j=1}^{L} M_j$. The probabilities of $Q(P|\lambda)$ are calculated using HMM's $\lambda^j = \left(\lambda_1^j, \ldots, \lambda_{M_j}^j\right\}$. If $Q_{max_{k_j}} = \underset{1 \le k \le M_j}{max} Q\left(P|\lambda_k^j\right)$, then label h_j makes communication with the increased probability available in group. Hence, it will receive the feature vector $f = (V_1, \ldots, V_s, \ldots, V_N\}$ with HMM λ_k^j, where h_j and f are ensemble to form a new feature vector $\overrightarrow{f} = (h_j, f\}$. Newly obtained feature vectors are normalized using $\widetilde{\overrightarrow{f}} = \frac{\overrightarrow{f}}{\|f\|^2}$, and it becomes essential for SVM classification.

So far, HMM feature extraction was discussed and the remaining discusses the learning of HMM parameter λ in generating the set $P = \left(P^1, P^2, \ldots .P^L\right\}$. To resolve the issues in generating $P = \left(P^1, P^2, \ldots .P^L\right\}$, Baum-Welch concept is utilized parameter estimation is shown in Eq.(6) and Eq.(7):

$$\bar{C}_{jk} = \frac{\sum_{L=1}^{L} \sum_{S=1}^{S_{L-1}} \varepsilon_s^l (j, k)}{\sum_{L=1}^{L} \sum_{S=1}^{S_{L-1}} \gamma_s^l (j)} \tag{6}$$

$$\bar{d}_k (j) = \frac{\sum_{L=1}^{L} \sum_{S=1, P_S^L = V_n}^{S_L} \gamma_s^l (j)}{\sum_{L=1}^{L} \sum_{S=1}^{S_L} \gamma_s^l (k)} \tag{7}$$

where joint-event is indicated as $\varepsilon_s^l (j, k)$, and the state-variable is indicated as $\gamma_s^l (j)$. $\varepsilon_s^l (j, k)$ and $\gamma_s^l (j)$ are associated with observation sequence L.

Consider a pair of new feature vectors that is received from output of HMM. SVM task is made to describe (i) class in a binary manner, (ii) classifier B_{jk} for every distinct classes of j and k. Positive labels are allocated to j-th class, and negative labels are allocated to k-th class. The classification that involves decision function is given by Eq.(8)

$$g_{jk}\left(\widetilde{\overrightarrow{f}}\right) = \sum_{m}^{\Lambda} x_m^{h_j h_k} \alpha_m^{h_j h_k} \phi\left(\widetilde{\overrightarrow{f}}^{h_j h_k}, \widetilde{\overrightarrow{f}}\right) + d^{h_j h_k}$$
$$j \ne k, j = 1, \ldots, N \tag{8}$$

where Λ represents total count of j-th and k-th classes in training data. When an unknown sample is given as a input and if the decision function predicts the sample to class j, then classifier B_{jk} provides one vote for class j, else the vote is provided to class k. Once after receiving all the votes from the classifier, then unknown sample will belong to class having increased votes.

3 Z-Alizadehsani Dataset

Z-Alizadehsani dataset holds the record of 303 patients with 54 features. Individual features in dataset are considered as the indicator of CAD symptom based on the medical history. Some particular features are not used in diagnosing the CAD by data mining. Features of Z-Alizadehsani dataset are categorized into 4 groups, which are: (a) demographic, (b) symptom and examination, (c) ECG, and (d) laboratory and echo. Every individual patient's falls in two categories, that is either Normal or CAD. If the diameter range is greater than or equal to 50% then patient is treated as CAD patient, else the patient is treated as normal. Few features are utilized to confirm the history of (a) hypertension, (b) Diabetes Mellitus, (c) smoking, (d) smoking of cigarettes formerly, and (v) confirmation of heart disease with patients first-degree relatives.

4 About MATLAB

Matlab R2013a is used to evaluate the proposed work against the baseline schemes. Matlab act as a major tool in business analytics, text mining, medical image mining, and for most machine learning based algorithms. Because of its ease of use and user friendliness, it is being utilized in many engineering applications, edification, research, and guidance of proposing new tools. By default, Matlab is built with the functionality of providing huge inbuilt mathematical functions, which aims to resolve scientific oriented problems. Mostly Matlab is used to design, explore, and solve many iteration-based problems. Inbuilt applications and tools available in Matlab make design of predictive models in an accurate and rapid manner.

5 Performance Measures

Specificity, Sensitivity, and Accuracy are the default performance measures used in data mining for validation. Positive rate of true and false values are mostly significant in medical field. Precision, Recall and F-Measure are also considered to check the performance of classifiers. Confusion matrix is treated as a special type of table to provide visualization to algorithms performance. By considering the dual class problems (i.e., Class 1 and Class 2), matrixes will receive rows and columns in 2 ways, it is identified to count true positives (TP), true negatives (TN). false positives (FP), and false negatives (FN). These measures [7] are defined as follows:

TP — Accurately classified class 1 samples based on statistical way of measuring.
TN — Accurately classified class 2 samples based on statistical way of measuring.
FN — Inaccurately classified class 1 as class 2 samples based on statistical way of measuring.
FP — Inaccurately classified class 2 as class 1 samples based on statistical way of measuring.

By make use of above-mentioned measures, this research work performs the calculation of result values using benchmark performance metrics and it is defined as below:

$$
\begin{array}{ll}
\bullet\ Sensitivity = \frac{TP}{TP+FN} & \bullet\ True\ Positive\ Rate = \frac{TP}{TP+FN} \\
\bullet\ Specificity = \frac{TN}{TN+FP} & \bullet\ Precision = \frac{TP}{TP+FP} \\
\bullet\ Accuracy = \frac{TP+TN}{TP+TN+FP+FN} & \bullet\ Recall = \frac{TP}{TP+FN} \\
\bullet\ False\ Positive\ Rate = \frac{FP}{FP+TN} & \bullet\ F-Measure = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}
\end{array}
$$

6 Results and Discussion

This section discusses the results obtained by HMM-SVM against the previous methods, namely SMO[16], ACO-SVM[17] and CPCA-SVM[18]. From Figures 1, 2, 3, 4, 5 and 6, X-axis is plotted with corresponding metrics and Y-axis is plotted with result values in percentage.

6.1 True Positive and True Negative Analysis

Figure 1 compares the performance of HMM-SVM towards predicting TP and TN against SMO[16], ACO-SVM[17] and CPCA-SVM[18]. It is clear that the performance of HMM-SVM is better than the other methods. HMM-SVM performs the classifica-

tion based on selected features. The other methods perform classification by considering the records of dataset in a sequential manner. Corresponding result values of Fig 1 is shown in Table 1.

Table 1. Result values of true positive and true negative analysis

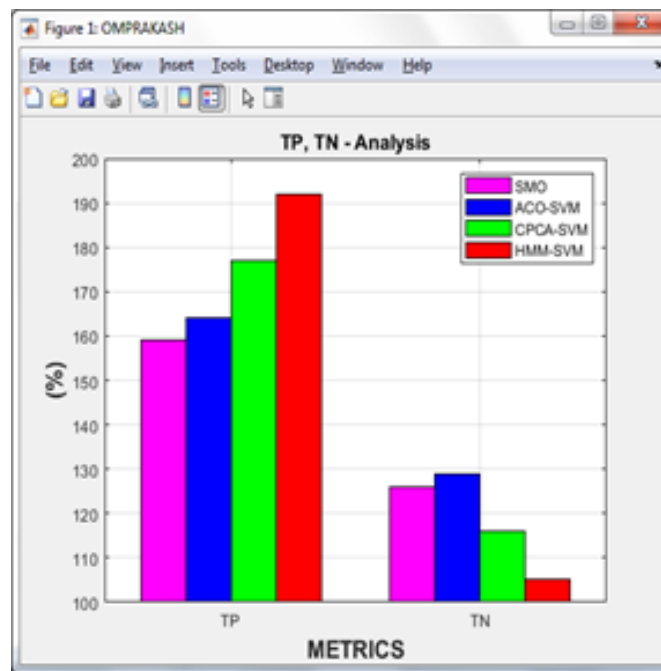| Metrics\ Classifiers	SMO[16]	ACO-SVM[17]	CPCA-SVM[18]	HMM-SVM
TP	159	164	177	192
TN	126	129	116	105

Fig 1. Analysis of TP and TN

6.2 False positive and false negative analysis

Figure 2 demonstrates that HMM-SVM has better performance than other considered methods. It is clear that SMO[16] has the worst performance where it provides more FP and FN. HMM-SVM too is providing FP and FN, but it is very low than SMO[16], ACO-SVM[17] and CPCA-SVM[18]. Because of considering the hidden state values, HMM-SVM has reduced FP and FN.

Table 2. Result values of false positive and false negative analysis

Metrics\ Classifiers	SMO[16]	ACO-SVM[17]	CPCA-SVM[18]	HMM-SVM
FP	8	4	3	2
FN	10	6	5	4

6.3 Sensitivity and specificity analysis

Figure 3 compares Sensitivity and Specificity results of HMM-SVM against SMO[16], ACO-SVM[17] and CPCA-SVM[18]. It is clear to understand that HMM-SVM has better results, where it gives enhanced results. *FoBa* process in HMM-SVM considers the previous state values in providing the results, hence the results are better than SMO[16], ACO-SVM[17] and CPCA-SVM[18].
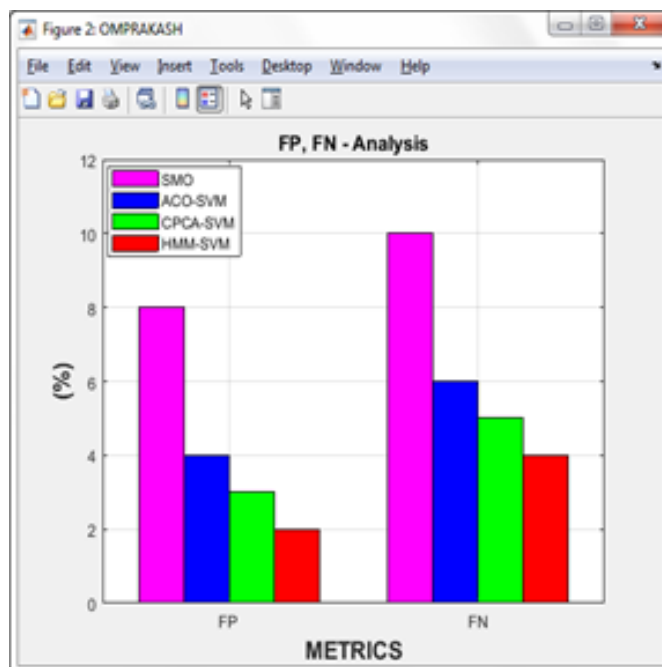
Fig 2. Analysis of FP and FN

Table 3. Result values of sensitivity and specificity analysis

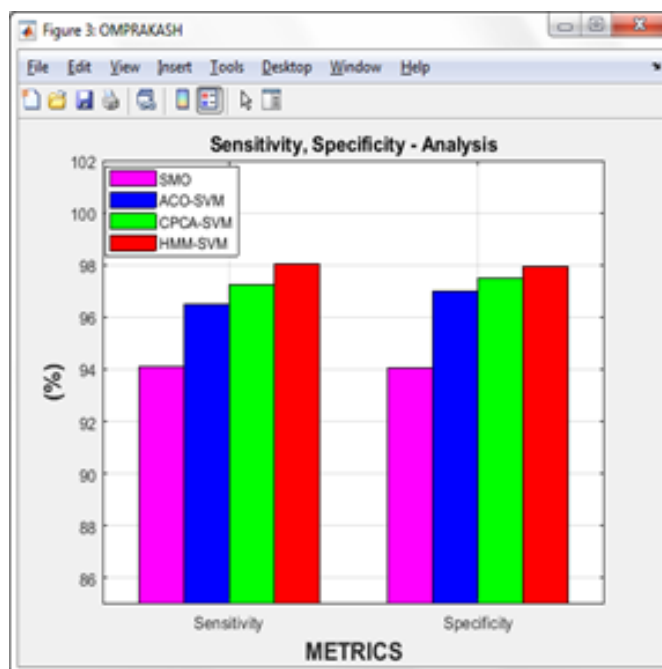| Metrics Classifiers	SMO [16]	ACO-SVM [17]	CPCA-SVM [18]	HMM-SVM
Sensitivity	94.08	96.47	97.25	98.02
Specificity	94.03	96.99	97.48	97.96

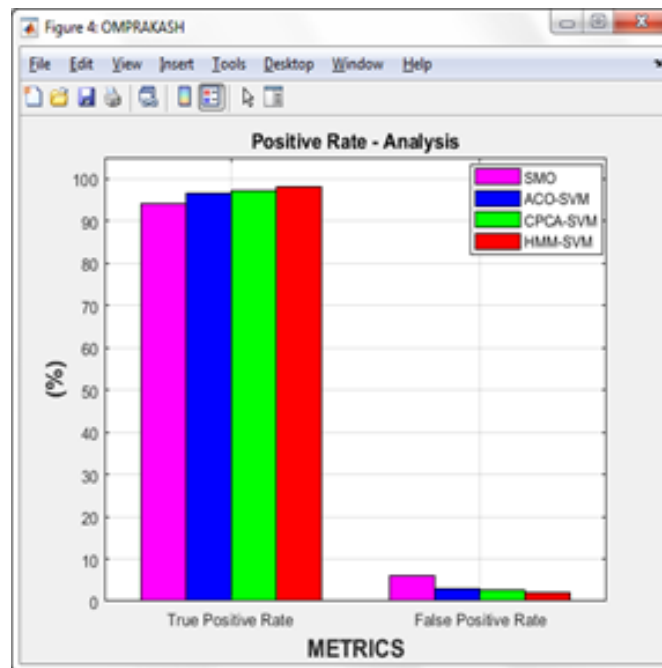**Fig 3.** Analysis of sensitivity and specificity

Fig 4. Analysis of positive rates

6.4 Positive rate analysis

Figure 4 illustrates the Positive Rate comparison of HMM-SVM against SMO[16], ACO-SVM[17] and CPCA-SVM[18]. It is evident that HMM-SVM has enhanced results than other methods. Utilizing the results obtained from newly derived features assist HMM-SVM to provide enhanced HMM-SVM. Other methods simply consider all features to find the results.

Table 4. Result values of positive rate analysis

Metrics Classifiers	SMO[16]	ACO-SVM[17]	CPCA-SVM[18]	HMM-SVM
True Positive Rate	94.08	96.47	97.25	97.96
False Positive Rate	5.97	3.01	2.52	1.87

6.5 Precision and recall analysis

Figure 5 compares the Precision and Recall results of HMM-SVM, SMO[16], ACO-SVM[17] and CPCA-SVM[18]. Expectation-maximization in HMM-SVM provides a way to give better results than other methods. Due to not giving priority to the expected results and processing the classification, SMO[16], ACO-SVM[17] and CPCA-SVM[18] is getting poor results in precision and recall.

Table 5. Result values of precision and recall analysis

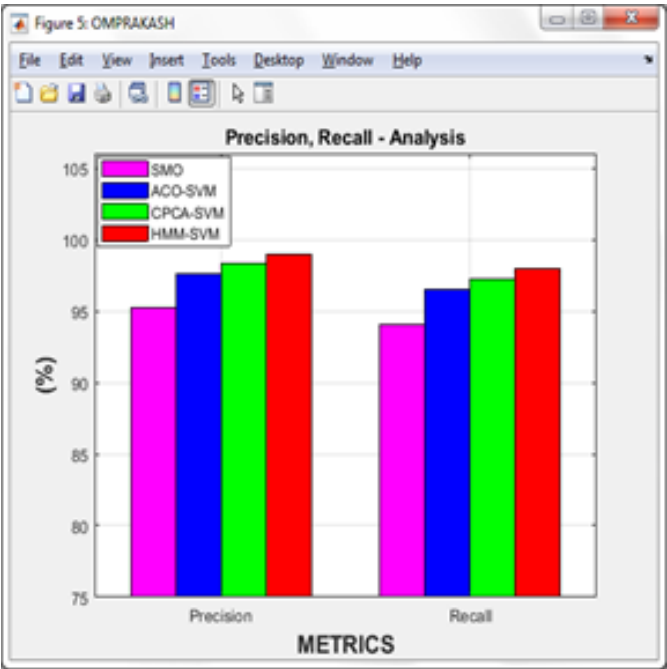| Metrics Classifiers	SMO[16]	ACO-SVM[17]	CPCA-SVM[18]	HMM-SVM
Precision	95.21	97.62	98.33	98.97
Recall	94.08	96.47	97.25	97.96

Fig 5. Analysis of precision and recall

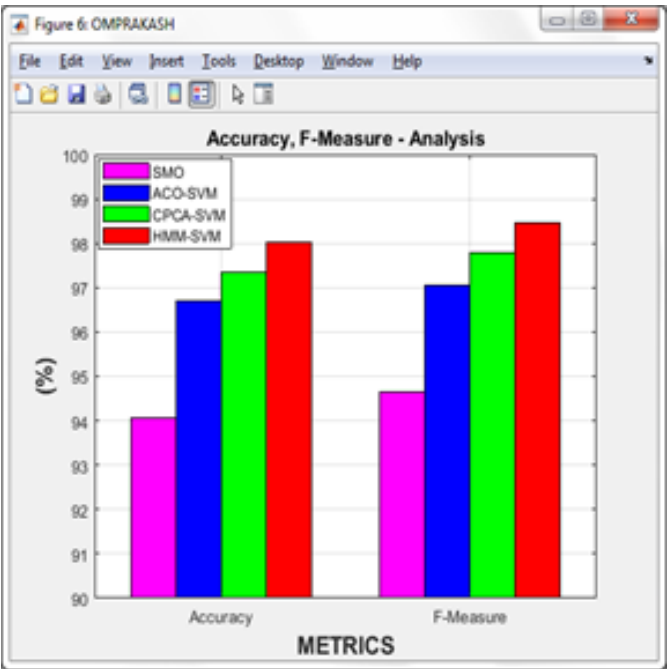**Fig 6.** Analysis of accuracy and F-Measure

6.6 Accuracy and F-Measure analysis

Figure 6 shows that HMM-SVM has out performing results than SMO[16], ACO-SVM[17] and CPCA-SVM[18], which indicates that HMM-SVM has better performance towards predicting CAD. It is because that HMM-SVM gives importance to the hidden state values and considering the features effectively, but due to not considering the hidden state values other methods are lacking in providing the better accuracy and f-measure.

Table 6. Result values of accuracy and F-Measure analysis

Metrics Classifiers	SMO[16]	ACO-SVM[17]	CPCA-SVM[18]	HMM-SVM
Accuracy	94.06	96.70	97.34	98.02
F-Measure	94.64	97.04	97.79	98.46

7 Conclusion

Novel classification approach ensembling HMM and SVM methods for medical data classification is proposed by us, and its core contribution is to ensemble the HMM and SVM to solve the issues that arise during classification. The proposed classifier is evaluated using Z-AlizadehSani dataset for classification accuracy to predict coronary artery disease. The result shows that the proposed classifier has outperformed the baseline schemes. Future enhancement of this research work can be focused bioinspired optimization based classification, which will result in achieving improved and increased classification accuracy.

References

1) Gupta D, Sundaram S, Khanna A, Hassanien AE, de Albuquerque VHC. Improved diagnosis of Parkinson's disease using optimized crow search algorithm. *Computers & Electrical Engineering*. 2018;68:412–424. doi:10.1016/j.compeleceng.2018.04.014.
2) Samanta P, Pathak A, Mandana K, Saha G. Classification of coronary artery diseased and normal subjects using multi-channel phonocardiogram signal. *Biocybernetics and Biomedical Engineering*. 2019;39(2):426–443. doi:10.1016/j.bbe.2019.02.003.
3) Malik SI, Akram MU, Siddiqi I. Localization and classification of heartbeats using robust adaptive algorithm. *Biomedical Signal Processing and Control*. 2019;49:57–77. doi:10.1016/j.bspc.2018.11.003.
4) Isler Y, Narin A, Ozer M, Perc M. Multi-stage classification of congestive heart failure based on short-term heart rate variability. *Chaos, Solitons & Fractals*. 2019;118:145–151. doi:10.1016/j.chaos.2018.11.020.
5) Mondéjar-Guerra V, Novo J, Rouco J, Penedo MG, Ortega M. Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. *Biomedical Signal Processing and Control*. 2019;47:41–48. doi:10.1016/j.bspc.2018.08.007.
6) Hu L, Yin C, Ma S, Liu Z. Rapid detection of three quality parameters and classification of wine based on Vis-NIR spectroscopy with wavelength selection by ACO and CARS algorithms. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2018;205:574–581. doi:10.1016/j.saa.2018.07.054.
7) Sannino G, Pietro GD. A deep learning approach for ECG-based heartbeat classification for arrhythmia detection. *Future Generation Computer Systems*. 2018;86:446–455. doi:10.1016/j.future.2018.03.057.
8) Potharaju SP, Sreedevi M, Ande VK, Tirandasu RK. Data mining approach for accelerating the classification accuracy of cardiotocography. *Clinical Epidemiology and Global Health*. 2019;7(2):160–164. doi:10.1016/j.cegh.2018.03.004.
9) Amutha AJ, Padmajavalli R, Prabhakar D. A novel approach for the prediction of treadmill test in cardiology using data mining algorithms implemented as a mobile application. *Indian Heart Journal*. 2018;70(4):511–518. doi:10.1016/j.ihj.2018.01.011.
10) Arabasadi Z, Alizadehsani R, Roshanzamir M, Moosaei H, Yarifard AA. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Computer Methods and Programs in Biomedicine*. 2017;141:19–26. doi:10.1016/j.cmpb.2017.01.004.
11) Thiyagaraja SR, Dantu R, Shrestha PL, Chitnis A, Thompson MA, Anumandla PT, et al. A novel heart-mobile interface for detection and classification of heart sounds. *Biomedical Signal Processing and Control*. 2018;45:313–324. doi:10.1016/j.bspc.2018.05.008.
12) Ata SK, Fang Y, Wu M, Li X, Xiao X. Disease gene classification with metagraph representations. *Methods*. 2017;131:83–92. doi:10.1016/j.ymeth.2017.06.036.
13) Sellami A, Hwang H. A robust deep convolutional neural network with batch-weighted loss for heartbeat classification. *Expert Systems with Applications*. 2019;122:75–84. doi:10.1016/j.eswa.2018.12.037.
14) Chabchoub S, Mansouri S, Salah RB. Detection of valvular heart diseases using impedance cardiography ICG. *Biocybernetics and Biomedical Engineering*. 2018;38(2):251–261. doi:10.1016/j.bbe.2017.12.002.
15) Shi H, Wang H, Huang Y, Zhao L, Qin C, Liu C. A hierarchical method based on weighted extreme gradient boosting in ECG heartbeat classification. *Computer Methods and Programs in Biomedicine*. 2019;171(10). doi:10.1016/j.cmpb.2019.02.005.
16) Alizadehsani R, Habibi J, Hosseini MJ, Mashayekhi H, Boghrati R, Ghandeharioun A, et al. A data mining approach for diagnosis of coronary artery disease. *Computer Methods and Programs in Biomedicine*. 2013;111(1):52–61. doi:10.1016/j.cmpb.2013.03.004.
17) Omprakash S, Ravichandran M. Ant Colony Optimization Based Support Vector Machine Towards Predicting Coronary Artery Disease. *International Journal of Recent Technology and Engineering*. 2019;7:210–215. Available from: https://www.ijrte.org/wp-content/uploads/papers/v7i5/E2044017519.pdf.
18) Omprakash S, Ravichandran M. Prediction of Coronary Artery Disease Using Core Principal Component Analysis Based Support Vector Machine. *International Journal of Scientific & Technology Research*. 2019;8:791–798.