

# Feature Selection based Least Square Twin Support Vector Machine for Diagnosis of Heart Disease

Divya Tomar and Sonali Agarwal

*Indian Institute of Information and Technology, Allahabad, India  
divyatomar26@gmail.com and sonali@iiita.ac.in*

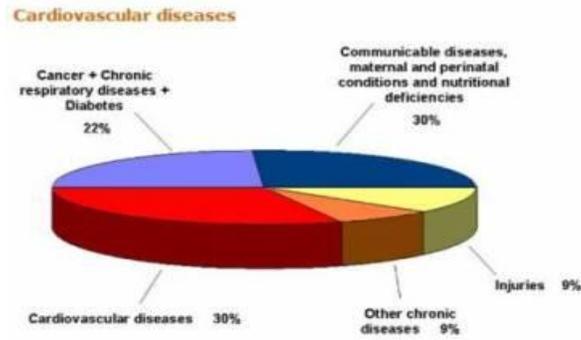
## **Abstract**

*It is evident from various researches that disease diagnosis using machine learning methods has been increasing rapidly. In this research work, feature selection based Least Square Twin Support Vector Machine (LSTSVM), which is a machine learning method, is used for diagnosis of heart diseases. In this approach F-score is used to calculate the weight of each feature and then features are selected according to their weight. The higher weight is assigned to the feature having high F-score. Grid search approach is also utilized to select the best value of classifier's parameters in order to enhance its performance. The heart-statlog disease dataset is used in this study, which is taken from the UCI repository. The performance of proposed model with different feature sets has been evaluated for different training-test datasets. The results indicate that LSTSVM model with 11 features has achieved highest accuracy. The results are very promising as compared to the other approaches proposed earlier.*

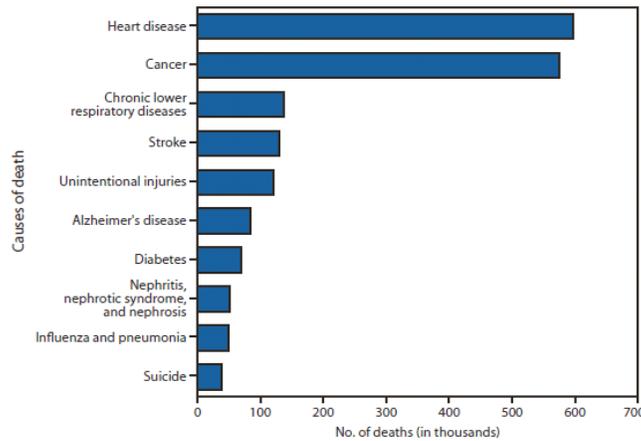
**Keywords:** Heart Disease, F-Score, LSTSVM

## **1. Introduction**

Cardiovascular Disease (CVD) is one of the important and crucial heart diseases which occur among both men and women and is also the leading cause of the deaths in the world. CVD may occur in several forms, for example abnormal blood pressure, valvular heart disease, cerebrovascular, coronary artery disease and heart failures. CVD affects the heart as well as its surrounding blood vessels and it can be caused due to abnormalities in the heart muscles. Fibrous elements and inflammatory cells increase the hardness and narrowness of the blood vessels. Fatty deposits are also harmful and responsible for CVD [1]. So, it is important to identify the root causes responsible for the heart diseases. So their remedy could also be planned by using suitable methods. Figure 1 represents the pie chart of various diseases released by World Health Organization in 2008. It was estimated that 30% of all global death occur due to heart problems. Figure 2 shows the number of deaths in thousands due to different diseases in United States (US). It is clearly visible that the heart diseases are the leading cause of the deaths in US.



**Figure 1. Status of Cardiovascular disease-2008 (World Health Organization) [1]**



**Figure 2. Number of deaths and ten leading causes recorded in United States, 2010 [2]**

In India, heart disease is observed as prime death cause in both rural and urban areas. Therefore it is urgently required to identify factors responsible for heart diseases and also to develop an effective system for heart disease diagnosis. In present era, data mining is becoming popular in medical field. Medical data mining has great capability of exploring unknown information or patterns from the dataset of medical domain, which is further utilized for disease diagnosis [3]. Keles et al. and Temurtas have developed a thyroid disease diagnosis system using Neuro Fuzzy and Neural Network approach respectively [4-5]. The research work consist a comparative analysis of the proposed system with other existing approaches. A diagnosis system for breast cancer diseases are also developed by using machine learning approaches [6-7]. Chen et al. have developed the breast cancer diagnosis model by using Support Vector Machine (SVM) and rough set based feature selection approach [6]. While the diagnostic model for breast cancer patients was developed by Polat et al. by using Least Square SVM [7]. Temurtas et al. also performed a comparative study on the diabetes diagnosis model [8]. It is noticed that in all existing research, proposed models have been validated by using confusion matrix, accuracy, sensitivity, specificity etc. This research proposes a heart disease diagnosis system by using Least Square Twin Support Vector Machine. For significant feature selection, this study has utilized F-score feature selection approach to identify the feature which responsible for heart diseases. The purpose of proposed model is to improve the performance of classifier as well as develop a less computationally

intensive model for the diagnosis of the heart diseases. To select best parameters for the LSTSVM model, grid search based approach has been used. There are various classification approaches such as Neural Network, SVM, Decision Tree, Naive Bayes, K-Nearest Neighbour etc. available in data mining which are used in the disease diagnosis [3]. Among them SVM is very popular due to its better learning ability and great performance. Vapnik et al. proposed SVM in 1990s which is based on the statistical learning theory and follows Structural Risk Minimization principle (SRM) [9-10]. SVM works well for both binary as well multi class problems [11-13]. It separates the data points of different classes by constructing hyper-planes that maximizes the separation between them. If the data are non-linearly separable then SVM maps the data points in high dimensional space using kernel function. In high dimension space, it constructs kernel surfaces to separate data points [14, 15]. Thus in this way it is suitable for both linearly separable and non-linearly separable data points. As compared to other machine learning approaches, SVM has several advantages. Firstly, for data classification it obtains unique global solution rather than local [12]. Secondly, it maximizes the generalization capability by minimizing the training error of data samples. Thus along with these advantages, researchers face various challenges with SVM. One challenge with SVM is its complexity because Quadratic Programming Problem (QPP) is difficult to solve. If 'l' indicates the size of training data samples then its computational complexity is  $O(l^3)$ , which is very expensive. Jayadeva et al. proposed a novel Twin Support Vector Machine (TSVM), a solution of binary classification to enhance the computational speed of SVM [16]. Rather than solving one complex QPP, TSVM solves two smaller sizes QPP and determines two non-parallel hyper-planes. The size of quadratic problem in TSVM is  $(l/2)$  and its computational complexity is  $O((l/2)^3)$  [16-18]. So, TSVM reduces the computational complexity by the factor of 4 i.e., it is 4 times faster as compared to the traditional SVM [16, 17]. This research work has developed a heart diseases diagnosis model using feature selection based Least Square Twin Support Vector Machine (LSTSVM). LSTSVM has many advantages such as better generalization capability and lesser computational time as compared to the traditional TSVM.

The paper is organized in following sections: the Section II gives an overview of Least Square Twin Support Vector Machine and Section III describes the proposed methodology. The Section IV and Section V discuss experiment results and conclusion of this research work respectively.

## 2. Least Square Twin Support Vector Machine

*For linearly separable Data:* Least Square TSVM (LSTSVM) was proposed by Kumar et al. to speed up the classification process [19]. It solves two linear equations rather than two complex quadratic programming problems. LSTSVM is faster than TSVM and has better generalization capability. Consider 'm' data points belong to positive class and 'n' data points belong to negative class. R represents training sample of 'k' dimension. Let  $X_1 \in R^{m \times k}$  is a matrix that represents the instances of positive class and  $X_2 \in R^{n \times k}$  is a matrix that represents the instances of negative class [18]. Two non-parallel hyper-planes in  $R^k$  are given below:

$$x^T w_1 + b_1 = 0 \quad \text{and} \quad x^T w_2 + b_2 = 0 \quad (1)$$

The primal problem of LS-TSVM is as follows:

$$\min(w_1, b_1, \xi) \quad \frac{1}{2} \|X_1 w_1 + e_1 b_1\|^2 + \frac{c_1}{2} \xi^T \xi$$

$$\text{s.t. } -(X_2 w_1 + e_2 b_1) + \xi = e_2 \quad (2)$$

and

$$\begin{aligned} \min(w_2, b_2, \eta) \quad & \frac{1}{2} \|X_2 w_2 + e_2 b_2\|^2 + \frac{c_2}{2} \eta^T \eta \\ \text{s.t. } & (X_1 w_2 + e_1 b_2) + \eta = e_1 \end{aligned} \quad (3)$$

where ' $\xi$ ' and ' $\eta$ ' are slack variable whereas  $e_1$  and  $e_2$  are the two vectors having all values as 1's with suitable dimensions. 'c1' and 'c2' are penalty parameters. Lagrangian of the above equation is given as:

$$L(w_1, b_1, \xi, \alpha) = \frac{1}{2} \|X_1 w_1 + e_1 b_1\|^2 + \frac{c_1}{2} \xi^T \xi - \alpha^T (-(X_2 w_1 + e_2 b_1) + \xi - e_2) \quad (4)$$

$$L(w_2, b_2, \eta, \beta) = \frac{1}{2} \|X_2 w_2 + e_2 b_2\|^2 + \frac{c_2}{2} \eta^T \eta - \beta^T ((X_1 w_2 + e_1 b_2) + \eta - e_1) \quad (5)$$

Where  $\alpha \in R^n$  and  $\beta \in R^n$  are the vectors of Lagrangian multiplier. KKT condition of above equation:

$$\frac{\partial L}{\partial w_1} = X_1^T (X_1 w_1 + e_1 b_1) + X_2^T \alpha = 0 \quad (6)$$

$$\frac{\partial L}{\partial b_1} = e_1^T (X_1 w_1 + e_1 b_1) + e_2^T \alpha = 0 \quad (7)$$

$$\frac{\partial L}{\partial \xi} = c_1 \xi - \alpha = 0 \quad (8)$$

$$\frac{\partial L}{\partial \alpha} = -(X_2 w_1 + e_2 b_1) + \xi - e_2 = 0 \quad (9)$$

Equation 6 and 7 when combined gives the following equation:

$$\begin{bmatrix} X_1^T \\ e_1^T \end{bmatrix} [X_1 \ e_1] \begin{bmatrix} w_1 \\ b_1 \end{bmatrix} + \begin{bmatrix} X_2^T \\ e_2^T \end{bmatrix} \alpha = 0 \quad (10)$$

Let  $H = [X_1 \ e_1]$  and  $G = [X_2 \ e_2]$ . Using equation 8, 9 and 10 the following values are obtained:

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = - \left( G^T G + \frac{1}{c_1} H^T H \right)^{-1} G^T e_2 \quad (11)$$

and

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = \left( H^T H + \frac{1}{c_2} G^T G \right)^{-1} H^T e_1 \quad (12)$$

After solving above two equations, we obtain weights and biases for the calculation of two non-parallel hyper-planes by using equation 1 and assign new data into a class as follows:

$$\text{class } i = \arg \min (j = 1, 2) \frac{|w_j^T x + b_j|}{\|w_j\|} \quad (13)$$

*For non-linear separable Data:* LSTSVM is also useful for the classification of non-linearly separable data samples. Kernel functions such as Polynomial, Gaussian etc. are used for this purpose. LSTSVM for non-linear data samples is obtained as follows:

$$\begin{aligned} \min(\mu_1, \gamma_1, \xi) \quad & \frac{1}{2} \|K(X_1, D^T)\mu_1 + e\gamma_1\|^2 + \frac{c_1}{2} \xi^T \xi \\ \text{s.t.} \quad & - (K(X_2, D^T)\mu_1 + e\gamma_1) = e - \xi \end{aligned} \quad (14)$$

and

$$\begin{aligned} \min(\mu_2, \gamma_2, \eta) \quad & \frac{1}{2} \|K(X_2, D^T)\mu_2 + e\gamma_2\|^2 + \frac{c_2}{2} \eta^T \eta \\ \text{s.t.} \quad & (K(X_1, D^T)\mu_2 + e\gamma_2) = e - \eta \end{aligned} \quad (15)$$

where  $D=[X_1 \ X_2]^T$ . After substituting  $P= [K(X_1, D^T) \ e]$ ,  $Q= [K(X_2, D^T) \ e]$ , We can further achieve :

$$\begin{bmatrix} \mu_1 \\ \gamma_1 \end{bmatrix} = -(Q^T Q + \frac{1}{c_1} P^T P)^{-1} Q^T e \quad (16)$$

$$\begin{bmatrix} \mu_2 \\ \gamma_2 \end{bmatrix} = (P^T P + \frac{1}{c_2} Q^T Q)^{-1} P^T e \quad (17)$$

Following are the kernel generated surfaces instead of planes:

$$K(x^T, D^T)\mu_1 + \gamma_1 = 0 \text{ and } K(x^T, D^T)\mu_2 + \gamma_2 = 0 \quad (18)$$

After solving equation 16 and 17, the value of  $\mu_1, \mu_2, \gamma_1$  and  $\gamma_2$  are obtained and further used to generate kernel generated surfaces. The new data point assigned to a particular class is evaluated as follows:

$$\text{class}(i) = \operatorname{argmin}(j = 1, 2) \frac{|x^T \mu_j + \gamma_j|}{\|\mu_j\|} \quad (19)$$

For every data point its distance from kernel surfaces is calculated and respectively assigned to a class having lesser distance.  $x_i$  and  $x_j$  are the vectors in the input space and  $K_G$  is a Gaussian kernel function which is represented as:

$$K_G = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (20)$$

Above stated equations are used to develop a LSTSVM classifier model. Feature selection and Grid Search approach are used with this existing classifier model which further enhances its performance for disease diagnosis.

### 3. Methodology and Experiments

#### 3.1 Heart statlog-Dataset

Heart-statlog data is collected from UCI machine repository database [20]. Researchers commonly used this data for heart disease diagnosis, so we can easily compare the performance of our proposed method with other existing methods. There are 270 data samples in Heart Statlog dataset which contains the information regarding presence and absence of

heart disease [20]. Heart-statlog dataset contains 13 features. The detail of dataset is given in Table 1.

**Table 1. Details of dataset**

Attributes	Type
Age	Real
Sex: Either male(0) or Female(1)	Binary
Chest Pain type	Nominal
Resting blood pressure	Real
Serum cholestoral in mg/dl	Real
Fasting blood sugar > 120 mg/dl	Binary
Resting electrocardiographic results (values 0,1,2)	Nominal
Maximum heart rate achieved	Real
Exercise induced angina	Binary
Oldpeak = ST depression induced by exercise relative to rest	Real
The slope of the peak exercise ST segment	Ordered
Number of major vessels (0-3) colored by flourosopy	Real
Thal: 3 = normal; 6 = fixed defect; 7 = reversable defect	Nominal
Class: Presence(1) and absence of heart disease	Binary

### 3.2 Feature Selection

Feature selection is one of the significant issues in the construction of classification model. Feature selection lessens the number of input features and selects the relevant features for a classifier to improve its predictive performance. Selection of relevant feature subset results lower test and diagnosis cost which play important role in the area of medical diagnosis. F-score is one of the simple and significant feature selection technique which is mostly used in machine learning [21]. It finds out how much the two sets of real numbers vary from each other. Let ‘m’ indicate positive samples and ‘n’ indicate the negative samples. Let  $x_k$  is any training vector. The F-score of  $i^{th}$  feature is [21][22]:

$$F(i) = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{m-1} \sum_{k=1}^m (x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n-1} \sum_{k=1}^n (x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \quad (21)$$

Where  $\bar{x}_i$ ,  $\bar{x}_i^{(+)}$  and  $\bar{x}_i^{(-)}$  represent the mean of the total  $i^{th}$  features, mean of positive  $i^{th}$  feature and mean of negative  $i^{th}$  feature respectively.  $x_{k,i}^{(+)}$  and  $x_{k,i}^{(-)}$  indicate  $i^{th}$  feature of  $k$ -positive and  $k$ -negative samples correspondingly. The larger value of F-score indicates that the corresponding feature is more discriminative or highly significant [21, 22].

### 3.3 Setting Model parameters

Setting of initial parameters for a classifier is one of the significant tasks because parameters also affect the performance of a classifier. LSTSVM has some initial parameters such as penalty parameters  $c_1$ ,  $c_2$  and kernel function parameter sigma which need to be initialized with some value. In this paper Gaussian Kernel function has been used. The proper setting of these parameters can enhance the accuracy of LSTSVM. The proposed research work has implemented the Grid Search in order to find out the appropriate value of the

parameters under consideration. It is also observed that Grid Search is an effective method for the best parameters selection in median size problem [21, 22]. Grid Search has been performed using k-fold cross validation method in order to find out the best value of penalty parameters and sigma. In this approach, the different values of the above mentioned parameters are tested. The selection is based on the basis of the value having best cross validation performance. The process of selecting the best parameters is given as follows:

Step 1: Set the grid search space of parameters with  $\log_2 C \in \{-5, -4, \dots, 10\}$  and  $\log_2 \text{sigma} \in \{-5, -4, \dots, 10\}$ .

Step 2: For each values of above mentioned parameters, k-fold cross-validation is performed.

Step 3: The parameter with the best classification accuracy is chosen and the predictive model is developed using this parameter.

### 3.4 LSTSVM model with Grid Search and Feature Selection

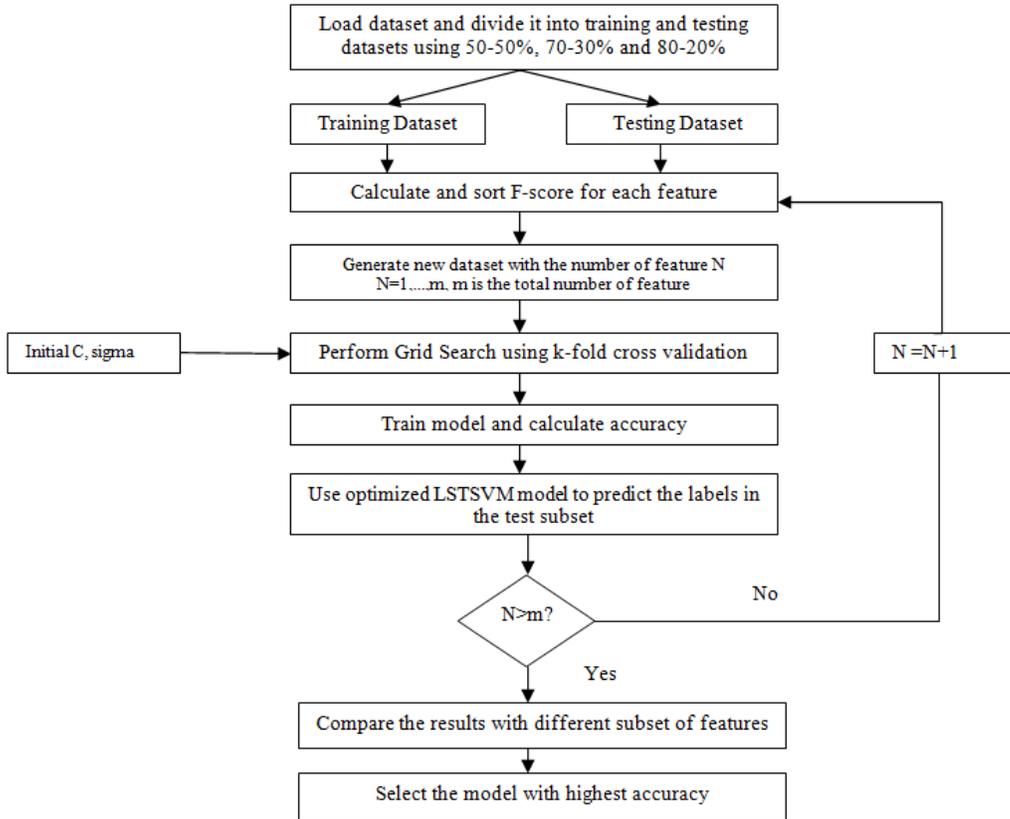
Figure 3 shows the block diagram of the proposed Feature Selection based LSTSVM model. This model uses F-score for selecting significant input features and Grid Search method to obtain the best parameters for predictive model. Initially the dataset is divided into training and testing dataset using 50-50%, 70-30% and 80-20% partitions. Then the F-score is calculated for each feature, these features are sorted as per their F-score. In the next stage, a new training subset is generated with a number of 'N' features, where  $N=1, \dots, m$  and  $m$  represents total number of features. After the subset generation, Grid Search is performed using 10-fold cross validation in order to get the best value of initial parameters. In the next stage, subset is trained using best parameter values. The class labels of the test subset are determined using this optimized LSTSVM model. These steps are carried out so as to obtain all feature as per their F-score values in the training subset. Following are steps of feature selection based LSTSVM model:

Step 1: Load dataset and divide the dataset into training and testing using 50-50%, 70-30% and 80-20% partitions.

Step 2: Calculate F-score for each feature and then sort them as per their F-score values.

Step 3: For N feature,  $N=\{1, 2, \dots, m\}$ , where 'm' indicates total number of features, keep first N features as per their sorted F-scores. Then obtain best parameters value using grid-search. Measure the accuracy of LSTSVM using 10-fold cross validation.

Step 4: Choose N feature for which model obtained high accuracy.



**Figure 3. Block Diagram of Proposed Model**

### 3.4 Measures for Performance Evaluation

The performance of different classifiers is evaluated using confusion matrix [17]. Confusion Matrix stores the details of actual and predicted class (estimated by the classifier) in tabular form as shown in Table 2.

**Table 2. Confusion Matrix**

Actual Class ↓	Predicted Class →	
	Yes	No
Yes	True Positive (TP)	False Negative (FN)
No	False Positive (FP)	True Negative (TN)

This paper evaluates the performance of proposed methodology using accuracy, specificity, sensitivity and geometric mean. The correct prediction made by a classifier in proportion to total number of prediction determines its accuracy which is formulated as:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (22)$$

Sensitivity of a classifier is evaluated by the percentage of positive values that are recognized accurately. It is also known as true positive rate and formulated as:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (23)$$

Specificity of a classifier is estimated by the percentage of negative values that are recognized correctly by the classifier. It is also known as true negative rate and formulated as:

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}) \quad (24)$$

Positive and Negative Predictive values are also used to measure the performance of a classifier and calculated as:

$$\text{Positive Predictive Value} = \text{TP} / (\text{TP} + \text{FP}) \quad (25)$$

$$\text{Negative Predictive Value} = \text{TN} / (\text{FN} + \text{TN}) \quad (26)$$

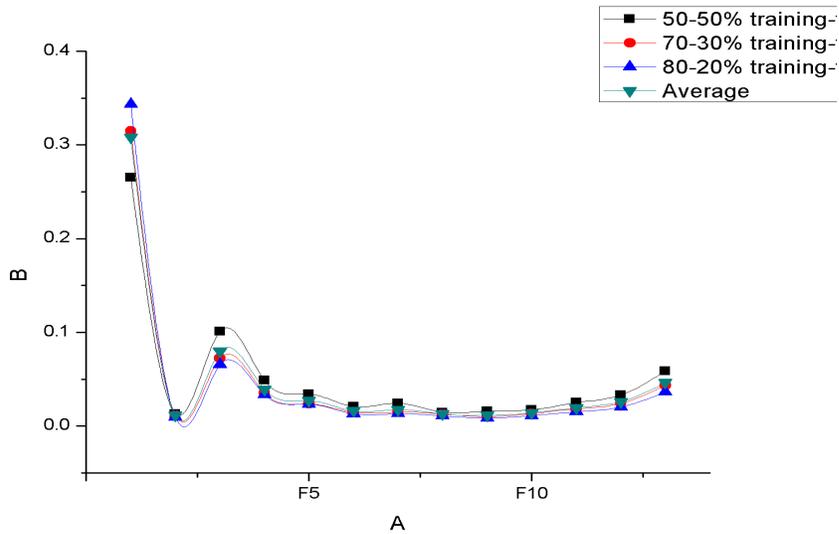
#### 4. Results and Discussion

The experiment has been performed on Heart-Statlog data which is taken from UCI repository in windows 7 64-bit operating system using matlab R2012a. The performance of proposed model is measured using equation 22-26. F-score is used to evaluate the importance of each feature of heart disease dataset and Grid Search approach is used to select best parameters for LSTSVM. F-score value of each feature is represented in Table 3 as follows:

**Table 3. F-score of features**

Feature No.	50-50% training-test partition	70-30% training-test partition	80-20% training-test partition	Average
F1	0.2653	0.3149	0.3434	0.3079
F2	0.0131	0.0116	0.0095	0.0114
F3	0.1012	0.0725	0.0655	0.0797
F4	0.0488	0.0353	0.0334	0.0392
F5	0.0343	0.0238	0.0235	0.0272
F6	0.0209	0.0148	0.0132	0.0163
F7	0.0244	0.0150	0.0133	0.0175
F8	0.0147	0.0129	0.0108	0.0128
F9	0.0161	0.0101	0.0087	0.0116
F10	0.0174	0.0134	0.0112	0.0140
F11	0.0251	0.0183	0.0154	0.0196
F12	0.0332	0.0243	0.0203	0.0260
F13	0.0589	0.0434	0.0365	0.0462

Figure 4 also indicates the importance of each feature on the basis of their F-score value. In figure, Y-axis 'B' indicates the F-score while X-axis 'A' indicates the features. F-score of each feature is calculated using 50-50%, 70-30% and 80-20% training test partitions.



**Figure 4. F-score value of each feature on the basis of different training-testing partition**

There are 13 features in heart-statlog dataset. The descending order of features sequence as per their F-score are F1, F3, F13, F4, F5, F12, F11, F7, F6, F10, F8, F9 and F2. In Table 4 thirteen models are shown which are generated with different number of features. We use them to obtain efficient LSTSVM model for heart disease diagnosis.

**Table 4. Thirteen LSTSVM classification models with different feature subset on the basis of F-score value**

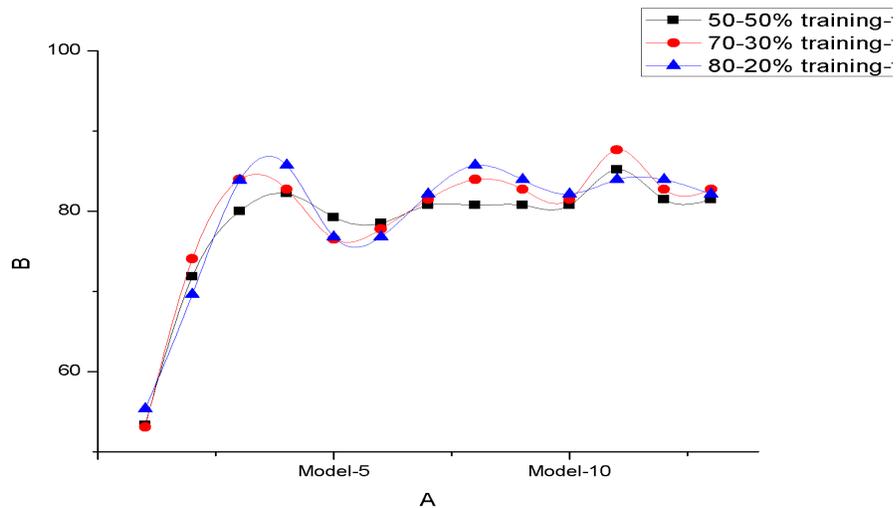
Model	Features
Model -1	F1
Model-2	F1, F3
Model-3	F1, F3, F13
Model-4	F1, F3, F13, F4
Model-5	F1, F3, F13, F4, F5
Model-6	F1, F3, F13, F4, F5, F12
Model-7	F1, F3, F13, F4, F5, F12, F11
Model-8	F1, F3, F13, F4, F5, F12, F11, F7
Model-9	F1, F3, F13, F4, F5, F12, F11, F7, F6
Model-10	F1, F3, F13, F4, F5, F12, F11, F7, F6, F10
Model-11	F1, F3, F13, F4, F5, F12, F11, F7, F6, F10, F8
Model-12	F1, F3, F13, F4, F5, F12, F11, F7, F6, F10, F8, F9
Model-13	F1, F3, F13, F4, F5, F12, F11, F7, F6, F10, F8, F9, F2

The performance of each model is evaluated for each training-test dataset partitions and is shown in Table 5 and Figure 5.

**Table 5. Comparison of accuracies of each model**

Model	Accuracy %		
	50-50% training-test partition	70-30% training-test partition	80-20% training-test partition
Model -1	53.33	53.09	55.36
Model-2	71.85	74.07	69.64
Model-3	80.00	83.95	83.83
Model-4	82.22	82.72	85.71
Model-5	79.26	76.54	76.79
Model-6	78.52	77.78	76.79
Model-7	80.74	81.48	82.14
Model-8	80.74	83.95	85.71
Model-9	80.74	82.72	83.93
Model-10	80.74	81.48	82.14
Model-11	85.19	87.65	83.93
Model-12	81.48	82.72	83.93
Model-13	81.48	82.72	82.14

It is clear from the Table 5 and Figure 5 that model-11 achieved high accuracy as compared to other models, *i.e.*, the performance of LSTSVM model using 11 features is better as compared to other models. That means only 11 features - age, chest pain, thal, blood pressure, cholestoral, major vessels, slope of the peak exercise ST segment, ECG, blood sugar, oldpeak and maximum heart rate are more relevant to the diagnosis of the heart disease in a patient.



**Figure 5. Accuracy Comparison of different models**

Other classifier's performance measurement parameters for model-11 are also indicated in Table 6.

**Table 6. Performance measurement of Model-11**

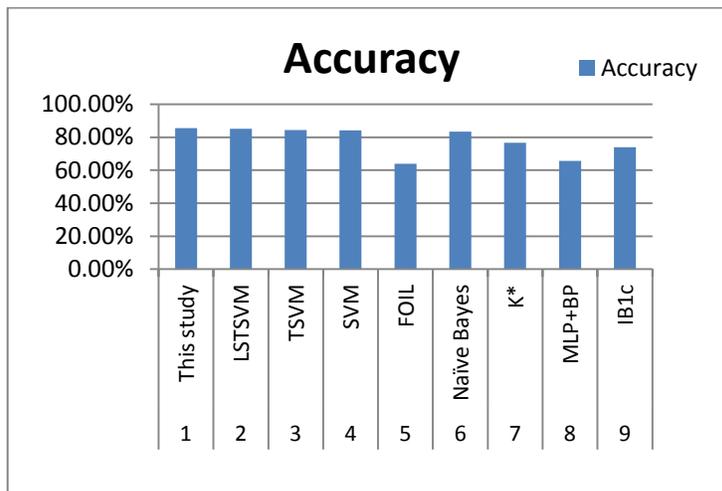
Parameters	50-50% training-test partition	70-30% training-test partition	80-20% training-test partition
Sensitivity	0.7759	0.8571	0.8462
Specificity	0.9091	0.8913	0.8333
Positive Predictive value	0.8654	0.8571	0.8148
Negative Predictive Value	0.8434	0.8913	0.8621

In this paper, we also analyzed the performance of proposed feature selection based LSTSVM model with other classifier model proposed earlier for Heart-statlog dataset. Table 7 and Figure 6 indicate the performance comparison of proposed model with other models.

**Table 7. Performance comparison of proposed model with other models**

S.No.	Heart Disease Diagnosis Model	Accuracy
1.	This study	<b>85.59%</b>
2.	LSTSVM	85.18%
3.	TSVM	84.44%
4.	SVM	84.07%
5.	FOIL	64.00%
6.	Naïve Bayes	83.36%
7.	K*	76.70%
8.	MLP+BP	65.60%
9.	IB1c	74.00%

The average accuracy of proposed model is 85.59% which is higher as compared to other models proposed earlier.



**Figure 6. Accuracy comparison of different models**

## 5. Conclusion

This research has developed a predictive model for heart disease diagnosis using Feature Selection based LSTSVM. The selection of significant features improves the accuracy of a classifier, so instead of taking all features we consider only the relevant features. This results in a classifier model which is more predictive and involves lesser computation. It is observed that model with 11 features has achieved the highest classification accuracy as compared to the other models. Heart statlog dataset contains two different classes: the presence and the absence of heart disease. So, proposed approach works well only for the classification of two classes. In future a classifier model for more than two classes could also be developed and its performance could be tested with high-dimensional healthcare data. Other Feature Selection and optimization approaches would also be utilized in future.

## References

- [1] Status of cardiovascular disease, <http://www.chemicalshealthmonitor.org/spip.php?rubrique10>, Accessed: (2013) December 26.
- [2] Number of deaths in United States from ten leading causes, <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6208a8.htm>, Accessed: (2013) December 26.
- [3] T. Divya and A. Sonali, "A survey on Data Mining approaches for Healthcare", *International Journal of Bio-Science and Bio-Technology*, vol. 5, no. 5, (2013), pp. 241-266.
- [4] A. Keleş and A. Keleş, "ESTDD: Expert system for thyroid diseases diagnosis", *Expert Systems with Applications*, vol. 34, no. 1, (2008), pp. 242-246.
- [5] F. Temurtas, "A comparative study on thyroid disease diagnosis using neural networks", *Expert Systems with Applications*, vol. 36, no. 1, (2009), pp. 944-949.
- [6] H. L. Chen, B. Yang, J. Liu and D. Y. Liu, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis", *Expert Systems with Applications*, vol. 38, no. 7, (2011), pp. 9014-9022.
- [7] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine", *Digital Signal Processing*, vol. 17, no. 4, (2007), pp. 694-701.
- [8] H. Temurtas, N. Yumusak and F. Temurtas, "A comparative study on diabetes disease diagnosis using neural networks", *Expert Systems with applications*, vol. 36, no. 4, (2009), pp. 8610-8615.
- [9] C. Cortes and V. Vapnik, "Support Vector Network", *Mach Learn*, vol. 20, (1995), pp. 273-297.
- [10] V. Vapnik, "The nature of statistical Learning, 2nd edn", Springer, New York, (1998).
- [11] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines, and other kernel-based learning methods", Cambridge University Press, (2000).
- [12] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines", Cambridge University Press, (2000).
- [13] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes", *Journal of Artificial Intelligence Research*, vol. 2, (1995), pp. 263-286.
- [14] B. Schölkopf, C. Burges and A. Smola, (Eds.), "Advances in Kernel Methods—Support Vector Learning", MIT Press, (1998).
- [15] C. W. Hsu, C. C. Chang and C. J. Lin, "A practical guide to support vector classification", Available from <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, (2003).
- [16] R. Jayadeva, R. Khemchandani and S. Chandra, "Twin Support vector Machine for pattern classification", *IEEE Trans Pattern Anal Mach Intell.*, vol. 29, no. 5, (2007), pp. 905-910.
- [17] Y. H. Shao, C. H. Zhang, X. B. Wang and N. Y. Deng, "Improvements on Twin Support Vector Machines", *IEEE transactions on Neural Network*, vol. 22, no. 6, (2011) June.
- [18] T. Divya and A. Sonali, "Predictive Model for diabetic patients using Hybrid Twin Support Vector Machine", 5th International Conferences on advances in communication Network and Computing (CNC-2014), In Press.
- [19] M. A. Kumar and M. Gopal, "Least squares twin support vector machines for pattern classification", *Expert Systems with Applications*, vol. 36, (2009), pp. 7535-7543.
- [20] Heart-statlog disease dataset, <http://archive.ics.uci.edu/ml/datasets/>, Accessed: (2013) December 1.
- [21] Y. W. Chen and C. J. Lin, "Combining SVMs with various feature selection strategies", Available from <http://www.csie.ntu.edu.tw/~cjlin/papers/features.pdf>, (2005).
- [22] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis", *Expert systems with applications*, vol. 36, no. 2, (2009), pp. 3240-3247.

## Authors



### **Divya Tomar**

She is a research scholar in Information Technology Division of Indian Institute of Information Technology (IIIT), Allahabad, India under the supervision of Dr. Sonali Agarwal. Her primary research interests are Data Mining, Data Warehousing especially with the application in the area of Medical Healthcare.



### **Dr. Sonali Agarwal**

**Dr. Sonali Agarwal** is working as an Assistant Professor in the Information Technology Division of Indian Institute of Information Technology (IIIT), Allahabad, India. Her primary research interests are in the areas of Data Mining, Data Warehousing, E Governance and Software Engineering. Her current focus in the last few years is on the research issues in Data Mining application especially in E Governance and Healthcare.