

SVM-PSO based Feature Selection for Improving Medical Diagnosis Reliability using Machine Learning Ensembles

¹Indrajit Mandal**, ²Sairam.N

^{1,2}School of Computing

SASTRA University, India-613401

¹indrajit@cse.sastra.edu, ²sairam@cse.sastra.edu

Tel: +918438366442, +919894835107

Abstract:

Improving accuracy of supervised classification algorithms in biomedical applications, especially CADx, is one of active area of research. This paper proposes construction of rotation forest (RF) ensemble using 20 learners over two clinical datasets namely lymphography and backache. We propose a new feature selection strategy based on support vector machines optimized by particle swarm optimization for relevant and minimum feature subset for obtaining higher accuracy of ensembles. We have quantitatively analyzed 20 base learners over two datasets and carried out the experiments with 10 fold cross validation leave-one-out strategy and the performance of 20 classifiers are evaluated using performance metrics namely accuracy (acc), kappa value (K), root mean square error (RMSE) and area under receiver operating characteristics curve (ROC). Base classifiers succeeded 79.96% & 81.71% average accuracies for lymphography & backache datasets respectively. As for RF ensembles, they produced average accuracies of 83.72% & 85.77% for respective diseases. The paper presents promising results using RF ensembles and provides a new direction towards construction of reliable and robust medical diagnosis systems.

Keywords:

Rotation forest, support vector machines, particle swarm optimization, ensembles, medical diagnosis systems.

1. Introduction

Machine learning ensembles are being employed successfully for designing CADx systems. They are initially trained with past diagnosed patient's data from a medical centre. While in testing stage, these algorithms are deployed to serve the medical physicians in performing diagnosis of new patients [1-5]. Therefore, in this regard, the success of decision analysis is dependent on the performance of algorithms to indicate the health status correctly.

The performance of Computer Aided Diagnosis systems may be substantially improved with higher accurate machine learning ensembles. The prediction accuracy of such strategic methods can be enhanced with two steps: (i) performing feature selection on the dataset (ii) construct multiple classifier systems to achieve better accuracy. The classification accuracy is reduced by the presence of irrelevant features in the dataset as it leads to over-fitting where in noise or irrelevant features and also due to finite size of the training data. In the literature, two kinds of used feature selection strategies: (i) filter approaches (ii) wrapper method. The wrapper technique finds feature subsets based on the performance of a preselected classification algorithm on a training data set, whereas the filter method depends on the properties of the features to be selected to form the best feature subset. The selection of a subset of relevant features, both strategies uses a search method such as individual ranking, genetic search, forward search and backward search etc [6-8].

One of the powerful methods in machine learning research to enhance predictive accuracy of base learner classifiers is to construct multiple classifier ensembles. An ensemble classifier may consist of n base classifiers that learn a targeted mathematical function by compounding their predictions mutually. Various ensemble learning strategies found in machine learning literature are composite classifier systems, mixture of experts, consensus aggregation, dynamic classifier selection, classifier fusion and committees of neural networks. Several Computer Aided Diagnosis system applications use classifier ensembles (especially Rotation Forest algorithm) to increase accuracy of convenient classifiers [8]. Besides the choice of base learner classifiers, the predictive performance of multiple classifier system is largely influenced by the degree of diversity of community of base learners constituting the ensemble. The combination of diverse classifiers having different configurations may lead to higher accurate decisions [9,10].

1.1. Introduction to the proposed system

This paper presents a qualitative investigation for CADx that can assist the designing of decision making medical systems with enhanced reliability. In order to achieve that the proposed method of suitable feature reduction technique based on support vector machines (SVM) and its optimization by particle swarm optimization (PSO) is obtained and then the ensembles are constructed based on rotation forest (RF) ensemble. The details of the 2 benchmark datasets and feature selection method are briefly discussed in data section. The details of 20 base learners and construction of RF ensembles approaches are presented in method section. The experimental test results are presented in section 4. The inference and remarks are discussed in section 5.

2. Data

In this study, two clinical benchmark data namely lymphography and schizophrenic datasets are deployed for benchmarking purpose. First, the datasets are randomized in such a way that there is even distribution of all the class levels in the training set. Secondly, all the missing values are substituted in case of nominal and numeric attributes with their mode and mean respectively in the dataset. A brief summary of each dataset is provided below:

Lymphography: The lymphography dataset was obtained from the University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia by Dr Zwitter and Dr. Soklic. It contains 148 instances and 19 attributes including class. The aim of the medical study is to diagnose the

patients as normal, metastases, malign lymph and fibrosis. It includes features like lymphatics, lymph size, changes in lymph nodes etc.

Backache: The dataset is compiled by Dr. Chris Chatfield and related to "backache in pregnancy". It consists of 180 instances, 33 attributes and a binary class. The aim of the medical study is to assess the onset of backache in pregnant women.

2.1. Feature Selection using SVM-PSO strategy

In this paper, the feature selection is based on the support vector machines (SVM) optimized by particle swarm optimization (PSO) [11-13]. SVM classifier is a supervised learning algorithm based on statistical learning theory, whose aim is to determine a hyper plane that optimally separates two classes by using train data sets. Assume that a training data set $\{x_i, y_i\}_{i=1}^n$, where x is the input vector, and $y \in \{+1, -1\}$ is class label. This hyper plane is defined as $w \cdot x + b = 0$, where x is a point lying on the hyper plane, w determines the orientation of the hyper plane and b is the bias of the distance of hyper plane from the origin. The aim is to determine the optimum hyper plane. Particle Swarm Optimization (PSO) is a population-based stochastic optimization technique based on the simulation of the behavior of birds within a flock. The swarm is a population of particles. Each particle represents a potential solution to the problem being solved. The personal best (pbest) of a given particle is the position of the particle that has provided the greatest success (i.e. the maximum value given by the classification method used). The local best (lbest) is the position of the best particle member of the neighborhood of a given particle. The global best (gbest) is the position of the best particle of the entire swarm. The leader is the particle that is used to guide another particle towards better regions of the search space. The velocity is the vector that determines the direction in which a particle needs to "fly" (move), in order to improve its current position. The inertia weight, denoted by W , is employed to control the impact of the previous history of velocities on the current velocity of a given particle. The learning factor represents the attraction that a particle has toward either its own success ($C1$ - cognitive learning factor) or that of its neighbors ($C2$ - social learning factor). Both, $C1$ and $C2$, are usually defined as constants. Finally, the neighborhood topology determines the set of particles that contribute to the calculation of the lbest value of a given particle [18]. The algorithm PSO-SVM is mentioned below:

Algorithm: Feature selection using SVM-PSO

Initialize parameters: inertia, correction factor, maxVelocity, minVelocity

Load trained linear SVM and Read input data

Initialize particles to random positions (x, y)

iteration = 0, repeat = 0

while repeat < R do

for each particle do

 Update position

 Evaluate SVM's result for this position

 Update pbest

end for

Update gbest's index

Update velocities for each particle

iteration++

if gbest's position has not changed and iteration > K then

repeat++

```

end if
end while
gbest's position = current particle positions

```

3. Methods

Classifier ensembles are generally more accurate compared to a single base classifier. There are miscellaneous classifier ensemble models in machine learning literature such as boosting and bagging [14,15]. In bagging, diversity is provided by building classifiers independent from one another using a randomized heuristic. Diversity in bagging is provided with further randomization yielding Random Forest ensemble model. This ensemble model uses decision trees trained on bootstrap samples from the data set and it improves diversity with randomizing the feature choice at each node while constructing trees. Similar to Random Forest, Rotation Forest (RF) is built with independent decision trees. However in RF each tree is trained with complete data set with a rotated feature space. As the algorithm builds classifiers using hyperplanes parallel to the feature axes and a small rotation of the axes lead to diverse trees [1,8,17]. More explicitly, the structure of the RF algorithm is given as follows:

Rotation forest algorithm

Let X be the training sample set. There are L base classifiers D_1, \dots, D_L in a Rotation Forest. The following steps are processed for each of the base classifier D_i :

Step 1: Splitting feature set into subsets. Assume there are n features in the dataset X . Feature set, F , is separated into K disjointed subsets randomly so each feature subset has $M = n / K$ features. It is not necessary to choose K as a factor of n .

Step 2: Generating coefficients matrix. i denotes for the iteration number of base classifier to be trained, D_i and j^{th} subset of features to train this classifier is F_{ij} . Let X_{ij} be the part of X having the data that corresponds to F_{ij} features. From each X_{ij} some subset of class labels is selected randomly, then 75% of remaining X_{ij} is selected randomly again in order to generate another dataset X'_{ij} . Then coefficients matrix, C_{ij} , is generated by operation of a linear transformation on X'_{ij} . Coefficients of this matrix are $a^{(1)}_{ij}, \dots, a^{(Mj)}_{ij}$.

Step 3: Constructing a rotation matrix. Coefficients generated in the previous step are used to obtain R_i :

Step 4: Generating rearrange matrix. R_i is arranged to the feature sequence of original X dataset to generate R_{ai} , so the actual rotation matrix is obtained. When the dataset is rotated by this rotation matrix, X is multiplied by R_{ai} , training set for D_i classifier is obtained as XR_{ai} .

As for classifying an instance, x , confidence of each class label is calculated. x is assigned to class label having largest confidence. First $x' = xR_{ai}$ is generated. Assume $w = \{w_1, \dots, w_c\}$ are class labels and $d_{ij}(x')$ is the probability that class label of x is w_j as determined by D_i classifier.

Each base learner uses a different set of axes. That is, each feature is considered as an axis, the RF algorithm selects the axes randomly to rotate according to the algorithm parameter K . Axes are rotated by principal component analysis (PCA) [16, 17], which is a statistical method for reducing dimensions with a covariance analysis between features. PCA rotated our data set into a

different configuration, which is easier to classify. In this way, data is simpler; relationships between features are more discriminative. Using PCA, it is possible to rotate the axes of our multi-dimensional space to new positions (principal axes). So data are defined differently, as if they were not the same as before. Using PCA, our aim is not to reduce dimensions, but rather to rotate the axes in order to define each example in the data set in a different way. For each classifier, this rotation is performed with different subsets of features. In other words, each classifier is trained with the whole data set with different extracted features. Also each base classifier takes different subsets of instances having the selected features so that diversity, an important property of ensemble methods, could be achieved. Another contribution to diversity is that each classifier is created apart from each other, with randomly chosen initial weights. All principal components are taken into account, so that the accuracy of the system is not ignored while achieving diversity. Usually the diversity of base classifiers conflicts with the accuracy. However, if the base classifiers are accurate, diversity among them is low. If there is not enough diversity among the base classifiers, their combination will not produce an effective output. Thus, the optimum results can be reached only by an ensemble consisting of highly correct classifiers that disagree as much as possible. RF can avoid this problem efficiently. In this study, diversity is provided by three separate techniques in order to create an ensemble consisting of classifiers that disagree on their predictions. Firstly, the data set is rotated by the transformation matrix gained by PCA. Secondly, the base classifiers are constructed including different initial weights. And finally, each network is trained by using different portions of the training set as a rule of the RF algorithm.

3.1. Performance evaluation metrics

Performance of any classifier needs to be tested with some metric, to assess the result and hence the quality of the algorithm. In our study, to evaluate the results of the experiments of 20 machine learning algorithms and corresponding RF classifier ensembles, we utilized four widely used metrics, i.e. classification accuracy (ACC), Kappa Error (KE), area under the Receiver Operating Characteristic Curve (AUC) and root mean square error (RMSE).

Most of the CADx problems deal with two class predictions to map data samples into one of the groups, i.e. benign or malignant. For such a two-class problem, the outcomes are labeled as positive (p) or negative (n). The possible outcomes with respect to this classification scheme is frequently defined in statistical learning as true positive (TP), false positive (FP), true negative (TN) and false negative (FN). These four outcomes are connected to each other with a table that is frequently called as confusion matrix. For a binary classification scheme, confusion matrix is used to derive most of the well known performance metrics such as sensitivity, specificity, accuracy, positive pre-diction value, F-measure, AUC and ROC curve [1]. These metrics are evaluated using the confusion matrix outcomes, i.e. TP, FP, TN and FN predictive values [8].

Accuracy (Acc) is a widely used metric to determine class discrimination ability of classifiers, and it is calculated using equation (1)

$$\text{Accuracy (\%)} = \frac{TP + TN}{P + N} \quad (1)$$

It is one of primary metrics in evaluating classifier performances and it is defined as the percentage of test samples that are correctly classified by the algorithm.

Area under the curve (AUC), an important classification performance measure, is widely used to measure classifier performances with relevant acceptance. AUC value is calculated from the area under the ROC curve. ROC curves are usually plotted using true positives rate versus false positives rate, as the discrimination threshold of classification algorithm is varied. In this aspect, since a ROC curve compares the classifiers' performance across the entire range of class distributions and error costs; an AUC value is accepted to be a good measure for comparative performances of classification algorithms. In terms of TP, FP, TN and FN predictive values, AUC is calculated using Eq. (2).

$$\text{AUC} = \frac{1}{2} \left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}} \right) \quad (2)$$

Since, it is relatively difficult to compare performances of 40 algorithms (20 base classifiers and 20 classifier ensembles) for each dataset in terms of ROC curves; we prefer AUC values to ROC curves for the sake of convenience.

KE or Cohen's Kappa Statistics value is used to support the comparative performances of classifiers. Performance comparisons depending only on accuracy based metrics might produce misleading results. The cost of error must also be taken into account in such assessments. KE, in this aspect, is a good measure to inspect classifications that may be due to chance. In general, KE takes values between (-1,1). As the Kappa value calculated for classifiers approaches to '1', then the performance of the classifier is assumed to be more realistic rather than by chance. Therefore, in the performance analysis of classifiers, KE is a recommended metric to consider for evaluation purposes [] and it is calculated with Eq. (3).

$$\text{K} = (P_0 - P_c) / (1 - P_c) \quad (3)$$

Where P_0 = Total agreement probability.

P_c = agreement probability due to chance.

Root Mean Square Error (RMSE): It indicates the difference between the actual and expected output in supervised learning. It is highly sensitive to large magnitude errors rather than minor errors, i.e. if the difference in one part is very large then it shows greater effect on RMSE, rather than if small errors occurs across the entire computation. The training seems to be accurate if its value approaches to 0 and it is calculated as shown in equation (5)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^N (E_i - \hat{E}_i)^2} \quad (4)$$

As a testing method, we used leave-one-out strategy of 10-fold cross validation. In this scheme, the classification accuracy is computed 10 times, each time leaving out one of the sub-samples from the computations and using that sub-sample as a test sample for cross-validation. In this structure, each sub-sample is used 9 times in the learning sample and just once as the test sample. The average of these two accuracies is calculated to find the general accuracy. This kind of cross validation is known as the "hold out" method.

4. Results

In this section, the results of the experiments for lymphography and backache datasets are discussed in table 1 and 2 respectively. The results obtained from base classifiers and their rotation forest ensembles are compared using four metrics like accuracy, KE, RMSE and ROC. Our proposed method of ensemble has provided improved results in this work. For the lymphography dataset, the average accuracy is 79.961% and that of RF ensemble is 83.72%, so there is an average increase of 3.759%. Even the KE and ROC values are better obtained for the ensembles classifiers as compared to the base learners. For the backache dataset, the average accuracy is 81.718% and that of RF ensemble is 85.774%, so there is an increase of 4.056%. Also the kappa value and ROC is more in case of our proposed method. The RMSE value is less for the ensembles in either datasets.

Table shows the results obtained for lymphography dataset. All computations are with 95% confidence level.

	With all the attributes				Our proposed method			
	Acc %	KE	RMSE	ROC	Acc	KSE	RMS	RO
Bayes Network	85.81	0.7321	0.2428	0.916	85.13	0.715	0.241	0.92
Naive Bayes	83.10	0.6772	0.2666	0.908	86.48	0.736	0.249	0.90
SVM	79.72	0.5955	0.3184	0.793	83.10	0.661	0.255	0.90
Logistics	73.64	0.5034	0.3585	0.83	81.75	0.646	0.250	0.92
Multilayer Perceptron	84.45	0.6959	0.2625	0.92	87.83	0.762	0.228	0.93
KStar	85.13	0.7115	0.2466	0.907	85.81	0.722	0.241	0.92
CHIRP	82.43	0.652	0.2964	0.821	75.86	0.513	0.318	0.79
Decision Table	77.02	0.5677	0.3017	0.818	81.08	0.624	0.300	0.87
DTNB	73.64	0.5024	0.3072	0.829	80.40	0.628	0.266	0.90
FURIA	83.78	0.6848	0.2664	0.875	84.45	0.694	0.235	0.93
RIPPER	77.70	0.5725	0.3108	0.795	85.13	0.708	0.249	0.94
NNge	78.37	0.5812	0.3288	0.792	82.43	0.654	0.253	0.91
PART	76.35	0.5495	0.327	0.804	87.83	0.763	0.230	0.94
BFTree	76.35	0.5444	0.3061	0.815	83.10	0.663	0.253	0.93
Decision tree	75.67	0.5316	0.3097	0.712	78.37	0.565	0.293	0.90
J48	77.02	0.5736	0.3151	0.785	85.13	0.711	0.238	0.93
Functional trees	86.48	0.7418	0.2551	0.89	88.51	0.778	0.220	0.93
Random forest	85.81	0.7238	0.2534	0.905	86.48	0.736	0.236	0.94
Random tree	75	0.5321	0.3401	0.793	83.10	0.670	0.236	0.93
Fuzzy Rough set rules	81.75	0.6454	0.309	0.885	82.43	0.650	0.265	0.93
AVERAGE	79.961	0.6159	0.2961	0.839	83.72	0.680	0.253	0.91

Table shows the results obtained from backache dataset. All computations are obtained with 95% confidence level.

	With all the attributes				Our proposed method			
	Acc %	KE	RMSE	ROC	Acc	KSE	RMS	RO
Bayes Network	77.77	0.2044	0.4167	0.706	85.55	0.042	0.347	0.60
Naive Bayes	78.33	0.1894	0.4189	0.68	81.66	0.292	0.375	0.75
SVM	86.11	0	0.3727	0.5	86.11	0.103	0.360	0.57
Logistics	68.33	-	0.5561	0.526	87.77	0.301	0.321	0.73
Multilayer	78.88	0.1729	0.4206	0.609	84.44	0.185	0.351	0.61
KStar	81.11	-0.032	0.4035	0.637	86.11	0.224	0.371	0.51
CHIRP	83.88	0.0114	0.4014	0.504	86.11	0.147	0.344	0.69
Decision Table	85.55	0.1349	0.3428	0.687	87.22	0.130	0.332	0.69
DTNB	82.77	0.0379	0.3788	0.509	86.66	0.201	0.332	0.71
FURIA	85.55	0.0429	0.3748	0.584	85.55	0.174	0.348	0.64
RIPPER	82.77	-	0.3694	0.495	88.33	0.247	0.326	0.55
NNge	82.22	0.0287	0.4216	0.511	82.22	0.107	0.416	0.55
PART	83.33	0.1276	0.389	0.572	85	0.122	0.337	0.63
BFTree	85	-0.021	0.3617	0.472	85.55	0.134	0.339	0.58
Decision tree	86.11	0	0.3301	0.626	86.11	0.102	0.333	0.68
J48	79.44	0.062	0.397	0.541	87.22	0.252	0.334	0.63
Functional trees	81.11	0.1829	0.4246	0.635	87.22	0.286	0.321	0.73
Random forest	85.55	-	0.3531	0.649	85	0.198	0.359	0.56
Random tree	75	0.0647	0.4955	0.542	84.44	0.185	0.376	0.51
Fuzzy Rough set	85.55	0.0429	0.3531	0.564	87.22	0.130	0.336	0.60
AVERAGE	81.718	0.0682	0.3990	0.577	85.77	0.178	0.348	0.63

5. Conclusion

The major findings in this work are summarized in this section. The multiple classifier system (MCS) is best suitable for the design of robust CADx systems as even a minute increase in the efficiency is valuable for such applications. In biomedical applications, obtaining reliable design is one of the hot area of research. In this regard, such applications require three step approaches: (1) a suitable feature reduction technique to obtain best feature subset (2) finding suitable base classifiers for the given diseases (3) construction of suitable ensembles using base classifiers.

In our study, we evaluated the effect of the feature selection strategy along with construction of RF ensembles over two benchmark datasets. The obtained results show that the ensemble algorithms have performed much better in both the datasets. In general, it requires an assessment study to obtain the high accurate classifier for achieving best performance. Though the construction of ensembles is computationally expensive still it's worth considering it's high efficiency.

The diversity among the ensembles is a vital issue. Here RF ensembles use PCA method as a method for introducing it. We analyzed the response of 20 base classifiers to RF ensembles applications in classification of two diseases as shown in table 1 & 2. The proposed feature selection strategy SVM-PSO is a new method as per our knowledge from the literature survey.

Another important criterion is the number of base classifiers in the ensemble. We have used 10 base classifiers in each ensemble. However it's a future work to study this effect on the overall

accuracy. Also the number of class labels is also a vital parameter. In this study we have used diverse class labels, i.e. 2 & 4. The study may be extended over other diseases also.

There is no ideal classifier that will give best performance for all diseases. So we need some assessment task over various classifiers to find the suitable ones. And further the performance can be enhanced by construction of ensembles using them.

Conflicts of interest

None

References

- [1] Mandal, I., Sairam, N. Accurate Prediction of Coronary Artery Disease Using Reliable Diagnosis System (2012) *Journal of Medical Systems*, pp. 1-21. Article in Press. DOI: 10.1007/s10916-012-9828-0
- [2] Mandal, I., Sairam, N. Enhanced classification performance using computational intelligence (2011) *Communications in Computer and Information Science*, 204 CCIS, pp. 384-391. DOI: 10.1007/978-3-642-24043-0_39
- [3] Mori, J., Kajikawa, Y., Kashima, H., Sakata, I. Machine learning approach for finding business partners and building reciprocal relationships (2012) *Expert Systems with Applications*, 39 (12), pp. 10402-10407.
- [4] Wang, D.H., Conilione, P. Machine learning approach for face image retrieval (2012) *Neural Computing and Applications*, 21 (4), pp. 683-694.
- [5] Mandal, I. Software reliability assessment using artificial neural network (2010) *ICWET 2010 - International Conference and Workshop on Emerging Trends in Technology 2010*, Conference Proceedings, pp. 698-699. DOI: 10.1145/1741906.1742067
- [6] Li, B., Meng, M.Q.-H. Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection (2012) *IEEE Transactions on Information Technology in Biomedicine*, 16 (3), art. no. 6138917, pp. 323-329.
- [7] Mandal, I. A low-power content-addressable memory (CAM) using pipelined search scheme (2010) *ICWET 2010 - International Conference and Workshop on Emerging Trends in Technology 2010*, Conference Proceedings, pp. 853-858. DOI: 10.1145/1741906.1742103
- [8] Ozcift, A., Gulten, A. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms (2011) *Computer Methods and Programs in Biomedicine*, 104 (3), pp. 443-451.
- [9] Connolly, J.-F., Granger, E., Sabourin, R. Evolution of heterogeneous ensembles through dynamic particle swarm optimization for video-based face recognition (2012) *Pattern Recognition*, 45 (7), pp. 2460-2477.
- [10] Woloszynski, T., Kurzynski, M., Podsiadlo, P., Stachowiak, G.W. A measure of competence based on random classification for dynamic ensemble selection (2012) *Information Fusion*, 13 (3), pp. 207-213.
- [11] Tan, J., Chen, X., Du, M. An internet traffic identification approach based on GA and PSO-SVM (2012) *Journal of Computers*, 7 (1), pp. 19-29.
- [12] Wu, P., Yi, X., Jin, K. A study on chinese output of timber prediction model based on PSO-SVM (2012) *Advances in Information Sciences and Service Sciences*, 4 (2), pp. 227-233.
- [13] Xin, J., Xiaofeng, H. A quantum-PSO-based SVM algorithm for fund price prediction (2012) *Journal of Convergence Information Technology*, 7 (3), pp. 267-273.
- [14] Kotsiantis, S. Combining bagging, boosting, rotation forest and random subspace methods (2011) *Artificial Intelligence Review*, 35 (3), pp. 223-240. Cited 5 times.
- [15] Rodríguez, J.J., Kuncheva, L.I., Alonso, C.J. Rotation forest: A New classifier ensemble method (2006) *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (10), pp. 1619-1630.

- [16] Poon, B., Amin, M.A., Yan, H. Performance evaluation and comparison of PCA Based human face recognition methods for distorted images (2011) *International Journal of Machine Learning and Cybernetics*, 2 (4), pp. 245-259.
- [17] Karabulut, E.M., Ibrikçi, T. Effective Diagnosis of Coronary Artery Disease Using The Rotation Forest Ensemble Method (2011) *Journal of Medical Systems*, pp. 1-8. Article in Press.
- [18] Wei, J., Jian-Qi, Z., Xiang, Z. Face recognition method based on support vector machine and particle swarm optimization (2011) *Expert Systems with Applications*, 38 (4), pp. 4390-4393.

Authors

Indrajit Mandal is working as research scholar at School of Computing, SASTRA University, India. Based on his research work, he has received two Gold medal awards in Computer Science & Engineering discipline from National Design and Research forum, The Institution of Engineers (India). He has won several prizes from IITs, NITs in technical paper presentations held at national level. He has published research papers in international peer reviewed journals and international conferences. His research interest includes Machine learning, Applied statistics, Computational intelligence, Software reliability.



Sairam N is working as Professor in School of Computing, SASTRA University, India and has teaching experience of 15 years. He has published several research papers at national and international journals and conferences. His research interest includes Soft Computing, Theory of Computation, Parallel Computing and Algorithms, Data Mining.

