

Integrated Framework for Prognosis of Coronary Artery Disease

D. Rajeswara Rao¹, J.V.R Murthy²

¹Department of CSE, K L E F, Vaddeswaram, Guntur, A.P, India

²Dept of CSE, J.N.T.U.H, J.N.T.U.K, Kakinada, India

*Corresponding author E-mail: rajeshpitam@gmail.com

Abstract

Coronary Artery Disease, commonly known as Heart Disease, has resulted in casualties all over the world. Currently, diagnosis of these conditions mainly relies on experience of doctors and partly on few decision support systems. Most decision support systems built for this domain have limitations. The major limitation is requirement of huge quantity of historical conditions data and appropriate related diagnosis labels. Proposed framework addresses this limitation by providing diagnosis comparable with existing systems while only taking available data in absence of labels. This is achieved by forming a novel framework using Self Organizing Networks and Learning Vector Quantization. Here, competitive learning paradigm of soft computing is primary focus. Proposed framework is compared with four well established existing systems namely Support Vector Machine, Decision Tree, Random Forest and Multi Layer Perceptron. After comparison and analysis it has been proved that proposed framework gives comparable results even without supplying labeled data. In future this framework can be extended to variegated applications in various domains.

Keywords: Heart Disease ; Soft Computing ; Cognitive Computing ; Competitive Learning;

1. Introduction

As per World Health Organization (WHO) statistics rate of patients with Heart problems is accelerating all over the world. Scientists are putting massive efforts to control this disease outbreak. One of the focal objective of these efforts is to have this disease prediagnosed in routine checkups. This will considerably increase chances of better cure. For such prediagnosed by the doctor an efficient setup is needed. Doctor needs to have historical records, relevant knowledge and probable estimations. This requires a doctor to be well established and experienced. Another option can be to have a decision support system which suggests doctor using historical diagnostic data. So, in this research focus is on developing a decision support system for a amateur doctor. Additionally, because of the uncertainties involved in heart disease it's difficult for only a doctor to manage and monitor all parameters. Thus these systems can be helpful even to various classes of doctors.

Competitive learning is selected as a basis of this system, because of well known performance of these techniques in various domains. Competitive learning techniques have lead to successful decision making systems. They have helped in domains like banking, finance, insurance, pathological diagnosis, chemical industries..

2. Literature Survey

Competitive learning is an important paradigm of soft computing and machine learning. Recently competitive learning is applied to various avenues. Work [1] applies competitive learning to cluster potential contestants for recommen- dations. Here competitive

learning is applied over multi layer perceptron. On other hand [2] applies this learning paradigm for real time data classification. Work [3] has novel application of this technique in reinforcement systems in- volving numerous agents. It is also being applied in image processing field by researchers in works like [4] Novel application in voting system is done by work [5].

Self Organizing map as proposed in [6] applies competitive learning. By nature it supports unsupervised learning. Later there are several improvements suggested by researchers. Work [7] contributes to Self Organizing Network by novel distance measure. This distance measure minimises divergence and gives improved performance for network planning applications. This work also claims new method to be scalable. Imbalanced evolution is applied to SOM in [8] to address issues in imbalanced learning. This work handles imbalance by using different SOMs for minority and majority class. Novel application of SOM on data of body cells is done in [9]. Here SOM with multi layer clustering is used to visualize hyper dimensional cytometry data in 2D flow.

Learning Vector Quantization (LVQ) is another architecture to implement competitive learning [10]. Latest review on recent eleven varieties of LVQ are presented in work [11]. Here all variations are categorised under Heuristic or margin maximization or likelihood ratio maximization approaches. Further work [12] gives approximation based robust approach for non-vector data. Here high computational requirement due to non-vector form of data is minimized by approximation functions. Approximation functions leads to sparseness in connections. Another work [13] applies LVQ for ordinal regression problems. Here intuitive cost function is proposed for LVQ which leads to intuitive pa- rameter updates.

All above works show how competitive learning is still state of the art method for various domains and applications. Also, it can be concluded that there needs to be a framework to handle data from medical domain in unsupervised manner. So, in this work a novel framework is proposed in domain of cardio sciences. This is based on cognitive model of human brain's decision making.

3. Cognitive Competitive Multi-Aspect Framework (CCMF)

Proposed integrated framework is novel formulation in field of cardio sciences using techniques based on competitive learning, as per author's knowledge. As, this is based on competition, it proves successful in variety of scenarios. Components of CCMF are Self Organizing Networks (SON) and Learning Vector Quantization (LVQ). Figure 1 shows components of CCMF. These components provide different aspects in decision making. Such architecture of having multiple components with different aspects integrated is inspired from human brain. In cognitive process multiple parts of human brain work in coordination for decision making. This framework is capable of handling both labeled and unlabeled data. So, it can be used for supervised or unsupervised or semisupervised learning scenarios. Thus, it addresses different currently dominant data issues in medical field.

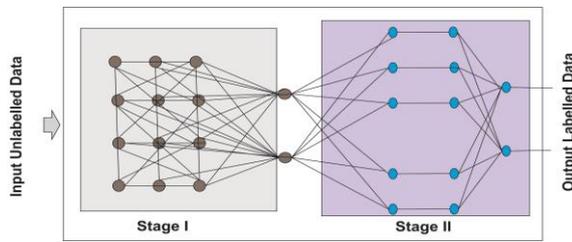


Figure 1. CCM Framework

4. Experimental Setup and Dataset

Table 1 shows details of datasets used for experiments. First entry is Heart Disease Dataset with all 75 independent attributes and one dependent attribute. It is taken from UCI [14]. Next entry is for Heart disease dataset with thirteen independent attributes. Here, no encoding is applied on data. Then, Heart disease dataset with One Hot encoding is taken. After encoding thirteen independent attributes are expanded to 24 attributes. Finally, Cardiocography dataset from UCI [15] is taken. It has 23 independent attributes. For all experiments Python 3.4 in Eclipse environment is used.

Table 1 Basic Statistics of Datasets

No	Data Set	Description	Value
1	Heart Disease	Samples	288
		Features	75
2	Heart Disease without Encoding	Samples	303
		Features	13
3	Heart Disease with Encoding	Samples	303
		Features	24
4	CTG Test	Samples	2126
		Features	23

5. Results and Discussion

Here results over all four datasets considered are presented. In this section detailed comparison of all models used is provided. Their performance is compared in a graphical form. Performance comparison parameters used are accuracy, positive predictive value (PPV), Sensitivity, dice similarity coefficient (DSC), and area under curve (AUC). Confusion matrix for each classifier is discussed. Ultimately, receiver operating characteristics (ROC)

curve presents visual comparison between all discussed classifiers with proposed framework.

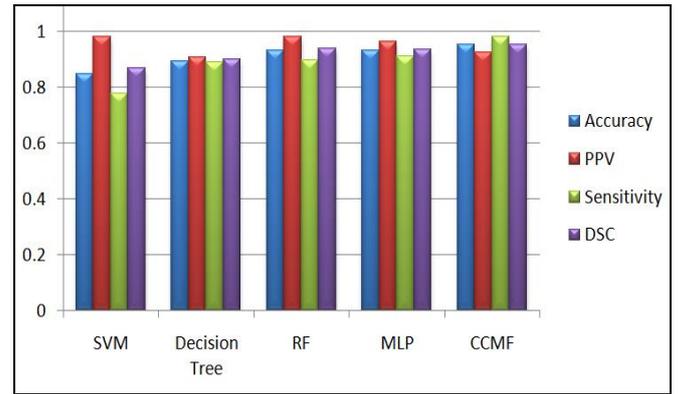


Figure 2a. Performance comparison on Dataset 1

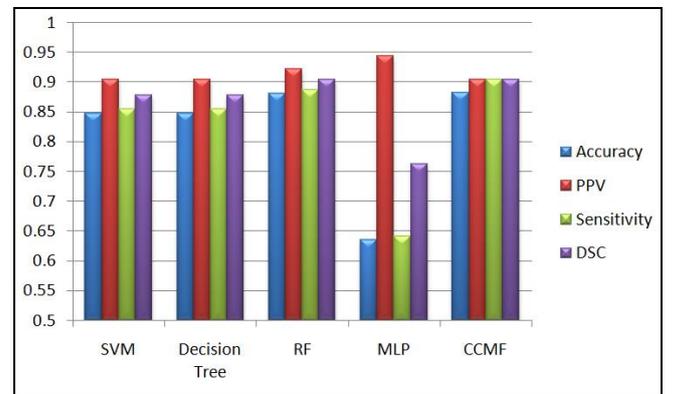


Figure 2b. Performance Comparison on Dataset 2

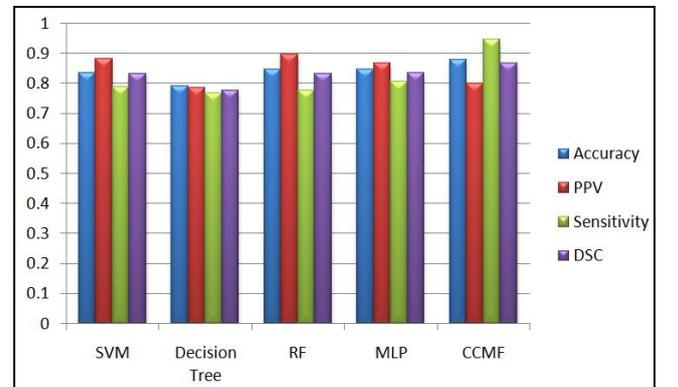


Figure 2c. Performance Comparison on Dataset 3

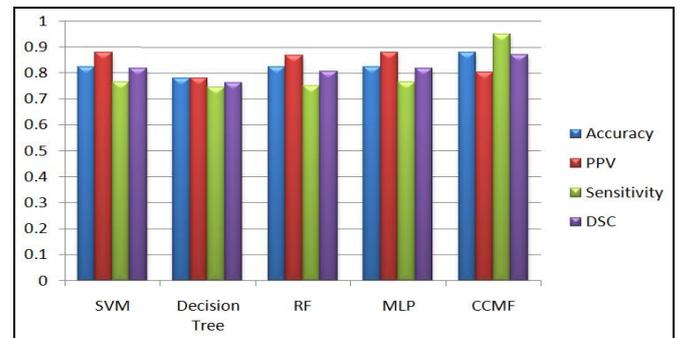


Figure 2d. Performance Comparison on Dataset 4

Results for all parameters are shown in figure 2a,2b,2c and 2d. Results for cost of models are shown in figure 3a, 3b, 3c and 3d. These results lead to following observations. SVM performance is

best on dataset 2. On other datasets SVM shows consistent performance. Even in cost SVM stands third compared to all others. Decision tree is unable to give desired performance on any considered datasets. Random forest performs well PPV and specificity over multiple datasets. For dataset 1 it gives best performance in DSC, AUC and DOR. Multi layer perceptron performs best in PPV for dataset 1 and 2. It has given second best cost for multiple datasets. As this is neural network based model, it's overall performance is better than others in terms of cost.

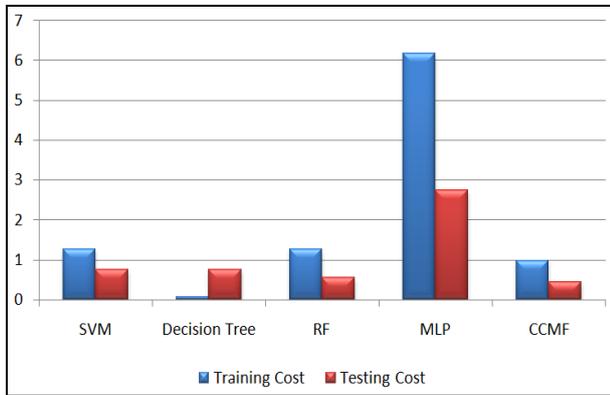


Figure 3a. Cost Comparison on Dataset 1

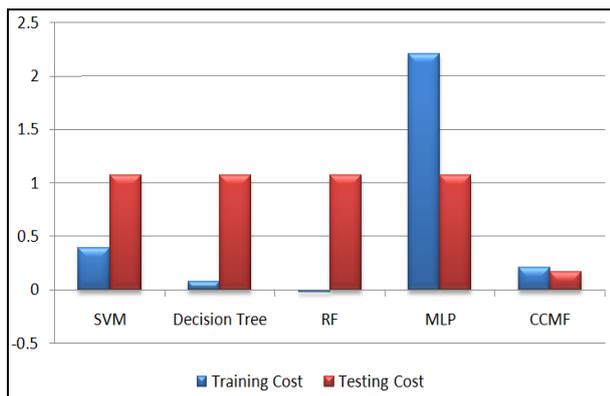


Figure 3b. Cost Comparison on Dataset 2

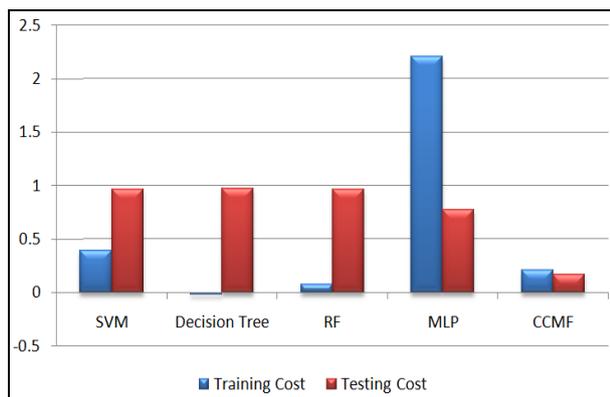


Figure 3c. Cost Comparison on Dataset 3

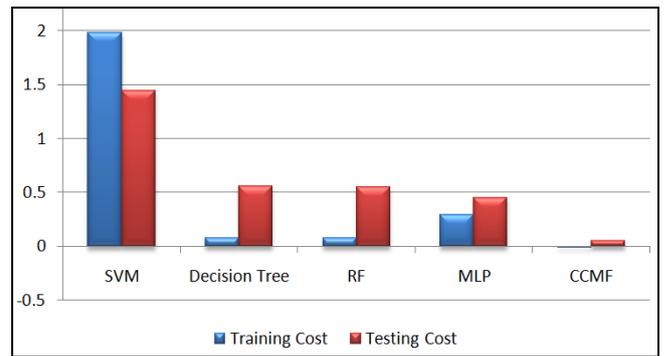


Figure 3d. Cost Comparison on Dataset 4

Proposed framework has performed well in multiple parameters in multiple datasets. Dataset wise, for dataset 1, proposed framework has accuracy of 88.24% with least cost, outperforming all others. Closest classifier is Random Forest with 88.1% accuracy with higher cost. For dataset 2 & 3 accuracy of proposed framework is 87.91%. Encoding done in dataset 3 resulted in better area under the curve and specificity. Cost of proposed framework for both dataset is 0.17, minimum amongst all other classifiers. In dataset 4, proposed framework gives highest accuracy of 95.24% along with best cost 0.07.

Parameter wise, proposed framework outperform others for all datasets in four parameters namely, accuracy, sensitivity, dice similarity coefficient, and cost. Only in two parameters positive predictive value and specificity it is overpowered by other classifiers.

Poor performance in positive predictive value and specificity is reasonable as the model still performs best in terms of cost. Less PPV leads to type I errors, intern causing reduced specificity. This behavior of system is preferred over type II errors. Proposed framework has minimal type-II errors which are desired generally and in medical applications. So, overall performance of the proposed framework is outstanding in terms of accuracy, minimal type-II errors and cost over all datasets.

6. Conclusion

Prognosis of Coronary Artery Disease in infants and adults is well handled in this work. Here four datasets about heart disease are used for testing proposed system. Three of the datasets are regarding heart disease in adults. Another dataset is about heart disease in babies before birth. They are found to be varying in parameters, data patterns and data distributions. Still, on all datasets proposed system performs well.

Proposed competitive learning based framework has excelled over other models on various datasets across parameters. This clearly concludes efficiency of proposed framework. Performance of the proposed system is found to be comparable with state of the art models like Support Vector Machine, Decision Tree, Random Forest and Multi layer Perceptron. This work details out how not just in accuracy but in multiple parameters proposed framework's performance is judged.

In future framework can be extended to address problems in various domains like diabetes detection, finance domain fraud detection, decision making in agriculture. Additionally, this work can further be extended to other forms of data like image processing, text processing, and bio-medical signal processing.

References

- [1] H. Almohri and R. B. Chinnam, "Deriving business recommendations for franchises using competitive learning driven mlp-based clustering," in International Conference on

- Engineering Applications of Neural Networks, pp. 490-497, Springer, 2017.
- [2] Z. Sarafraz, H. Sarafraz, and M. R. Sayeh, "Real-time classifier based on adaptive competitive self-organizing algorithm," *Adaptive Behavior*, vol. 26, no. 1, pp. 21-31, 2018.
- [3] Z. Sarafraz, H. Sarafraz, and M. R. Sayeh, "Real-time classifier based on adaptive competitive self-organizing algorithm," *Adaptive Behavior*, vol. 26, no. 1, pp. 21-31, 2018.
- [4] H. Xie, X. Luo, C. Wang, S. Liu, X. Xu, and X. Tong, "Multispectral remote sensing image segmentation using rival penalized controlled competitive learning and fuzzy entropy," *Soft Computing*, vol. 20, no. 12, pp. 4709-4722, 2016.
- [5] A. Tamvakis, G. E. Tsekouras, A. Rigos, C. Kalloniatis, C.-N. Anagnostopoulos, and G. Anastassopoulos, "A methodology to carry out voting classification tasks using a particle swarm optimization-based neuro-fuzzy competitive learning network," *Evolving Systems*, vol. 8, no. 1, pp. 49-69, 2017.
- [6] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning," *Cognitive science*, vol. 9, no. 1, pp. 75-112, 1985.
- [7] O. M. Abusaid and F. M. Salem, "Kullback-leibler divergence minimization for competitive learning of self-organizing maps," in *Engineering and Technology (ICET), 2017 International Conference on*, pp. 1-6, IEEE, 2017.
- [8] Q. Cai, H. He, and H. Man, "Imbalanced evolving self-organizing learning," *Neurocomputing*, vol. 133, pp. 258-270, 2014.
- [9] S. Van Gassen, B. Callebaut, M. J. Van Helden, B. N. Lambrecht, P. Demeester, T. Dhaene, and Y. Saeys, "Flowsom: Using self-organizing maps for visualization and interpretation of cytometry data," *Cytometry Part A*, vol. 87, no. 7, pp. 636-645, 2015.
- [10] T. Kohonen, "Improved versions of learning vector quantization," in *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*, pp. 545-550, IEEE, 1990.
- [11] D. Nova and P. A. Estevez, "A review of learning vector quantization classifiers," *Neural Computing and Applications*, vol. 25, no. 3-4, pp. 511-524, 2014.
- [12] D. Hofmann, A. Gisbrecht, and B. Hammer, "Efficient approximations of robust soft learning vector quantization for non-vectorial data," *Neurocomputing*, vol. 147, pp. 96-106, 2015.
- [13] F. Tang and P. Tino, "Ordinal regression based on learning vector quantization," *Neural Networks*, vol. 93, pp. 76-88, 2017.
- [14] J. Andras, S. William, P. Matthias, and D. Robert, "Heart disease data set." <https://archive.ics.uci.edu/ml/datasets/cardiocography>, July 1988.
- [15] J. Marques de S, J. Bernardes, and D. Ayres de Campos, "Cardiography data set." <https://archive.ics.uci.edu/ml/datasets/cardiocography>, Sept 2010.