Robust Cluster-then-label (RCTL) Approach for Heart Disease Prediction

Jyostna Devi Bodapati[1*], Venkata Rama Krishna Sajja[1], Nirupama Bhat Mundukur[1], Naralasetti Veeranjaneyulu[2]

[1] Department of CSE, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, AP 522213, India
[2] Department of IT, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, AP 522213, India

Corresponding Author Email: jyostna.bodapati82@gmail.com

ABSTRACT

Heart attack or stroke occurs due to the narrowed or blocked blood vessels that are present in the heart. If not identified at the early stages of the disease, there is a high probability to loose life. An algorithm that takes the advantage of the inferences given by unsupervised learning is proposed. Our proposed framework leverages the unsupervised information that is hidden in the data to improve the accuracy of the classifier. By using clustering, a label is obtained and this cluster label is included as one of the features. This augmented feature set fed as input to the classifier to predict the disease. For clustering to avoid uncertainty, k-means and spectral clustering are used. If the cluster label predicted by both the methods match, then only that information is included as one of the features. To evaluate the effectiveness of the proposed algorithm two heart disease benchmark datasets, Framingham and stat-log are used. For the classification task Logistic regression, Naïve Baye's and SVM are being used to make the model simple yet effective. Based on the experimental results we claim that by augmenting the feature set with the cluster label using the proposed method significantly improves the performance of the heart disease prediction task.

1. INTRODUCTION

Heart disease is the most common reason for human deaths in the recent past [1]. As per review of WHO (World Health Organization), 17 million all out worldwide passings are because of coronary illness [2]. The root cause of heart disease is mostly because of the work over-burden, Mental Stress and numerous different issues [3].

There are two major lines of treatment for cardiac diseases. Initially, an individual will be given medication to treat the heart condition using medications. If the medication will not give any positive impact, then surgical choices are suggested to correct the difficulty [4].

It would be helpful to save lives if this disease is predicted at the early stages of it [4]. Different Machine learning algorithms can be used to accomplish task [5, 6]. But none of them tried to explore the usage of underlying unsupervised information that is hidden in the data.

In this work we propose a robust algorithm for automatic prediction of heart disease called robust cluster-then-label (RCTL) approach. Our proposed framework leverages the unsupervised information that can be extracted from the dataset and then the feature set is augmented with the extracted unsupervised information. In simple words the whole dataset is initially grouped into k clusters. As we know the number of classes in-advance we set k as the number of classes. k-means and spectral clustering are used to cluster the data independently. We decide on the cluster label for an example, if both the algorithms generate same label for that example. The examples for which the clustering label generated by both the clustering techniques mismatch then it is not assigned to any of the clusters. The feature set is augmented with the obtained cluster label. A supervised Classification technique is then applied on the dataset to predict the heart disease. To show the effectiveness of the proposed algorithm we use different benchmark heart disease datasets like Cleveland, Framingham and statlog which are available in UCI repository [7]. For clustering we use k-means and for classification we use Logistic regression, Naïve Baye's and SVM. Based on our investigations on benchmark datasets we claim that by augmenting the feature set with the cluster label significantly improves the generalization ability of the classification.

Major contributions of the work:

• Proposed RCTL method leverages the use of underlying unsupervised information that is hidden in the data.

• Extensive studies on different benchmark datasets show the effectiveness of the decision rule used for deciding the cluster label.

In the succeeding sections we initially introduce motivation behind the work, followed by the details of the proposed RCTL approach. In section 2, we try to explore different techniques that are available in the literature. Section 3 elaborates the different stages of the proposed RCTL approach. Section 4 investigates the results

2. RELATED WORK

In recent past significant number of deaths are due to the heart disease [8]. As per the statistics, in the U.S., at least one person incorporates a coronary failure each forty seconds, and a minimum of one person dies per minute from an incident regarding heart issues [9]. In most of the cases the patient dies just because of the delay in identifying the heart disease. The major goal is to automatically predict heart strokes in advance based on the given features like age, intensity of smoking,

gender etc. This section presents an overview of different existing machine learning algorithms available in the literature to predict heart disease.

In the recent past large number of researchers worked towards the identification of the disease at the early stages [10]. In [4, 9, 11] different machine learning techniques are explored for heart disease prediction. In [9] in addition to that a support system is created to assist medical professionals to accurately predict the disease. In [12], a hybrid algorithm is proposed that is a combination of KNN and ID3 algorithms are used. In [13] data mining techniques are used to predict heart disease and to suggest suitable treatment.

An interactive web based application [14] is designed to assist people diagnosing heart disease. Naïve Baye's algorithm is used to model the data. In [15] Mega Trend Diffusion Function is applied for each example of each class in dataset. Then PCA is applied to reduce the dimensionality [16]. In [17-25] data mining techniques are being explored to extract patterns related to heart disease based on the previous medical history of a patient. In [26], a combination CFP techniques and statistical features are used together to predict heart failures in advance. In [27] heart failure events are predicted in the early stages based on the data collected by telemonitoring study. LSTM based sequential model is proposed in [28] based to EHS records and patients health records. Neural network based models are proposed in [29, 30] and SVM based models are explored in [31].

In all these above models either different classification algorithms [4, 9, 13] are being explored or a combination of classification techniques are combined to introduce a hybrid method [12, 17]. None of the existing methods explored the use of unsupervised methods for the early prediction of heart stroke.

Major objective of this work is to propose a robust cluster-then-label-approach for heart disease prediction. Our proposed framework leverages the unsupervised information that can be extracted from the dataset and then the feature set is augmented with the extracted unsupervised information. We propose a novel method to predict the cluster label for all the examples in the dataset. In the next step this obtained cluster label is added as one of the features to the data, then a supervised classifier is used to classify the data. Based on our investigations on benchmark datasets we claim that by augmenting the feature set with the cluster label significantly improves the performance of classification.

The book size will be in A4 (8.27 inches x 11.69 inches). Do not change the current page settings when you use the template.

The number of pages for the manuscript must be no more than ten, including all the sections. Please make sure that the whole text ends on an even page. Please do not insert page numbers. Please do not use the Headers or the Footers because they are reserved for the technical editing by editors.

3. PROPOSED WORK

The major objective of this work is to predict heart failure at early stages to avoid loss of life. Prediction of the heart stroke is done based on the past history of the patient's records (features). In all the traditional supervised methods the knowledge available in the form of features alone is used and none of the existing methods focus on how the extract the additional features based on the available features to improve classification accuracy.

A robust cluster-then-label (RCTL) is proposed to predict the disease early in the life. Initially cluster label for all the examples is identified in the dataset. In the next step this obtained cluster label is added as one of the features to the data, then a supervised classifier trained to accurately predict the test data. As we are including cluster label as one of the features we claim it as feature augmentation. To predict accurate cluster label we use two clustering algorithms. If the cluster label produced by both the clustering schemes match then only we confirm it otherwise we assign 0 as the cluster label indicating that it is not belonging to any cluster. To cluster the data two different clusters are applied. To check the robustness of the proposed method different classification methods are being used.

In traditional machine learning algorithms major steps involved are: collect the dataset, extract features from the data, apply supervised classifier to obtain label prediction. Figure1 shows the steps involved in the proposed approach: collection of dataset, preprocess the data, cluster the data to predict the label, add this cluster label as one of the feature to augment features of the data, then apply supervised classifier to obtain the label prediction. To predict accurate cluster label, two different simple clustering schemes are being used. If both schemes agree on the cluster prediction then the assignment is being considered.

K-means clustering: Once the dataset is available, k-means unsupervised clustering is applied to partition the entire dataset into k disjoint sets. As the dataset contains 2 classes, k value is set as 2. Each cluster is designated by a cluster center μ called centroid of the cluster. The objective of k-means algorithm is to choose cluster centers such that the following objective is minimized:

$$\sum_{i=0}^{n} \min_{\mu_j \in C} (||x_i - \mu_j||^2) \tag{1}$$

Spectral Clustering: This part of this section gives the working details of spectral clustering algorithm. Later we will give the details of how we adopted it to our work. Let the given dataset contains n samples $\{x_1, x_2, x_3, \ldots x_n\}$.

Step1: Compute a weighted matrix, W, based on the similarity between the data points. W is a square matrix of size nxn.

Step2: Based on W, Compute a graph laplacian L as follows:

$$L = D - W \tag{2}$$

In equation (2), D refers to the diagonal degree matrix of size nxn and is computed as follows:

$$D[i,j] = \sum_{j=1}^{n} w[i,j] \tag{3}$$

Step 3: Now compute the eigen vectors, v, of Laplacian matrix, L by solving an Eigenvalue problem, such as $Lv=\lambda v$.

Step 4: Select k eigenvectors $\{v_i, i = 1, 2 \ldots k\}$ corresponding to the k significant eigenvalues $\{\lambda_i, i=1, 2 \ldots k\}$ to define a k-dimensional subspace $P^t LP$. Form clusters in this subspace using any clustering algorithm.

Following are the specific details of the spectral clustering that we have adopted for our experiments.

a. Gaussian kernel is used in step1 to compute the similarity between the data points x_i and x_j. Similarity between the

samples is represented using d(x$_i$, x$_j$) and is measured as follows:

$$d(x_i, x_j) = \exp(-(\|x_i - x_j\|^2 / 2\sigma^2)) \qquad (4)$$

b. To generate W as sparse as possible, similarity is measured only for the k nearest samples from a sample and to the rest of samples similarity is set to 0. The value of each cell of W is computed is follows:

$$W[i,j] = d(x_i, x_j) \text{ if } x_i \in knn(x_j) \qquad (5)$$

W[i,j] = 0 otherwise.
As W is sparse, it leads to efficient computation.

c. In step 4, to form clusters in subspace, k-means clustering is used.

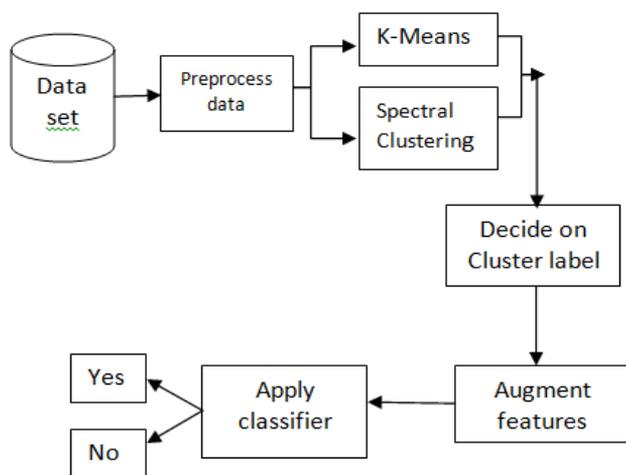We have used different bench mark datasets for heart disease prediction.

Figure 1. Block diagram of the proposed approach

Justification to use two clustering methods: K-means is computationally efficient even with large data when the number of clusters is known in advance and when k value is small. For our case both the conditions are satisfied. But k-means fails with clusters of different densities [32].
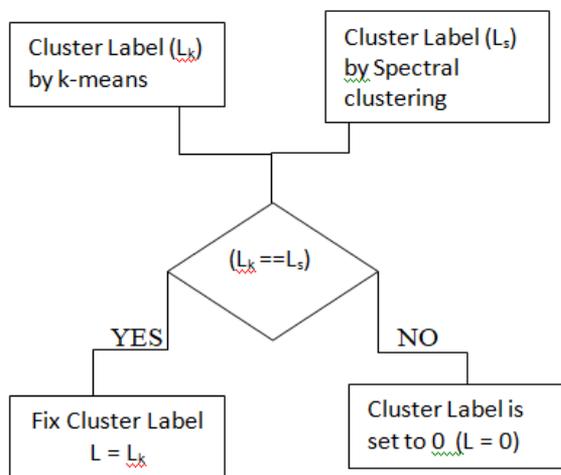

Figure 2. Approach to decide on cluster label

Spectral Clustering is semi-convex and the clusters are formed in the projected space not in the original space [33]. Hence can produce better clusters compared to k-means but with large data sets spectral clustering fails [34-35]. To leverage the benefit of both k-means and spectral clustering we use both the methods to get the final cluster label. On the entire dataset both the clustering techniques are used and cluster labels are produced.

Let L$_k$ and L$_s$ be the cluster labels produced by k-means and spectral clustering respectively. Figure 2 shows the approach we followed to fix the cluster label. For an example x$_i$, we fix decide the cluster label, if both the clustering approaches agree on the label produced by them. That is if L$_k$ and L$_s$ match then we fix the cluster label for x$_i$ as L$_k$ (or as L$_s$). Otherwise we set the cluster label as 0. Cluster label 0 indicates that the example x$_i$ is not assigned to any of the clusters as we are not confident about the cluster to which it belongs to.

Once the cluster label is decided for all the examples, in the dataset then this label is added as one of the features to the feature set. This is how we are augmenting the features by providing additional information that would help for better classification. To show the effectiveness of the proposed approach the proposed RCTL approach is applied for heart disease prediction task. Based on the results obtained on two bench mark datasets we claim that the proposed robust cluster – then – label method gives better performance than the method that do not use the cluster label. The performance is also compared with the other baseline methods to who that our approach is on par with the state of the art methods.

4. EXPERIMENTAL RESULTS

The objective of this work is to introduce a novel efficient machine learning algorithm that predicts the heart disease in the early stages based on the profile of the patient. We introduce a novel, simple and robust cluster – then – label method that makes use of the unsupervised clustering information to improve the generalization ability of the classifier. To show the efficiency of the proposed approach we apply the proposed method for heart disease prediction task. Two of the datasets from UCI repository are selected for the task and the details of them are summarized in Table 1.

Table 1. Summary of heart disease benchmark datasets

Dataset Name	Attributes	No Of instances
Framingham	Gender, Age, Education, Current smoking status, cigars per day, BPMeds, Prelevent stroke, Prevalent Hyp, Diabetes, totchol, sysBP, diaBP, BMI, Heart rate, glucose, Tenyear CHD.	4241
Stat log	age, sex, chest pain type, resting blood pressure, serum cholestoral, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise induced, in an old peak slope, number of major vessels, thal.	270

We have selected k-means and spectral clustering to cluster

the entire data and the method explained in Figure2 is adopted to decide on the cluster label. After fixing the decision on cluster label, this label is included as one of the features. That is in case of the original datasets Framingham has 13 attributes and Statlog dataset has 14 attributes. After including the cluster label as one of the labels the datasets Framingham and Statlog contain 14 and 15 features respectively. We then split the entire dataset into train and test datasets following 80-20 rule. To check the efficiency of the proposed robust cluster-then-label (RCTL) approach we use different approaches for classification task. To make the model simple we use shallow classifiers like Logistic regression, Naïve Baye's and SVM based classifier. Accuracy of the model is used to evaluate the performance that model.

In the result section we compare the proposed method's performance with different models. Below are the different models used for comparison.

Basic Model: The vanilla supervised model without applying any clustering technique. This model is used as the baseline and the objective is to improve the performance of the basic model.

RCTL: Proposed robust cluster – then label approach, that uses 2 clustering methods to obtain robust label as described in section3.

K means + Basic model: The model that uses k-means alone to predict the cluster label, then the feature set is augmented with cluster label and then supervised classifier is used.

Spectral + Basic model: The model that uses spectral clustering alone to predict the cluster label, then the feature set is augmented with cluster label and then supervised classifier is used.

Table 2. Performance of the proposed RCTL method on Framingham dataset

Classifier	Accuracy	
	Basic Model	RCTL (proposed)
Logistic Regression	81.48	**81.48**
Naive Bayes	79.62	**83.33**
SVM	79.62	**85.18**

Table 3. Performance of the proposed RCTL method on Stat-log dataset

Classifier	Accuracy	
	Basic Model	RCTL (proposed)
Logistic Regression	83.84	**86.32**
Naive Bayes	81.11	**82.42**
SVM	84.08	**85.37**

Table 3 presents the results on Stat-log dataset, one of the benchmark dataset for heart disease prediction. The performance obtained by the proposed model is compared with the basic model. The results exhibit that the proposed method either improves the performance or in the worst case it retains the performance of the basic model. The proposed model never degrades the performance of the basic model. By using Logistic regression the performance is retained and by using Naïve baye's and SVM the accuracy is increased. The

reason for the improvement could be the cluster labels predicted by the proposed RCTL are accurate and adds more discrimination to the data. Hence the classifier that works on the augmented features performs better on this data.

Table 4. Performance of the proposed RCTL method on Framingham dataset

Classifier	Accuracy	
	Basic Model	RCTL (proposed)
Logistic Regression	83.84	**86.32**
Naive Bayes	81.11	**82.42**
SVM	84.08	**85.37**

Similar to Table 3, Table 4 presents the results on Framingham dataset, another benchmark dataset for heart disease prediction. The results exhibit that the proposed method significantly improves the performance of the basic model. In all the cases more than 2 % improvement is achieved. The reason for the improvement could be the cluster labels predicted by the proposed RCTL are accurate and adds better discrimination to the data.

The second experiment is done to show the importance of using two different clustering methods to obtain cluster label and also the importance of taking the confidence labels into consideration and ignoring other labels.
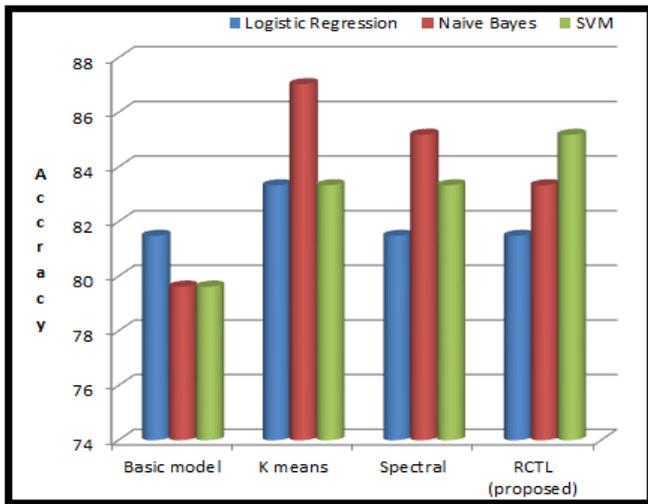
Table 5. Comparison of different cluster-then-label approaches on stat-log dataset

Classifier	Basic Model	Clustering used to obtain cluster label		
		K means + Basic model	Spectral + Basic model	RCTL (proposed)
Logistic Regression	81.48	83.33	81.48	81.48
Naive Bayes	79.62	87.03	85.18	83.33
SVM	79.62	83.33	83.33	85.18

Table 6. Comparison of different cluster-then-label approaches on Framingham dataset

Classifier	Accuracy %			
	Basic Model	Clustering used to obtain cluster label		
		K means	Spectral	RCTL (proposed)
Logistic Regression	83.84	84.08	84.08	86.32
Naive Bayes	81.11	81.36	81.13	82.42
SVM	84.08	83.37	83.84	85.37

Table 5 shows the comparison of using different clustering methods to obtain cluster label on stat-log dataset. We have done the experiments using k-means and spectral clustering for fair comparison. When either of these methods are used the accuracy of the basic model is improved. We might feel that when single method is used the model achieves better performance. Graph1 clearly visualize that the proposed RCTL method is better compared to other models.

Figure 3. Comparison of different cluster-then-label approaches on Stat-log dataset

To claim the robustness of the proposed model the experiment is repeated on Framingham dataset. Table 5 and Graph 2 shows the results obtained on Framingham dataset. Based on the results shown in Table6, we can understand that using single clustering method to obtain cluster label may degrade the performance of the basic model. Especially when used with SVM as classifier and k-means as the classifier performance of the basic model is not retained and degraded significantly. The results shown in Table 5 also show that the proposed method achieves accuracy better than using single clustering approach for cluster labeling. Graph 4 visualizes the comparison between the proposed model and existing approaches.
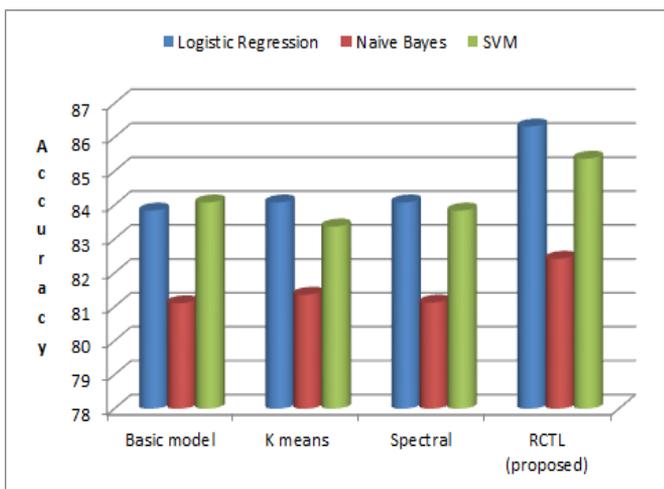
Figure 4. Comparison of different cluster-then-label approaches on Framingham dataset

As fair comparison is not possible we did not compare the proposed model with the performance of the existing models available in the literature. The proposed model is simple and robust.

5. CONCLUSIONS

The objective of this work is to predict heart disease at the early stages to save human lives. In this work a robust prediction model is proposed for the task that makes use of the hidden unsupervised information. Based on experimental studies we claim that cluster label obtained by the proposed model preserves discriminative information and when this feature is added as one of the feature the model's performance is improved significantly. In the future we would like to explore and experiment combination of different clustering models and also use neural network based models for classification.

REFERENCES

[1] Resul, D., Turkoglu, I., Sengur, A. (2009). Diagnosis of valvular heart disease through neural networks ensembles. Computer Methods and Programs in Biomedicine, 93(2): 185-191. https://doi.org/10.1016/j.cmpb.2008.09.005

[2] Resul, D., Turkoglu, I., Sengur, A. (2009). Effective diagnosis of heart disease through neural networks ensembles. Expert Systems with Applications, 36(4): 7675-7680. https://doi.org/10.1016/j.eswa.2008.09.013

[3] Latha, P., Subramanian, R. (2008). Intelligent heart disease prediction system using CANFIS and genetic algorithm. International Journal of Biological, Biomedical and Medical Sciences, 3(3): 157-160.

[4] Thomas, J., Theresa Princy, R. (2016). Human heart disease prediction system using data mining techniques. International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, India. https://doi.org/10.1109/ICCPCT.2016.7530265

[5] Chitra, R., Seenivasagam, V. (2103). Heart disease prediction system using supervised learning classifier. Bonfring International Journal of Software Engineering and Soft Computing, 3(1): 01-07. https://doi.org/10.9756/BIJSESC.4336

[6] Lakshmi, B.N., Indumathi, T.S., Ravi, N. (2016). A study on C. 5 decision tree classification algorithm for risk predictions during pregnancy. Procedia Technology, 24: 1542-1549. https://doi.org/10.1016/j.protcy.2016.05.128

[7] Frank, A., Asuncion, A. (2010). UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, Calif, USA.

[8] Salma Banu, N.K., Swamy, S. (2016). Prediction of heart disease at early stage using data mining and big data analytics: A survey. International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, (ICEECCOT). https://doi.org/10.1109/ICEECCOT.2016.7955226

[9] Alladoumbaye, N., Lei, L., Wang, H.Z. (2016). Comparative study of data mining techniques on heart disease prediction system: A case study for the Republic of Chaz. International Journal of Science and Research, 5(5): 1564-1571.

[10] Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J.J., Sandhu, S., Guppy, K.H., Lee, S., Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. The American Journal of Cardiology, 64(5): 304-310. https://doi.org/10.1016/0002-9149(89)90524-9

[11] Gandhi, M., Singh, S.N. (2015). Predictions in heart disease using techniques of data mining. International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp.

520-525.
https://doi.org/10.1109/ABLAZE.2015.7154917

[12] Pramanik, S.K., Pramanik, S., Pramanik, B.K., Islam Molla, M.K., Md. Hamid, E. (2012). Hybrid classification algorithm for knowledge acquisition of biomedical data. International Journal of Advanced Science and Technology, 44: 99-112.

[13] Mai, S., Turner, T., Stocker, R. (2012). Using data mining techniques in heart disease diagnosis and treatment. Japan-Egypt Conference on Electronics, Communications and Computers, Nagercoil, India. https://doi.org/10.1109/JEC-ECC.2012.6186978

[14] Medhekar, D.S., Bote, M.P., Deshmukh, S.D. (2013). c. International J Enhanced Res Sci Technol Eng, 2(3).

[15] Nirmala, K., Singh, R.M. (2015) Analysis and detection of heart related issues: A proposed algorithm. International Journal of Computer Science and Information Technologies, 6(2): 1040-1043.

[16] Bodapati, J.D., Veeranjaneyulu, N., Kishore, K.V.K. (2010). A novel face recognition system based on combining eigenfaces with fisher faces using wavelets. Procedia Computer Science, 2: 44-51. https://doi.org/10.1016/j.procs.2010.11.007

[17] Dewan, A., Sharma, M. (2015). Prediction of heart disease using a hybrid technique in data mining classification. In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 704-706. IEEE.

[18] Sharma, Purushottam, Kanak Saxena, and Richa Sharma (2016). Heart disease prediction system evaluation using C4. 5 rules and partial tree. Computational Intelligence in Data Mining, 2: 285-294. https://doi.org/10.1007/978-81-322-2731-1_26

[19] Carlos, O. (2016). Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology in Biomedicine, 10(2): 334-343. https://doi.org/10.15680/IJIRSET.2016.0510114

[20] Mai, S., Turner, T., Stocker, R. (2011). Using decision tree for diagnosing heart disease patients. Proceedings of the Ninth Australasian Data Mining Conference. Vol. 121.

[21] Muhammed, L.A. (2012). Using data mining technique to diagnosis heart disease. International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, Malaysia. https://doi.org/10.1109/ICSSBE.2012.6396533

[22] Akhil, J.M., Deekshatulu, B.L., Chandra, P. (2013). Heart disease prediction using lazy associative classification. International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s). https://doi.org/10.1109/iMac4s.2013.6526381

[23] Wilson, A., Wilson, G., Likhiya, J.K. (2014). Heart disease prediction using the data mining techniques. International Journal of Computer Science Trends and Technology (IJCST), 2(1).

[24] Aljaaf, A.J., Al-Jumeily, D., Hussain, A.J., Dawson, T., Fergus, P., Al-Jumaily, M. (2015). Predicting the likelihood of heart failure with a multi-level risk assessment using decision tree. Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), Beirut, Lebanon. https://doi.org/10.1109/TAEECE.2015.7113608

[25] Srinivas, K., Kavihta Rani, B., Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering (IJCSE), 2(02): 250-255.

[26] Afef, M., Ismael, B.R., Khalil, C., Abid, L., Freisleben, B. (2017). CEP4HFP: Complex event processing for heart failure prediction. IEEE Transactions on Nanobioscience, 16(8): 708-717. https://doi.org/10.1109/TNB.2017.2769671

[27] Jorge, H., Carvalho, P., Paredes, S., Rocha, T., Habetha, J., Antunes, M., Morais, J. (2014). Prediction of heart failure decompensation events by trend analysis of telemonitoring data. IEEE Journal of Biomedical and Health Informatics, 19(5): 1757-1769. https://doi.org/10.1109/JBHI.2014.2358715

[28] Jin, B., Che, C., Liu, Z., Zhang, S., Yin, X., Wei, X. (2018). Predicting the risk of heart failure with ehr sequential data modeling. IEEE Access, 6: 9256-9261. https://doi.org/10.1109/ACCESS.2017.2789324

[29] Desai, S.D., Dessai, I.F., Kulkarni, L. (2013). Intelligent heart disease prediction system using probabilistic neural network. International Journal on Advanced Computer Theory and Engineering (IJACTE) 2(3): 2319-2526.

[30] Durairaj, M., Revathi, V. (2015). Prediction of heart disease using back propagation MLP algorithm. International Journal of Scientific & Technology Research, 4(8).

[31] Xu, Y., Pan, X., Zhou, Z., Yang, Z., Zhang, Y. (2015). Structural least square twin support vector machine for classification. Appl Intell, 42(3): 527-536. https://doi.org/10.1007/s10489-014-0611-4

[32] Jain, A.K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8): 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

[33] Ulrike, V.L. (2007). A tutorial on spectral clustering. Statistics and Computing, 17(4): 395-416. https://doi.org/10.1007/s11222-007-9033-z

[34] Ng, A.Y., Jordan, M.I., Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems.

[35] Veeranjaneyulu, N., Raghunath, A., Jyostna Devi, B., Mandhala, V.N. (2014). Scene classification using support vector machines with LDA. Journal of Theoretical & Applied Information Technology, 63(3).