Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

Noise Removal Using Adaptive Damping Based Affinity Propagation Clustering Algorithm in Coronary Artery Disease

[1*|] Ramasamy MaheshKumar, [2] Sundaram Veni

[1]Ph.D. Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, 641021
[2]Assistant Professor & Head, Department of Computer Science, Karpagam Academy of Higher Education Coimbatore, 641021
*Corresponding authors Email: mahe83.r@gmail.com

Abstract: The health business generates enormous amounts of data, and by using this massive quantity of data, many illnesses may be recognized, predicted, and even treated at early stage. Humans face significant danger from coronary artery disease, cancer, and tumor illness. Predicting Coronary Artery Disease (CAD) is a difficult and time-consuming process in the medical sector. Early prediction is a virtuoso skill in the medical area, particularly in the cardiovascular sector. Prior research on developing early prediction model provided a grasp of modern strategies for detecting variance in medical imaging. Cardiovascular disease prevention may be accomplished with a diet plan established by the concerned physician after early diagnosis. This study aims to forecast CAD utilizing the suggested approach by creating noise reduction in CAD using the Adaptive Damping based Affinity Propagation (ADAP) clustering algorithm. This kind of knowledge-based identification is critical for accurate prediction. Despite the lack of supporting evidence, this substantial strategy positively influences determining variance in medical disciplines. Additionally, this publication has minimized the use of noisy data to aid in illness identification. This article discusses novel adaptive image-based clustering algorithms and compares them to established classification methods for predicting CAD early and with better accuracy.
Keywords: CAD, Clustering, ADAP, Noise Removal, HD, Prediction

1. Introduction

Human Heart Disease (HD) is one of the most prevalent illnesses worldwide. According to the poll, 18 million individuals die each year from cardiovascular disease. The heart is responsible for essentially all component activation [1]. HD is a serious condition that must be managed and anticipated early. Early detection may significantly reduce the number of people who die from HD complications [2].

Human beings are under increased strain and stress in this modern period due to their hectic job schedules, and individuals tend to live unhealthy lifestyles [3]. Individuals are increasingly drawn to cigarette and alcohol intake to cope with the stressful work environment. This unhealthy activity creates an environment conducive to scary disorders such as coronary sickness [4]. Coronary disease is a leading cause of mortality worldwide. The number of people affected by HD is rapidly growing.

Arrhythmia is a kind of cardiac illness that results in an irregular heart rhythm. Atherosclerosis is the term used to describe the hardening of the wall. In most cases, the estimated status of HD is uncertain [6][7]. The method for CAD is shown in Figure 1. In the field of clinical health, recently, health care institutions have been hastened in their efforts to predict HD early in the intricate and difficult analysis. In recent years, the medical imaging and healthcare industries have experienced a boom in data mining [7]. To monitor heart function activities, the prediction model often uses age, gender, blood pressure, chest pain factor, and electrocardiogram (ECG) test [8].

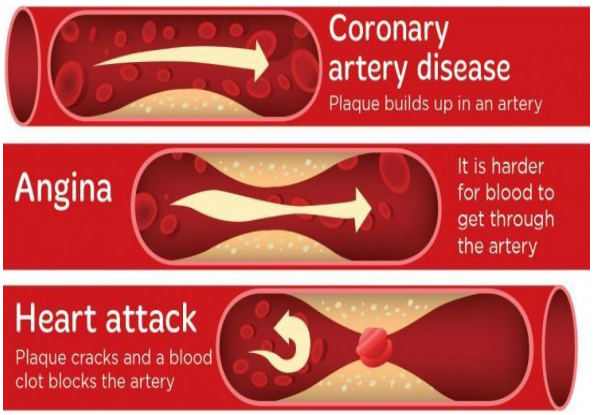

Fig 1: Coronary Artery Disease

Serum cholesterol levels, peak heart rate, depression level, blood sugar level, smoking history, eating habits, and fundamental measurements such as Body Mass Index (BMI) report are all analyzed at medical institutes to diagnose HD [9]. An abnormal heartbeat may indicate the presence of an arrhythmia. Cardiovascular arrest illnesses are marked by abrupt changes in or loss of heart function [10]. High blood pressure is a condition in which the blood pressure is increased due to constricted blood vessels in the branch [8]. CAD is caused by blood vessel injury. A lack of blood flow hampers the heart's ability to pump blood. As a result, the heart's supporting muscles weaken. The last stage is Arrhythmias illness, which is just the heart's failure to operate properly [12].

Data mining is largely concerned with classifying and collecting data patterns [13]. Medical information extraction is a critical and difficult operation that must be carried out with precision and quality in the health industry. It tries to extract diagnostic information to solve real-world health concerns related to illness diagnosis and treatment. In this paper, we predict cardiac illness using K-mean and A-priori algorithms. This enables clinicians to evaluate the patient's status and accurately locate and diagnose the ailment. Additionally, you may use patient information systems to store particular patient information and maintain a patient record [14].

Numerous these techniques are beneficial for forecasting cardiac diseases. However, these approaches are inefficient in predicting cardiovascular illnesses. Additionally, specificity is low, which is necessary for any predicting method for HD .

This documentation established a new protocol for the accurate diagnosis of coronary diseases, therefore mitigating problems. The suggested noise reduction approach was built utilizing MATLAB programs and the ADAP clustering algorithm.

The remainder of the manuscript is structured as follows: The next section discusses the related work. The proposed work in section III encompasses dataset description, preprocessing, and noise reduction models. Section IV contains experimental analysis, which details the parameters used to test the presence of the algorithm. Finally, section V concludes the article.

2. Background Study

Dimmer, C. et al. [2] When only individuals with normal ventricles are studied, significant coronary atherosclerosis is not related to altered autonomic tone. When an aberrant HRV value is identified, and concurrent disorders are ruled out, it is reasonable to conclude that CAD was worsened by infarction or left ventricular dysfunction.

Jabbar, M. A., & Samreen, S. [3] used the HNB classifier to identify cardiac disease. The authors used HNB in the cardiac stalog data set and evaluated its performance. The experimental results indicate that the HNB model outperforms alternative approaches.

Krishnani, D. et al. [4] The authors present an intensive preprocessing effort. Random Forest is the most suitable candidate for the prediction model and provides the greatest performance measure compared to K-Nearest Neighbor and Decision Tree. Although K-Nearest Neighbour takes the longest to execute, the

performance metrics are close to the Decision Tree. Thus, given a comparable environment to the one in the utilized dataset, if all characteristics are preprocessed to create a normal distribution, Random Forest is an excellent choice for obtaining a robust prediction model.

Kavitha, M. et al. [5] HD is a leading cause of death globally. The disease is jeopardized by changing lifestyles and a lack of physical activity. The medical business offers a variety of diagnostic procedures. However, machine learning is believed to be the most accurate option. This study makes use of a TkInter Python application for the prediction of heart illness. As a hybrid model, this system predicts cardiac disease using Decision Tree and Random Forest combinations.

Mischie, N., & albu, A. [6] The quantity of medical data that must be analyzed to judge a patient grows. As a result, doctors need computerized decision-support tools to assist them in their decision-making process. Artificial neural networks are an extremely useful tool for assisting the medical system. The program was developed to forecast whether or not a person would get HD in the following ten years based on medical data. If the prediction is positive, physicians may recommend that patients give up specific addictions or take particular drugs to greatly lower their chance of getting coronary HD.

Sengur, A., & Turkoglu, I. [11] The authors examined the utility of an AIS-based fuzzy k-NN approach for identifying cardiac aortic and mitral valve problems. Numerous experimental experiments were conducted to this end. Additionally, many statistical validation indices were employed to evaluate the efficacy of this technique.

Wang, J. et al. [13] The model parameters used by the authors are not ideal. The authors concentrate primarily on the method for finding models using tenfold cross-validation. Changes in model parameters will also significantly affect the final findings. Second, training numerous models at each level is time-consuming and the strategy is incapable of rapidly narrowing the search results.

Zhang, Y. et al. [14] The SVM technique was utilized in this article to translate nonlinear divisible CHD. The ideal plane is then utilized to perform linear classification. Data normalization and feature extraction resolved the uneven data's falling classification accuracy.

According to Zoubida Alaoui Mdaghri et al. [15], data mining may analyze health data for various reasons and investigations. Coordination and analysis of many forms of healthcare data across time may provide answers to a vast variety of emerging healthcare concerns. Data mining technologies may assist professionals in the area in obtaining a second opinion on the majority of results, especially to guarantee that the illness is not overlooked during analysis. Similarly, by applying predictive models, high-risk individuals may be detected earlier in their disease progression, providing greater care to the patient and lowering healthcare expenditures via intervention and counteractive action plans.

a) Research Gap

Day by day, inadequate blood flow impairs cardiac function, clogging the arteries with plaque. This is referred to as the pre-CAD stage. Plaque is composed of cholesterol and chemicals. Gradually, the oxygen level required for cardiac function decreases. The valve and wall will contract during this period, giving the reference picture an odd look. Our suggested approach is utilized to detect CAD more accurately and early.

3. Proposed Model

Techniques from Data Mining (DM) have been used to diagnose cardiac problems. They are, however, constrained by several data quality issues, including inconsistencies, noise, missing data, outliers, excessive dimensionality, and unbalanced data. Thus, Data Preparation (DP) methods were utilized to prepare data to increase the performance of DM-based prediction systems for HD. The Adaptive Damping-based affinity Propagation Clustering technique is utilized in this study to do noise reduction and clustering. The performance was compared to that of other methods.

3.1 Dataset collection

The dataset has collected from https://www.kaggle.com/ronitf/heart-disease-uci link with 14 parameters, and 300 records are contained. Like Age, Gender, Cp, Trestbps, Chol, Fbs, Restecg, Thalach, Exang, Oldpeak, Slope, Thal, Target

3.2 Affinity propagation clustering algorithm

There is no need to provide the cluster count in Affinity Propagation. Each data point sends signals to every other place, informing them of the attractiveness of the sender to their intended recipients. Based on the appeal of messages received from all other senders, senders are told when a target is available for affiliation, and each target responds. Senders reply to targets by sending messages informing them of changes in their updated relative attractiveness to that target when messages from all targets are available. The procedure of message transmission is repeated until an agreement is reached. Once a sender and one of its targets are associated, the target acts as an illustration for the point. All points that share an example are grouped.

Uses four matrices to cluster, i.e.,

1. Similarity matrix,
    - Minimum possible similarity
    - Maximum possible similarity
    - Median Possible similarity.
2. Availability matrix,
3. Responsibility matrix,
4. Criterion matrix

Our technique determines the data's similarity index using Median Possible similarity. Data points convey responsibility messages to other data points based on available information from other data points (via the example) (to exemplar).

The exemplars may be found by adding up the responsibilities and availability of all data points. Following the exemplar identification, the information points are allocated to the exemplar to build the clusters.

3.3 Similarity matrix

Each column in the similarity matrix is produced by subtracting the squares of the participants' differences. By negating the distances between objects, the similarity matrix is generated. Typically, these distances are determined by adding the squares of the differences between the variables that comprise the items.

3.4 Responsibility matrix

Data points are used to communicate information through message forwarding. The process generates two distinct messages, each addressing a distinct kind of competition. The first is referred to as "responsibility" (I-k) sent from information point, I to potential exemplar point k.

3.5 Availability matrix

It collects data points in order to assess the quality of each prospective example. It stretches from Point K of the Candidate Rep. to Point 1 as evidence that we should use Point K as an example. Based on evidence from previous points, point k is used as an example.

3.6 Exemplar

Following iterative communication passing, exemplars can be recognized by manipulative the utmost $(i, k) + (i, k)$ for point i. If $k=i$, point i is preferred as an exemplar, or position k is the exemplar of position.

3.6.1 PROCESS:

Step 1: Find the similarities of the data
Step 2: Initialize the availability as zero
Step 3: Update the availabilities and responsibilities
Step 4: Identify the exemplars for data point i by adding responsibilities and availabilities.
Step 5: If Exemplars didn’t modify for a flat quantity of iterations, go to step (6). Else, go to Step (1)
Step 6: Allocate the information points to Exemplars based on utmost similarity to discover clusters

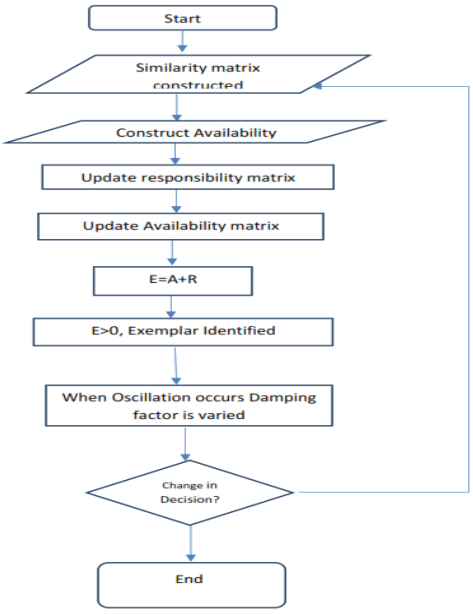

Figure 2: Proposed Flow chart for Noise Removal and clustering

3.7 AFFINITY PROPAGATION PSEUDO-CODE:

1. Calculate the similarity matrix using Median Possible similarity.

$$Dist(i\ ,\ j)= -||xi-xj||^2$$

2. Calculate Responsibility *R(i,k)*

$$R(i,k)=s(i,k)-max(a(i,k)-s(i,k))$$

s(i,k)->Similarity Matrix
a(i,k)-> Availability Matrix

3. Add the damping factor.

$$Rf= lam*Rold$$
$$R=(1-lam)*R+ Rf$$

Lam-> Damping Factor (lam=0.5)
R ->R_{t+1}(i,k)
Rold ->R_t(i,k)

4. Calculate the Availability

$$A\ (i,\ k)\ =min\ \{0,\ r(k,\ k)+\ \sum max(0,r(i,\ k)\}$$

5. Add the damping factor for accessibility

$$A=(1-lam)*A+lam*Aold$$

Lam-> Damping Factor (lam=0.5)
A ->R_{t+1}(i,k)
Aold -> R_t(i,k)

6. Calculate exemplar

$$E=(Diag(A)+Diag(R))>0$$

A->Availability
R->Responsibility

7. From step 2 to step 6 process continues to reach the maximum iteration.
8. Calculate the convergence from the exemplar
9. Identify the clusters

AP is a novel and effective method, for example, learning. It is a quick clustering technique, particularly useful when dealing with large numbers of clusters, and it has many benefits, including speed, universal application, and high performance. When oscillations arise, AP is unable to eradicate them automatically. An adjustable damping factor is developed to reduce oscillations and enhance clustering performance to meet this need.

4. Results And Discussion

The proposed system was developed using the MATLAB programming language. According to World Health Organization (WHO) studies, CAD is the most frequently known disease that causes death globally, especially in underdeveloped countries. To overcome these obstacles, anticipating the occurrence of the disorder is critical. Thus, the results are compared to those obtained using the K-means method, the expectation-maximization algorithm, the farthest first algorithm, the density-based clustering algorithm, and the hierarchical clustering.

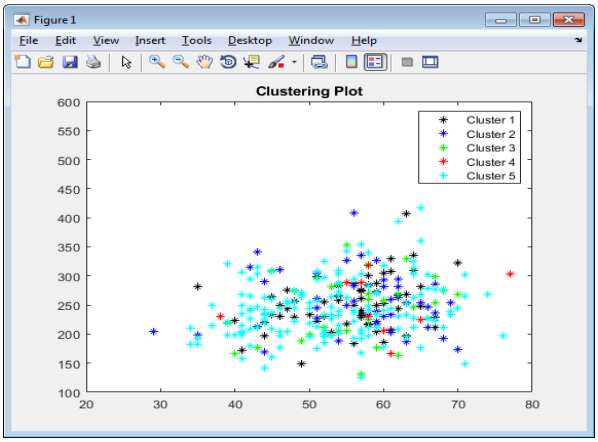

Fig 3: Clustering Plots

Figure 3 indicates cluster 1 to 5 clusters has plotted in each section

Table 1: FM index level

Sno	Algorithm	FM Index
1	K-means	0.78
2	Expectation Maximization	0.73
3	Farthest First	0.80
4	Density Bas Cluster	0.79
5	Hierarchical cluster	0.71
6	Affinity Propagation	0.90

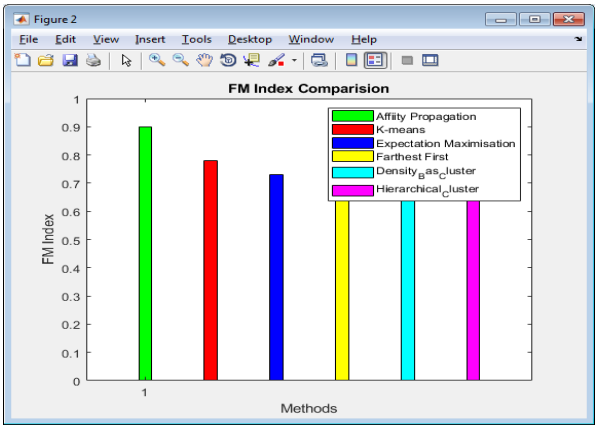

Fig 4: FM index Comparison chart

Table 1 indicates the FM index comparison value, and figure 4 indicates the comparison chart. The X-axis indicates the various methods and the proposed method, and Y-axis indicates the FM index level.

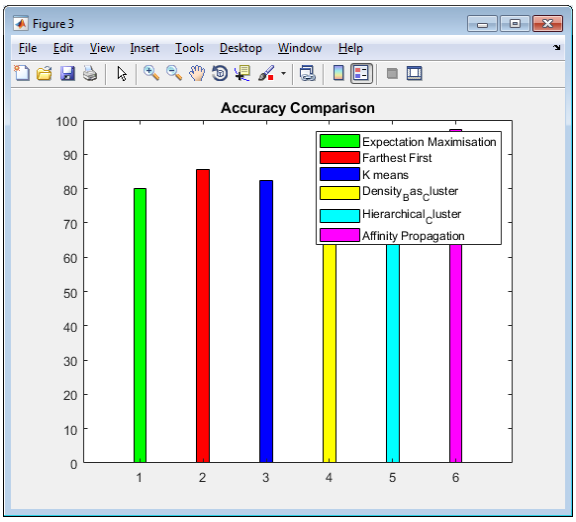

Fig 5: Accuracy Comparison chart

Figure 5 indicates the accuracy comparison chart for various algorithm comparisons. In x-axis represents the different ways and Y-axis indicates the accuracy percentage.

Table 2: Error Rate value

Sno	Algorithm	Error Rate
1	K-means	0.17
2	Expectation Maximization	0.20
3	Farthest First	0.14
4	Density Bas Cluster	0.16
5	Hierarchical cluster	0.27
6	Affinity Propagation	0.02

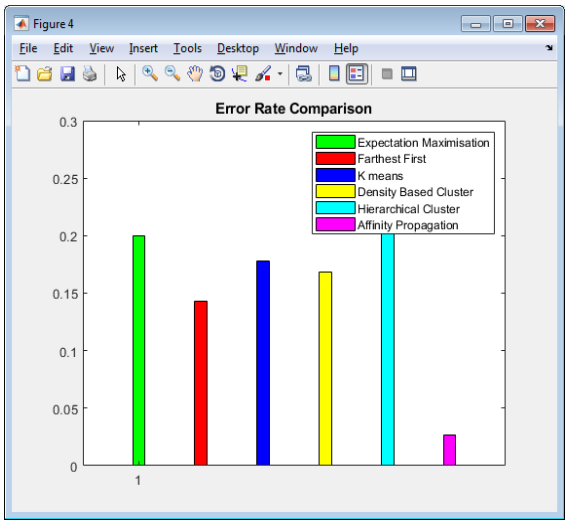

Fig 6: Error Rate Comparison

Table 2 indicates the error rate values, and figure 6 indicates the error rate comparison chart. The x-axis shows the different algorithms, while the y-axis shows the Error rate value level.

Table 3: Performance time comparison

Sno	Algorithm	Time (ms)
1	K-means	0.60
2	Affinity propagation	0.14

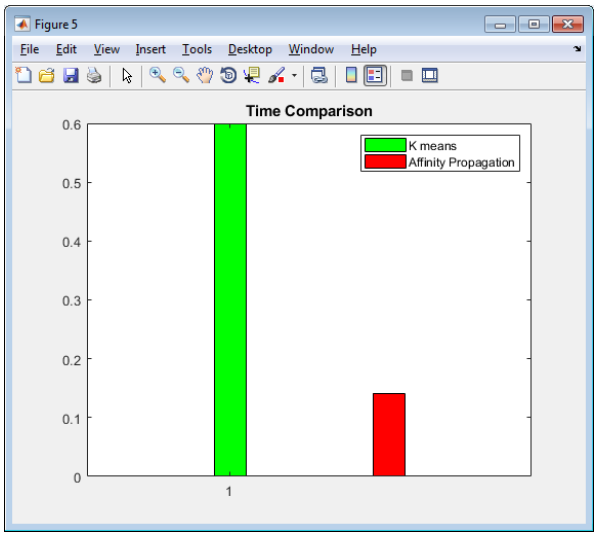

Fig 7: Time Comparison

Table 3 indicates the execution time in the proposed and existing algorithm, and the chart is displayed in figure 7.

Table 4: Performance metrics

Sno	Metrics	Accuracy (%)
1	Accuracy	97.3
2	Sensitivity	92.3
3	Specificity	97.7
4	Precision	92.2

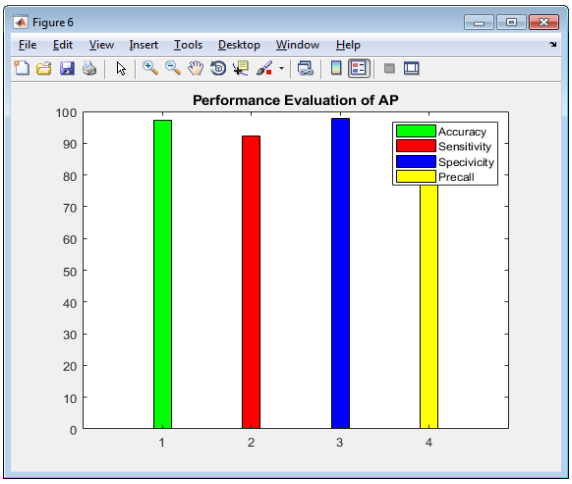

Fig 8: Performance Evaluation of Affinity propagation

Table 4 indicates the performance evaluation in the affinity propagation method. Figure 8 indicates the graphical chart for performance evaluation. X-axis measures performance while Y-axis shows accuracy as a percent.

5. Conclusion

The primary objective of this study is to quantify the prevalence of CAD more precisely. Data mining is a daunting task that requires fewer qualities than a series of tests. The ADAP clustering method is proposed in this study to address one of the fundamental drawbacks of basic noise reduction approaches that need the input of several clusters. This structure may also be used for future tasks. Additionally to the items listed above, it may involve other medical characteristics. Text mining, available via the healthcare sector archive, is a technique for mining massive volumes of unstructured data.

Conflicts of Interest

The authors declare no conflicts of Interest

Author Contribution

This work is the contribution of Authors: “Conceptualization , Ramasamy MaheshKumar and Sundaram Veni; Methodology, Ramasamy MaheshKumar; software, Ramasamy MaheshKumar; Validation, Ramasamy MaheshKumar and Sundaram Veni; Formal Analaysis, Ramasamy MaheshKumar; writing—Original Draft Preparation, Ramasamy MaheshKumar; Writing—Review and Editing, Ramasamy MaheshKumar and Sundaram Veni.

References

[1] Ang, Q., Liu, Z., Wang, W., & Li, K. (2010). Explored research on data preprocessing and mining technology for clinical data applications. *2010 2nd IEEE International Conference on Information Management and Engineering*. doi:10.1109/icime.2010.5477660

[2] Dimmer, C., Gregoire, J.-M., Tavernier, R., & Jordaens, L. (n.d.). Autonomic tone assessed by heart rate variability in uncomplicated coronary artery disease. *Computers in Cardiology 1996.* doi:10.1109/cic.1996.542463

[3] Jabbar, M. A., & Samreen, S. (2016). Heart disease prediction system based on hidden naïve bayes classifier. *2016 International Conference on Circuits, Controls, Communications, and Computing (I4C).* doi:10.1109/cimca.2016.8053261

[4] Krishnani, D., Kumari, A., Dewangan, A., Singh, A., & Naik, N. S. (2019). Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*. doi:10.1109/tencon.2019.8929434

[5] Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021). Heart Disease Prediction using Hybrid machine Learning Model. *2021 6th International Conference on Inventive Computation Technologies (ICICT).* doi:10.1109/icict50816.2021.93585

[6] MISCHIE, N., & ALBU, A. (2020). Artificial Neural Networks for Diagnosis of Coronary Heart Disease. *2020 International Conference on e-Health and Bioengineering (EHB).* doi:10.1109/ehb50910.2020.9280271

[7] M.A.Jabbar,"Heart Disease Prediction System using Associative Classification and Genetic Algorithm", *ICECIT*, pp 183-192, Elsevier, vol 1(2012)

[8] Pathak, A., Samanta, P., Mandana, K., & Saha, G. (2020). An improved method to detect coronary artery disease using phonocardiogram signals in noisy environment. *Applied Acoustics, 164, 107242*. doi:10.1016/j.apacoust.2020.10724

[9] Rao, K., Gopal, P. R., & Lata, K. (2021). Computational Analysis of Machine Learning Algorithms to Predict Heart Disease. *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence).* doi:10.1109/confluence51648.2021

[10] Runjing, Z., & Keyang, L. (2011). Fisher classifier in diagnosis of coronary heart disease. *2011 4th International Congress on Image and SignaProcessing.* doi:10.1109/cisp.2011.6100787

[11] Sengur, A., & Turkoglu, I. (2008). A hybrid method based on artificial immune system and fuzzy k-NN algorithm for diagnosis of heart valve diseases. *Expert Systems with Applications, 35(3), 1011–1020.* doi:10.1016/j.eswa.2007.08.003

[12] Terzi, M. B., & Arikan, O. (2019). Coronary Artery Disease Detection by using Support Vector Machines and Gaussian Mixture Model. *2019 Medical Technologies Congress(TIPTEKNO).* doi:10.1109/tiptekno.2019.889495

[13] Wang, J., Liu, C., Li, L., Li, W., Yao, L., Li, H., & Zhang, H. (2020). A stacking based model for non-invasive detection of coronary heart disease. *IEEE Access, 1–1*. doi:10.1109/access.2020.2975377

[14] Zhang,Y., Liu, F., Zhao, Z., Li, D., Zhou, X., & Wang, J. (2012). Studies on Application of Support Vector Machine in Diagnose of Coronary Heart Disease. *2012 Sixth International Conference on Electromagnetic Field Problems and Applications.* doi:10.1109/icef.2012.6310380

[15] Zoubida Alaoui Mdaghri, Mourad El Yadari, Abdelillah Benyoussef, Abdellah El Kenz "Study and analysis of Data Mining for Healthcare", *2016 4th IEEE International Colloquium on Information Science and Technology.*