

Enhanced Prediction of Heart Disease Using Particle Swarm Optimization and Rough Sets with Transductive Support Vector Machines Classifier



M. Thiyagaraj and G. Suseendran

Abstract Over the last decade heart disease has significantly increased and it has emerged to be the primary reason behind the mortality in people living in many nations across the world. The computer-assisted systems act as a tool for the doctors in the prediction and diagnosis of heart disease. In the medical domain, Data Mining yields a variety of techniques that are extensively employed in the medical and clinical decision support systems that has to be quite useful in diagnosing and predicting the heart diseases with less time and good accuracy to improve their health. The previous system designed a radial basis function with support vector machine for heart disease prediction. However it does not provides a satisfactory classification result. To solve this problem the proposed system designed a Particle Swarm Optimization and Rough Sets with Transductive Support Vector Machines (PSO and RS with TSVM) based prediction is performed. In this proposed work, the dataset of the heart disease is collected from UCI repository. In order to reduce data redundancy and improve data integrity, the data normalization is performed by using Zero-Score (Z-Score). Then Particle Swarm Optimization (PSO) algorithm and Rough Sets (RS) based attribute reduction technique is used for selecting the optimal subset of attributes that, in turn, minimizes the computational hurdles and improves the performance of the prediction system. Finally, the Radial Basis Function-Transductive Support Vector Machines (RBF-TSVM) classifier is used for heart disease prediction. The results obtained from the experiments indicate that the system proposed accomplishes a superior performance in comparison.

Keywords Particle swarm optimization (PSO) · Rough sets (RS) · Radial basis Function-Transductive support vector machines (RBF-TSVM)

M. Thiyagaraj (✉)

Department of Information and Technology, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India

e-mail: mthyagarajphd@gmail.com

G. Suseendran

Department of Information and Technology, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India

e-mail: suseendar_1234@yahoo.co.in

© Springer Nature Singapore Pte Ltd. 2020

N. Sharma et al. (eds.), *Data Management, Analytics and Innovation*,

Advances in Intelligent Systems and Computing 1016,

https://doi.org/10.1007/978-981-13-9364-8_11

1 Introduction

“Data Mining refers to the sensitive extraction of inherent, earlier unknown and potentially helpful knowledge about data” [1]. Briefly said, it is defined as the process of evaluating the data from various points of view and collecting. Data mining offers a number of approaches for the discovery of hidden patterns out of the data. One important problem encountered in the Healthcare domain is associated with quality of service. Quality of service indicates the correct diagnosis of the disease and yielding treatments that are effective to patients. Incorrect diagnosis can result in dangerous results that cannot be accepted [2].

As per the survey of WHO, an overall of 17 million deaths globally are because of heart attacks and strokes. The deaths occurring because of heart disease in several nations are the result of work overload, emotional stress, and several other issues. Overall, it has been observed to be the major cause behind the death in adult people [3]. The diagnosis is a complex and significant task, which has to be executed with accuracy and efficiency. The diagnosis is frequently done, on the basis of the experience and knowledge of the doctor. This results in incorrect outcomes and expensive medical treatments given to patients [4].

Cardiovascular disease is a type of critical health-endangering and often occurring disease. The world health organization has made an estimation that 12 million deaths happen across the entire world, each year the reason being the cardiovascular disease. Progress made in the medical field over the last few decades facilitated the recognition of risk factors, which may have contributed to the cardiovascular diseases [5]. The most typical reason of heart disease includes the narrowing or blockage of the coronary arteries, the blood vessels, which are responsible for supplying blood to the heart itself. This is known as coronary artery disease and it occurs gradually as time passes. It's the primary reason that people suffer from heart attacks. A blockage, which does not get treatment within a couple of hours makes the affected heart muscle to face death. Nearly 30% of all those who suffered from heart attacks experienced no symptoms. But, apparent symptoms of the attack stay in the bloodstream for many days. Medical diagnosis is a vital but sophisticated task, which has to be performed with accuracy and efficiency and its automation would prove to be very helpful. Unfortunately, all the doctors do not have equal skills in all the subspecialties and in several regions. With these many factors used for the analysis of the heart attacks' diagnosis, physicians usually diagnose by analyzing the current test results of the patient [6]. The physicians also investigate the earlier diagnoses done on other patients with the same kind of results. These sophisticated procedures are nontrivial. Hence, a physician has to be experienced and hugely expertized for diagnosing the patients' heart attacks. Subsequently, the undertakings made to utilize the learning and experience of different experts and the clinical screening information of patients aggregated in databases for enabling the analysis procedure is respected to be a useful framework, which is the mix of clinical choice help and PC based patient records and could limit the therapeutic mistakes, enhance the safety of patient, reduce unnecessary rehearse contrasts, and improve the patient outcomes.

The rest of the paper is organized as given below: Section 2 discusses the various methods that have been used for heart diseases prediction. Section 3 focuses on our proposed methodology of classification method, and PSO algorithm that is used for attribute reduction was performed. Section 4 provides the experimental results that were conducted, and finally Sect. 5 concludes the work.

2 Literature Review

Dangare et al. [7] assessed the forecast frameworks utilized for Heart ailment making use of a broad number of data attributes. The structure applies remedial terms like sex, circulatory strain, cholesterol-like 13 characteristics for foreseeing likelihood being influenced by illness. Till now, 13 properties are used for forecast purposes. This exploration work has included another two characteristics, which are corpulence and smoking. The information mining arrangement strategies.

Anooj et al. [8] portrayed a weighted fluffy control based clinical choice emotionally supportive network (CDSS) for diagnosing the coronary illness, via naturally procuring the learning from the clinical information of the patient. The recently presented clinical choice emotionally supportive network utilized for hazard expectation of heart patients involves two phases, (1) robotized plot for creating the weighted fluffy tenets, and (2) plan of a fluffy govern based choice emotionally supportive network. In the main stage, the mining approach, quality choice and trait weight age system are used forgetting the weighted fluffy principles. After this, the fluffy framework is planned by the weighted fluffy guidelines and chose characteristics. Finally, the investigation is performed on the recently presented framework utilizing framework as far as exactness, affectability, and specificity.

Tahseen et al. [9] utilized information mining approaches. It was assumed that the reliant variable was taken to be the conclusion—having dichotomous qualities showing the diseases existence or nonattendance. Twofold relapse has been connected to the components of the reliant variable. The informational collection has been procured from two diverse cardiovascular doctor's facilities in Karachi, Pakistan. An aggregate of sixteen factors out of which one was thought to be reliant and the rest of the 15 free factors are utilized. Information Reduction methodologies, for example, rule segment examination was utilized for having a superior execution of the relapse demonstrate in the forecast of Acute Coronary Syndrome. Only 14 out of sixteen variables have been viewed as in light of the results of information decrease.

Making use of neural network, Shantakumar et al. [10] presented a brilliant and effective heart assault expectation framework. For extricating the critical examples from the coronary illness databases for doing heart assault forecast, an efficient technique has been presented. At point when the preprocessing is done, the coronary illness vault was bunched with the assistance of the K-means grouping calculation that will play out the extraction of the information suitable for heart assault from the store. Hence, with the assistance of the MAFIA calculation, the incessant examples that are a match to coronary illness are mined from the information extricated. Heart

assault expectation proficiently, the neural system was prepared with the picked noteworthy examples. Utilizing the Back-spread as the preparation calculation, the Multi-layer Perceptron Neural Network has been utilized. Outcomes accordingly accomplished have demonstrated that the new expectation framework was proficient in the forecast of the heart assault with effectiveness.

Parthiban et al. [2] drawn closer a coactive neuro-fluffy surmising framework (CANFIS) for coronary illness forecast. The recently presented CANFIS demonstrate coordinated the neural system versatile potential and the fluffy rationale subjective strategy that, thus, is joined with hereditary calculation for diagnosing the presence of the ailment. The exhibitions of the CANFIS display were evaluated as far as the preparation exhibitions and order correctnesses and the outcomes uncovered that the CANFIS show proposed has amazing capability with respect to the coronary illness forecast.

Anbarasi et al. [11] portrayed about the forecast of the nearness of heart disease with more exactness using limited number of properties. As a matter of fact, thirteen qualities were utilized in the forecast of the coronary illness. This exploration work, Genetic calculation is used for deciding the traits that help more in the diagnosing the heart sicknesses that in a roundabout way restricts the quantity of tests required. Ascribes constrained characteristics with the assistance of hereditary hunt. Next, three classifiers, for example, Naïve Bayes, Classification by bunching and Decision Tree are used for anticipating the finding of patients as precisely as it is accomplished before the minimization of the quantity of properties. What's more, the deductions demonstrate that the execution of Decision Tree information mining procedure is vastly improved than other two information mining approaches once the highlight subset choice is consolidated with significantly high model development time. The execution of Naïve Bayes is reliable earlier and after the decrease in the number of traits with a similar time taken for display development. The execution of Classification through bunching is poor in examination with other two methods.

Soni et al. [12] developed a GUI based Interface for entering the patient record and for anticipating if the patient is experiencing Heart contamination or not utilizing Weighted Association poisonous Classifier. The longing is done through the mining of certain information or information storeroom of the patient. It has just been built up that the execution of the Associative Classifiers is great in correlation with conventional classifiers methods including choice tree and manage acceptance.

3 Proposed Methodology

In PC helped coronary illness finding systems, the information is obtained from a couple of different sources and is analyzed using PC based applications. PCs have for the most part been used for building learning based clinical choice emotionally supportive networks that made utilization of the data acquired from therapeutic expert, and the exchange of this information into PC calculations was performed with manual intercession. This procedure devours much time and is basically dependent on

the supposition of the restorative master that might be abstract. In order to deal with this issue, machine learning approaches have been designed to obtained knowledge by automatic means from examples or unprocessed data.

The proposed system designed a Particle Swarm optimization algorithm with Rough set and Radial Basis Function based Transductive Support Vector Machines (TSVM) for heart diseases prediction. The elaborate design of heart disease diagnosis system comprises of three important stages: Normalization, attribute reduction, feature extraction, and classification (Fig. 1).

3.1 Normalization Using Z-Score Normalization

All the input and output data underwent normalization prior training and the testing processes so as to guarantee that there is no overwhelming of data in terms of distance metric.

In Z-Score standardization, ordinarily it is helpful. For the standardization of an arrangement of scores making utilization of the standard deviation, each score is isolated by the standard deviation of this arrangement of scores. Appearance of this, the vast majority of the occasions, the mean of the scores is subtracted from each score before isolating by the standard deviation. This standardization is called as Z-scores [13]. Mathematically, each set of N scores is represented by Y_n and whose mean is equivalent to M and whose standard deviation is equivalent to \hat{S} gets transformed in Z-scores to be

$$Z_n = \frac{Y_n - M}{\hat{S}} \tag{1}$$

Using basic algebraic expressions, it can be proven that a set of Z-score has a mean equivalent to zero and a standard deviation of one. Hence, Z-scores have a unit free measure that can be utilized for comparing the observations that are measured with various units. Once the normalization is done, the transformed datasets are applied for attribute reduction technique.

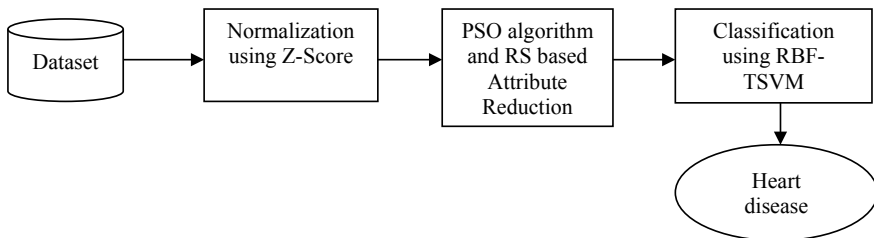


Fig. 1 Block diagram of the proposed methodology

3.2 Particle Swarm Optimization (PSO)

Fundamentally a streamlining calculation in the swarm insight field. Accept in D measurements seek space, n particles are introduced in irregular [14, 15]. The PSO considers the present data of the populace to be its examination arrangement to be exploration field, a few times of data cycles and exchange, it will get the current worldwide ideal arrangement. Each molecule at that point refreshes its speed and position amid each cycle in view of Eqs. (2) and (3)

$$v_{id}^{t+1} = v_{id}^t + c_1 R_1 * (p_{id}^t - X_{id}^t) + c_2 R_2 * (p_d^t - X_{id}^t) \quad (2)$$

$$X_{id}^{t+1} = X_{id}^t + v_{id}^{t+1} \quad (3)$$

In above conditions, the nearby and worldwide optima. R1 and R2 stand for an irregular number somewhere in the range of 0 and 1. C1 and C2 specify a positive consistent that is known as quickened factor utilized for altering the progression length in the nearby and worldwide ideal heading and generally its esteem is somewhere in the range of 0 and 4.

$$v_{ij}(t+1) = \omega * v_{ij}(t) + c_1 R_1 * (p_{best}(t) - X_{ij}(t)) + c_2 R_2 * (p_{best}(t) - X_{ij}(t)) \quad (4)$$

$$X_{ij}(t) = \begin{cases} 1, & \rho < s(v_{ij}(t)), \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The calculation will perform better when ω that is known as weight factor esteems goes somewhere in the range of 0.4 and 0.9 [16]. ρ refers to an irregular number somewhere in the range of 0 and 1. $s(v_{ij}(t))$ refers to a fluffy capacity by and large utilized in neural system, and its capacity recipe is as communicated as underneath:

$$s(v_{ij}(t)) = \frac{1}{1 + e^{-v_{ij}(t)}} \quad (6)$$

p_{best} and g_{best} refers to the individual extremum and global optimal solution correspondingly that are represented by Eqs. (7) and (8):

$$p_{best} = \max(p_{best}, \text{fitness}(i)) \quad (7)$$

$$g_{best} = \max(p_{best}, g_{best}) \quad (8)$$

Well-being goes about as the sole marker for reflecting and managing the swarm molecule to keep moving toward the ideal arrangement. Since the decrease calculation is subject to positive area decrease calculation [17], it modifies the wellness work for accomplishing a base decrease with various outcomes much

advantageously Eq. (9).

$$\text{fitness}(i) = \begin{cases} |i|, & \text{if } (\text{pos}'_{|i|}(D)) = U'_{\text{POS}} \\ |c|, & \text{if } (\text{pos}'_{|i|}(D)) \neq U'_{\text{POS}} \end{cases} \quad (9)$$

3.2.1 Rough Set Algorithm for Attribute Reduction Through Particle Swarm Optimization

Twofold division encoding mapping technique is utilized to unite the obnoxious set and the molecule swarm figuring in this examination work. Qualities estimation of the molecules are mapped specifically and the measurement is set “0” or “1” where “0” shows its particular quality ought to be decreased in a definitive outcome when “1” won’t. From there on, it considers the BPSO calculation to be the heuristic data and tosses it into the positive area quality decrease calculation. In this way, the calculation gives arrangement to the issue all the more effortlessly and productively, and the outcomes incorporate different learning decrease, and the calculation will step if the most extreme number of cycle is reached [18].

3.3 Classification Using Radial Basis Function-Transductive Support Vector Machines (RBF-TSVM)

Classification of RBF based TSVM support to use for heart diseases prediction. In previous work briefly discussed about the RBF method.

As TSVM algorithm makes utilization of the idea of transductive learning with effectiveness, it can combine the showed dispersion data having a place with unlabeled examples and preparing tests much better. Subsequently, in correlation with the traditional help vector machine calculation, TSVM algorithm offers more noteworthy arrangement precision. Be that as it may, there are a couple of disadvantages with TSVM calculation, similar to the number N of positive name tests present in the unlabeled examples for TSVM algorithm must be physically indicated, however N esteem is for the most part hard to make an important gauge [19].

TSVM algorithm makes utilization of a basic technique for assessing the estimation of N that shows the estimation of the proportion of tests with positive names to all the unlabeled examples in view of the proportion of tests with positive marks and all the named tests. Be that as it may, if the quantity of tests with marks is less, it turns out to be difficult for the method to get a more precise estimation of N evaluated. At the point when the pre-characterized estimation of N is very dissimilar from the first number of tests having positive names, the execution of the TSVM calculation will have a tendency to end up exceptionally temperamental, and also, the order exactness of the calculation can’t be guaranteed to be successful [20].

Before long, the major immense work of transductive end in the zone of assistance vector purchasing consolidates Transductive Support Vector Machine (TSVM) that will be in no time clarified in whatever is left of this division. Given a course of action of self-ruling, correspondingly scattered named points of reference.

$$(x_1, y_1), \dots, (x_n, y_n), x_i \in R^m, y_i \in \{-1, +1\} \quad (10)$$

and another set of unlabeled examples from the identical sharing,

$$x_1^*, x_2^*, x_3^*, \dots, x_k^*$$

:

$$(y_1^*, \dots, y_k^*, w, b, \xi_1, \dots, \xi_n, \xi_1^*, \dots, \xi_k^*)$$

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + C^* \sum_{j=1}^k \xi_j^* \quad (11)$$

Subject to:

$$\begin{aligned} \forall_{i=1}^n : y_i t [w \cdot x_i + b] &\geq 1 - \xi_i \\ \forall_{j=1}^k : y_j^* [w \cdot x_j^* + b] &\geq 1 - \xi_j^* \\ \forall_{i=1}^n : \xi_i &\geq 0 \\ \forall_{j=1}^k : \xi_j^* &\geq 0 \end{aligned}$$

The Radial premise work (RBF) based piece otherwise called the Gaussian part does the mapping of the lower dimensional component space onto an unending dimensional space. At the point when directly indistinguishable highlights are mapped onto higher dimensional space they oftentimes turn out to be straight differentiable.

The RBF based TSVM serves the purpose of classification is used to find the heart diseases prediction.

4 Experimental Results

Efficient Heart Disease Prediction is the most significant method to detect heart diseases. Here proposed PSO and RBF-TSVM approach and existing system IT2FLS [21] and modified FA and RBF-SVM are compared in terms of Sensitivity, Specificity, and Accuracy.

False Positive Rate (FPR)

Defined as the percentage of cases in which an image was segmented to tumor portion, but actually did not.

$$FPR = \frac{FP}{FP + TN} \tag{12}$$

False Negative Rate (FNR)

Defined as the percentage of cases in which an image was segmented to non-tumor portion, but actually it did.

1. Accuracy

The weighted percentage of tumor parts in images that is segmented correctly is measured by the metric accuracy. It is expressed as,

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \tag{13}$$

Figure 2 illustrates that the proposed PSO and RBF-TSVM based classification approach is compared with the existing IT2FLS and MFA and RBF-SVM based classification approach in terms of accuracy [22]. The size of dataset is taken as X-axis and accuracy is taken in Y-axis. In order to achieve high accuracy the proposed system used PSO is used for attribute reduction. It concludes that the PSO and based RBF-TSVM classification approach has show the high accuracy results for all size of dataset compared with the existing method.

2. Sensitivity

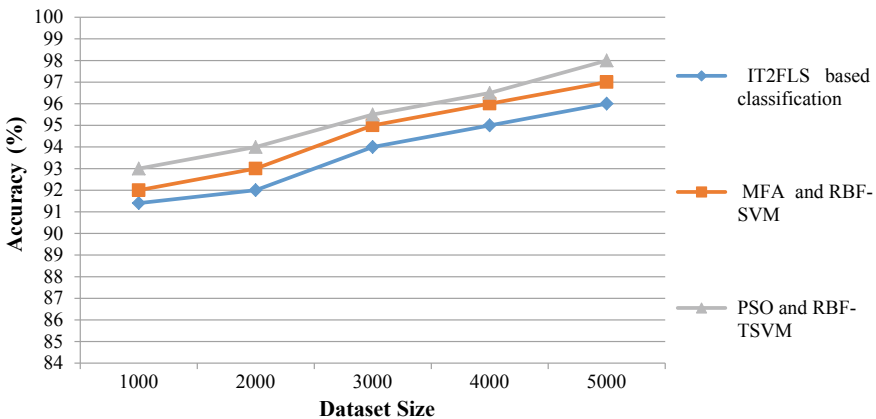


Fig. 2 Accuracy comparison

The sensitivity measure is defined as the ratio of actual positives that are properly identified. It is associated with the capability of test to identify positive results.

$$\text{Sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} \times 100 \quad (14)$$

Figure 3 illustrates that the proposed PSO and RBF-TSVM based classification approach is compared with the existing IT2FLS and MFA and RBF-SVM based classification approach in terms of accuracy. The size of dataset is taken as X-axis and sensitivity is taken in Y-axis. Z-Score algorithm is used for normalization purpose. And also an efficient classification is done by using RBF-TSVM. It improves the true positive rate. For all size of dataset the proposed PSO and based RBF-TSVM classification approach has shown the high sensitivity results compared with the existing system.

3. Specificity

The specificity measure is defined as the ratio of negatives that are properly identified. It is associated with the capability of test to identify negative results.

$$\text{Specificity} = \frac{\text{Number of true negatives}}{\text{Number of true negtaives} + \text{number of false positives}} \times 100 \quad (15)$$

Figure 4 illustrates that the proposed PSO and RBF-TSVM based classification approach is compared with the existing IT2FLS and MFA and RBF-SVM based classification approach in terms of accuracy [23]. The size of dataset is taken as X-axis and specificity is taken in Y-axis. For all size of dataset the proposed PSO and based RBF-TSVM classification approach has shown the higher specificity results compared with the existing system.

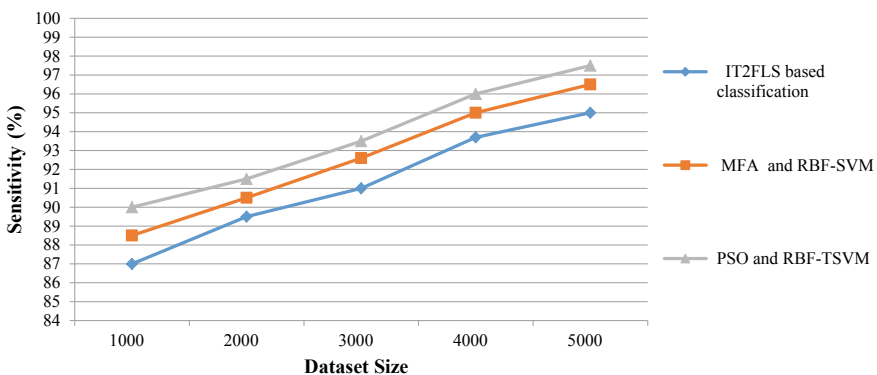


Fig. 3 Sensitivity comparison

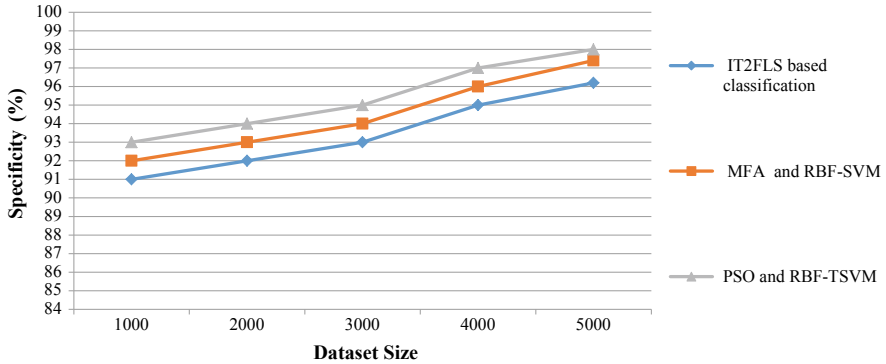


Fig. 4 Specificity comparison

5 Conclusion

In this research work, the proposed system designed a PSO algorithm and RBF based TSVM approach for predicting the heart disease with intelligence and efficiency, and to get over manual labor. After the normalization process, in turn, minimizes the redundancy and improves the performance of classifier and attribute features are also extracted through. Finally the classification is performed by using RBF-TSVM, it predicts the heart diseases. The truth is that humans cannot be replaced by computers and through the comparison of the computer-assisted detection results obtained from the pathological observations, doctors can be informed about the best means of evaluating the areas, which is focused by computer-assisted detection. The results of experiments indicate that the newly introduced system accomplishes a superior execution as far as precision, affectability, and specificity.

References

1. Senthil, D., Suseendran, G. (2017). Data mining techniques using time series analysis. In *Proceedings of the 11th INDIACom; INDIACom-2017; IEEE Conference ID: 40353 2017 4th International Conference on "Computing for Sustainable Global Development* (pp. 2864–2872), March 01st–03rd, 2017 BharatiVidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA) ISSN 0973-7529; ISBN 978-93-80544-24-3.
2. Parthiban, L., & Subramanian, R. (2008). Intelligent heart disease prediction system using CANFIS and genetic algorithm. *International Journal of Biological, Biomedical and Medical Sciences*, 3(3).
3. Thiyyagaraj, M., Suseendran, G. (2017). Review of chronic kidney disease based on data mining. In *Proceedings of the 11th INDIACom; INDIACom-2017; IEEE Conference ID: 40353 2017 4th International Conference on Computing for Sustainable Global Development* (pp. 2873–2878), March 01st–03rd, 2017 BharatiVidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA) ISSN 0973-7529; ISBN 978-93-80544-24-3.

4. Thiagaraj, M., Suseendran, G. (2017). Survey on heart disease prediction system based on data mining techniques. *Indian Journal of Innovations and Developments*, 6(1), 1–9.
5. Senthil, D., Suseendran, G. (2018). Efficient time series data classification using sliding window technique based improved association rule mining with enhanced support vector machine. *International Journal of Engineering & Technology*, 7(2.33), 218–223. <https://doi.org/10.14419/ijet.v7i2.33.13890>.
6. Rohini, K., Suseendran, G. (2016). Aggregated K means clustering and decision tree algorithm for spirometry data. *Indian Journal of Science and Technology*, 9(44).
7. Dangare, C. S., & Apte, S. S. (2012). Improved study of heart disease prediction system using data mining classification techniques. *International Journal of Computer Applications*, 47(10), 44–48.
8. Anooj, P. K. (2012). Clinical decision support system: Risk level prediction of heart disease using decision tree fuzzy rules. *Int J Res Rev ComputSci*, 3(3), 1659–1667.
9. Jilani, T. A., Yasin, H., Yasin, M., Ardil, C. (2009). Acute coronary syndrome prediction using data mining techniques—An application. *World Academy of Science, Engineering and Technology*, 59.
10. Patil, S. B., Kumaraswamy, Y. S. (2009). Intelligent and effective heart attack prediction system using data mining and artificial neural network. *European Journal of Scientific Research*, 31(4), 642–656.
11. Anbarasi, M., Anupriya, E., & Iyengar, N. C. S. N. (2010). Enhanced prediction of heart disease with feature subset selection using genetic algorithm. *International Journal of Engineering Science and Technology*, 2(10), 5370–5376.
12. Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Intelligent and effective heart disease prediction system using weighted associative classifiers. *International Journal on Computer Science and Engineering*, 3(6), 2385–2392.
13. Al Shalabi, L., Shaaban, Z., & Kasasbeh, B. (2006). Data mining: A preprocessing engine. *Journal of Computer Science*, 2(9), 735–739.
14. Kennedy, & Eberhart, R. (1995). Particle swarm optimization. In *IEEE International Conference on Neural Networks* (pp. 129–132).
15. Abraham, A., Guo, H., & Liu, H. (2006). Swarm intelligence: Foundations, perspectives and applications. In *Swarm intelligent systems* (pp. 3–25). Heidelberg: Springer.
16. Liu, H., Abraham, A., & Clerc, M. (2007). Chaotic dynamic characteristics in swarm intelligence. *Applied Soft Computing*, 7(3), 1019–1026.
17. Guan, W., & Bell, D. A. (1998). Rough computational methods for information. *Artificial Intelligence*, 105(98), 77–103.
18. Thyagaraj, M., & Suseendran, G. (2018). An efficient heart disease prediction system using modified firefly algorithm based radial basis function with support vector machine. *International Journal of Engineering & Technology*, 7(2.33), 1040–1045.
19. Bruzzone, L., Chi, M., & Marconcini, M. (2006). A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11), 3363–3373.
20. <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>.
21. Long, N. C., Meesad, P., & Unger, H. (2015). A highly accurate firefly based algorithm for heart disease prediction. *Expert Systems with Applications*, 42(21), 8221–8231.
22. Zheng, X., Zeng, B., & Liu, S. (2008, December). Rough set based attribute reduction and extension data mining. In *Second International Symposium on Intelligent Information Technology Application* (Vol. 2, pp. 109–112), IITA'08. IEEE.
23. Kennedy, J., & Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. In *1997 IEEE International Conference on Systems, Man, and Cybernetics* (Vol. 5, pp. 4104–4108). Computational Cybernetics and Simulation.