Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 12, December 2017, pg.151 – 156

PERFORMANCE ANALYSIS OF SOFT COMPUTING TECHNIQUES TOWARDS HEART DISEASE DIAGNOSIS SYSTEM

S.Nithya
Assistant Professor, Department of Information Technology, KG College of Arts and Science, KGiSl Campus, Coimbatore

Abstract: Data mining is the extraction of concealed prescient data from expansive databases furthermore a capable of new innovation with incredible potential to examine critical data in their data warehouses. Data mining algorithms anticipate future patterns and behaviors, permitting organizations to make proactive and knowledge driven choices. The computerized, forthcoming analyses offered by data mining move beyond investigations of past occasions gave by review tools commonplace of choice emotionally supportive networks. Data mining algorithms can answer business addresses that customarily were excessively prolonged to determine. They scour databases for concealed examples, discovering the prediction of disease that specialists may miss in light of the fact that it lies outside their desires. Manual checking is highly impossible to diagnose for this disease. To predict heart disease several approaches have been carried out. This comparative study paper provides a thorough analysis of various algorithms made towards disease prediction. Several data mining and soft computing approaches are studied. This study concludes that the performance of various algorithms comparison of accuracy, sensitivity and specificity of several algorithms and approaches.

Keywords: Heart Diseases, Weighted Fuzzy Rule, K- Nearest Neighbor, Genetic, Scoring System, PRAA, SVM classifier, Support Vector Machines, Particle Swarm Optimization

1. Introduction:

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve [Han. J, Kamber.M (2006)]. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. The process of data mining consists of three stages: (1) The initial exploration, (2) Model building or Pattern identification with validation/verification, and (3) deployment (i.e., the application of the model to new data in order to generate predictions). [9]

Data mining commonly involves four classes of tasks:

- ✓ **Clustering** - is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.
- ✓ **Classification** - is the task of generalizing known structure to apply to new data. For example, an email program might attempt to classify an email as legitimate or spam. Common algorithms include decision tree learning, nearest neighbor, naive Bayesian classification and neural networks.
- ✓ **Regression** - Attempts to find a function which models the data with the least error.

- ✓ **Association rule learning** - Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes as market basket analysis.

Data mining is the methodology of discovering beforehand obscure patterns and patterns in databases and utilizing the data to fabricate prescient models. In social insurance, data mining is getting to be progressively prominent, if not if not progressively more essential. Medicinal services industry today produces expansive measure of complex Data about patients, medicinal centers assets, sickness analysis, electronic patient records, therapeutic gadgets, and so on. The extensive measure of data is a key asset to be handled and investigated for learning extraction that empowers help for expense funds and choice making. Data mining gives a set of devices and strategies that can be connected to this transformed data to find shrouded examples furthermore gives medicinal services experts an extra wellspring of learning for deciding.

According to the World Health Organization heart disease is the first leading cause of death in high and low income countries and occurs almost equally in men and women [World Health Organization.,2011]. By the year 2030, about 76% of the deaths in the world will be due to non-communicable diseases (NCDs) [World Health Organization.,2005]. Cardiovascular diseases (CVDs), also on the rise, comprise a major portion of non communicable diseases. In the year 2010, of all projected worldwide deaths, 23 million are expected to be because of cardiovascular diseases.

This paper discusses about the data mining algorithms used to predict the heart diseases which was proposed by various authors.

2. Weighted fuzzy rules

P.K. Anooj,2012 manages weighted fuzzy rule-based clinical decision support system (CDSS) for computer-aided diagnosis of the coronary illness. The weighted methodology presented in the proposed work is extra preference for learning of the fuzzy system. It utilized a weighted fuzzy rule to foresee the coronary illness. The experimentation was done on the UCI machine learning repository and the outcomes in danger forecast guaranteed that the proposed clinical decision support system enhanced fundamentally contrasted and the system based system as far as exactness, affectability and specificity. The creator have tried his weighted fuzzy rule on distinctive information set and given the outcome (shown in Table 1).

Table 1. Performance of the CDSS in risk prediction

Data sets	Class	Metric	Proposed System	
			Training	Testing
Cleveland	<50%	Ac	0.509901	0.623529
		Se	0.724771	0.765957
		Sp	0.258065	0.447368
	>50%	Ac	0.509901	0.623529
		Se	0.258065	0.447368
		Sp	0.724771	0.765957
Hungarian	<50%	Ac	0.715045	0.469388
		Se	0.8	0.31746
		Sp	0.540984	0.742857
	>50%	Ac	0.715045	0.469388
		Se	0.540984	0.742857
		Sp	0.8	0.31746
Swizerland	<50%	Ac	0.364706	0.512195
		Se	0.625	0.333333
		Sp	0.337662	0.526316
	>50%	Ac	0.364706	0.512195
		Se	0.337662	0.526316
		Sp	0.625	0.333333

3. Scoring System

Scoring systems have been generally utilized as a part of intensive care units (ICUs) to foresee clinical results and survey the seriousness of disease. Nan Liu et al.,2012 deals with risk score forecast framework with HRV parameters and vital signs, in which geometric separation serves as the key segment. The analysis of scoring framework in patient populace is spoken by a situated of peculiarity vectors, from which hazard scores are determined in light of geometric separation calculation and help vector machine. They utilized forecast model named "scoring framework" to figure a danger score on a quiet's clinical result, using both HRV parameters and vital signs. The scoring framework was based on the calculation of geometric separations among a set of feature vectors got from the records of various patients.

Table 2. Risk Prediction using Feature Extraction

Measure	Proposed	SVM - LIN	SVM - RBF	GLM
Sens	78.80%	73.01%	61.05%	63.05%
Spec	80.80%	80.01%	80.08%	80.08%

Table 3 shows the comparison of their proposed algorithm with SVM with linear kernel (SVM-LIN), SVM with RBF kernel (SVM-RBF) and generalized linear model (GLM).

4. K- Nearest Neighbor and Genetic Algorithm

M.Akhil jabbar et al.,2013 deals with grouping the coronary illness. It consolidates the methodology of KNN and hereditary calculation to enhance the order exactness of coronary illness data set. The creators utilized hereditary hunt as a decency measure to prune repetitive and insignificant ascribes, and to rank the qualities which contribute more towards grouping. Slightest positioned qualities are evacuated, and arrangement calculation was based in light of assessed properties. This classifier was prepared to group coronary illness information set as either healthy or debilitated.

Table 3. Accuracy Comparison

Data Set Name	Accuracy Without GA (Knn only)	Accuracy With Ga (Knn + GA)
Heart Disease A.P	95	100

5. Predictive Risk Assessment of Atherosclerosis (PRAA)

Table 4. Performance Comparison of The PRAA with C4.5, SVM, NB, AND MLP

Dataset	Method	Accuracy %	SE	SP	AUC
DS1	PRAA	98.04	93.75	100	0.94
	C4.5	70.59	6.25	100	1
	NB	52.94	56.25	51.43	0.53
	SVM	35.29	100	5.71	0.53
	MLP	66.67	0	97.14	0.97
DS2	PRAA	99.73	99.35	100	1
	C4.5	50	55.48	45.97	0.52
	NB	61.2	48.39	70.62	0.58
	SVM	45.63	92.26	11.37	0.56
	MLP	57.92	0.65	100	1

V. Sree Hari Rao and M. Naresh Kumar.,2013 deals with PRAA and inherent peculiarities for ascription of MVs that can be connected on datasets wherein the characteristic qualities are either ordinary and/or profoundly skewed having either clear cut and/or numeric characteristics and recognizable proof of danger variables utilizing wrapper-based peculiarity choice. The creators

have assigned the present machine-learning approach as predictive risk assessment of atherosclerosis (PRAA) philosophy all through their study.

The Table 4 shows that the PRAA methodology has outperformed when compared with the classifiers C4.5, SVM, MLP, and NB in terms of sensitivity (SE), specificity (SP), and AUC performance metrics.

6. Genetic-SVM

E. Avci.,2009 deals the wavelet entropy processing in the discrete wavelet change layer of GSVM which can be performed for strong gimmick extraction against to clamor from DHS signals.

An insightful framework in view of genetic-support vector machines (GSVM) methodology was proposed for order of the Doppler signs of the heart valve illnesses. This canny framework manages blend of the peculiarity extraction and order from measured Doppler sign waveforms at the heart valve utilizing the Doppler ultrasound. GSVM is utilized as a part of this study for conclusion of the heart valve infections. The GSVM chooses of most suitable wavelet channel sort for issue, wavelet entropy parameter, the ideal piece capacity sort, bit capacity parameter, and delicate edge consistent C punishment parameter of support vector machines (SVM) classifier. The execution of the GSVM framework is assessed in 215 specimens. The test outcomes demonstrate that this GSVM framework is successful to distinguish Doppler heart sounds. The found the middle value of rate of right grouping rate was around 95%. The creators have analyzed the execution of Genetic-SVM with Turkoglu et al.,2003.

Table 5. Performance Comparison of intelligent system.

The Kernel Function Types		Value of Kernel Function Parameter	Value of C Parameter	The average recognition (%)	
				Normal	Abnormal
GSVM Model -1	RBF	36.8	0.1	90	93.15
GSVM Model -2	RBF	18.3	0.1	88	89.04
GSVM Model -3	RBF	24.5	0.1	92	94.52
GSVM Model -4	ERBF	2.2	0.1	96	94.52
GSVM Model -5	ERBF	2.2	10,000	94	94.52
WNN used in Turkoglu et al., (2003)				95.9	84

7. Least Square Support Vector Machines

Davut Hanbay.,2009 deals the elucidation of the DHS signs utilizing pattern recognition..

A specialist system in view of least squares support vector machines (LS-SVM) for determination of valvular heart disease (VHD) was introduced. Wavelet packet decomposition (WPD) and fast-Fourier transform (FFT) systems are utilized for peculiarity removing from Doppler signs. LS-SVM is utilized as a part of the classification stage. Triple cross-approval technique is utilized to assess the master system execution. The exhibitions of the created systems were assessed in 105 specimens that contain 39 typical and 66 unusual subjects for mitral valve disease. The outcomes demonstrated that this system is powerful to distinguish Doppler heart sounds. The average correct classification rate was around 96.13% for ordinary subjects and anomalous subjects.

Table 6. Threefold test performance of LS-SVM model

Data set (105)		Correct Classified	Wrong Classified	Performance (%)
Training Sets	Test Sets			
Set-1, set-2 (70)	Set-3 (35)	35	0	100
Set-1, set-2 (70)	Set-2 (35)	33	2	94.2
Set-1, set-2 (70)	Set-1(35)	33	2	94.2
Average Performance				96.13

The target of the classification is to show the viability of the peculiarity extraction technique from the DHS signals. For this reason, the gimmick vectors were connected as the info to a LS-SVM classifier. The KULeuven's LS-SVMlab MATLAB/C

Tool compartment was utilized with the end goal of preparing and testing. RBF bit is utilized. Network look calculation is utilized to tune the c regularization steady and r width of RBF part parameters. The decided ideal c worth is 2.8439 and ideal r quality is 29.316 for predicting mitral valve diseases. Triple cross-acceptance technique was connected to the 105 trial information sets for registering the approval of LS-SVM model. In k-fold cross-approval technique, the information set is isolated into k subsets, and the holdout strategy is rehashed k times. At every time, k - 1 subsets are utilized for preparing and kth subset is utilized for testing. At that point the average slip over all k trials is figured. Consequently, every information point becomes in a test set precisely once, and becomes in a preparation set k-1 times. Diverse assessment routines were utilized for figuring the execution of the master system. The best test execution of LS-SVM model is demonstrated in the underneath table.

LS-SVM model predicts the deliberate qualities at a high precision rate. Triple test execution of LS-SVM model is indicated in Table 6. The average correct classification rate is 96.13%.

8. Modified binary particle swarm optimization (MBPSO)

Susana M. Vieiraa et al.,2013 manages Modified binary particle swarm optimization (MBPSO) strategy for peculiarity choice with the concurrent optimization of SVM part parameter setting, connected to mortality forecast in septic patients.

An improved rendition of binary particle swarm optimization, intended to adapt to untimely union of the BPSO calculation. MBPSO control the swarm variability utilizing the speed and the comparability between best swarm arrangements. It uses support vector machines in a wrapper approach, where the portion parameters are improved in the meantime. The methodology is connected to foresee the result (survived or perished) of patients with septic stun. Further, MBPSO is tried in a few benchmark datasets and is contrasted and other PSO based algorithms and genetic algorithms (GA). Their trial results demonstrate that this methodology can accurately select the segregating information peculiarities furthermore accomplish high order precision, particularly when contrasted with other PSO based algorithms. At the point when contrasted with GA, MBPSO is comparative as far as precision; however the subset arrangements have less chosen characteristics.

The modified binary PSO includes the SVM model parameters in the encoding of the particles and is described in the following.

Table 7. Performance Comparison of MBPSO

Method	Accuracy	NF	Sensitivity	Specificity
No-FS	89.0 ± 1.7	28	76.1 ± 5.4	95.6 ± 1.7
BPSO	94.0 ± 1.5	6 ± 1	89.5 ± 5.6	96.1 ± 2.0
IBPSO	94.2 ± 1.1	6 ± 1	90.4 ± 5.3	95.9 ± 1.8
GA	95.7 ± 1.4	7 ± 1	94.3 ± 1.2	96.5 ± 2.1
MBPSO	94.4 ± 1.2	6 ± 1	90.2 ± 5.1	96.5 ±1.9

The authors have compared their MBPSO with existing techniques and showed that MBPSO outperforms in feature selection used in the prediction of heart diseases.

Conclusion:

Data mining is very beneficial in the field of medical domain. This comparative study paper thoroughly analyzed the various algorithms made towards disease prediction. Several data mining and soft computing approaches are studied. In this paper, we have presented effective heart disease prediction strategies by utilizing data mining methods. In most papers authors had chosen the metrics as accuracy, sensitivity and specificity to check the performance of their proposed algorithms. This study concludes that the future scope of research can be done in risk assessment among diabetic patients those who are developing heart diseases.

References:

1. Davut Hanbay, "An expert system based on least square support vector machines for diagnosis of the valvular heart disease", Expert Systems with Applications 36 (2009) 4232–4238
2. E. Avci, “A new intelligent diagnosis system for the heart valve diseases by using genetic-SVM classifier”, Expert Systems with Applications 36 (2009) 10618–10626
3. M.Akhil jabbar, B.L Deekshatulua Priti Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm", International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013, India, Elsevier, Procedia Technology 10 (2013) 85 – 94
4. Nan Liu, Zhiping Lin, Jiuwen Cao, Zhixiong Koh, Tongtong Zhang, "An Intelligent Scoring System and Its Application to Cardiac Arrest Prediction", IEEE Transactions On Information Technology In Biomedicine, Vol. 16, No. 6, November 2012

5. P.K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University – Computer and Information Sciences (2012), Elsevier, 24, 27–40.
6. Susana M. Vieiraa, Luís F. Mendonc, Gonc¸ J. Farinhaa, Joao M.C. Sousaa, "Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients", Applied Soft Computing 08/2013; 13(8):3494–3504.
7. Turkoglu, I., Arslan, A., & ve Ilkay, E. (2003). A wavelet neural network for the detection of the heart valve diseases. Expert Systems: The International Journal of Knowledge Engineering and Neural Networks, 20(1), 1–7.
8. V. Sree Hari Rao and M. Naresh Kumar, "Novel Approaches for Predicting Risk Factors of Atherosclerosis", IEEE Journal of Biomedical And Health Informatics, Vol. 17, No. 1, January 2013.
9. Nirkhi, S., "Potential use of Artificial Neural Network in Data Mining," Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on , vol.2, no., pp.339,343, 26-28 Feb. 2010.
10. Krissna Priya.R., "A Improved Classification of Network Traffic using Adaptive Nearest cluster Based Classifier", International Journal of Computer Trends and Technology ISSN:2231-2803,Vol.18, Issue No.1 , January 2015.

