JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AN EFFICIENT DATA MINING ALGORITHM FOR CARDIAC ARRHYTHMIA DETECTION USING ECG DATA

[1]D.DEVIPRIYA, [2]Dr.V.S MEENAKSHI

[1]M.PHIL RESEARCH SCHOLAR, [2]ASSISTANT PROFESSOR

[1]PG & RESEARCH DEPARTMENT OF COMPUTER SCIENCE,

[1]CHIKKANNA GOVERNMENT ARTS COLLEGE,

TIRUPUR,

TAMILNADU,

INDIA

ABSTRACT

An automated Artificial Neural Network (ANN) based classification system for cardiac arrhythmia using multi-channel ECG recordings has been proposed. In this study, it basically keen on delivering high certain arrhythmia arrangement results to be material in analytic choice decision support systems. Neural network model with back propagation calculation is utilized to group arrhythmia cases into typical and strange classes. Networks models are trained and tested for MIT-BIH arrhythmia. The various designs of ANN have been prepared by combination of arrhythmic and non arrhythmic information patient. The classification performance is evaluated using measures; Sensitivity, Specificity, Classification accuracy, Mean Squared Error (MSE), Receiver Operating Characteristics (ROC) and Area Under Curve (AUC).

Keywords: Feature Extraction, ECG arrhythmia, Arrhythmia Classification, Artificial Neural Networks.

1. INTRODUCTION

In today's world death rate is very high because of heart disease is a significant danger to people. Arrhythmia is such a kind of coronary illness which is expanding quickly and significant passing in heart patients are because of arrhythmia. The time stretch between progressive heart beats is called as mood and when the musicality is not ordinary it is called as arrhythmia. Side effects of arrhythmia are quick or slow heart beat, skipping thumps, chest torment and so on that should be visible in the beginning phase. Information mining is the most common way of dissecting information according to alternate points of view and summing up it into helpful data [1]. It permits clients to investigate information from various aspects or points, classify it, and sum up the connections distinguished. Actually, it is the method involved with tracking down connections or examples among huge number of gathered information. It is the method which gains from conduct of information.

Information mining is assuming essential part in numerous applications particularly in medical care. The information produced in medical services framework is tremendous and complex, so it turns out to be undeniably challenging to deal with it [2]. This paper presents arrhythmia and predicts its sort progressively. As arrhythmia is caused because of unusual pulse it tends to be handily determined to have the assistance of ECG signals. This framework at first spotlights on pre-handling and element extraction. Then information mining calculation is utilized to gain from the gathered information of ECG and structures the principles which help to anticipate the sort of arrhythmia.

II. RELATED WORK

There are several methods for automatic detection and classification of cardiac arrhythmia like Artificial Immune Recognition system, Neural Network, Fuzzy Neural Network, Data Mining etc. Artificial Immune Recognition framework is a method which utilizes a notable brain network engineering named Multi-Layered Perception (MLP) with back engendering preparing calculation, and another Fluffy Clustering NN design (FCNN) for early finding [3]. A fluffy brain network was carried out utilizing a multithreading approach for discovery of atrial fibrillation, bigeminy, and ordinary sinus beat in the MIT-BIH Arrhythmia Database. The information to the brain network comprised of nine data sources. Seven bordering RR stretches, their normal and their standard deviation. Wavelet-based calculation is utilized for arrhythmia separation which examinations the Electrocardiograph (ECG) signal by utilizing the consistent wavelet change and its standard in various sizes of variety and it can naturally recognize arrhythmia.

This method based on Fuzzy Neural Network (FNN) is developed to create fuzzy membership functions for classification of cardiac arrhythmia [4]. The proposed system is based on Data Stream Mining techniques which consist of different algorithms for classifying arrhythmia into seven types. They are namely Normal Beat (NB), Left Bundle Branch Block Beat (LBBB), Right Bundle Branch Block Beat (RBBB), Premature Ventricular Contraction (PVC), Fusion of Ventricular and normal beat (FUSION), Atrial Premature Contraction (APC) and Paced Beat (PACE) [4]. Earlier there was K-mean algorithm used for stream data clustering which use single phase model but this algorithm was not able to detect the priority of data and it gave same priority to recent and historic data.

Likewise, K-mean calculation was unequipped for recognizing erratic states of group and commotion in signals. To conquer these downsides Aggarwal et al. proposed 2-gradually work plot which processes the crude information in miniature groups in web-based mode and gives outline which is utilized by disconnected parts. This plan prompts Clu-Stream framework. Another essential grouping calculation called Density based bunching calculation which is more broad as it can recognize erratic shapes bunches, handle clamor and sweep information just a single time so it very well may be applied to enormous volume of information stream. In this paper, different time space highlight are removed from ECG and HRV flags then got information is grouped with assistance of Data Mining calculation called D-stream calculation [5].

III. SIGNIFICANCE OF THE SYSTEM

The principle worry in the proposed framework is to recognize the arrhythmia and its sort. However there are different kinds of illnesses, the overall outcomes [13] displayed in figure 1 have showed that more demise rates are caused because of the cardiovascular sicknesses and in that arrhythmia results are markable.

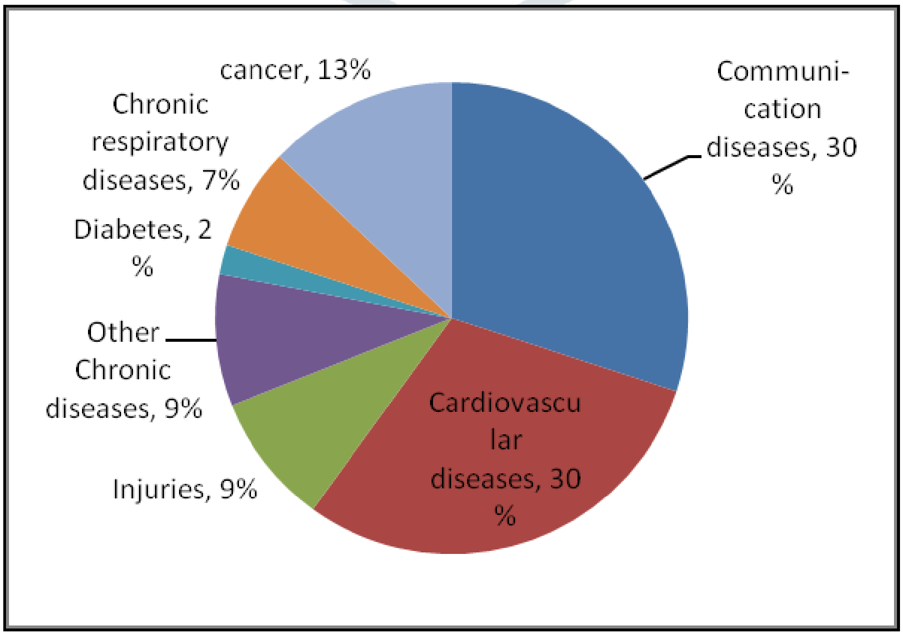

Figure 1: Principle causes of death worldwide

IV. METHODOLOGY

In this study, the back-propagation learning algorithm is used since it is the most popular supervised learning algorithm [6].

4.1 DATA PREPROCESSING

Unique ECG contains unpredictable distance between tops, sporadic pinnacle structure, and presence of low-recurrence part in ECG because of patient breathing and so on. To settle the errand the handling pipeline ought to contain specific stages to decrease impact of those elements.

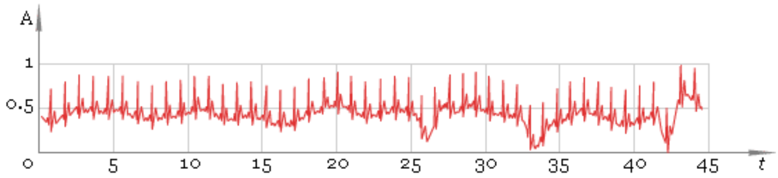

Figure 2: Raw ECG data

(This Figure contains raw ECG data, which is unfiltered and contains clamor which is expected to be taken out before additional tasks) The information eliminated low-recurrence part [7]. It has applied direct Fast Fourier Transformation eliminate low frequencies and re-establish ECG with the assistance of backwards FFT.

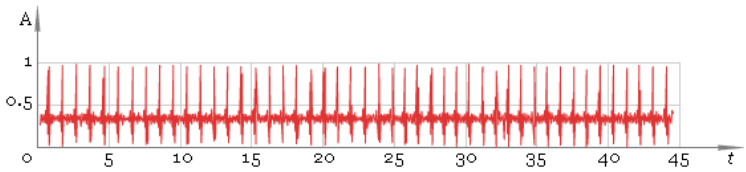

Figure 3: FFT filtered ECG.

(Information which was gotten subsequent to applying Fast Fourier Transformation, which eliminates low frequencies parts of information, then low recurrence parts were taken out and reverse FFT was applied to get back fixed ECG information)

After benchmark noise decrease, division of ECG beat was finished. In this step, the ceaseless ECG signals were changed into individual ECG beats. The width of individual beat was approximated to 300 example information and the extricated beat is based on R top [8] [13]. For this reason, use the explanation given by the information base to do the change. This utilizes the R top explanation as the turn point for each beat. For every R-top, cut off the constant sign for each beat start at R-150 pos until R+149 pos, consequently it will get a beat with 300 example information in width.

4.2 FEATURE EXTRACTION

To deal with the Multi channel information, summation of information from the two channels was finished to set up an information vector as the contribution to Artificial Neural Network. Hence basically diminishing the possibilities of False Positive cases, when effort might make a strange ECG signal in a specific channel [9]. Accordingly, adding information from the two channels limits the possibilities of inaccurate distinguishing proof of Arrhythmia. In this paper one ECG beat relates to one example of 300 data sources, which covers the entire ECG beat. The contributions for the organizations were chosen considering two significant focuses [10]:

a) The inputs must be of a standard size such that it is neither too small to cover up one ECG cycle and nor two high to increase the number of beats required to analyse the signal, thus increasing the hardware requirements.

b) The input must be so arranged that the R peak in the QRS complex must be at the centre of the signal cycle under considerations.

The first condition was achieved by setting up an arbitrary value of 300 samples of MLII lead data obtained from the database in which the 150 samples were on the left side and 149 samples on the right side of the 151 sample value, which in turn is the detected R peak. Now, Similar was done for other lead, and generated samples were added to create an input of 300 samples.

Thus the input becomes a matrix of 300 x no of samples and ready to be used in Python. The same process was repeated to make all the inputs of all the kinds of beats that are normal, fusion and ventricular premature [15]. The second condition was achieved by allowing the 150st sample to be the best value of both channels obtained from the database for particular conditions. E.g. If a number (2250- 150=) 2100 to sample number (2251+150=) 2401 will be the input data.

4.3 TRANING AND TESTING

In this work is dedicated to thought of various neural networks to decide their exactness in ID and detachment of classifications or classes. Among all neural networks Feed Forward back propagation has been chosen based on the below mentioned reasons [11].

1. It has 2 hidden layers including input layer and output layer shown in fig 4.
2. As result of this fact that the numbers of existed neurons in hidden layer is an effective parameter for improvement of learning results, neuron numbers was chosen in order to achieve the optimum number based on output results. So, 1 hidden layer has 3 neurons and 5 neurons in second layer.
3. Tansig and Purelin function and also their combination function have been compared as transfer function of network neurons and finally, the effective one has been chosen [13].
4. For training utilized of BP algorithm and trained function.
5. Lr parameter 1 has been chosen.
6. For testing of mentioned neural network, Mean Squared Error (MSE) or goal parameter criterion was utilized in which error of 0.0001 was the stopping point of testing and maximum repetitions was 1000 times [8].

4.3.1 Training of ANN

Structures of ANN were trained using abnormal and normal patients. If the value of node output of output layer was logic- 1, that interpreted this as arrhythmia [14]. If the value was logic-0, this was considered as normal. If $y(i) \geq 0.5$, accept as logic-1 and it used $h(i)=|1-y(i)|$ in the error calculation.

If $y(i)<0.5$, it was considered as logic-0 and we used $h(i) = | 0-y(i)|$. Trained ANN architectures is given in Fig 4. Normal sinus arrhythmia and abnormal were mixed in sequence. The length of the training pattern was 21200 samples (106 sets). The test pattern was done similarly [12]. Both patterns, used in training of ANN, and used in testing trained ANN were occurred from MIT-BIH ECG database and normal sinus rhythm database obtained from MIT-BIT NSRDB database Learning rate (ε) was 0.01 and momentum coefficient (α) was 0.2. Whereas training error was found 0.1% after 1000 iteration, test error became 3.79% [16].

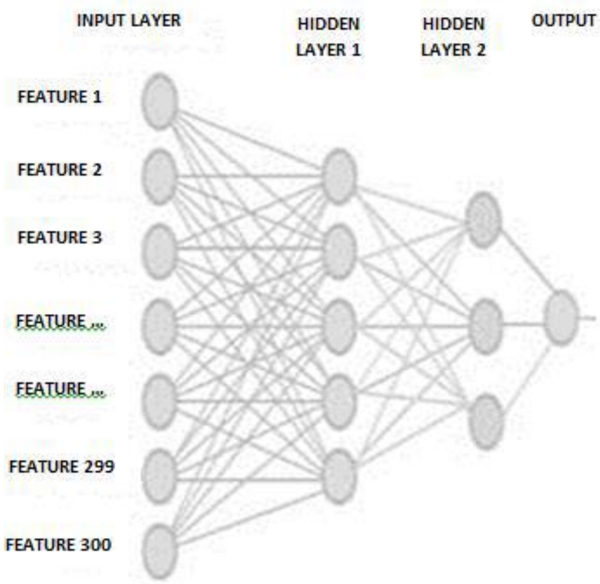

Figure 4: Architecture of the Artificial Neural Network

(Visual representation of 300:5:3:1 neural network architecture)

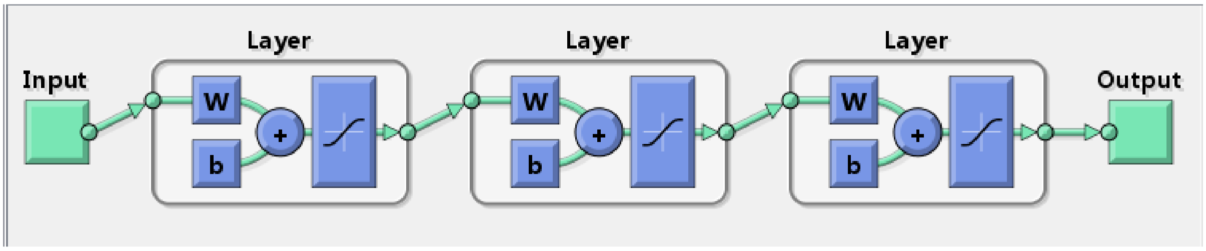

Figure 5: Model of Neural Network

4.3.2 Testing

Testing phase use mixture of MIT –BIT Arrhythmia, QT database and NSR database. They are randomly mixed. Simulating the resultant network with the test dataset, the results are summarized in Table 1. A detection accuracy of 96.21% was obtained.

V. EXPERIMENTS AND RESULTS

5.1 PERFORMANCE MEASURES

The evaluation of the performance of the classification algorithms using six measures; Sensitivity, Specificity, Classification accuracy, Mean Squared Error (MSE), Receiver Operating Characteristics (ROC) [9,10]. These measures are defined using True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP decision occurs when an arrhythmia detection of the classifier coincided with a decision of the physician. TN decision occurs when both the classifier and the physician suggested the absence of arrhythmia. FP occurs when the system labels a healthy case as an arrhythmia one. Finally, FN occurs when the system labels an arrhythmia case as healthy (Figure 6 and Figure 7).

1) Classification Accuracy: Classification accuracy is defined as the ratio of the number of correctly classified cases and is equal to the sum of TP and TN divided by the total number of cases N.

Accuracy = (TP + TN)/N [4]

2) Classification Sensitivity: Sensitivity refers to the rate of correctly classified positive and is equal to TP divided by the sum of TP and FN. Sensitivity may be referred as a True Positive Rate.

Sensitivity = TP/(TP + FN) [5]

3) Classification Specificity: Specificity refers to the rate of correctly classified negative and is equal to the ratio of TN to the sum of TN and FP. False Positive Rate equals (100 − Specificity).

Specificity = TN/(FP + TN) [6]

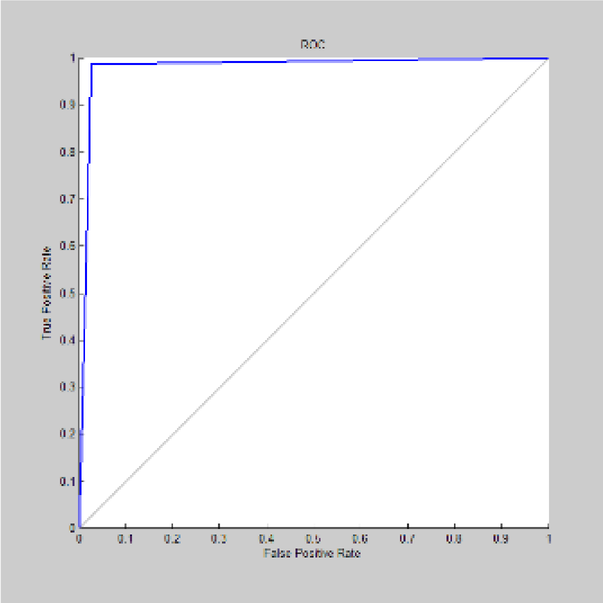

Figure: 6 ROC Plot after Training stage (MIT- BIH data)

(Reverse operating characteristic plot which is graphical relationship diagram between tue positive rate vs. False positive rate, showing results of training phase)

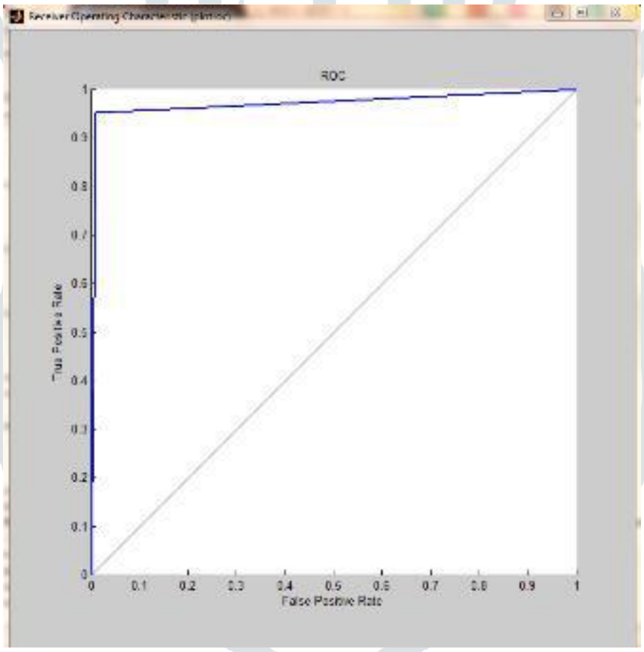

Figure: 7 ROC plot Mixed -data

(Reverse operating characteristic plot which is graphical relationship diagram between true positive rate VS false positive rate, showing results of training phase)

TABLE I -- RESULT OF VARIOUS DATABASES

Cardiac signal Database	Data set Testing	Correctly Classified	Sets Misqualified	% Accuracy
MIT BIH	248	240	8	96.77
NSRDB	27	25	2	94
Mixed	275	265	10	96.21

TABLE II – COMPARISON FROM PREVIOUS RESEARCH WORK

Accuracy	Accuracy from [4]	Accuracy from [8]	Accuracy from [14]	Accuracy from [15]
96.21	92.5	95.7	90.56	95.00

VI. CONCLUSION AND FUTURE SCOPE

This paper proposes a successful robotized ANN based framework for multi class Cardiac Arrhythmia grouping from ECG signal information. Each ANN has been tried and contrasted and the most well-known conventional ECG analyzers on proper data sets. Accordingly, in light of the outcomes, the ANN's methodology is demonstrated to be equipped for managing the vague idea of the ECG signal.

The vital job of information pre-handling and post handling emerges, either for decreasing the info space aspect or for all the more suitably portraying the info highlights. Obviously the assessed feed forward ANN with error back propagation algorithm work as a great classifier for given heart arrhythmia informational index. Accordingly our future degree will be further involving grouping calculations for tracking down much better exactness for arrhythmia.

REFERENCES

[1] Yixin Chen, Li Tu", Density-Based Clustering for Real- Time Stream Data",in KDD'07, August 12–15, 2007, San Jose, California, USA, ACM 978-1-59593-609-7/07/0008

[2] B. Anuradha, K. Suresh Kumar and V. C. Veera Reddy " Classification Of Cardiac Signals Using Time Domain Method" in ARPN Journal of Engineering and Applied Sciences, VOL. 3, NO. 3, JUNE 2008 ,ISSN 1819-6608

[3] Manish Verma, Mauly Srivastava, Neha Chack, Atul Kumar Diswar, Nidhi Gupta" A Comparative Study of Various Clustering Algorithms in Data Mining". In GLNA Institute of Technology, Mathura , Vol. 2, Issue 3, May-Jun 2012, pp.1379-1384

[4] Y. S. Tsai, B. N. Hung, and S. F. Tung, ―An experiment on ECG classification using back-propagation neural networks,‖ In Pederson PC, Onaral B(eds): Proceedings of the Annual International Conference of the IEEE Eng. of Med. and Biology Society, 12, pp. 1463, 1990.

[5] D. E Rumelhart, G. E Hinton, and Williams R. J. Hinton, ―Learning internal representations by error propagation. In Parallel distributed

Processing: explorations in the microstructure of cognition,‖ MIT Press Camhridge, Massachusetts, J. L., 1, pp. 3 18462,1986.

[6] Y. Ozhay, ―Fast Recognition of ECG Arrhythmias,‖ Ph.D. Paper, University of Selcuk, Konya, Turkey, 1999.

[7] Bortolan, G. Jekova, I. Christov, ―Comparison of four methods for premature ventricular contraction and normal beat clustering,‖

Computers in Cardiology, September 2005 pp. 921 – 924

[8] Dayong Gao, Michael Madden, ―Bayesian ANN Classifier for ECG Arrhythmias Diagnostic System,‖ Proceeding of IEEE International Joint Conference on Neural Network 2003 pp. 2454-2459.

[9] Douglas A. Coast, Richard M. Sten, Stanley A. Briller, "An Approach to Cardiac Arrhythmia Analysis Using

Hidden Markov Models"in IEEE Transactions On Biomedical Engineering. Vol. 37. No. 9. September 1990, Pp- 001 8-9294/90/0900-0826

[10] V.Ilankumaran ,Dr.S.ThamaraiSelvi," Ventricular Arrhythmias Detection using Wavelet Decomposition",in International Journal of Computer Applications (0975 – 8887)Volume 20– No.1, April 2011.

[11] Miad Faezipour, Adnan Saeed, Suma Chandrika Bulusu,Mehrdad Nourani,, Hlaing Minn," A Patient- Adaptive Profiling Scheme for ECG Beat Classification"in IEEE Transactions On Information Technology In

Biomedicine, Vol. 14, No. 5, September 2010,PP- 1089- 7771.

[12] Mohd Afzan Othman, Norlaili Mat Safri, Rubita Sudirman "Characterization of Ventricular Arrhythmias in Electrocardiogram Signal using Semantic Mining Algorithm"in 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation,PP- 978-0-7695-4062-7/10

[13] Jimena Rodríguez, Alfredo Goñi, and Arantza Illarramendi," Real-Time Classification of ECGs on a PDA" in IEEE Transactions On Information Technology In Biomedicine, Vol. 9, No. 1, March 2005, Pp-1089-7771

[14] B. Anuradha and V. C. Veera Reddy , ―ANN FOR CLASSIFICATION OF CARDIAC ARRHYTHMIAS‖ in ARPN Journal of Engineering and Applied Sciences, vol. 3, no. 3, june 2008.

[15] Fortis Escorts Hospital, Jaipur 302017 India, "Burden of Cardiovascular Diseases in India".

[16] Claude Seidman. "Data Mining with Microsoft SQL Server 2000 Technical Reference", ISBN: 0-7356-1271 4,amazon.com/Mining-Microsoft-Server-Technical Reference / dp / 0735612714.