Electrocardiogram Beat Classification using Probabilistic Neural Network

Banupriya C.V
M.phil Research scholar
PSGR Krishnammal College for Women
Coimbatore, India

Karpagavalli.S
Associate Professor
GR Govindarajalu School of Applied
Computer Technology, Coimbatore, India

ABSTRACT
The Electrocardiogram (ECG) plays significant role in assessing patients with abnormal activity in their heart. ECG recordings of the patient taken to analyze abnormality and classify type of disorder present in the heart functionality. An Electrocardiogram is a bioelectrical signal that records the heart's electrical activity versus time. It is used to measure the rate and regularity of heartbeats, as well as the size and location of the chambers, the occurrence of any damage to the heart, and the effect of drugs or devices used to regulate the heart. An electrocardiogram recording of a patient is important clinical information for the medical experts to diagnose the heart functionality of the patient or to assess the patient before any surgery. The interpretation of ECG signal is an application of pattern recognition. There are several classes of heart disorders including Premature Ventricular Contraction (PVC), Atrial Premature beat (APB), Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), Paced Beat (PB), and Atrial Escape Beat (AEB). To analyze ECG various feature extraction methods and classification algorithms are used. The planned work employed discrete wavelet transform (DWT) in feature extraction on ECG signals obtained from MIT-BIH Arrhythmia Database. The Machine Learning Technique, Probabilistic Neural Network (PNN) has been used to classify four types of heart beats that consist of PVC, LBBB, RBBB and Normal.

Keywords
Electrocardiogram, Wavelet, Probabilistic Neural Network, Premature Ventricular Contraction, Left Bundle Branch Block, Right Bundle Branch Block.

1. INTRODUCTION
Electrocardiography represents the electrical activity of the heart. Electrocardiogram is a record of the origin and propagation of electrical potential through cardiac muscles obtained using sensors at limb extremities of the human. It provides useful and important information to cardiologists about the rhythm and functioning of the heart so as to perform diagnostic analysis with maximum accuracy. The classification of the ECG beats is an important task in the coronary care unit, and essential tool for diagnosis of heart diseases. The electrical signals are measured by the electrocardiogram where each heart beat is displayed as a series of electrical waves characterized by peaks and valleys. Therefore it is important to record the ECG signal since it allows for the detection of abnormalities or diseases in the heart. The recording of the ECG usually takes place by placing electrodes to measure potentials on the surface of specific body parts usually the arms, legs and the chest. There are three basic techniques used in clinical electrocardiography. The most common is the standard clinical electrocardiogram that utilizes 12 different potential differences called ECG leads [1]. The standard 12-lead ECG is based on 3 limb leads (I, II, III), 3 augmented leads (limb potential relative to a modified Wilson terminal, aVR, aVL, aVF) and 6 leads placed across the front of the chest and referenced to the Wilson terminal (Lead V1, V2, V3, V4, V5, V6). The typical structure of the ECG is shown in Figure1.

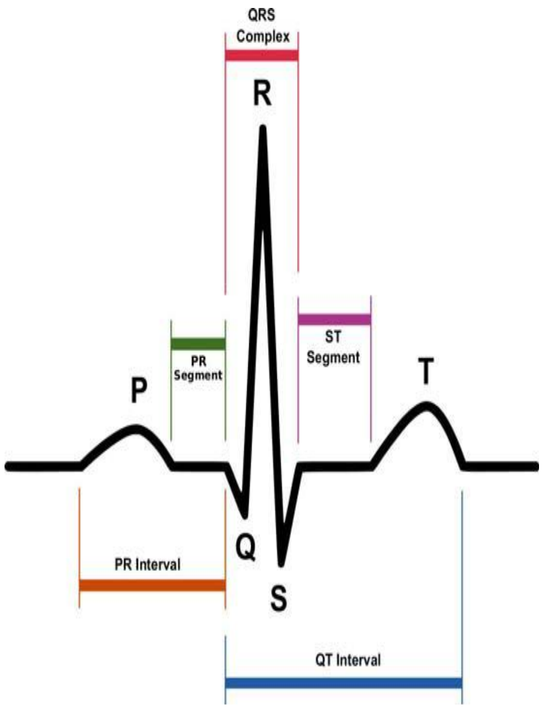


Fig 1: Structure of the ECG signal

The P wave, the QRS complex, and the T wave are three characteristic features of the waveform which are easily identified. These waves are associated with the activation of the atria, activation of the ventricles, and re-polarization of the ventricles. A patient may have different ECG waveforms and in a single ECG different types of beats may present that are unlike each other [2]. The beats may be normal beat, Premature Ventricular Contraction (PVC) beat, Left Bundle Branch Block beat (LBBB), Right Bundle Branch Block beat (RBBB), Atrial Premature beat (APB) and Paced beat (PB).

Researchers have developed different techniques and algorithms for automated processing of ECG signals for various medical applications. Some of them used time domain and some use frequency domain for depiction.

F.de Chazal et.al, [3] investigates the design of an efficient system for recognition of the premature ventricular contraction from the normal beats and other heart diseases. This classification comprises three main modules namely denoising part, feature extraction part and classifier part. In the denoising module of the classification, it proposed the stationary wavelet transform for noise reduction of the electrocardiogram signals. In the feature extraction of the ECG module a proper combination of the morphological-based features and timing interval-based features are extracted. In their work, many supervised classifiers are investigated, that include: a number of Multi-Layer Perceptron neural networks with different number of layers and research algorithms, support vector machines with different kernel types, radial basis function (RBF) and probabilistic neural networks. They used 12 files from the MIT–BIH arrhythmia database and achieved about 97.14% for classification of ECG beats.

Thaweesak, et al. performed the classification of ECG using SVM to classify the 3 classes, premature ventricular contraction (PVC), Normal and Atrial Premature Contraction heart diseases [4]. In their work the ECG signal is obtained from MIT-BIH Arrhythmia database. They used 6 files from the arrhythmia database which contain enough beats of normal, PVC and APC arrhythmia and they used only the MLII lead as source data for the experiment and achieved more accuracy.

Wisnu Jatmiko, et al. employed Back-Propagation Neural Networks and Fuzzy Neuro Learning Vector Quantization (FLVQ) as classifier in ECG classification [5]. In their work, they used only the MLII lead as source data. The classes that are considered are Left Bundle Branch Block beat (LBBB), Normal beat (NORMAL), Right Bundle Branch Block beat (RBBB), Premature Ventricular Contraction (PVC). They used training classification methods namely Back-propagation and FLVQ for their experiment. It produces an average accuracy 99.20% using Back- Propagation and 95.50% for FLVQ. The result shows that back-propagation leading than FLVQ but, back-propagation has disadvantages to classified unknown category beat but not for FLVQ. FLVQ has stable accuracy although contain unknown category beat.

Maedeh Kiani Sarkaleh, et al. [6], proposed a Neural Network (NN) based algorithm for classification of Paced Beat (PB), Atrial Premature Beat (APB) arrhythmias as well as the normal beat signal. They applied Discrete Wavelet Transform (DWT) for feature extraction and used it along with timing interval features to train the Neural Network. About 10 recordings of the MIT-BIH arrhythmias database have been used for training and testing the neural network based classifier. The model results show that the classification accuracy is 96.5%.

Karpagachelvi.S, et al. [7], a novel ECG beat classification system using RVM is proposed and applied to MIT/BIH arrhythmia database to classify five kinds of abnormal waveforms and normal beats. In exacting, the sensitivity of the RVM classifier is tested and that is compared with ELM. The obtained results clearly confirm the superiority of the RVM approach when compared to traditional classifiers.

In proposed work, efficient models are built using Probabilistic Neural Network to classify the four different heart beats, Normal, Premature Ventricular Contraction (PVC), Left Bundle Branch Block (LBBB) and Right Bundle Branch Block (RBBB) in ECG. The ECG signals are obtained from MIT-BIH Arrhythmia Database. This work involves the various tasks that include preprocessing, feature extraction, building models through training and testing the models.

2. FEATURE EXTRACTION

2.1 Preprocessing

Preprocessing of ECG signals need to be performed for effective feature extraction. In preprocessing [8] signal extension, cutting the normal and abnormal beats, de-noising and decomposition operations are performed. The number of samples in ECG record has been extended to 50,000 samples. In denoising, the signals are decomposed using Daubechies Wavelet with decomposition level 4 (db4).

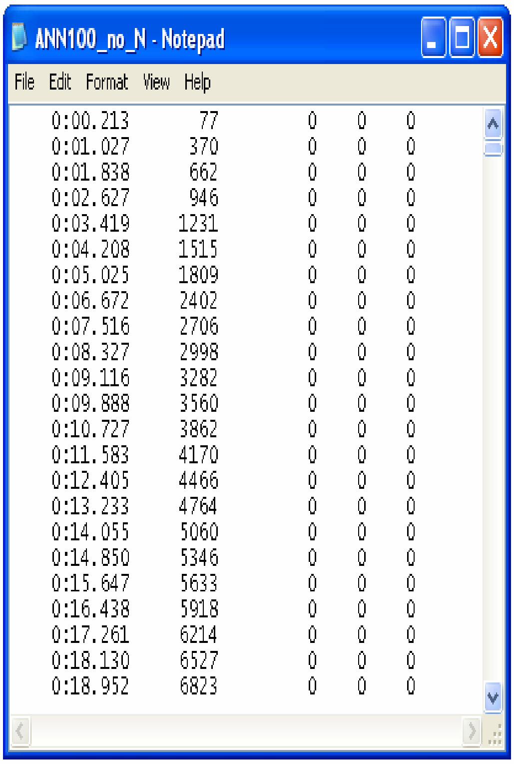
2.2 Cutting the Normal & Abnormal Beats

In order to cut the normal and abnormal beats, code is implemented in MATLAB that takes ECG signal file in .mat format and annotation file as input. The .mat file contains the signal with the sample numbers and their amplitudes.

For this specific run of the code, record 100 was used. The signals' samples are located in ys_100.mat which is present in the MATLAB workspace, and ANN100_no_N.txt is the annotation file that includes the locations of the normal beats. The annotation file is shown in figure 2.The code examines the second column of the annotation file which corresponds to the sample number of the normal beat. Since our signals were cut up-to 50,000 beats only, the code checks for annotation files with sample numbers greater than 50,000 in order to ignore the samples.

The code then creates a matrix with number of rows equals the number of normal beats. Afterwards, the code uses the matrix containing all the samples of the record and will save the 40 samples surrounding the normal beat in addition to the sample of the beat itself into its final matrix. Figure 3 shows the first cut normal beat of record 100 which corresponds to the plot of the first row of the final matrix obtained from record 100.

Similarly, same procedure followed to cut the normal and abnormal beats of all signals with the same code using the PVC, LBBB and RBBB annotated files.


ANN100_no_N - Notepad

File Edit Format View Help

0:00.213	77	0	0	0
0:01.027	370	0	0	0
0:01.838	662	0	0	0
0:02.627	946	0	0	0
0:03.419	1231	0	0	0
0:04.208	1515	0	0	0
0:05.025	1809	0	0	0
0:06.672	2402	0	0	0
0:07.516	2706	0	0	0
0:08.327	2998	0	0	0
0:09.116	3282	0	0	0
0:09.888	3560	0	0	0
0:10.727	3862	0	0	0
0:11.583	4170	0	0	0
0:12.405	4466	0	0	0
0:13.233	4764	0	0	0
0:14.055	5060	0	0	0
0:14.850	5346	0	0	0
0:15.647	5633	0	0	0
0:16.438	5918	0	0	0
0:17.261	6214	0	0	0
0:18.130	6527	0	0	0
0:18.952	6823	0	0	0

Fig 2: Input annotated text file

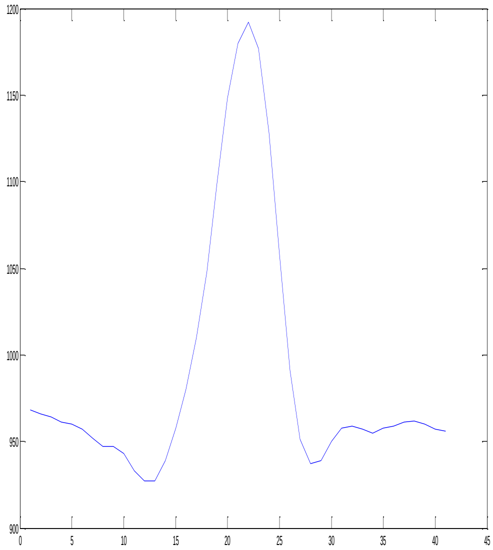


Fig 3: Normal Beat

Figure 4 shows the first cut Premature Ventricular Contraction beat of record 106 which corresponds to the plot of the first row of the final matrix from record 106. Figure 5 shows the first cut Left Bundle Branch Block beat of record 109 which corresponds to the plot of the first row of the final matrix from record 109. Figure 6 shows the first cut Right Bundle Branch Block beat of record 118 which corresponds to the plot of the first row of the final matrix from record 118.

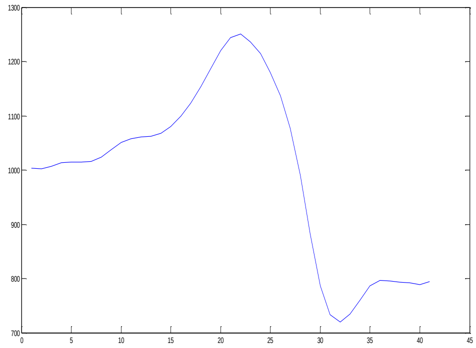


Fig 4: Premature Ventricular Contraction

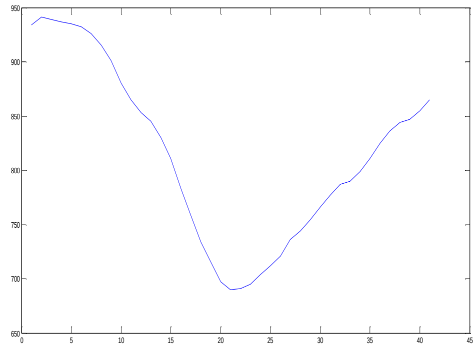


Fig 5: Left Bundle Branch Block

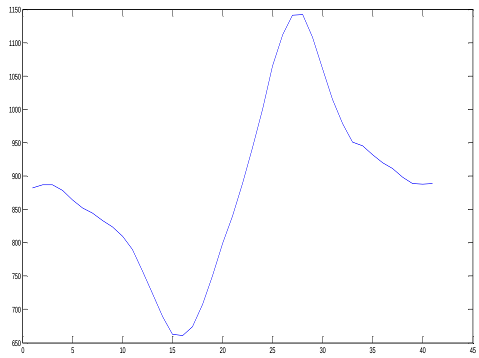


Fig 6: Right Bundle Branch Block

After cutting the ECG beats decomposition and reconstruction of the signal using wavelet transform is performed [9]. Wavelet analysis is performed using daubechies wavelet. Approximation coefficients are computed at level N using wavelet decomposition structure. Wavelet decomposition can be regarded as projection of the signal on the set of wavelet basis vectors. All wavelet coefficients can be computed as the dot product of the signal with the corresponding basis vector. Then, reconstruction is performed [10]. The approximation and detail coefficients at every level are up sampled by two passed through the both synthesis filters and then added [11]. After performing baseline wandering and noise suppression of that signal, features are extracted for every 41 samples of normal and abnormal beats.

2.3 Feature Extraction

In this proposed work suitable combination of morphological features and temporal features has been extracted from the ECG beats. Actual feature extraction carried out in two phases. The first phase involves the cutting of the Normal, PVC, LBBB and RBBB beats by making use of the annotation files which exist in MIT-BIH arrhythmia database. The annotation file provides beat type and sample number (R), from that R sample R+20 and R-20 a total of 41 samples is cut from the continuous ECG signal for each type of beat. The second phase involves identification of the peaks and locations in 41 samples of each type of beat. The morphological features that describe the basic shape of the beats are: amplitude of P-peak (Pamp), amplitude of Q-valley (Qamp), amplitude of R-peak (Ramp), amplitude of S-valley (Samp) and amplitude of T-peak (Tamp).

Features that describe the position of waves in the window of beat are: position of P-peak (Ploc), position of Q-valley (Qloc), position of R-peak (Rloc), position of S-valley (Sloc) and position of T-peak (Tloc). Along with the 10 morphological features, the temporal feature QRS duration has been calculated using starting position of the Q wave and end of the S wave.

To mark P, Q, R, S, T wave locations in the ECG signal, the following algorithm has been used. After reading the ECG signal, decomposition and reconstruction using wavelet is done. Then, R peak location is detected by keeping 60% of the signal as threshold. P, Q, S and T points are detected with reference to R locations. Then QRS Complex duration is calculated which plays an important role in Abnormality detection [12].

2.4 Algorithm used to calculate QRS Complex Duration

Step 1: Read the ECG signal.
Step 2: Detect the QRS complex duration waveform.
Step3: Perform the wavelet analysis using Daubechies Wavelet.
Step 4: Compute the approximation and detailed coefficients using wavelet decomposition.
Step 5: Detect R peak location in the signal keeping 60% of the signal value as threshold.
Step 6: Detect Q point by finding the smallest value in the range Rloc-50 to Rloc-10.

Step 7: Detect S point by finding the smallest value in the range Rloc+5 to Rloc+50.

Step 8: Detect T point by finding the highest value in the range Rloc+25 to Rloc+100.

Step 9: Calculate QRS complex duration using the following step:

QRS (i, j) = ceil ((SOFF (i, j)-QON (i, j)))

Step 10: Find X=QRS.

The actual signal of normal ECG is shown in figure 7. The wavelet decomposition is applied to normal signal and reconstructed signal is shown in figure 8 and 9. The base line wandering is removed from the normal signal is shown in figure 10.

Detected R-Peak in actual signal is shown in figure 11. Feature points of normal ECG signal is pointed in figure 12.

In the same procedure is adopted for abnormal beats of PVC, LBBB and RBBB and necessary features are extracted.

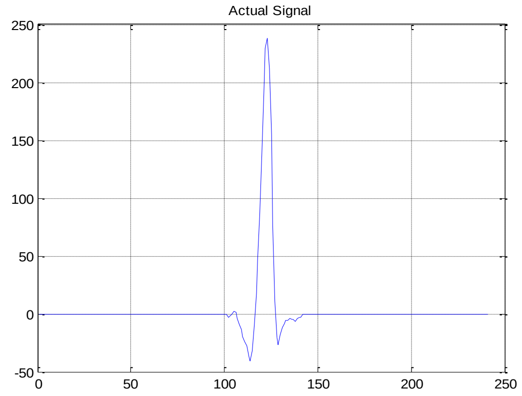


Fig 7: Normal ECG Signal

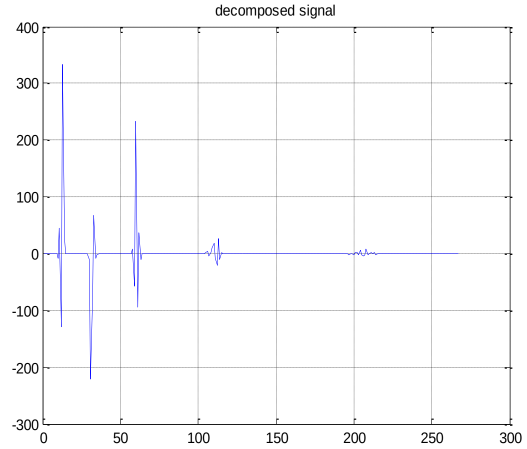


Fig 8: Decomposed Signal

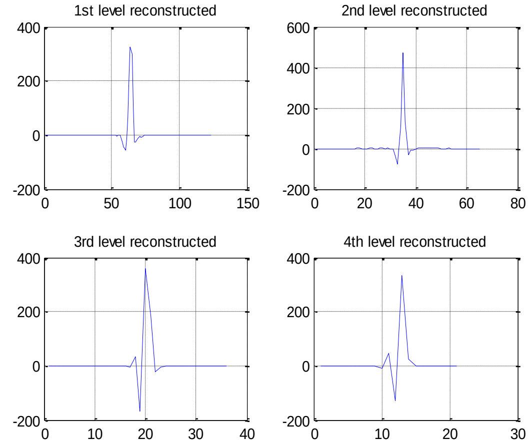


Fig 9: Reconstructed Signal

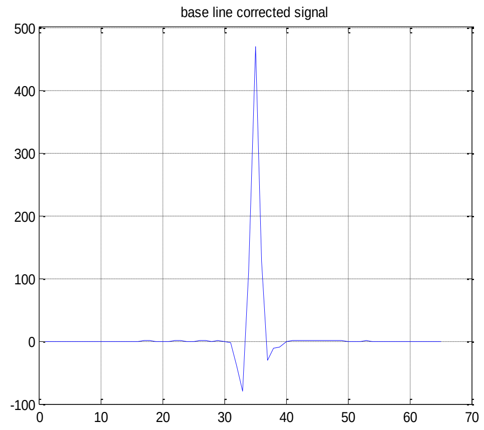


Fig 10: Base line Corrected signal

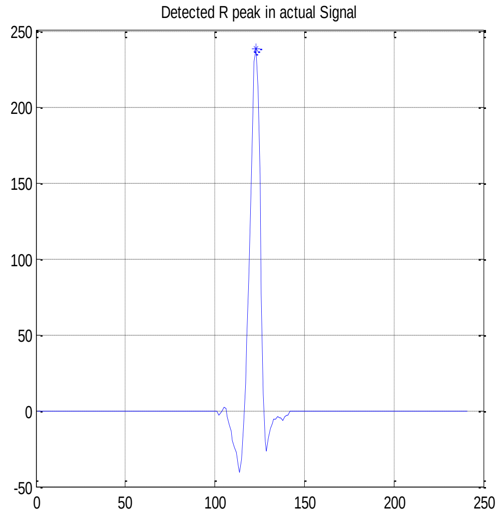


Fig 11: Detected R peak in Normal Signal

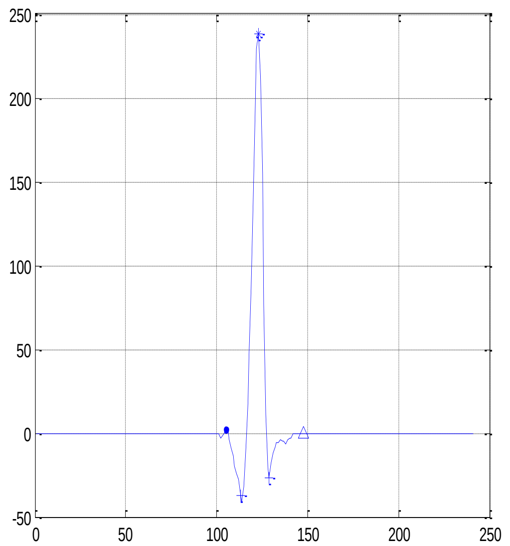


Fig 12: Feature Points of Normal ECG Signal

3. PROBABILISTIC NEURAL NETWORK

Artificial neural networks have been used to solve a wide variety of tasks that are hard to solve using ordinary rule-based programming. In this work, Probabilistic Neural Network (PNN) was used for classification [13] [14]. A probabilistic neural network (PNN) is a feed-forward neural network, derived from the Bayesian network and a statistical algorithm called Kernel Fisher Discriminant analysis. In a PNN, the operations are organized into a multilayered feed-forward network with four layers namely Input layer, Pattern layer, Summation layer and Decision layer as shown in figure 13.

There is one neuron in the input layer for each predictor variable value. The input neurons then supply the values to each of the neurons in the pattern layer. Pattern layer has one neuron for each case in the training data set. The neuron stores the values of the predictor variables for the case beside with the target value.

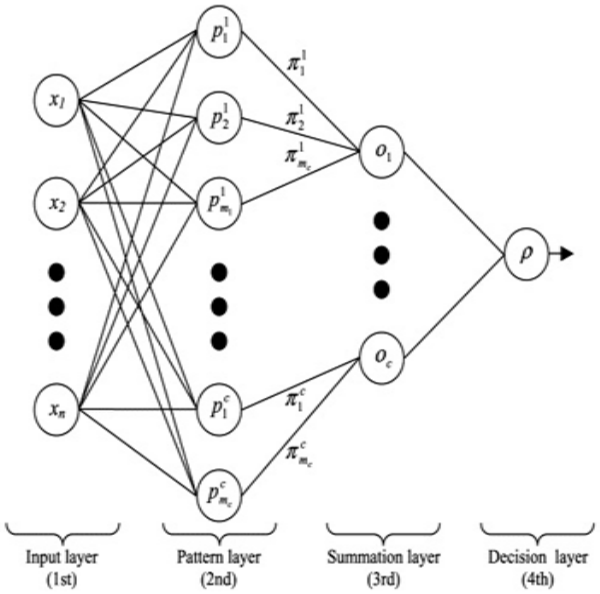


Fig 13: Architecture of Probabilistic Neural Network

On receiving a pattern x from the input layer, the neuron x_{ij} of the pattern layer computes its output

$$\emptyset_{ij}(x) = \frac{1}{(2\pi)^{\frac{d}{2}}\,\sigma^d} \exp\left[-\frac{\|(x - x_{ij})\|^2}{2\sigma^2}\right]$$

Where, d denotes the dimension of the pattern vector x, σ is the smoothing parameter and x_{ij} is the neuron vector stored in pattern layer neuron. The summation layer neurons compute the maximum likelihood of pattern x being classified into c_i by summarizing and averaging the output of all neurons that belong to the same class

$$P_i(x) = \frac{1}{(2\pi)^{\frac{d}{2}}\,\sigma^d}\frac{1}{N_i}\sum_{j=1}^{N_i} \exp\left[-\frac{\|(x - x_{ij})\|^2}{2\sigma^2}\right]$$

Where, N_i denotes the total number of samples in class C_i. If the apriori probabilities for each class are the equal, and the losses associated with making an incorrect decision for each class are the equal, the decision layer unit classifies the pattern in accordance with the Bayes's decision rule based on the output of all the summation layer neurons

$$\hat{C}(x) = argmax\{P_i(x)\}, \qquad i = 1,2, \ldots \ldots . m$$

Where, $\hat{C}(x)$ denotes the estimated class of the pattern x and m is the total number of classes in the training samples [15]. Design parameter of probabilistic neural network is spread of radial basis transfer function. Small or no training is required for probabilistic neural network except spread optimization [16].

4. EXPERIMENT AND RESULTS

The experiment has been carried out using the ECG signals from the MIT–BIH arrhythmia database [17]. The database comprise 48 recordings, each one of 30 minutes duration and includes two leads, the modified limb lead II (MLII) and one of the modified leads V1, V2, V4 or V5. The 48 recordings of ECG signals cover all the 16 different heart beat types including normal beat. In this experiment, a total of six ECG records with numbers: 100, 106, 119, 208, 109 and 118 are used from the database to collect different type of beats. A total of 400 beats are extracted where 100 beats for each type i.e., Normal, PVC, LBBB, RBBB. To collect Normal and PVC beats, MLII lead is used and for LBBB and RBBB beats, V1 lead is used.

In the dataset, a cross-validation procedure is applied to provide better generalization of neural network classifiers. A cross-validation procedure is applied to provide better generalization of neural network classifiers. To carry out the cross-validation procedure input data is partitioned into 3 sets called training, validation and test data sets. The training dataset is used to train the network. The validation dataset is used to permit the network, to adjust network design parameters. The test dataset is used to test the generalization performance of the selected design of neural network. The separating of input data is performed randomly with a ratio 70:15:15 of input entities to be stored as training dataset, validation dataset and test dataset respectively.

To obtain the optimal performance of the probabilistic neural network spread factor is adjusted. The method of adjusting the spread is organized as a search. The search is performed in 3 iterations with different range of varying parameter and different search step.

For each significance value of spread the probabilistic neural network is simulated for the train and validation datasets, the values of correct classification function for the train dataset and for the validation dataset are stored. The values of the correct classification function are plotted versus the spread.

The significance value of spread factor that ensures the best generalization is chosen. The probabilistic neural network is build and simulated on joined training and validation datasets, the value of correct classification function is calculated. Finally, the overview performance of the network is tested on the test dataset. The performance of the classifier is tabulated in Table I.

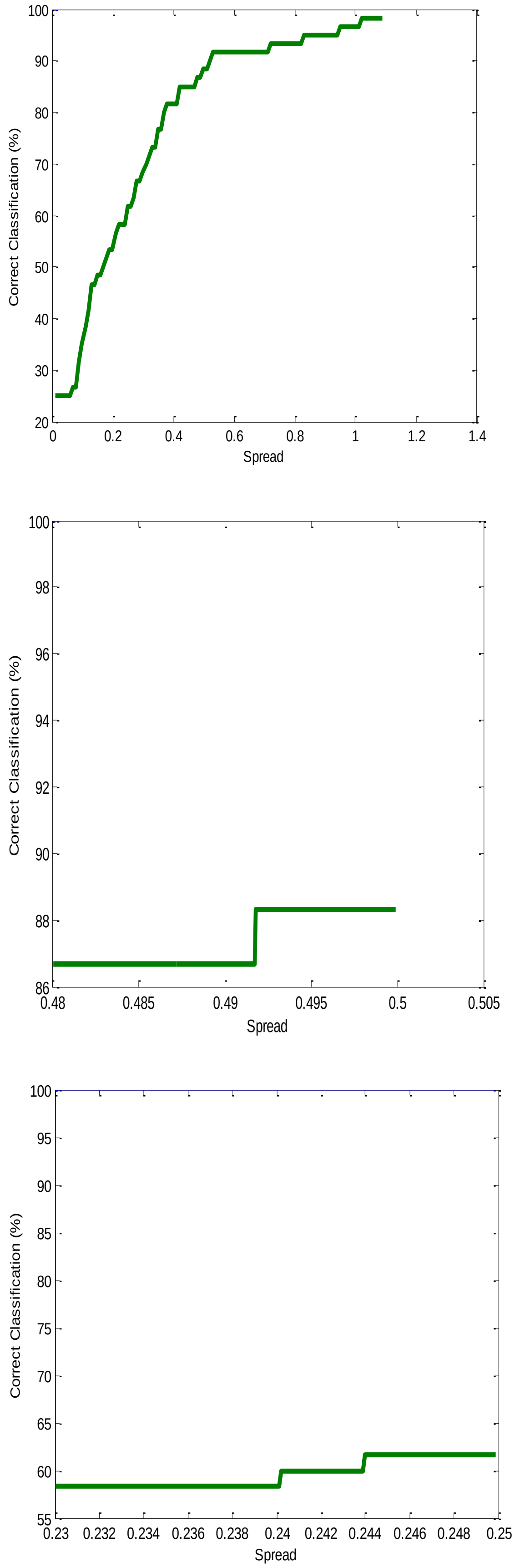


Fig 14: Classification Accuracy for different spread (σ) values
Blue line – training set; Green line – validation set

Table 1. Performance of the Classifier

Spread Parameter (σ)	Predictive Accuracy (%)
0.742	95.00
0.752	96.67
0.772	98.33

5. CONCLUSION

An electrocardiogram (ECG) recording of a patient is important clinical information for the medical experts to diagnose the heart functionality of the patient or assess the patient before any surgery. The proposed work automated the ECG beat classification task using powerful classifier namely Probabilistic Neural Network. Automated systems to classify the ECG will enable the doctors in decision-making process and to take effective decisions for the patients with heart problems.

6. REFERENCES

[1] W. J. Tompkins, "Biomedical Digital Signal Processing", Prentice Hall, New Jersey, 1993.

[2] Atul Luthra, "ECG Made Easy". Japee Brothers Publishers, 2007.

[3] F. de Chazal and R. B. Reilly, "A patient adapting heart beat classifier using ECG morphology and heartbeat interval features" ,IEEE Trans. Biomedical Engineering Volume 53, Issue 12, pp. 2535–2543, December 2006.

[4] Thaweesak, Yingthawornsuk "Classification of Cardiac Arrhythmia via SVM", 2012 2nd International Conference on Biomedical Engineering and Technology, IPCBEE vol. 34 (2012) © (2012) IACSIT Press, Singapore.

[5] Wisnu Jatmiko, Nulad W. P., Elly Matul I.,I Made Agus Setiawan, P. Mursanto," Heart Beat Classification Using Wavelet Feature Based on Neural Network ," Wseas Transactions on Systems, ISSN: 1109-2777 Issue 1, Volume 10, January 2011.

[6] Maedeh Kiani Sarkaleh and Asadollah Shahbahrami, "Classification of ECG Arrhythmias using Discrete Wavelet Transform and Neural Networks", International Journal of Computer Science, Engineering and Applications (IJCSEA) Volume 2, Issue 1, February 2012.

[7] Karpagachelvi.S, Dr.M.Arthanari and Sivakumar M, "Classification of Electrocardiogram Signals with Extreme Learning Machine and Relevance Vector Machine", International Journal of Computer Science Issues, Volume 8, Issue 1, January 2011 ISSN (Online): 1694-0814.

[8] Ali Khazaee, "Automated Cardiac Beat Classification using RBF Neural Networks", International Journal of Modern Education and Computer Science, Voloum 3, pp. 42-48, 2013.

[9] Amit kumar manocha, Mandeep singh, "An Overview of Ischemia detection techniques, International Journal of Scientific and Engineering Research", Volume 2, Issue 11, 2011.

[10] N.C Griswold, Somit Shah Mathur, Mark Yeary, "Ronald G. Spencer, Wavelet Decomposition or Reconstruction of images via direct products", Journal of Electric Imaging Volume 9, Issue 1, pp.61-71 2000.

[11] Misiti, Y. Misiti and G. Oppenheim, J-M. Poggi, "Wavelet toolbox for use with MATLAB", 2006.

[12] P. M. Okin, M. J. Roman, R. B. Devereux, and P. Kligfield, "Time voltage area of the QRS for the identification of Left Ventricular Hypertrophy", The American Heart Association, Inc. , volume 27, pp. 251-258, 1996.

[13] D. F. Specht, "Probabilistic neural networks", Neural Networks, Vol. 3, No. 1, pp. 109 – 118, 1990.

[14] K. Z. Mao, K. C. Tan and W. Ser, "Probabilistic Neural-Network Structure Determination for Pattern Classification", IEEE Transactions on Neural Networks, Volume 11, Issue 4, pp. 1009 – 1016, 2000.

[15] "Enhancements to the probabilistic neural networks," in Proceedings IEEE International Joint Conference of Neural Networks, Baltimore, MD, 1992, pp. 761–768.

[16] R. L. Streit and T. E. Luginbuhl, "Maximum likelihood training of probabilistic neural network," IEEE Trans. Neural Networks, volume 5, pp.764–783, 1994.

[17] R. Mark and G. Moody, MIT-BIH Arrhythmia Database 1997 [Online]. Available: http://ecg.mit.edu/dbinfo.html.