Optimization of a multi-class MLP ECG classifier using FCM

R. B. Ghongade¹ and A. A. Ghatol²

¹Vishwakarma Institute of Information Technology, Pune- 411048, India
²Dr. Babasaheb Ambedkar Technological University, Lonere- 402103, India
rbghongade@gmail.com

Abstract
ECG pattern classification using MLP is an effective and robust technique. Due to the inherent structure of MLP and training algorithm, MLPs tend to be slow and bulky in terms of the hidden layer neurons. This condition is aggravated further if the input data dimension is large as in the case of ECG. The paper addresses this problem by optimizing the MLP using an additional clustering of the feature extracted data. Considerable reduction in the size of MLP was recorded (max. 67.9%) with a reduction in training time (maximum of 59.81%). The experimentation used benchmark arrhythmia database from Physionet Massachusetts institute of technology-Beth Israel hospital (MIT-BIH). Four feature extraction methodologies were subjected to fuzzy c-means clustering for obtaining optimized MLPs. Ten statistical morphological features were also considered for designing the MLP classifiers.

Keywords: ECG, MLP, DCT, DWT, FCM, morphological features.

Introduction
Electrocardiogram pattern of classification using multilayer perceptron is now a fairly established paradigm (Bortolan et al., 1996). Average classification accuracy up to 99%, has been already reported (De Chazel et al., 2000). Although high accuracies are obtained using MLP, they suffer from disadvantages like, they train slowly, require lots of training data (typically three times more training samples than network weights), network sizes (network weights) grow with the dimensions of the input data and output classes and the requirement of a large amount of training time. This paper proposes the use of fuzzy c-means clustering as an additional data conditioning after feature extraction, for compacting the dimensions of the input data. Different ten-class ECG beat classifiers are designed using DCT, biorthogonal wavelet (bior2.2), daubechies wavelet (db9), coiflet wavelet (coif1) as feature extraction schemes. These classifiers were tested and in the second phase optimized using FCM.

Materials and methods
Arrhythmia benchmark database from Physionet Massachusetts institute of technology-Beth Israel hospital (MIT-BIH) was used for this work. Record numbers 102, 107,109, 111, 124, 200, 203, 205, 207, 208, 212, 213, 214, 217, 231, 232 were used for training, cross-validation and testing purposes. A total of 37,797 heartbeats were extracted and 29,363 heartbeats were selected. In this work arrhythmia is the general class of heart disease under consideration. Arrhythmia is visible in the form of morphological changes in the QRS complex of the individual heartbeat. The types of arrhythmia to be diagnosed are: Premature ventricular contraction (PVC), left bundle branch block beat (LBBBB), right bundle branch block beat (RBBBB), atrial premature contraction (APC), fusion of normal heartbeat and PVC in addition to normal and paced beats. Since PVC (also LBBBB & RBBBB) may appear in two different morphological forms in order to improve the classification accuracy (reduce the complexity of the decision boundaries), we have further divided these in TYPE I and TYPE II categories. Thus there are 10 total types of heartbeats to be classified: APC, FUSION, LBBBB I, NORMAL, PACED, RBBBB I, PVC I, PVC II, LBBBB II and RBBBB II. The platform used for experimentation was a Pentium-IV based machine running on Microsoft Windows XP SP3 at 2.66 GHz, 2GB RAM. Feature extraction and training was done using MATLAB® R2008b software.

The methodology for the evaluation and search of the best network was the same as in (Ghongade & Ghatol, 2009). The complete scheme includes various steps like the extraction of equal length ECG signals beat wise; mean adjust to remove the base line wander and dc shift, feature extraction and finally training of the ANN (Sternickel, 2002).

Data pre-processing
Since each record is a continuous waveform it is necessary to extract only a single heartbeat. This is done by considering the R peak and extracting 90 samples on either side of this R peak (Vargas et al., 2002). Since the MIT/BIH database comes with annotations for each heartbeat it is necessary to associate the categories with individual extracted beats. A total of 6000 beats were selected with 600 beats belonging to each category viz. APC, FUSION, LBBBB I, NORMAL, PACED, RBBBB I, PVC I, PVC II, LBBBB II and RBBBB II.
These features are: R-peak in addition to the R-R interval. Supportive for good classification features were found to be. On computing and morphological feature set is input vector. Hence an optimal restrain the dimensionality of the features has to be limited to features, but again the number of large number of possible (Engin et al., 2007) suggests a QRS morphology. Literature these features are specific and distinct for each type of complex contains important morphological features. In addition to the transform domain features, the QRS coif1.1 and bior2.2 in this work.母亲 wavelets were used for feature extraction; db9, coefficients can be used as the feature vector. Three out at three levels and only the level 3 approximation as the mother wavelet. The decomposition can be carried to a single heart beat can be subjected to DWT with db9 2004). For example, the 180 sample signal corresponding comparing it with the original normal waveform (Vargas et al., 2002). Discrete wavelet transform also offers an effective method for data reduction (Liang-Yu Shyu et al., 2008).

Feature extraction
For feature extraction, we can use the energy compaction property of the DCT to reduce the dimensions of the data. Only those dominant components are retained, which contribute most to the signal energy. First 30 coefficients were found to contribute to about 99.49% of the total energy. In addition to this the percent root mean difference (PRD) was calculated by reconstructing the normal waveform from only 30 components and comparing it with the original normal waveform (Vargas et al., 2004). For example, the 180 sample signal corresponding to a single heart beat can be subjected to DWT with db9 as the mother wavelet. The decomposition can be carried out at three levels and only the level 3 approximation coefficients can be used as the feature vector. Three mother wavelets were used for feature extraction; db9, coif1.1 and bior2.2 in this work.

In addition to the transform domain features, the QRS complex contains important morphological features. These features are specific and distinct for each type of QRS morphology. Literature (Engin et al., 2007) suggests a large number of possible features, but again the number of features has to be limited to restrain the dimensionality of the input vector. Hence an optimal morphological feature set is desired. On computing and graphical representation, ten features were found to be supportive for good classification in addition to the R-R interval. These features are: R-peak amplitude, mean power spectral density, the Q-S distance, signal energy, QRS area, singular value decomposition, area under auto-correlation curve, Q-R slope , R-S slope, R-R interval (Ghongade & Ghatol, 2008). MLP
Multilayer perceptrons (MLPs) are layered feed-forward networks typically trained with static back propagation (Ghongade & Ghatol, 2009). This work uses one hidden layer MLP with momentum learning and tan-sigmoidal activation functions for hidden as well as output layers.

Experimentation
The feature vector of 30 DCT coefficients plus 10 features per heartbeat is used as input data and the corresponding annotation forms the target. Thus the dataset consists of 6000 input-target pairs. Similarly, 37 coefficients corresponding to three-level decomposition by db9, 26 coefficients for coif1 and bior2.2 along with ten morphological features form the input data.

Table 1. Performance of 10 class classifiers.

<table>
<thead>
<tr>
<th>Feature extraction</th>
<th>Average accuracy %</th>
<th>Normal SPE</th>
<th>PVC I SPE</th>
<th>PVC II SPE</th>
<th># of HLN</th>
<th>Training time (S)</th>
<th>Feature vector size</th>
<th>Total ADJ parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOR2.2</td>
<td>99.14</td>
<td>0.999444</td>
<td>0.996341</td>
<td>0.999556</td>
<td>70</td>
<td>265</td>
<td>26+10=36</td>
<td>3300</td>
</tr>
<tr>
<td>DB9</td>
<td>99.14</td>
<td>0.999444</td>
<td>0.99645</td>
<td>0.999334</td>
<td>45</td>
<td>202</td>
<td>37+10=47</td>
<td>2620</td>
</tr>
<tr>
<td>COIF1</td>
<td>99.12</td>
<td>0.999444</td>
<td>0.996561</td>
<td>0.99445</td>
<td>80</td>
<td>300</td>
<td>26+10=36</td>
<td>3770</td>
</tr>
<tr>
<td>DCT</td>
<td>99.10</td>
<td>0.999333</td>
<td>0.997003</td>
<td>0.999223</td>
<td>80</td>
<td>316</td>
<td>30+10=40</td>
<td>4090</td>
</tr>
</tbody>
</table>

Table 2. Performance of various schemes with level 2 feature extraction scheme.

<table>
<thead>
<tr>
<th>Scheme</th>
<th># of HLN</th>
<th>APC accuracy %</th>
<th>Fusion accuracy %</th>
<th>Normal accuracy %</th>
<th>Paced accuracy %</th>
<th>RBBB accuracy %</th>
<th>PVC II accuracy %</th>
<th>PVC I accuracy %</th>
<th>LBBBBII accuracy %</th>
<th>RB BBB accuracy %</th>
<th>Average accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIOR2.2 + FCM</td>
<td>60</td>
<td>98.66667</td>
<td>98.33333</td>
<td>100</td>
<td>99.5</td>
<td>99.6667</td>
<td>99.6667</td>
<td>99.5000</td>
<td>99.6667</td>
<td>99.5000</td>
<td>99.216667</td>
</tr>
<tr>
<td>DCT+ FCM</td>
<td>60</td>
<td>98.33333</td>
<td>98.16667</td>
<td>100</td>
<td>99.5</td>
<td>99.3333</td>
<td>97.1667</td>
<td>99.3333</td>
<td>99.5000</td>
<td>99.6667</td>
<td>99.100000</td>
</tr>
</tbody>
</table>

Note: # of HLN means the number of hidden layer neurons.

Fig. 1. Block schematic of the proposed novel scheme for optimization of feature vector size.
Classifier design

A multilayer perceptron is employed with back propagation and momentum learning. The first experiment was to find optimum number of hidden layer neurons thus deciding the network configuration. For this, the number of exemplars for each class was kept equal and the dataset was divided into training, cross-validation and testing sets with a ratio of 50%, 25% and 25% respectively. "One-hot encoding" of the classes is again employed for supervised training. The datasets formed were then used for training a single hidden layered MLP, employing momentum weight update with a momentum rate = 0.7. The best networks were again searched for, by varying the number of hidden layered neurons (15 to 80) and training the network 5 times with termination on cross-validation MSE. Table 1 depicts the performance; time required (time per exemplar per epoch) and the total adjustable parameters (weights & biases) of the best networks found.

Classifier optimization

From Table 1 it can be seen that the feature size and the best MLP configuration contribute to the size of the network, i.e. the total adjustable parameters (weights) and the training time. In an attempt to optimize the classifier in terms of efficiency and speed the number of adjustable parameters and the feature vector size were reduced further. A novel scheme was devised by the author for optimization of the classifier. Fuzzy c-means clustering offers an effective method of clustering data (Ozbayet et al., 2006; Ceylan et al., 2007). A second level feature extraction in the form of FCM was introduced. This reduces the dimensionality of the feature vector and thus reduced the number of processing elements, number of weights in effect improving the training speed. The level 2 feature extraction was employed only for the transform part of the feature vector, keeping the statistical features intact. Such optimized datasets were created for BIOR2.2, DB9, COIF1 and DCT only. The scheme is depicted in block schematic of Fig. 1.

Results

Datasets, according to the optimization scheme discussed above, were generated. It was found that 19 clusters were sufficient to produce acceptable classification accuracy. The datasets so formed are then again trained and tested as per previous phases to get the best performing networks and schemes. Here the number of hidden layer neurons was again varied from 15 to 80.
level feature extraction scheme (Table 4). Fig. 2 and 3 depict the comparison between the un-optimized and optimized schemes with respect to number of adjustable parameters (weights) and training time, respectively. Fig. 4 and 5 show the percent saving between the un-optimized and optimized schemes with respect to number of adjustable parameters (weights) and training time, respectively. Thus, the usage of FCM as second level feature extraction reduced the network adjustable parameters and the training time without compromising the average classification accuracy. Training time for optimized classifier takes into consideration the additional overhead of time required for clustering.

**Table 4. Comparison of single level feature extraction scheme with two level feature extraction scheme.**

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Average accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without optimization</td>
</tr>
<tr>
<td>BIOR2.2</td>
<td>99.14</td>
</tr>
<tr>
<td>DB9</td>
<td>99.14</td>
</tr>
<tr>
<td>COIF1</td>
<td>99.12</td>
</tr>
<tr>
<td>DCT</td>
<td>99.10</td>
</tr>
</tbody>
</table>

**Conclusion**

MLP-based ten-class classifiers designed with DCT, bior2.2, db9 and coif1 as feature extraction schemes and trained with back propagation, achieved average classification accuracies greater than 99.1%. These classifiers can be effectively optimized for adjustable parameters (network size) and training time with the use of FCM as a second level input dimension compactor. The reduction in adjustable weights ranged from 26.97% to 67.90% and the reduction in training time was from 18.25% to 59.81% without any deterioration in the average classification accuracies. The time required for fuzzy clustering was also considered as an overhead, in spite of the reduction in overall training time. Clearly, the fuzzy clustering of the transformed data proves an effective method for data dimensionality reduction.

**References**