J Med Syst (2012) 36:2705–2711
DOI 10.1007/s10916-011-9746-6

ORIGINAL PAPER

Classification of Juvenile Myoclonic Epilepsy Data Acquired Through Scanning Electromyography with Machine Learning Algorithms

Imran Goker · Onur Osman · Serhat Ozekes · M. Baris Baslo · Mustafa Ertas · Yekta Ulgen

Received: 23 March 2011 / Accepted: 2 June 2011 / Published online: 17 June 2011

Abstract In this paper, classification of Juvenile Myoclonic Epilepsy (JME) patients and healthy volunteers included into Normal Control (NC) groups was established using Feed-Forward Neural Networks (NN), Support Vector Machines (SVM), Decision Trees (DT), and Naïve Bayes (NB) methods by utilizing the data obtained through the scanning EMG method used in a clinical study. An experimental setup was built for this purpose. 105 motor units were measured. 44 of them belonged to JME group consisting of 9 patients and 61 of them belonged to NC group comprising ten healthy volunteers. k-fold cross validation was applied to train and test the models. ROC curves were drawn for k values of 4, 6, 8 and 10. 100% of detection sensitivity was obtained for DT, NN, and NB classification methods. The lowest FP number, which was obtained by NN, was 5.

Keywords Scanning electromyography · Juvenile myoclonic epilepsy · Feed-forward neural networks · Support vector machines · Decision trees · Naïve bayes

I. Goker
Faculty of Economics and Administrative Sciences, Department of Management Information Systems, Okan University, Istanbul, Turkey
e-mail: imran.goker@okan.edu.tr

O. Osman (✉)
Faculty of Engineering and Architecture, Department of Electrical & Electronics Engineering, Istanbul Arel University, Istanbul, Turkey
e-mail: onurosman@arel.edu.tr

S. Ozekes
Faculty of Engineering and Architecture, Department of Computer Engineering, Istanbul Arel University, Istanbul, Turkey
e-mail: serhatozekes@arel.edu.tr

M. B. Baslo
Istanbul University Capa Medical Faculty, Istanbul, Turkey
e-mail: mbbaslo@istanbul.edu.tr

M. Ertas
Anadolu Health Center, Istanbul, Turkey
e-mail: mustafaertas@superonline.com

Y. Ulgen
Institute of Biomedical Engineering, Bogazici University, Istanbul, Turkey
e-mail: ulgeny@bun.edu.tr

Introduction

Scanning EMG is an experimental technique developed to study the spatial and temporal properties of a motor unit (MU) electrical activity in order to better understand the topography of a motor unit by Stalberg et al. [1, 2]. Motor unit is the basic functional and anatomical unit of the skeletal muscle [3]. Motor unit action potential (MUAP) which is the electrical activity generated by the MU and detected by the conventional electromyography (EMG) is utilized in the diagnosis of the neuromuscular disease. However, conventional EMG provides information only 5–10% of electrical activity of a MU [4]. On the other hand, scanning EMG method introduces new parameters such as length of cross section (LCS) and maximum amplitude in addition to amplitude and duration provided by the conventional EMG. These new parameters ensure to have insight about spatial and temporal characteristics of the electrical activity of a MU [4, 5]. LCS is measured from the 3-D maps of the electrophysiological cross-section of

the MU. Either LCS or maximum amplitude is considered to be used in the differentiation of the neuromuscular and/or neurological diseases.

In a previous experimental clinical study [6], the differences and the similarities between three subjects groups including Juvenile Myoclonic Epilepsy (JME) patients, healthy volunteers considered as normal controls (NC) and Spinal Muscular Atrophy (SMA) patients were investigated using these two parameters. The results of this study can be supplemented via a classification established by using different classification methods.

Classification is a method which is used in computer-aided categorization of several cases in terms of related parameters. Classification can be described as either supervised learning algorithm or unsupervised method in the machine learning process. The classification algorithms used in this study are: decision trees [7], neural networks [8], support vector machines [9] and naïve bayes. These algorithms are used in literature in various classification studies such as decision trees in gastric cancer diagnosis by Su et al. [10], a classification and regression tree model in the mass-to-charge ratio and peak heights of proteins identified by mass spectroscopy by Markey et al. [11], the classification of ambulatory ECG arrhythmic events via artificial neural networks (ANN) by Silipo et al. [12], in the recognition of malignant melanoma versus dysplastic naevus using support vector machines algorithm by Maglogiannis et al. [13], in the prediction of transmembrane protein topology using SVM by Nugent et al. [14], in the prediction of palmitoylation site using naïve bayes algorithm by Xue et al. [15], in the classification of microarray data using sequential feature extraction approach for naïve bayes by Fan et al. [16].

The data used in this study was acquired by means of an experimental setup designed for scanning EMG described in a previous study from the subject groups [6, 17]. These data were used either to construct the 3-D maps of the electrophysiological cross-section of MUs from which LCS of the MUs were measured and/or to extract the maximum amplitude within the MU territory. The acquired data are used in this study to determine the variables namely the length of cross section, maximum amplitude, the minimum amplitude, the variance, the standard deviation, the median, the mode, and the peak-to-peak voltage within the MU territory which are represented as LCS, max, min, var, std, mean, median, mode, v_{ppmax} respectively.

Four different classifiers, which were based on decision trees (DTs), neural networks (NNs), support vector machines (SVMs) and naïve bayes (NB) were trained and tested using k-fold cross validation, and several experiments were performed with different k values. The detection performances of DT, NN, SVM, and NB in classification tasks were compared by showing the results.

Materials and methods

The subject groups

Three subject groups which were referred as JME and NC were constructed. JME group contained nine Juvenile Myoclonic Epilepsy (JME) patients. NC group comprising ten healthy volunteers was the Normal Control group. Five to ten measurements were made in each patient. While 44 MUs were measured in JME group, 61 measurements were established in NC group. This study was approved by the

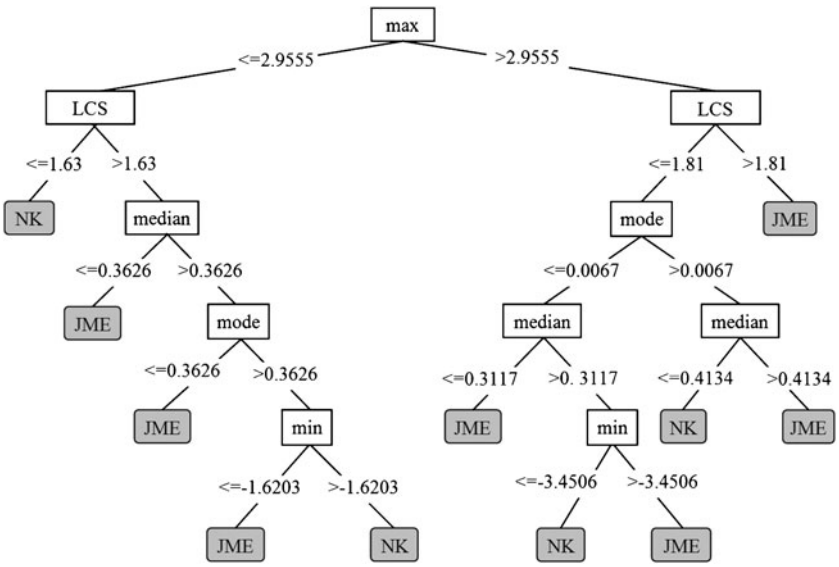

Fig. 1 The constructed decision tree

Table 1 k-fold cross validation results of DT classifier for different k values

k values	TP	FN	TN	FP	sensitivity	Specificity	AUC
4	42	2	26	35	0.955	0.426	0.987
6	42	2	31	30	0.955	0.508	0.986
8	42	2	29	32	0.955	0.475	0.967
10	44	0	28	33	1.000	0.459	0.992

Local Ethical Committee of Istanbul Faculty of Medicine. The measurements were made from the biceps brachii muscles of the subjects by applying an experimental protocol which the details are described in a recent publication [6].

Data acquisition

An experimental setup was built to establish the scanning EMG method. This setup included an EMG instrument to amplify and to filter the MUAPs measured from the biceps muscle through the concentric needle EMG electrodes, a linear actuator to move the record electrode across the MU territory, a data acquisition system to digitalize the amplified and filtered signals to be stored in a computer. These stored data were used to construct 3-D maps of the electrophysiological cross-section of the MU territories. EMG signals were acquired via a concentric needle EMG (CNEMG) electrode in each step of the upward movement of this electrode in the form Motor Unit Action Potential (MUAP) versus time. This experimental setup was described with details in recent publications [6, 17].

Features of electrophysiological cross-section

In this study the features of Electrophysiological Cross-Section are determined as length of cross section, maximum amplitude, the minimum amplitude, the variance, the standard deviation, the median, the mode, and the peak-to-peak voltage within the MU territory which are represented as shortly LCS, max, min, var, std, mean, median, mode, v_{ppmax} respectively. These features are used as the input of the classifiers which are briefly explained as below.

Decision tree classifier

The decision tree is made up of decision nodes, branches and leaves. Decision node determines the test to be realized. The result of this test causes the branching without losing data. In every node, testing and branching are realized consecutively and this branching is dependent on upper level. Every branch of the tree is a candidate to complete the classification. If the classification cannot be realized, a decision node is formed at the end of the branch. But if a certain class is formed at the end of the branch, there is a leaf [18]. This leaf is one of the classes to be determined from the data. The operation of the decision tree starts from the root nodes and, follows the consecutive nodes from top to bottom until reaching the leaf.

Decision tree classification depends on the fact that constructing a decision tree by using the training set is chosen from the dataset. Also the quality of the decision tree depends on the size of the tree and the accuracy of its classification [19]. At this stage the determination of nodes in the decision tree is very important. It should be designated that which fields of the dataset will be used in which order to construct the tree [20]. For this purpose the most commonly used measure is entropy. The entropy is also used in information technologies. The higher the entropy of an attribute, the more uncertainty there is with respect to its outcomes. Thus we would wish to select attributes in order of increasing entropy, where the root node of our tree would correspond to the attribute, A_k, is given as [19]:

$$E(C|A_k) = \sum_{j=1}^{M_k} p(a_{k,j}) \times \left[-\sum_{i=1}^{N} p(c_i|a_{k,j}) \log_2 p(c_i|a_{k,j}) \right] \quad (1)$$

Table 2 k-fold cross validation results of NN classifier for different k values

k values	TP	FN	TN	FP	sensitivity	Specificity	AUC
4	41	3	56	5	0.932	0.918	0.954
6	41	3	53	8	0.932	0.869	0.903
8	42	2	53	8	0.955	0.869	0.938
10	44	0	51	10	1.000	0.836	0.908

Table 3 k-fold cross validation results of SVM classifier for different k values

k values	TP	FN	TN	FP	sensitivity	Specificity	AUC
4	41	3	39	22	0.932	0.639	0.828
6	39	5	45	16	0.886	0.738	0.796
8	37	7	46	15	0.841	0.754	0.823
10	38	6	52	9	0.864	0.852	0.824

where

$E(C\|A_k)$	entropy of the classification property of attribute A_k
$p(a_{k,j})$	probability of attribute k being at value j
$p(c_i\|a_{k,j})$	probability that the class value is c_i when attribute k is at its jth value
M_k	total number of values for attribute a_k; j=1,2,..,M_k
N	total number of different classes; i=1,2,.., N
K	total number of attributes; k=1,2,.., K

The term in the brackets is called the information. Thus as Eq. 1 implies, entropy is the expected information that is the sum of the information in the several possible outcomes multiplied by their probability. Logarithms are generally taken to base 2, so that the information is measured in bits.

If a set S of records is partitioned into classes C_1, C_2, C_3, . . ., C_i, then the information needed to identify the class of an element of S is denoted by:

$$I(S) = -(p_1\log_2(p_1) + p_2\log_2(p_2) + \ldots + p_i\log_2(p_i)) \tag{2}$$

where p_i is the probability distribution of the partition C_i [19]. The term in the brackets in Eq. 1 is similar to Eq. 2. Thus entropy in Eq. 1 can be written as [19]:

$$E(A) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times I(S_i) \tag{3}$$

Thus the information gain in performing a branching with attribute A can be calculated with this equation [20]:

$$Gain(A) = I(S) - E(A) \tag{4}$$

The information gain computed for each attribute is used to choose the test attribute in each node of the decision tree. The attribute with highest information gain is chosen as the test attribute for the current node. This attribute minimizes the information necessary for the classification of the data and the problems which may occur during branching.

For each variable in the dataset, the information gains were computed using Eqs. 1, 2, 3 and 4, and the decision tree was constructed as seen in Fig. 1. The number of leaves is 11 and the size of the tree is 21.

Neural network classifier

The classification of the JME and NC data was also performed by feed-forward neural network. The studies on neural networks were emerged by the inspiration from the psychology and the neuroscience. A neural network is a set of connected input/output units where each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class of the input samples. In this study the back propagation algorithm performed learning on a multilayer feed-forward neural network. The input layer of the network consisted of the features of the electrophysiological cross-section as explained above. The hidden layer included 50 neurons and the output layer had one neuron.

Support vector machine based classifier

The classification of the JME and NC data was also achieved using C-support vector classification (C-SVC). Given training vectors $x_i \in R^n, i = 1, \ldots, l$ in two classes, and a vector $y \in R^l$ such that $y_i \in \{1,-1\}$, C-SVC [21–23] training involves the minimization of the error function

$$\min_{w,b,\xi} \left[\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i\right] \tag{5}$$

subjected to the constraints: $y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i$ and $\xi_i \geq 0, i = 1, \ldots, l$

Table 4 k-fold cross validation results of NB classifier for different k values

k values	TP	FN	TN	FP	sensitivity	Specificity	AUC
4	41	3	50	11	0.932	0.820	0.921
6	39	5	50	11	0.886	0.820	0.877
8	44	0	47	14	1.000	0.770	0.901
10	45	1	53	8	0.978	0.869	0.923

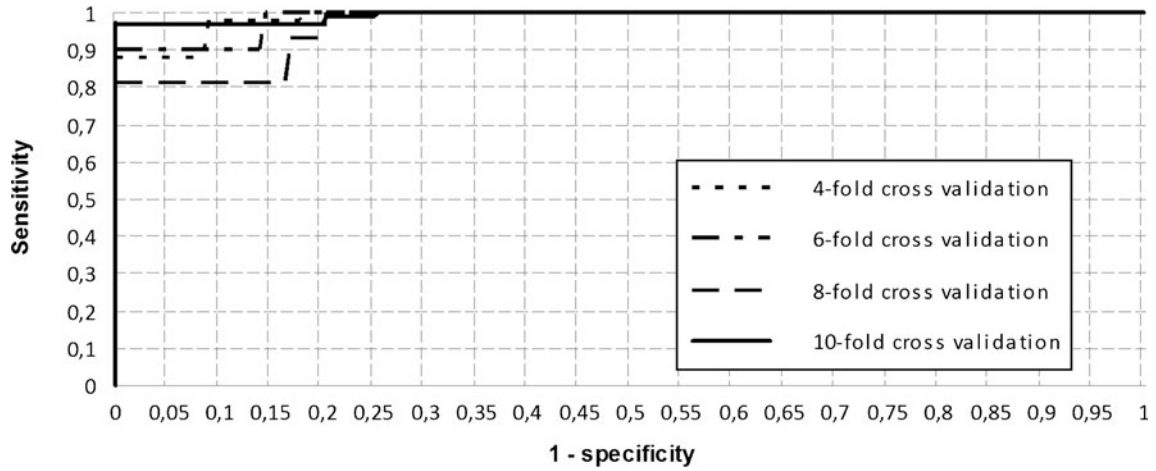

Fig. 2 ROC curves for DT classifier

where $C=1$ is the capacity constant determining the tradeoff, w is the vector of coefficients, b is a constant representing the bias of the hyperplane of SVM and ξ_i are parameters for handling non separable data (inputs).

Its dual is $\min_{\alpha}\left[\frac{1}{2}\alpha^T Q\alpha - e^T\alpha\right]$, subject to $\mathbf{y}^T\alpha = 0,\ 0 \leq \alpha_i \leq C$,

where e is the vector of all ones, $C>0$ is the upper bound, Q is an $l \times l$ positive semidefinite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \phi(x_i)^T\phi(x_j)$ is the kernel. Here training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function ϕ.

The decision function can be formulated as

$$\mathrm{sgn}\left(\sum_{i=1}^{l} y_i\alpha_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{6}$$

where b *is* the parameter determined by SVM's learning algorithm. Those samples x_i with nonzero parameters α_i are called "support vectors".

In this study radial basis function (RBF) kernel was used in C-SVC model. RBF is seen in

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\gamma\|\mathbf{x} - \mathbf{x}_i\|^2\right) \tag{7}$$

where γ is a constant in kernel function and is set to 1/ number of attributes.

Naïve bayes based classifier

The classification of the JME and NC data was also performed using the simple probabilistic classifier called as naïve bayes classifier. The posterior probability of each class, C_i, is obtained by the naïve bayes classifier using bayes rule. The classifier makes the simplifying assumption that the attributes, A, are independent given the class, so the likelihood can be obtained by the product of the individual conditional probabilities of each attribute given the class

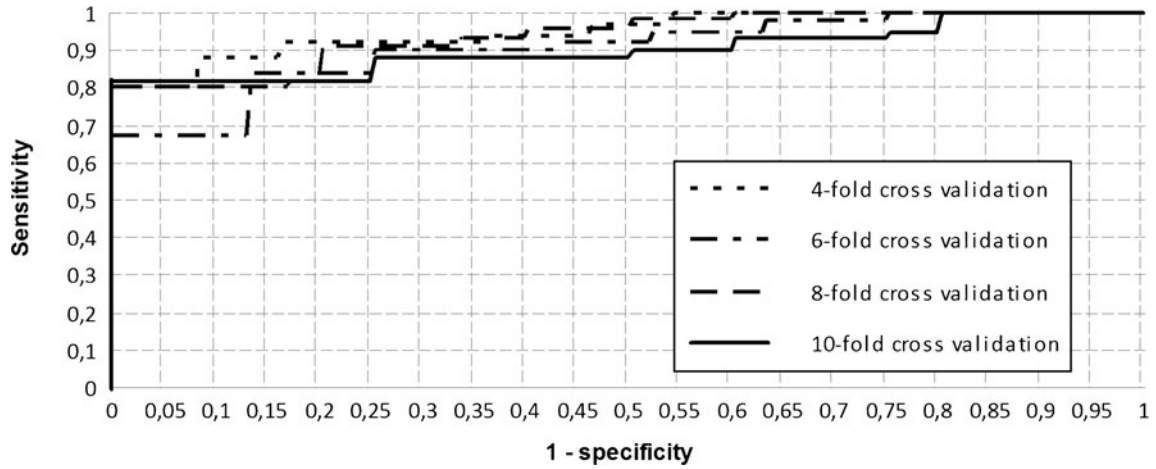

Fig. 3 ROC curves for NN classifier

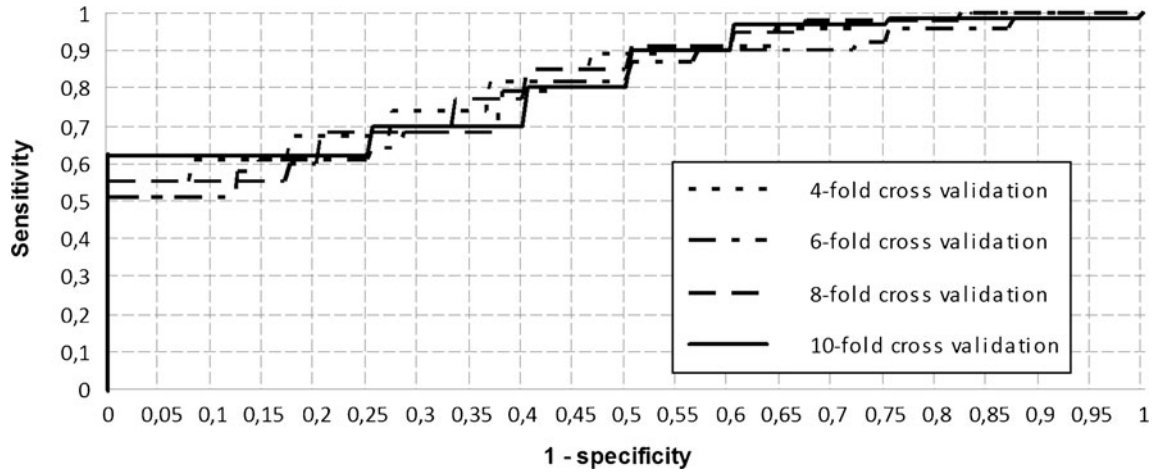

Fig. 4 ROC curves for SVM classifier

[24]. Thus, the posterior probability, $P(C_i \mid A_1, \ldots, A_n)$, can be given by:

$$P(C_i \mid A_1, \ldots, A_n) = P(C_i)P(A_1 \mid C_i) \ldots P(A_n \mid C_i)/P(A) \quad (8)$$

Results

In the previous clinical study [6], 105 motor units were measured. 44 of them belonged to JME group consisting of nine patients and 61 of them belonged to NC group comprising ten healthy volunteers.

The classification task was carried out using the models based on DT, NN, SVM and NB. k-fold cross validation was performed to train and test each classifier. Different results were obtained using various k values. The true positive (TP), false negative (FN), true negative (TN), false positive (FP), sensitivity, specificity and area under the ROC curve (AUC) were summarized in Tables 1, 2, 3 and 4 according to the k values. Also ROC curves can be seen in Figs. 2, 3, 4 and 5.

DT based classifier reached to 100% detection sensitivity with 33 FPs in 10-fold cross validation. For the rest of the k values 95.5% detection sensitivity was achieved by the DT based method. The AUCs were between 0.992 and 0.967. 100% detection sensitivity was obtained with 10 FPs using the NN based method and 10-fold cross validation. The sensitivity of the NN classifier was 95.5% in 8-fold, and 93.2% in 4-fold and 6-fold cross validation. The number of FPs was only 5 and the AUC was 0.954 when the classifier was NN and the k value was 4. The detection sensitivity of the SVM classifier was 93.2% with 22 FPs and the AUC was 0.828 in 4-fold cross validation. In SVM classifier the FP number was only 9 in 10-fold cross validation. 100% detection sensitivity was obtained with 14 FPs using the NB based method and 8-fold cross validation. In NB classifier 8 FPs were achieved with 97.8% sensitivity and the AUC was 0.923 in 10-fold cross validation.

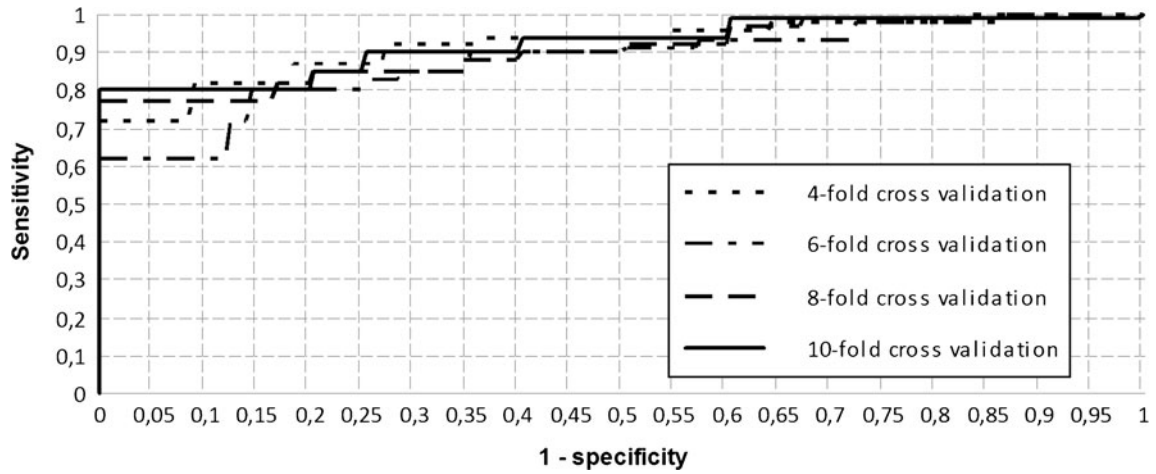

Fig. 5 ROC curves for NB classifier

Discussion

When the results given Tables 1, 2, 3 and 4 are referred, it is observed that DT classifier has the 100%-sensitivity value at 10-fold cross validation. Its FP value is 33. NN classifier has also 100%-sensitivity value at 10-fold cross validation. Its FP value is 10 being lower than that of DT. SVM is not able to reach 100%-sensitivity value. NB has 100%-sensitivity value at 8-fold cross validation. Its FP value is 14 at this sensitivity level being higher than that of NN. It seems that NN has the best sensitivity value with the lowest FP value.

AUC can be considered as another performance criterion. DT classifier has the highest AUC value in 10-fold cross validation compared to other classifiers. It seems that this classifier has the best classification performance in terms of AUC.

Conclusion

We have tested the effects of different learning based classification methods on JME detection. This study includes 105 motor units where 44 belong to JME group and 61 belong to NC groups.

To classify the JME patients and the individuals in NC group, DT, NN, SVM and NB based models were applied separately. We presented the effectiveness of our system by applying it to data obtained from JME and NC groups. k-fold cross validation was applied to train and test the models. The results were recorded for various k values.

For example in 10-fold cross validation the detection sensitivity was 100% for DT and NN classifiers, in 8-fold cross validation it was again 100% for NB classifier. The best detection sensitivity for SVM classifier was 93.2% in 4-fold cross validation. The lowest FP numbers in DT, NN, SVM and NB based methods were 30, 5, 9 and 8, respectively for various k values. To demonstrate the system performance, ROC curves were drawn for k values of 4, 6, 8 and 10. When Tables 1, 2, 3 and 4 are examined, and when both the sensitivity and the FP values are considered, it can be demonstrated that the JME and NC classification performance of neural network based model is better than other models. Thus, this model can be integrated with a scanning EMG system for the classification of such neurological and neuromuscular pathologies to be used as a clinical testing procedure in differential diagnosis of these pathologies.

References

1. Stålberg, E., and Eriksson, P. O., A scanning electromyographic study of the topography of human masseter single motor units. *Arch Oral Biol* 32:793–797, 1987.
2. Stålberg, E., and Falck, B., The role of EMG in neurology. *Electroencephalogr Clin Neurophysiol* 103:579–598, 1997.
3. Preston, D. C, and B. E. Shapiro, *Electromyography and neuromuscular disorders: clinical-electrophysiologic correlations*, (Butterworth-Heinemann, Philadelphia, 2005).
4. Diószeghy, P., Scanning electromyography. *Muscle Nerve Suppl* 11:S66–S71, 2002.
5. Gootzen, T. H. J. M., Vingerhoest, D. J. M., and Stegeman, D. F., A study of motor unit structure by means of scanning EMG. *Muscle and Nerve* 15:349–357, 1992.
6. Goker, I., Baslo, B., Ertas, M., and Ulgen, Y., Large motor unit territories by scanning electromyography in patients with juvenile myoclonic epilepsy. *Journal of Clinical Neurophysiology* 27:212–215, 2010.
7. Quinlan, J. R., *C45: programs for machine learning*. Morgan Kaufmann, San Mateo, 1993.
8. R. Lippmann, An Introduction to computing with neural nets, *IEEE ASSP Magazine*. 22, 1987.
9. V. N. Vapnik, *The Nature of Statistical Learning Theory.* Springer, 1998.
10. Su, Y., Shen, J., Qian, H., Ma, H., Ji, J., Ma, H., Ma, L., Zhang, W., Meng, L., Li, Z., Wu, J., Jin, G., Zhang, J., and Shou, C., Diagnosis of gastric cancer using decision tree classification of mass spectral data. *Cancer Sci* 98:37–43, 2007. doi:10.1111/j.1349-7006.2006.00339.x.
11. Markey, M. K., et al., Decision tree classification of proteins identified by mass spectrometry of blood serum samples from people with and without lung cancer. *Proteomics* 3:1678–1679, 2003.
12. Silipo, R., Gori, M., Taddei, A., Varanini, M., and Marchesi, C., Classification of arrhythmic events in ambulatory electrocardiogram, using artificial neural networks. *Comput Biomed Res* 28:305–318, 1995.
13. Maglogiannis, I. G., and Zafiropoulos, E. P., EP (2004) Characterization of digital medical images utilizing support vector machines. *BMC Med Inform Decis Making* 4:4, 2004.
14. Nugent, T., and Jones, D. T., Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* 10:159, 2009.
15. Xue, Y., Chen, H., Jin, C., Sun, Z., and Yao, X., NBA-Palm: prediction of palmitoylation site implemented in Naïve Bayes algorithm. *BMC Bioinformatics* 7:458, 2006.
16. Fan, L., Poh, K. L., and Zhou, P., A sequential feature extraction approach for Naïve bayes classification of microarray data. *Expert Systems with Applications* 36:9919–9923, 2009.
17. Goker, I., Baslo, B., Ulgen, Y., and Ertas, M., Design of an experimental system for scanning electromyography method to investigate alterations of motor units in neurological disorders, *Digest Journal of Nanomaterials and Biostructures* 4:133–139, 2009.
18. Berson, A., Smith, S., and Thearling, K., *Building Data Mining Applications for CRM*. McGraw-Hill Professional Publishing, New York, 2000.
19. Han, J., and Kamber, M., *Data mining concepts and techniques, the morgan kaufmann series in data management systems*, 2nd edition. Elsevier Inc, San Francisco, 2006.
20. Agrawal, R., Imielinski, T., and Swami, A., Database mining: a performance perspective. *IEEE Transactions on Knowledge and Data Engineering* 5:914–925, 1993.
21. B. E. Boser, I. Guyon, and V. Vapnik, A training algorithm for optimal margin classifiers, *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory,* (1992), 144–152.
22. Cortes, C., and Vapnik, V., Support-vector network. *Machine Learning* 20:273–297, 1995.
23. C.C. Chang, C. J. Lin, LIBSVM : A library for support vector machines, (2008).
24. M. Martínez, L. E. Sucar, Learning an optimal Naïve Bayes classifier, *Proc. IEEE Inter. Conf. on Pattern Recognition (ICPR)*, China, (2006).