Machine Learning Algorithms for Characterization of EMG Signals

Bekir Karlık, Member, IACSIT

I. INTRODUCTION

The electromyographic (EMG) signal provides information about the performance of muscles and nerves. In the other words, the EMG signal meters electrical currents generated in muscles during its convulsion representing neuromuscular activities. It can be detected from the skin surface by using surface Ag/AgCl bipolar electrodes easily. Surface EMG signals recorded from skin surface have been widely used in different fields such as prosthesis control [1]-[25], analysis of functional electrical stimulation (FES) [26], [27], human-machine interaction [28], [29], pathological tremor [30], and muscle fatigue analysis [31], [32]. EMG signal is a type of random signals. Hence, it is of very importance to accurately extract the signature information from surface EMG signals.

Recently novel signal processing techniques and mathematical models have made it practical to develop advanced EMG detection and analysis methods [33]. Different mathematical and machine learning techniques such as Artificial Neural Networks (ANN), Fuzzy Systems, Probabilistic model algorithms, Metaheuristic and Swarm intelligence algorithms, and some hybrid algorithms are used for characterization of EMG signals.

This article firstly presents a brief explanation about machine learning algorithms. Secondly, each part of myoelectric control of human arm prosthesis has been described such as EMG signal analysis, useful feature extraction and classifiers techniques for EMG signal have been defined. Finally, some literature applications of EMG signal characterization for human arm prosthesis have been comprised and discussed.

II. MACHINE LEARNING ALGORITHMS

Learning algorithm is an adaptive method by network computing units self-organizes to realize the target (or desired) behavior. Machine learning is about learning to predict from samples of target behaviors or past observations of data. Machine learning algorithms are classified as [34].

1) Supervised learning where the algorithm creates a function that maps inputs to target outputs. The learner then compares its actual response to the target and adjusts its internal memory in such a way that it is more likely to produce the appropriate response the next time it receives the same input.

2) Unsupervised learning (clustering, dimensionality reduction, recommender systems, self-organizing learning) which models a set of inputs. There is no target outputs (no any labeled examples). The learner receives no feedback from environment.

3) Semi-supervised learning where the algorithm creates both labeled and unlabeled examples a special function.

4) Reinforcement learning is learning by interacting with an environment. The learner receives feedback about the appropriateness of its response.

5) Learning to learn where the algorithm learns its own inductive bias based on previous experience. It calls as inductive learning.

A. Artificial Neural Networks

ANN can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques with their remarkable ability to derive meaning from complicated or imprecise data. ANN is an information processing system. It is composed of a large number of interconnected parallel processing elements (called as neurons) working in unison to solve different problems. The other advantages of ANN include [35]:

- Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
- Self-Organization: ANN generates its own organization or representation of the information it receives during learning time.
- Real Time Operation: ANN computations may be carried out in parallel and special hardware devices are being designed and manufactured which take advantage of this capability.
- Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network harm.

Well-known and useful ANN algorithms are; Learning
Vector Quantization (LVQ), Back-Propagation (BP), Radial Basis Function (RBF), Recurrent Neural Network, and Kohonen self-organizing network.

B. Fuzzy System

If intelligent systems are to mimic human beings, these should also be able to handle the same imprecision and uncertainty of human thought processes. Fuzzy logic is used in these systems - a generalization of stiff Boolean logic. It uses fuzzy sets that are a generalization of crisp sets in classical set theory. The main deviation is the appropriation of fuzzy membership functions for each set. Thus, whereas in classical set theory, an object could just be either a member of set or not at all, in fuzzy set theory, a given object is said to be of a certain “degree of membership” to the set [36]. A fuzzy system consists of a fuzzy rule base, a fuzzification module, an inference engine, and a defuzzification module. The fuzzification module pre-processes the input values submitted to the fuzzy expert system. The inference engine uses the results of the fuzzification module and accesses the fuzzy rules in the fuzzy rule base to infer what intermediate and output values to produce. The final output of the fuzzy expert system is provided by the defuzzification module.

An efficient way for cluster analysis defined by Bezdek [37]. This methodology, namely fuzzy logic based c-means (FCM), is a data-clustering method where in each data point belongs to a cluster to some degree that is specified by a membership grade. Based on an iterative procedure, FCM minimizes an objective function that represents the distance from any given data point to a cluster centre weighted by that point’s membership grade. The use of the FCM clustering method is motivated by the fact that traditional clustering approaches create partitions where in each pattern belongs to one and only one cluster. In this way, the clusters in a hard clustering are disjointed.

C. Probabilistic Model Algorithms

Bayesian networks are well known representative probabilistic graphical models algorithms. Maximum entropy is another general method for estimating probability distributions from data. Each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. The overriding principle in maximum entropy is that when nothing is known, the distribution should be as uniform as possible, that is, have maximal entropy. Labeled training data is used to derive a set of constraints for the model that characterize the class-specific expectations for the distribution [38].

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes’ theorem [39]. According to the precise nature of the probability model, naive Bayes classifiers can be trained efficiently in a supervised learning setting. One of advantages of the Naive Bayes classifier is requirement small size of training data to predict the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the full compliment covariance matrix [40].

Linear Discriminant Analysis (LDA) and the related Fisher's linear discriminant are simple methods used in statistics and machine learning to find the linear combination of features. These features separate two or more classes of object LDA works when the measurements made on each observation are continuous quantities [41].

The k-nearest neighbor algorithm (k-NN) a non parametric lazy learning algorithm which is an instant-based learning algorithm that classified objects based on closest feature space in the training set. The training sets are mapped into multi-dimensional feature space. The feature space is partitioned into regions based on the category of the training set. A point in the feature space is assigned to a particular class if it is the most frequent class among the k nearest training data [42]. Generally Euclidean Distance is used in computing the distance between the vectors.

A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian that has been used [43]. GMM not only provides a smooth overall distribution fit, its components can, if required, clearly detail a multimodal density. GMM parameters are predicted from training data using the iterative Expectation-Maximization algorithm or Maximum A Posteriori estimation from a well-trained prior model. It has shown noticeable performance in many applications, such as bioinformatics, biomedical, text and speech recognition, and has been a tool in pattern recognition problems.

Polynomial Classifier (PC) is universal approximators to the optimal Bayes classifier [44]. It is based on statistical methods or minimizing a mean-squared error (MSE) criterion. PC is linear or second order classifier. Hence it has some limitations.

Support Vector Machines (SVM) is a group of supervised learning algorithm that can be applied to classification or regression. It is theoretically well motivated algorithm: defined from Statistical Learning Theory by Vapnik and Chervonenkis since the 60s [45]. SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each labeled as depending on one of two categories, an SVM training algorithm builds a model that designates new examples into one category or the other. This method is a representation of the examples as points in space, mapped so that the examples of the different classes are divided by a clear gap that is as wide as possible. SVM are the data points that lie closest to the decision surface. New examples are then mapped into that same space and estimated according to a class based on which side of the gap they fall on [46]. SVM has empirically good performance and successful applications in many fields (bioinformatics, text, pattern recognition, etc.)

D. Hybrid Algorithms

Some hybrid classifier algorithms such as Adaptive Neuro-Fuzzy Inference System (ANFIS), Fuzzy Clustering Neural Network (FCNN) are also used to solve pattern recognition problems.

ANFIS is integration both Fuzzy system and artificial neural network. Algorithm was defined by Jang in 1992 [47]. It creates a fuzzy decision tree to classify the data into one of \(2^p \) (or \( p^{th} \)) linear regression models to minimize the sum of squared errors (SSE). Its inference system corresponds to a
set of fuzzy IF–THEN rules that have learning capability to approximate nonlinear functions. ANFIS uses other cost function (rather than SSE) to represent the user’s utility values of the error (error asymmetry, saturation effects of outliers, etc.). It can also use other type of aggregation function (rather than convex sum) to better handle slopes of different signs. Fig. 1 shows the architecture of ANFIS.

Fuzzy Clustering Neural Networks (FCNN) is a hybrid learning algorithm which integrates both Fuzzy C-means clustering and neural networks. FCNN was defined and used by Karlik [7], [48]-[50]. When one encounters fuzzy clustering, membership design includes various uncertainties such as ambiguous cluster membership assignment due to choice of distance measure, fuzzifier, prototype, and initialization of prototype parameters, to name a few. Proper management of uncertainty in the various parameters that are used in clustering algorithms is essential to the successful development of algorithms to further yield improved clustering results. The idea of fuzzy clustering is to divide the data into fuzzy partitions, which overlap with each other. Therefore, the containment of each data to each cluster is defined by a membership grade in [0, 1]. Then, a novel fuzzy clustering neural network structure was used for the training consists of two stages. At the first stage, inputs and outputs values of feed-forward type neural network are found using Fuzzy C-means clustering algorithm. At the second stage, these clustering data is applied as desired values of MLP, which has one hidden layers [51].

Number of data points was reduced using fuzzy c-means clustering before inputs are presented to a neural network system. Therefore, training period of the neural network is decreased.

III. MYOELECTRIC CONTROL OF PROSTHESIS

Fig. 3 shows that the block diagram of myoelectric control of human arm prosthesis. Surface EMG signals are recorded by standard Ag/AgCl disposable bipolar electrodes which are accompanied by miniature pre-amplifiers to differentiate small signals. The EMG electrodes are put for recording the muscle activities of the biceps, triceps, wrist flexors, and wrist extensors which are most useful. Signals are then amplified, filtered (using 2nd order Butterworth filter), performed sampling and segmentation.

The feature extraction module presents preselected features for a classifier. Features, instead of raw signals, are fed into a classifier for improving classification efficiency. The classification module recognizes EMG signal patterns, and classifies them into predefined categories. Because of the complexity of EMG signals, and the influence of physiological and physical conditions, the classifier should be adequately robust and intelligent. So, it needs machine learning algorithms to solve this complexity of EMG signals. Controller generates output commands based on signal patterns and control schemes. The participants felt the prosthesis should automatically continue holding an object once grasped or tracing his/her arm. This allows the user to focus on moving the object with arm movements [33].

IV. FEATURE EXTRACTION METHODS

After the data recording, the acquired samples are converted into features which are used for classification. There are many feature extraction methods are applied on raw EMG to carry out actual EMG signal such as time series analysis (AR, MA, ARMA), Wavelet Transform (WT), Discrete Wavelet Transform (DWT) Wavelet Packet Transform (WPT), Fast Fourier Transform (FFT), Discrete Fourier Transform (DFT) etc.

A. Time-Series Modeling

A time series is a chronological sequence of observations of a particular variable of the amplitude of the raw EMG signal. The time series depend on the modeling of a signal to estimate future values as a linear combination of its past values and the present value. A model depends only on the previous outputs of the system is called an autoregressive model (AR). AR models are constructed using a recursive filter. AR method is the most frequently used parametric method for spectral analysis. By a rational system, the model-based parametric methods are established on modeling the data sequence \(x(n)\) as the output of a linear system characterized and the spectrum estimation procedure consists of two steps. The parameters of the method are calculated given data sequence \(x(n)\) that is \(0\leq n\leq N−1\). Then from these approximatives the he power spectral density...
(PSD) estimate is computed. AR models such as selection of the optimum estimation method (or selection of the model order) the length of the signal which is modeled, and the level of stationary of the data [52].

A model depends only on the inputs to the system is called a moving average model (MA). A model depends on both the inputs and on the outputs is considered autoregressive and moving average model which is called as ARMA. The model is usually then referred to as the ARMA (p, q) model where p is the order of the autoregressive part and q is the order of the moving average part. ARMA model is generally considered good practice to find the smallest values of p and q which provide an acceptable fit to the data. For a pure AR model the Yule-Walker equations may be used to provide a fit [53]. The method of moments gives good estimators for AR models, but less efficient ones for MA or ARMA processes. Hence, AR model is more useful than the other time series models.

Table I shows that comparison of some application results used machine learning classification algorithms based on time series modeling for characterization of EMG signals. These studies are listed in chronological order (from 1975 to 2010). In this table, the first column describes the authors and the second column describes type of machine learning algorithms that are used.

### TABLE I: COMPARISON OF MACHINE LEARNING ALGORITHMS APPLIED TO TIME SERIES MODELING FOR CHARACTERIZATION OF EMG SIGNALS

<table>
<thead>
<tr>
<th>Author</th>
<th>Method</th>
<th>Features</th>
<th>Class</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graupe &amp; Cline</td>
<td>NNC</td>
<td>ARMA</td>
<td>4</td>
<td>95</td>
</tr>
<tr>
<td>Doerschuk et al.</td>
<td>NNC</td>
<td>ARMA</td>
<td>4</td>
<td>95</td>
</tr>
<tr>
<td>Karlik et al.</td>
<td>MLP-BP</td>
<td>AR-1,P</td>
<td>6</td>
<td>84</td>
</tr>
<tr>
<td>Karlik et al.</td>
<td>MLP-BP</td>
<td>AR-2,P</td>
<td>6</td>
<td>92</td>
</tr>
<tr>
<td>Karlik et al.</td>
<td>MLP-BP</td>
<td>AR-3,P</td>
<td>6</td>
<td>95</td>
</tr>
<tr>
<td>Karlik</td>
<td>MLP-BP</td>
<td>AR-4,P</td>
<td>6</td>
<td>96</td>
</tr>
<tr>
<td>Lamoumier et al.</td>
<td>MLP-BP</td>
<td>AR-4</td>
<td>6</td>
<td>96</td>
</tr>
<tr>
<td>Soares et al.</td>
<td>MLP-BP</td>
<td>AR-10</td>
<td>4</td>
<td>95</td>
</tr>
<tr>
<td>Soares et al.</td>
<td>MLP-BP</td>
<td>AR-4</td>
<td>4</td>
<td>96</td>
</tr>
<tr>
<td>Karlik et al.</td>
<td>FCNN</td>
<td>AR-4</td>
<td>6</td>
<td>98</td>
</tr>
<tr>
<td>Chank&amp;Englehart</td>
<td>HMM</td>
<td>AR-6</td>
<td>6</td>
<td>95</td>
</tr>
<tr>
<td>Nils et al.</td>
<td>MLP-BP</td>
<td>MA</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>Farrell &amp; Weir</td>
<td>LDA</td>
<td>AR-3</td>
<td>6</td>
<td>90</td>
</tr>
<tr>
<td>Huang et al.</td>
<td>GMM</td>
<td>AR-6</td>
<td>6</td>
<td>97</td>
</tr>
<tr>
<td>Al-Assaf</td>
<td>PC</td>
<td>AR-3</td>
<td>5</td>
<td>95</td>
</tr>
<tr>
<td>Hargrove et al.</td>
<td>LDA/MLP</td>
<td>AR-6</td>
<td>6</td>
<td>97</td>
</tr>
<tr>
<td>Khezri &amp; Jahed</td>
<td>ANFIS</td>
<td>AR-4</td>
<td>6</td>
<td>95</td>
</tr>
<tr>
<td>Oskoei &amp; Hu</td>
<td>SVM</td>
<td>AR-6</td>
<td>6</td>
<td>96</td>
</tr>
<tr>
<td>Karlik et al.</td>
<td>FCNN</td>
<td>AR-4</td>
<td>4</td>
<td>89</td>
</tr>
<tr>
<td>Zhou et al.</td>
<td>LDA</td>
<td>AR-6</td>
<td>11</td>
<td>81</td>
</tr>
<tr>
<td>Khokhar et al.</td>
<td>SVM</td>
<td>AR-4</td>
<td>19</td>
<td>88</td>
</tr>
<tr>
<td>Khokhar et al.</td>
<td>SVM</td>
<td>AR-4</td>
<td>13</td>
<td>96</td>
</tr>
</tbody>
</table>


According to Table I, we can say that hybrid FCNN is better than the other machine learning algorithms for characterization of EMG signals used AR model parameters. Regarding the AR model, the authors observed that a fourth order model can adequately represent the EMG signal.

### B. Wavelet Transform

Wavelet transform (WT) reveals data aspects that other techniques miss, such as trends, breakdown points, discontinuities in higher derivatives, and self-similarity. Furthermore, WT can often compress or de-noise a signal, without appreciable degradation. There is a correspondence between scale and frequency in wavelet analysis: a low scale shows the rapidly changing details of a signal with a high frequency and a high scale illustrates slowly changing coarse features, with a low frequency. The most important advantage of the wavelet transform method is for the large low-frequency, high frequency which is changed to be narrow for the window size [54]. Therefore, WT acts as a ‘‘mathematical microscope’’, in which one can monitor different parts of a signal by just adjusting focus. As a generalization of WT, a wavelet packet transform (WPT) allows the ‘‘best’’ adapted analysis of a signal in a timescale domain [15]. WPT provides adaptive partitioning; a complete set of partitions are provided as alternatives, and the best for a given application is selected.

Discrete wavelet transform (DWT) is a special form of wavelet transform and provides efficient processing of the signal in time and frequency domains. In the DWT, each level is computed by passing only the previous wavelet approximation coefficients through discrete-time low and high pass filters. WPT is a wavelet transform where the discrete-time (sampled) signal is passed through more filters than DWT [55].

Table II describes comparison of some application results used machine learning classification algorithms based on wavelet transform (WT), discrete wavelet transform (DWT), and wavelet packet transform (WPT) for characterization of EMG signals. These studies are listed in chronological order.

### TABLE II: COMPARISON OF MACHINE LEARNING ALGORITHMS APPLIED TO WAVELET TRANSFORM FOR CHARACTERIZATION OF EMG SIGNALS

<table>
<thead>
<tr>
<th>Author</th>
<th>Method</th>
<th>Features</th>
<th>Class</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Englehart et al.</td>
<td>LDA</td>
<td>WPT</td>
<td>6</td>
<td>97</td>
</tr>
<tr>
<td>Englehart et al.</td>
<td>MLP-BP</td>
<td>WPT</td>
<td>6</td>
<td>93</td>
</tr>
<tr>
<td>Koçyiğit&amp;Korürek</td>
<td>FKNN</td>
<td>WT</td>
<td>4</td>
<td>96</td>
</tr>
<tr>
<td>Chu et al.</td>
<td>MLP-BP</td>
<td>WPT</td>
<td>9</td>
<td>97</td>
</tr>
<tr>
<td>Arvetti et al.</td>
<td>MLP-BP</td>
<td>WT</td>
<td>5</td>
<td>97</td>
</tr>
<tr>
<td>Khezri et al.</td>
<td>ANFIS</td>
<td>WT</td>
<td>6</td>
<td>97</td>
</tr>
<tr>
<td>Liu &amp; Luo</td>
<td>LVQ</td>
<td>WPT</td>
<td>4</td>
<td>98</td>
</tr>
<tr>
<td>Karlik et al.</td>
<td>MLP</td>
<td>DWT</td>
<td>4</td>
<td>97</td>
</tr>
<tr>
<td>Karlik et al.</td>
<td>FCNN</td>
<td>DWT</td>
<td>4</td>
<td>98</td>
</tr>
<tr>
<td>Khezri &amp; Jahed</td>
<td>MLP-BP</td>
<td>AR/DWT</td>
<td>6</td>
<td>87</td>
</tr>
<tr>
<td>Khezri &amp; Jahed</td>
<td>ANFIS</td>
<td>AR/DWT</td>
<td>6</td>
<td>92</td>
</tr>
</tbody>
</table>

**Accuracy:** percentage of correctly classified muscle-activation patterns, Class: The number of class of arm movements (depending on elbow, wrist, and grasp), ANFIS: adaptive neurofuzzy inference system, FCNN: fuzzy clustering neural network, FKNN: fuzzy k-nearest neighbor classifier, LDA: linear discriminant analysis, LVQ: learning vector quantization neural network, MLP-BP: multilayer perceptron with Backpropagation training algorithm, WPT: wavelet packet transform, WT: wavelet transforms, DWT: discrete wavelet transform, AR/DWT: combination of both AR and DWT models.

According to Table II, we can say that WPT and DWT are better feature extraction method than WT. Moreover, both hybrid models (FCNN and ANFIS) show more accuracy than.
the other machine learning algorithms for characterization of EMG signals.

V. CONCLUSION

This review article has presented comparison different machine learning algorithms used characterization of EMG signals for myoelectric control of human arm prosthesis. The EMG signals are modeled via time series models and wavelet transform models. These model coefficients are used as input for used machine learning classifiers. The outputs of classifiers are used as control data for the arm prosthesis.

Literatures results show that near perfect performance (95% to 98% rate of success) can be achieved when using the described machine learning methods. With respect to EMG signal feature extraction, it has been observed that the classifiers have successfully achieved the segmentation of AR coefficients into both four and six distinct pattern classes with very high rates of success. DWT is also very useful feature extraction method for EMG signals. But, the calculation of the AR coefficients is very faster than calculation of the DWT coefficients. Moreover, AR model does not require a lot of computing resources and the model did not have its performance reduced by variations of the shape (amplitude and phase) of the EMG signal.

For future work, the use of prosthesis to respond to the classified EMG signals can be used as a simulation environment to study new designs and control strategies.

REFERENCES

B. K. Arık received his BS, MS, and PhD degrees from Yildiz Technical University respectively. He has been working at Department of Computer Engineering, Engineering Faculty, Selcuk University in Konya, Turkey. His research work focuses on soft computing, machine learning, pattern recognition, biometric recognition, computer vision, biomedical data processing, and telemedicine. Prof. Karlık authored and co-authored more than 200 research papers in national or international academic journals and conferences, as well as 2 books and 3 edited books. He has been editor-in-chief and of advisory board of 2 journals, the editorial board member of 10, and reviewers of 33 journals. He has supervised 12 graduate students, and 2 PhD students, the program for those 11 of who are on-going now. He is recipient of several awards, and referee or reviewer for more than hundreds journals and numerous conferences. He also served on dean, head of departments, and several faculty and institute committees. He served Honorary Chair, Co-chair, and Session Chair, as well as TPC member or Keynote Speaker for a number of conferences. Prof. Karlık is senior member of IACSIT and member of Association for Scientific Research.