Regent University College of Science and Technology, April 7, 2015

# Machine Learning Applications in Medicine

by

**PROF. DR. BEKIR KARLIK**
**Selcuk University**
**Department of Computer Engineering**

# Overview

- Definition of ML Classifier
- Artificial Neural Networks
  - Back-Propagation (BP)
  - RBF
- Fuzzy Models
- Probabilistic Model Algorithms
  - K-NN
  - Naïve Bayes
  - SVM
  - The Others
- Hybrid Algorithms
- Applications

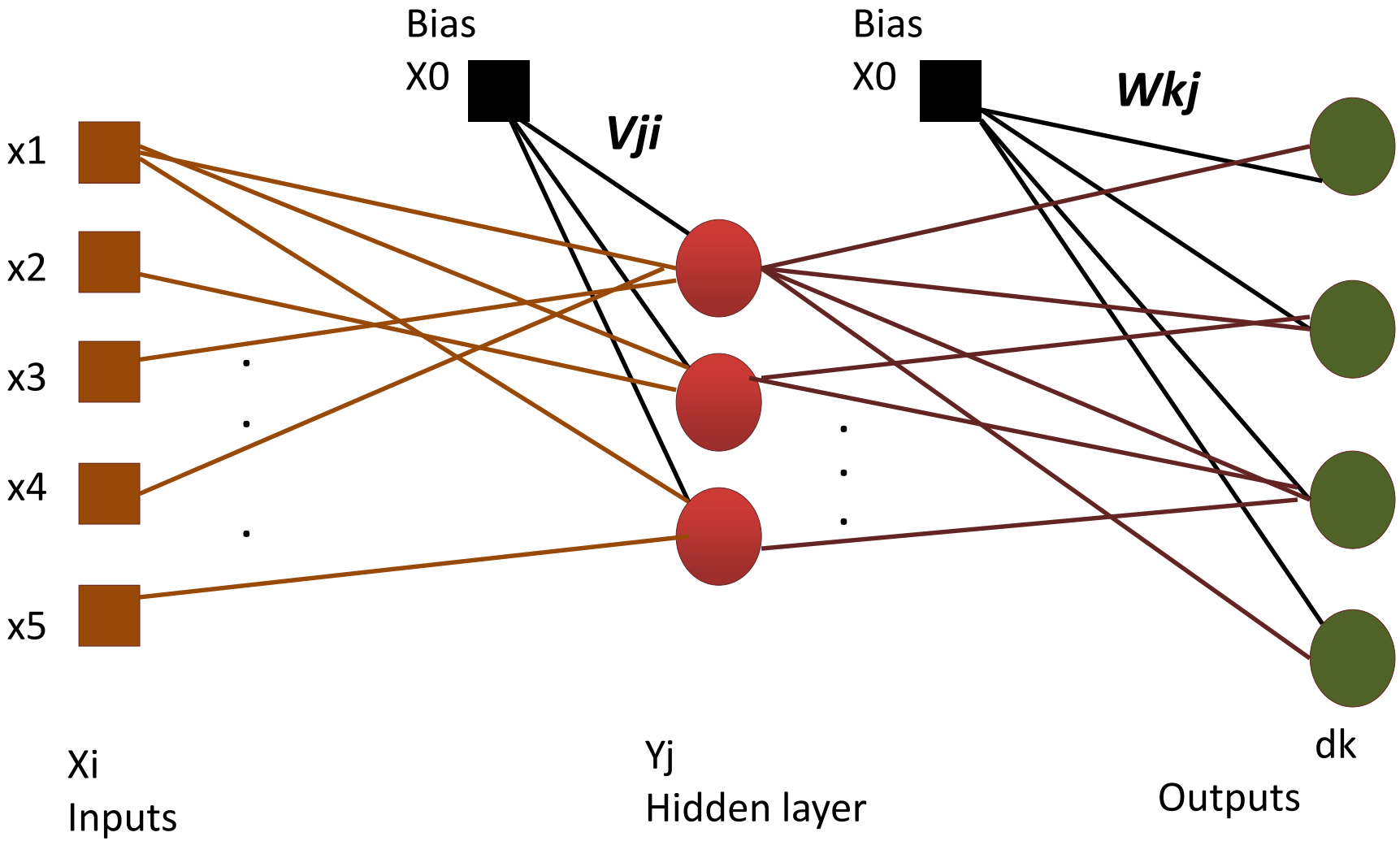
Learning algorithm is an adaptive method by network computing units self-organizes to realize the target (or desired) behavior. Machine learning is about learning to predict from samples of target behaviors or past observations of data.

Machine learning algorithms are classified as;

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**.

Artificial Neural Networks (ANN) is an information processing model, implemented in hardware or software that is modeled after biological process of the brain studied. Artificial neural network has ability to derive meaning from imprecise or complicated data to extract patterns and to detect trends that are not easily to recognize by humans or other computer techniques. ANN has been widely used to examine the complex relationships between input and output variables in many scientific and technological areas including biomedical and bioinformatics . 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.

# Multi-Layers Perceptron



## Back-Propagation

***The algorithm of Back-propagation*** used generalized delta learning rule is an iterative gradient algorithm designed to minimize the root mean square error between the actual output of a multilayered feed-forward ANN and a desired output. Each layer is fully connected to the previous layer, and has no other connection. The algorithm of Back-propagation classifier can be described as;

- Initialization: Set all the weights and biases to small real random values.
- Presentation of input and desired outputs: Present the input vector x(1), x(2),…,x(N) and corresponding desired response d(1),d(2),…,d(N), one pair at a time, where N is the number of training patterns.
- Calculation of actual outputs: Use Equation given below to calculate the output signals.

$$y_1, y_2, \ldots, y_{N_M} \quad y_i = \varphi\left(\sum_{j=1}^{N_{M-1}} w_{ij}^{(M-1)} x_j^{(M-1)} + b_i^{(M-1)}\right), \; i = 1, \ldots, N_{M-1}$$

- Adaptation of weights (*wij*) and biases (*bi*):

$$\Delta w_{ij}^{(l-1)}(n) = \mu . x_j(n) . \delta_i^{(l-1)}(n)$$

$$\Delta b_i^{(l-1)}(n) = \mu . \delta_i^{(l-1)}(n)$$

where

$$\delta_i^{(l-1)}(n) = \begin{cases} \varphi'(net_i^{(l-1)})\left[d_i - y_i(n)\right], & l = M \\ \varphi'(net_i^{(l-1)}) \sum_k w_{ki} . \delta_k^{(l)}(n), & 1 \le l \le M \end{cases}$$

## Radial Basis Function

**Radial basis function** (RBF) neural network is based on supervised learning. RBF's are embedded in a two layer neural network, where each hidden unit implements a radial activated function. The output units implement a weighted sum of hidden unit outputs. All hidden units simultaneously receive the n-dimensional real valued input vector X. Hidden-unit output is obtained by closeness of the input X to an n-dimensional parameter vector associated with the jth hidden unit. The response characteristics of the jth hidden unit ( j = 1, 2, …, J) is assumed as

$$Z_j = K(\| x - \mu_j \| / \sigma_j^2)$$

*where K* is a strictly positive radials symmetric function (kernel) with a unique maximum at its '**center**' and which drops off rapidly to zero away from the center. is the width of the receptive field in the input space from unit j. This implies that has an appreciable value only when the distance is smaller than the width .Given an input vector X, the output of the RBF network is the L-dimensional activity vector Y, whose lth component (l = 1, 2…L) is given by,

$$Y_l(x) = \sum_{j=1}^{j} w_{lj} Z_j(x)$$

**The Fuzzy system** model is the knowledge-based model with linguistic rules. Fuzzy sets are defined for all input and output variables and the set of rules. Fuzzy logic provides the means to process this knowledge and compute output values for given input data. The major problem of this approach is to find a suitable set of linguistic rules that describe the system to be modeled. Fuzzy systems is represented in the form of if-then rules or fuzzy conditional statements are expression of the form IF A THEN B, where A and B are labels of the fuzzy sets. The set of rules should be complete and provide an answer for every input value.

Fuzzy system consist of three steps as the fuzzification, fuzzy inference and the defuzzification. 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. Fuzzification is the transformation of numerical variables into linguistic variables and the corresponding allocation of the grade of membership (a scalar between 0 and 1) to the diverse membership functions. The linguistic combination of the traits was carried out in the fuzzy inference sytem (FIS). There are two FIS approaches which are Mamdani and Takagi-Sugeno models.

## Fuzzy C- Means

**Fuzzy c-means** (FCM) clustering algorithm is often used as an initial step for fuzzy system to find membership values of each training data vector in each cluster. These membership values are assumed to represent best partitions of given dataset. Formally, clustering an unlabeled data X = {x1, x2, . . . , xN} ⊂ Rh, where N represents the number of data vectors and h the dimension of each data vector, is the assignment of c partition labels to the vectors in X. c-partition of X constitutes sets of (cN){uik} membership values that can be conveniently arranged as a (c × N) matrix U = [uik]. The problem of fuzzy clustering is to find the optimum membership matrix U. The most widely used objective function for fuzzy clustering is the weighted within-groups sum of squared errors Jm, which is used to define the following constrained optimization problem.

$$\min\{j_m(U,V,X) = \sum_{k=1}^{N}\sum_{i=1}^{c}(u_{ik})^m \|x_k - v_i\|_A^2$$

where

$$U \in M_{fcn} = \left\{ U \in \Re^{cN} \left| \begin{array}{l} 0 \leqslant u_{ik} \leqslant 1 \ \forall ik \ \& \ \forall k, u_{ik} > 0 \exists i \\ 0 < \sum_{k=1}^{N} u_{ik} > \eta \ \forall i \ \& \ \sum_{i=1}^{c} u_{ik} = 1 \ \forall k \end{array} \right. \right\}$$

Different statistical classification algorithms can also use to solve bioinformatics problems such as K- Nearest Neighbors, Naïve Bayes and Support Vector Machines.

**K-Nearest Neighbor** (K-NN) is an simple non parametric algorithm which is a method for classifying cases based on their similarity to other cases. Similar cases are near each other and dissimilar cases are distant from each other. Thus, the distance between two cases is a measure of their dissimilarity. Training a nearest neighbor model involves computing the distances between cases based upon their values in the feature set. The nearest neighbors to a given case have the smallest distances from that case. The distance is calculated using one of the following measures:

- Euclidean Distance
- Minkowski Distance
- Mahalanobis Distance

Simple K-NN algorithm consists of following steps:

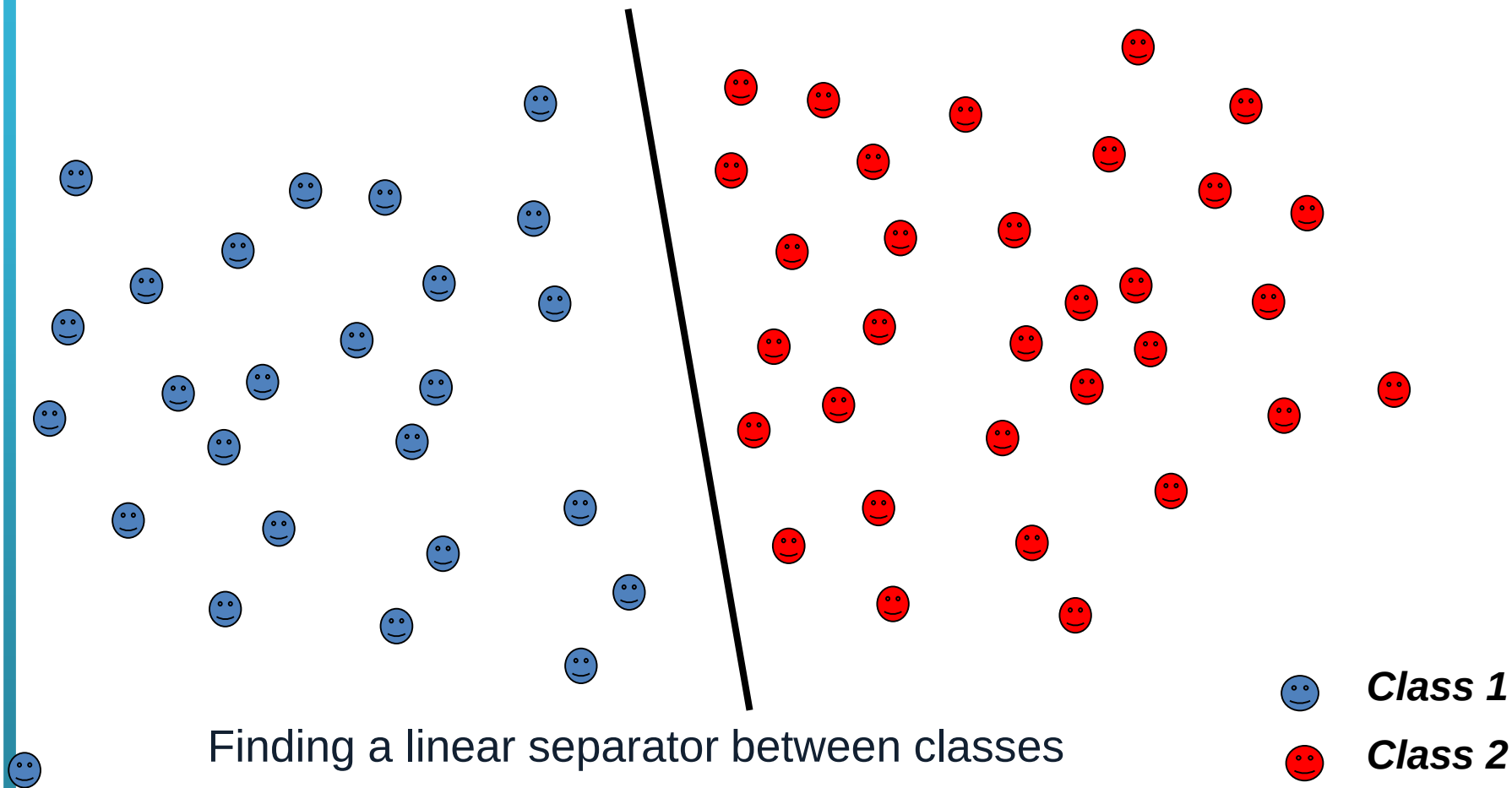
- For each training example <x,f(x)>, add the example to the list of training_examples,
- Given a query instance xq ¨ Given a query instance x to be classified, q to be classified, Let x1, x2….xk denote the k instances from training_examples that are nearest to xq. Then, return the class that represents the maximum of the k instances.

**A Naïve Bayes** classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be independent feature model. Depending on the precise nature of the probability model, Naïve Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for Naïve Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods. In spite of their naive design and apparently over-simplified assumptions, Naïve Bayes classifiers often work much better in many complex real-world situations than one might expect[36]. An advantage of Naïve Bayes classifier is that it requires a small amount of training data to estimate 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 entire covariance matrix. Naïve Bayes classifier combines this model with a decision rule. The corresponding classifier is the function classify defined as follows:

$$classify(f_1, \ldots, f_n) = \arg\max_c \; p(C = c) \prod_{i=1}^{n} p(F_i = f_i \mid C = c)$$

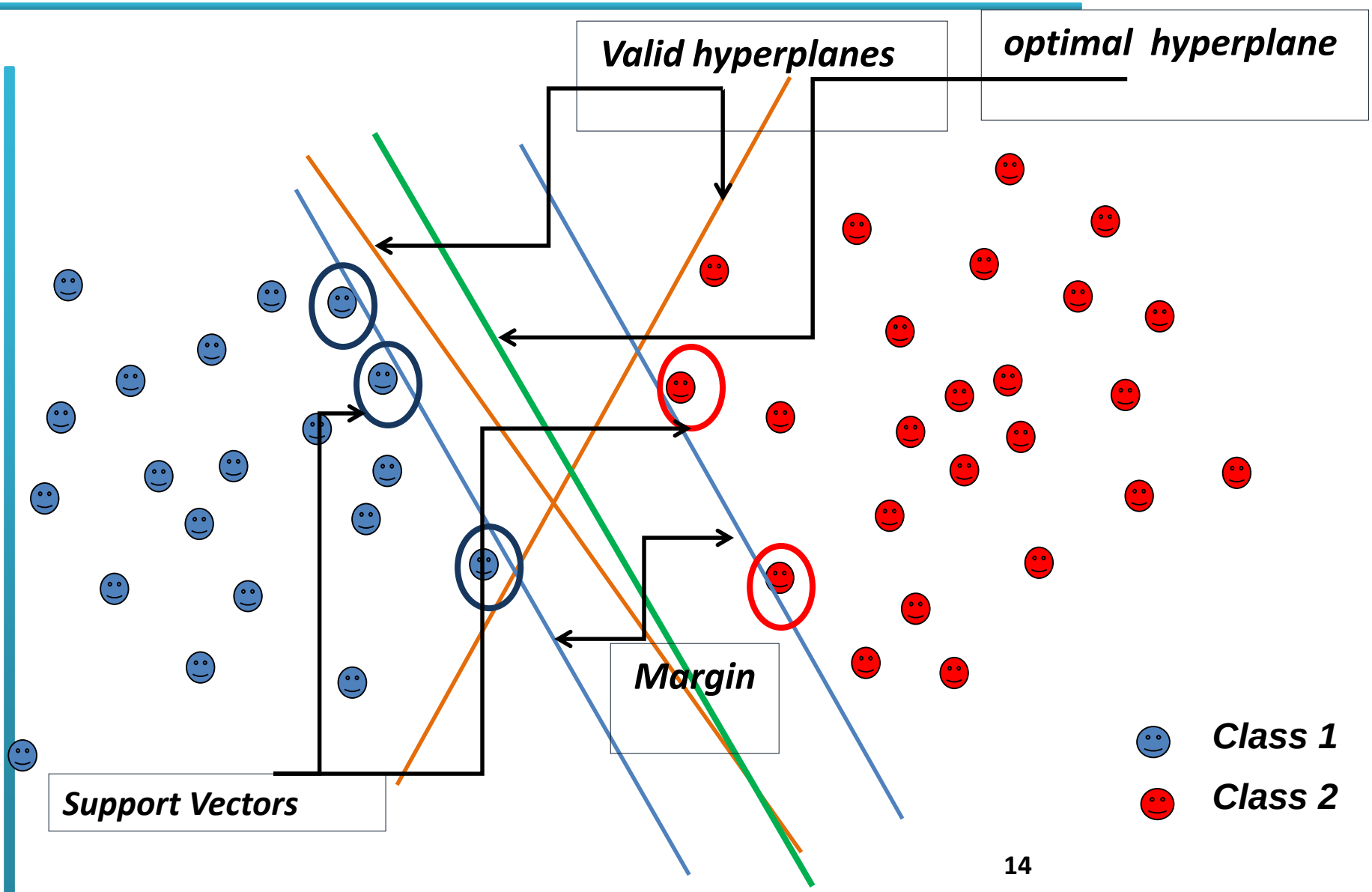
**Support Vector Machines** (SVM) is specifically formulated to solve a binary classification problem in a supervised manner and the learning problem is formulated as a quadratic optimization problem where the error surface is free of any local minimum and has global optimum. SVM is to build an optimal separating hyper plane in such a way that the margin of separation between two classes is maximized. The machine achieves this desirable property on the basis of the principle of structural risk minimization principle. To develop the SVM based classifiers for linearly separable patterns, let us consider a training set represented by $\{(x_i,y_i)\}$ $(i=1,\ldots, N)$, where $x_i$ is the n-dimensional input feature vector and $y_i$ represents the target output. The input patterns represented by the target output $y_i = 1$ constitute the positive group and the target output $y_i = -1$ constitute the negative group.

The machine is assumed to be deterministic: for a given input x, and choice of α, it will always give the same output $f(x; \alpha)$. A particular choice of α generates what we will call a "trained machine." Thus, for example, a neural network with fixed architecture, with α corresponding to the weights and biases, is a learning machine in this sense.

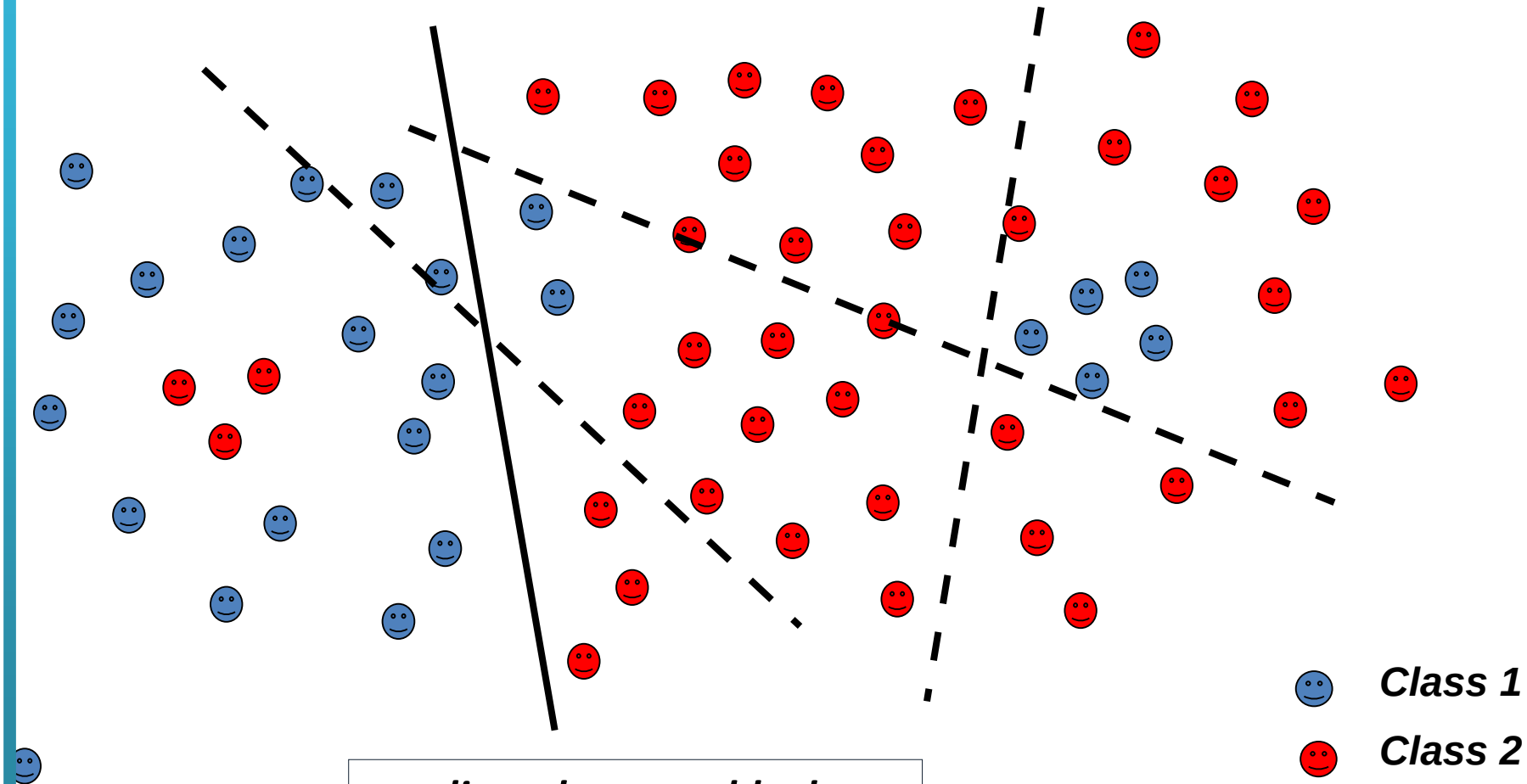



Finding a linear separator between classes

# Support Vector Machines



**Transforms data points to a higher dimensional space where the problem is linearly separable using *kernel functions*.**




***non linearly separable data***

## Probabilistic Model Algorithms

**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.

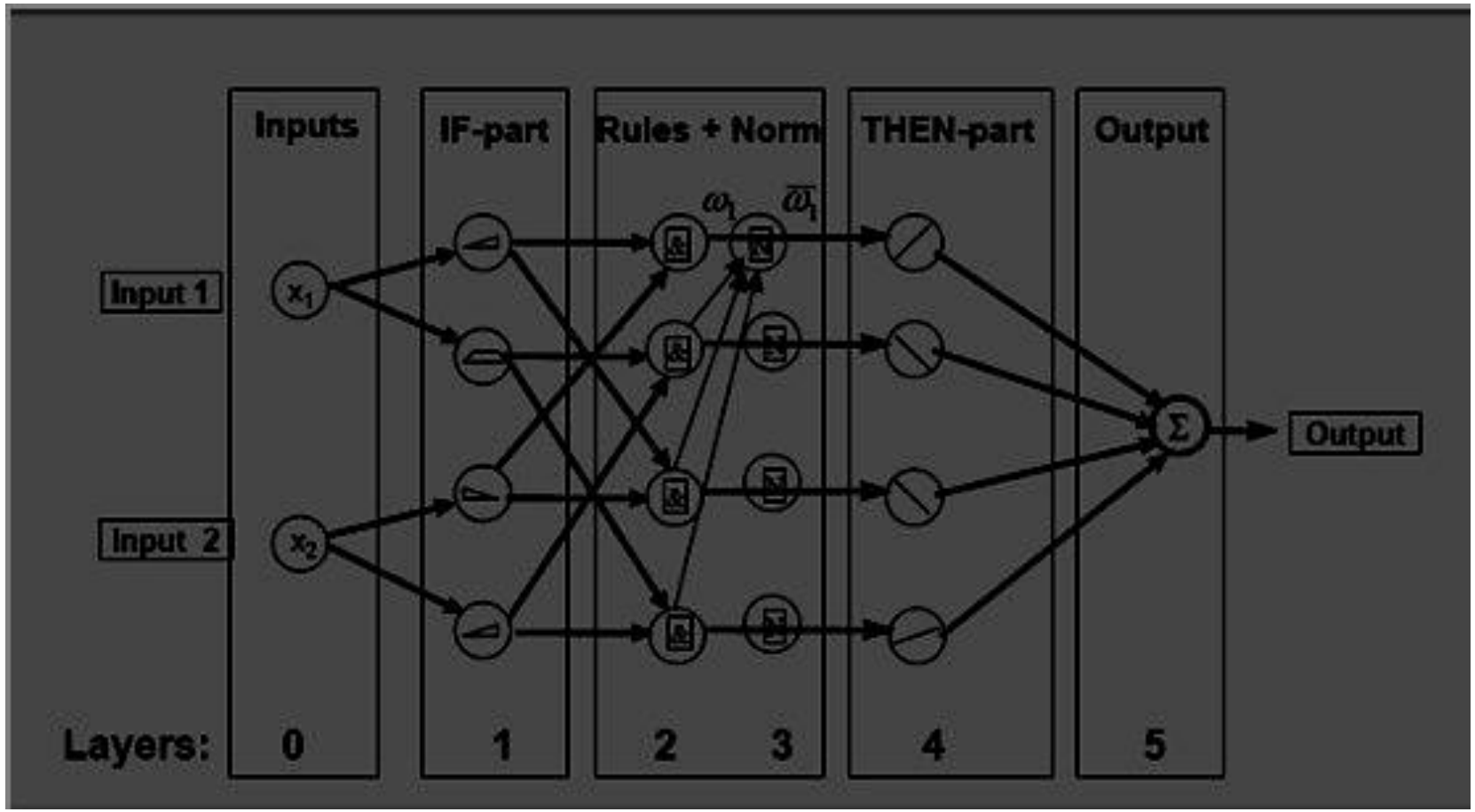
**A Gaussian Mixture Model** (GMM) is a parametric probability density function represented as a weighted sum of Gaussian that has been used. GMM not only provides a smooth overall distribution fit, its components can, if required, clearly detail a multimodal density. GMM parameters are predicted from 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. 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.

## Hybrid Algorithms

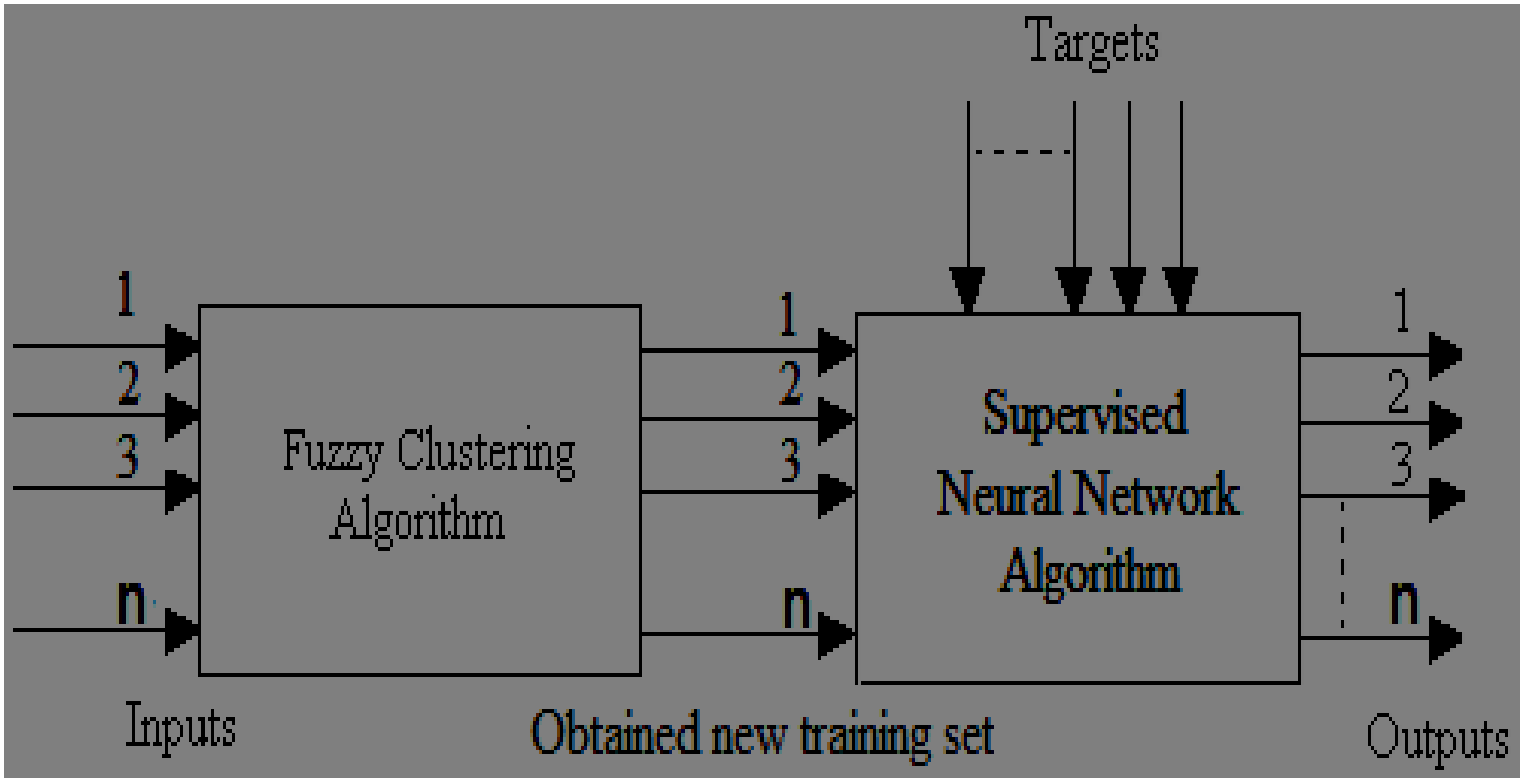
Some hybrid classifier algorithms such as **Adaptive Neuro-Fuzzy Inference System** (ANFIS), Fuzzy Clustering Neural Network (FCNN) are also used to solve pattern recognition problems.

ANFIS is integration both Fuzzy system and artificial neural network. Algorithm was defined by Jang in 1992. It creates a fuzzy decision tree to classify the data into one of 2n (or pn) 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. Next slayt shows the architecture of ANFIS.




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 Karlık. 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 of these data. The architecture of FCNN 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.

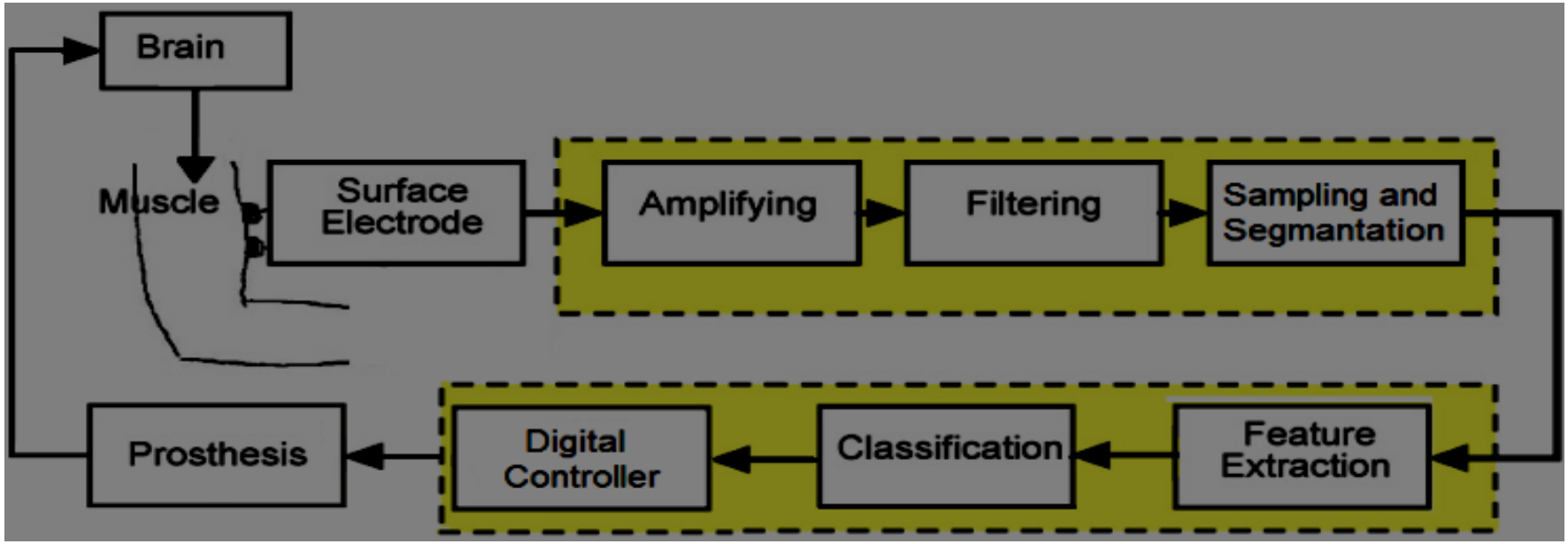



The Architecture of FCNN

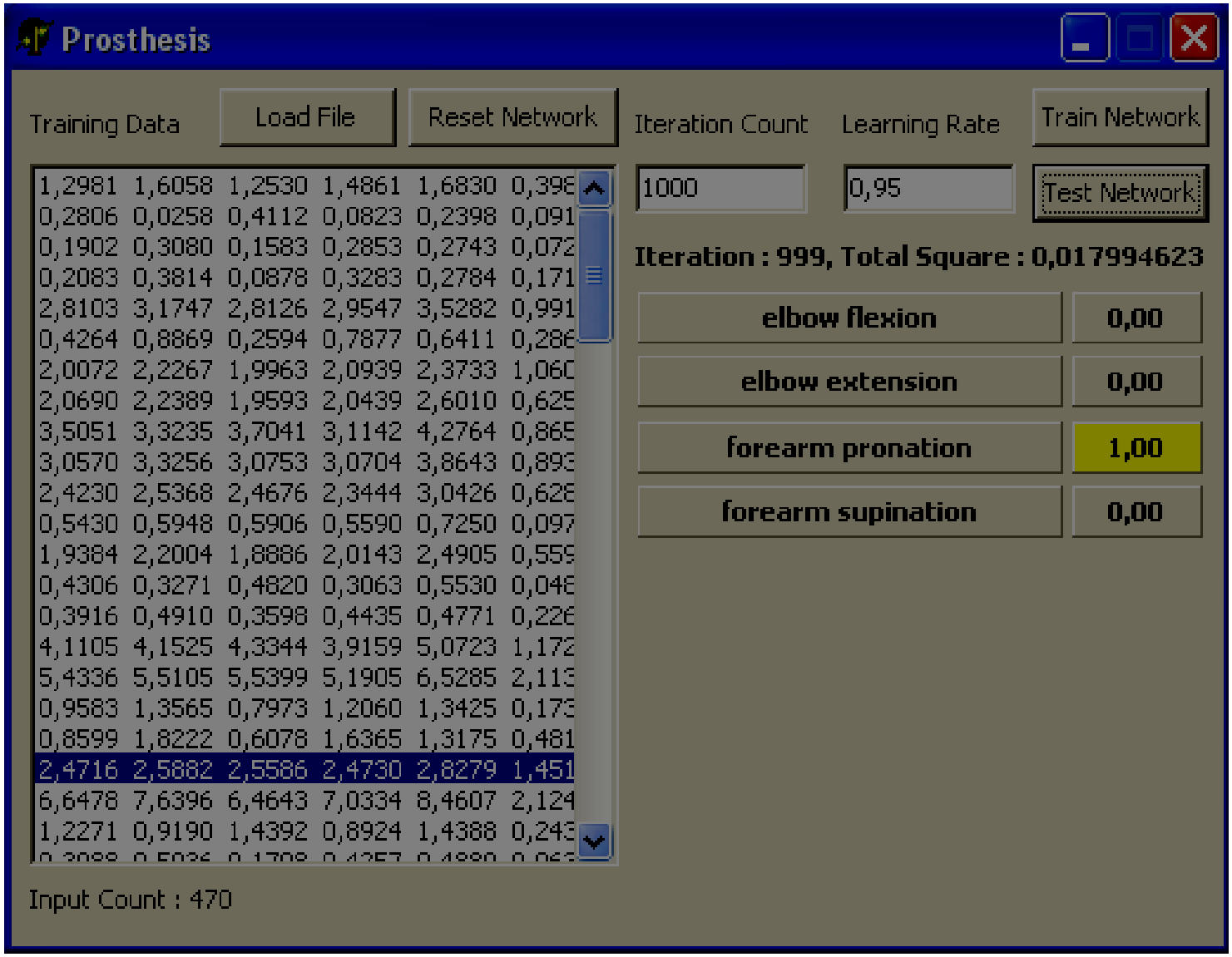
1. Application:

# Machine Learning Algorithms for Characterization of EMG Signals

Figure shows that the block diagram of myoelectric control of human arm prosthesis. Surface EMG signals are recorded by standard Ag/AgCl 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, performed sampling and segmentation.

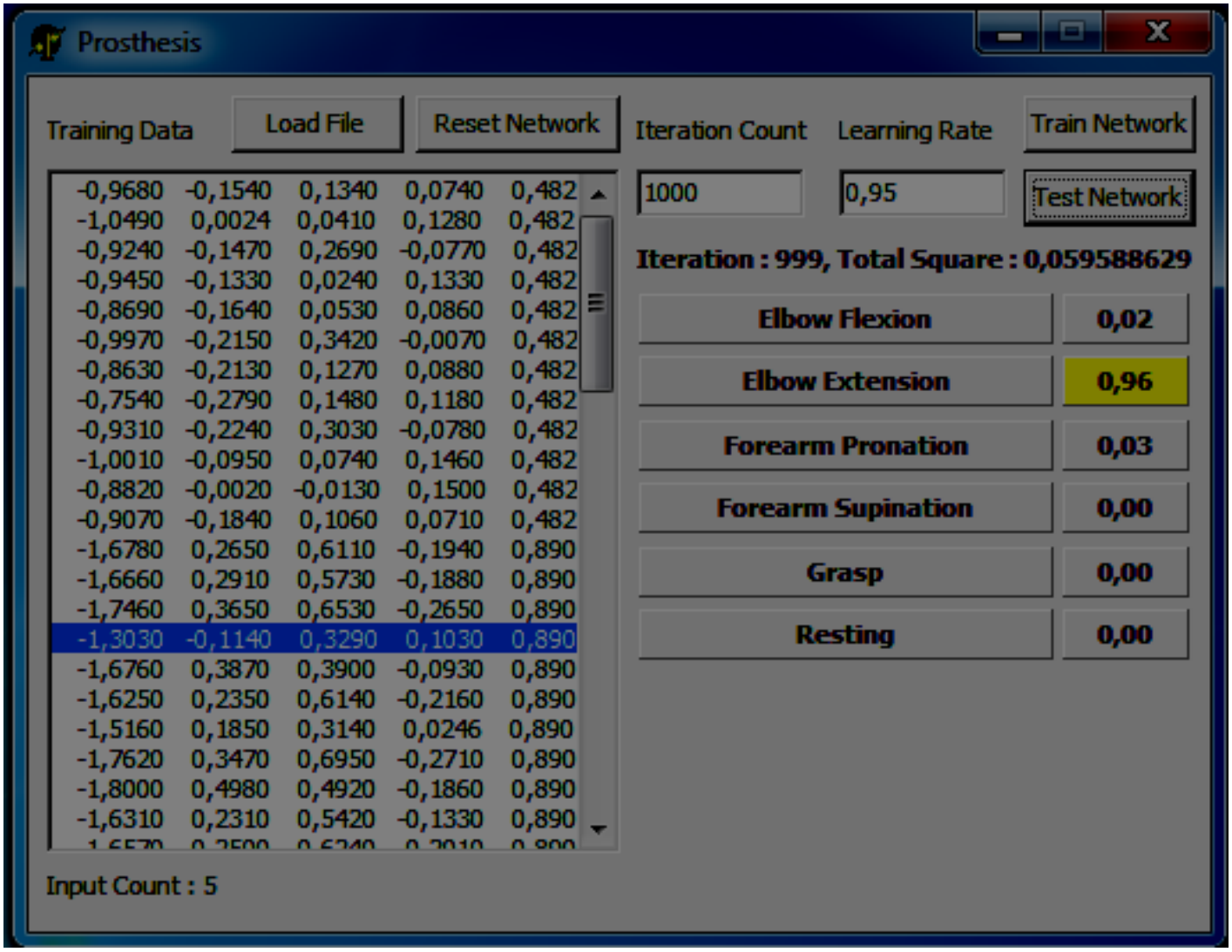


# Application Used Wavelet




| | | | | | |
|---|---|---|---|---|---|
| 1,2981 | 1,6058 | 1,2530 | 1,4861 | 1,6830 | 0,398 |
| 0,2806 | 0,0258 | 0,4112 | 0,0823 | 0,2398 | 0,091 |
| 0,1902 | 0,3080 | 0,1583 | 0,2853 | 0,2743 | 0,072 |
| 0,2083 | 0,3814 | 0,0878 | 0,3283 | 0,2784 | 0,171 |
| 2,8103 | 3,1747 | 2,8126 | 2,9547 | 3,5282 | 0,991 |
| 0,4264 | 0,8869 | 0,2594 | 0,7877 | 0,6411 | 0,286 |
| 2,0072 | 2,2267 | 1,9963 | 2,0939 | 2,3733 | 1,060 |
| 2,0690 | 2,2389 | 1,9593 | 2,0439 | 2,6010 | 0,625 |
| 3,5051 | 3,3235 | 3,7041 | 3,1142 | 4,2764 | 0,865 |
| 3,0570 | 3,3256 | 3,0753 | 3,0704 | 3,8643 | 0,893 |
| 2,4230 | 2,5368 | 2,4676 | 2,3444 | 3,0426 | 0,628 |
| 0,5430 | 0,5948 | 0,5906 | 0,5590 | 0,7250 | 0,097 |
| 1,9384 | 2,2004 | 1,8886 | 2,0143 | 2,4905 | 0,559 |
| 0,4306 | 0,3271 | 0,4820 | 0,3063 | 0,5530 | 0,048 |
| 0,3916 | 0,4910 | 0,3598 | 0,4435 | 0,4771 | 0,226 |
| 4,1105 | 4,1525 | 4,3344 | 3,9159 | 5,0723 | 1,172 |
| 5,4336 | 5,5105 | 5,5399 | 5,1905 | 6,5285 | 2,113 |
| 0,9583 | 1,3565 | 0,7973 | 1,2060 | 1,3425 | 0,173 |
| 0,8599 | 1,8222 | 0,6078 | 1,6365 | 1,3175 | 0,481 |
| 2,4716 | 2,5882 | 2,5586 | 2,4730 | 2,8279 | 1,451 |
| 6,6478 | 7,6396 | 6,4643 | 7,0334 | 8,4607 | 2,124 |
| 1,2271 | 0,9190 | 1,4392 | 0,8924 | 1,4388 | 0,243 |

# Application Used AR Model




| | | | | |
|---|---|---|---|---|
| -0,9680 | -0,1540 | 0,1340 | 0,0740 | 0,482 |
| -1,0490 | 0,0024 | 0,0410 | 0,1280 | 0,482 |
| -0,9240 | -0,1470 | 0,2690 | -0,0770 | 0,482 |
| -0,9450 | -0,1330 | 0,0240 | 0,1330 | 0,482 |
| -0,8690 | -0,1640 | 0,0530 | 0,0860 | 0,482 |
| -0,9970 | -0,2150 | 0,3420 | -0,0070 | 0,482 |
| -0,8630 | -0,2130 | 0,1270 | 0,0880 | 0,482 |
| -0,7540 | -0,2790 | 0,1480 | 0,1180 | 0,482 |
| -0,9310 | -0,2240 | 0,3030 | -0,0780 | 0,482 |
| -1,0010 | -0,0950 | 0,0740 | 0,1460 | 0,482 |
| -0,8820 | -0,0020 | -0,0130 | 0,1500 | 0,482 |
| -0,9070 | -0,1840 | 0,1060 | 0,0710 | 0,482 |
| -1,6780 | 0,2650 | 0,6110 | -0,1940 | 0,890 |
| -1,6660 | 0,2910 | 0,5730 | -0,1880 | 0,890 |
| -1,7460 | 0,3650 | 0,6530 | -0,2650 | 0,890 |
| -1,3030 | -0,1140 | 0,3290 | 0,1030 | 0,890 |
| -1,6760 | 0,3870 | 0,3900 | -0,0930 | 0,890 |
| -1,6250 | 0,2350 | 0,6140 | -0,2160 | 0,890 |
| -1,5160 | 0,1850 | 0,3140 | 0,0246 | 0,890 |
| -1,7620 | 0,3470 | 0,6950 | -0,2710 | 0,890 |
| -1,8000 | 0,4980 | 0,4920 | -0,1860 | 0,890 |
| -1,6310 | 0,2310 | 0,5420 | -0,1330 | 0,890 |

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 to 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.

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.

## Time-series modeling

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≤n≤N−1. Then from these approximatives the he power spectral density (PSD) estimate is computed. AR model, given by

$$S_k = -\sum_{i=1}^{p} a_i S_{k-i} + e_k$$

where; Sk : denoting the recorder signal ( kth discrete time), ai : being the AR parameters, p : being the order of the AR model, ek : being white noise.

**Table:** AR parameters of elbow extension

| a1 | a2 | a3 | a4 |
|---|---|---|---|
| -2.2914128854E+00 | 1.4157880401E+00 | 1.5643946565E-01 | -2.7576257484E-01 |
| -2.2236665563E+00 | 1.1925184658E+00 | 3.7804077194E-01 | -3.4387288667E-01 |
| -2.5605990742E+00 | 2.1284516167E+00 | -4.7985836909E-01 | -8.5217132090E-02 |
| -2.1855094142E+00 | 1.1668151587E+00 | 3.5497629656E-01 | -3.0421272700E-01 |
| -2.1335049591E+00 | 1.0162656173E+00 | 4.8782226231E-01 | -3.6814453276E-01 |
| -2.3205685494E+00 | 1.3983241604E+00 | 2.4926231425E-01 | -3.2406063327E-01 |
| -2.2736460701E+00 | 1.3090069187E+00 | 3.0759887510E-01 | -3.4078144846E-01 |
| -2.1544811453E+00 | 1.0935288607E+00 | 3.8948655289E-01 | -3.2407480865E-01 |
| -2.2177809049E+00 | 1.1889682963E+00 | 4.0943813260E-01 | -3.7655125553E-01 |
| -2.3595552024E+00 | 1.5141832312E+00 | 1.3451033751E-01 | -2.8448465373E-01 |
| -2.3105310227E+00 | 1.4117335132E+00 | 2.0540127137E-01 | -3.0442982559E-01 |
| -2.0866797895E+00 | 9.1354732122E-01 | 5.1311636760E-01 | -3.3697516577E-01 |

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.

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. 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:** Comparison of MLAs applied time series modeling for characterization of EMG signals

| Author | Method | Features | Class | % Accuracy |
|---|---|---|---|---|
| Graupe & Cline [1] | NNC | ARMA | 4 | 95 |
| Doerschuk et al. [2] | NNC | ARMA | 4 | 95 |
| Karlık et al. [3] | MLP-BP | AR-1,P | 6 | 84 |
| Karlık et al. [3] | MLP-BP | AR-2,P | 6 | 92 |
| Karlık et al. [3] | MLP-BP | AR-3,P | 6 | 95 |
| Karlık [4] | MLP-BP | AR-4,P | 6 | 96 |
| Lamounier et al. [5] | MLP-BP | AR-4 | 4 | 96 |
| Soares et al. [6] | MLP-BP | AR-10 | 4 | 95 |
| Soares et al. [6] | MLP-BP | AR-4 | 4 | 96 |
| **Karlık et al. [7]** | **FCNN** | **AR-4,P** | **6** | **98** |
| Chan&Englehart [8] | HMM | AR-6 | 6 | 95 |
| Nilas et al. [9] | MLP-BP | MA | 8 | 60 |
| Farrell & Weir [10] | LDA | AR-3 | 6 | 90 |
| Huang et al. [11] | GMM | AR-6 | 6 | 97 |
| Al-Assaf [12] | PC | AR-5 | 5 | 95 |
| Hargrove et al. [13] | LDA/MLP | AR-6 | 6 | 97 |
| Khezri & Jahed [14] | ANFIS | AR-4 | 6 | 95 |
| Oskoei & Hu [15] | SVM | AR-6 | 6 | 96 |
| Karlık et al. [16] | FCNN | AR-4 | 4 | 89 |
| Zhou et al. [17] | LDA | AR-6 | 11 | 81 |
| Khokhar et al. [18] | SVM | AR-4 | 19 | 88 |
| Khokhar et al. [18] | SVM | AR-4 | 13 | 96 |

## Wavelet modeling

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. As a generalization of WT, a wavelet packet transform (WPT) allows the ‘‘best’’ adapted analysis of a signal in a timescale domain. 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.

**Table:** Comparison of MLAs applied wavelet transform for characterization of EMG signals

| Author | Method | Features | Class | %Accuracy |
|---|---|---|---|---|
| Englehart et al. [1] | LDA | WPT | 6 | 97 |
| Englehart et al. [2] | MLP-BP | WPT | 6 | 93 |
| Koçyiğit & Korürek [3] | FKNN | WT | 4 | 96 |
| Chu et al. [4] | MLP-BP | WPT | 9 | 97 |
| Arvetti et al. [5] | MLP-BP | WT | 5 | 97 |
| Khezri et al. [6] | ANFIS | WT | 6 | 97 |
| Liu & Luo [7] | LVQ | WPT | 4 | 98 |
| Karlık et al. [8] | MLP | DWT | 4 | 97 |
| Karlık et al. [9] | FCNN | DWT | 4 | 98 |
| Khezri & Jahed [10] | MLP-BP | AR/DWT | 6 | 87 |
| Khezri & Jahed [11] | ANFIS | AR/DWT | 6 | 92 |

## Conclusion

This review article has presented comparison different machine learning algorithms used characterization of EMG signals for myoelectic 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.

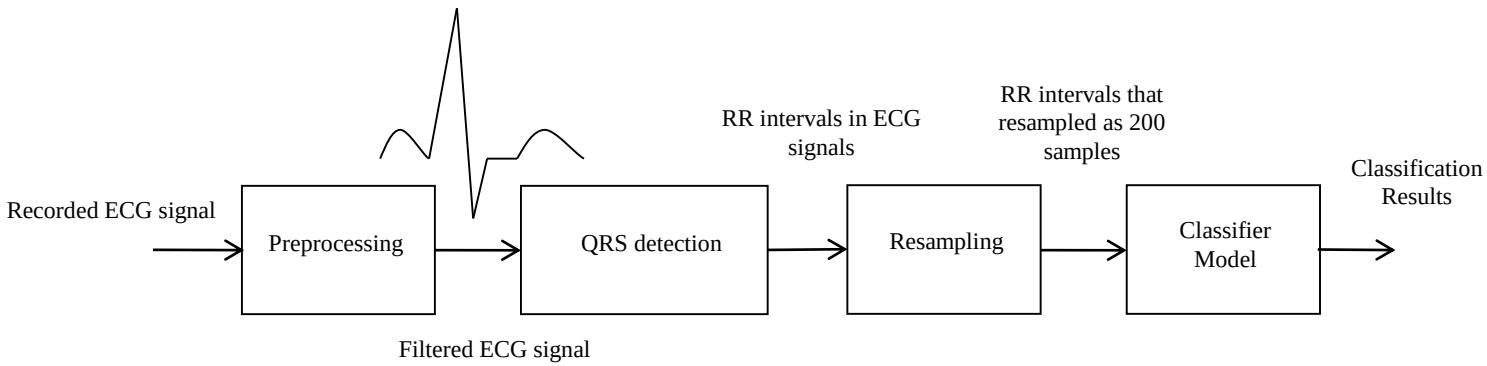
2. Application:

# Machine Learning Algorithms for ECG arrhythmias

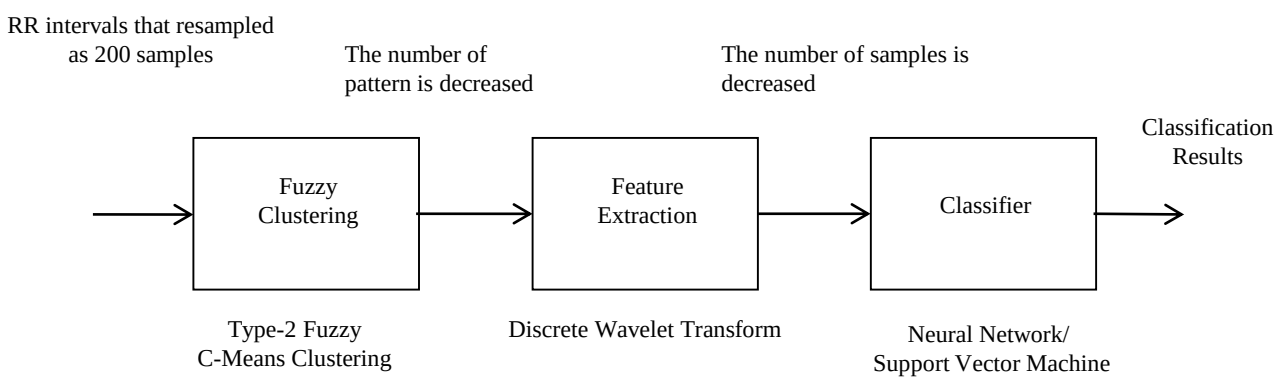
Electrocardiography is a valuable tool and it uses to detect of cardiovascular diseases. As known, electrocardiogram (ECG) demonstrates electrical and physical activity of the heart. On the other hand, ECG signal takes some information about physiology of heart and its activity [1]. The right and fast classification of ECG arrhythmias is considerable process for patients in the intensive care unit [2]. Up to now, several techniques have been used to develop computer-aided diagnostic (CAD) systems for classification of arrhythmias. These techniques have composed from multivariate statistics, decision trees, fuzzy logic, expert systems and hybrid approaches.

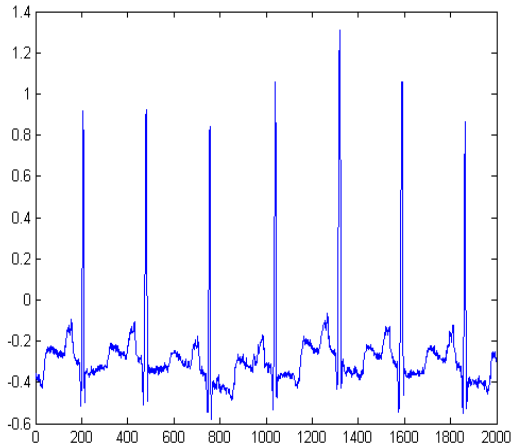
We have utilized electrocardiography arrhythmia signals obtained from MIT-BIH ECG Arrhythmia Database for both training and testing of the proposed models. The first form of these signals is unsuitable for classification models. Hence, QRS detection process on these ECG signal is implemented. Each of extracted RR intervals by QRS detection algorithm is considered as a pattern. In this way, both training and testing sets are composed by mixing different patterns taken from different patients.

## The signal processing flow in the proposed ECG classifier models

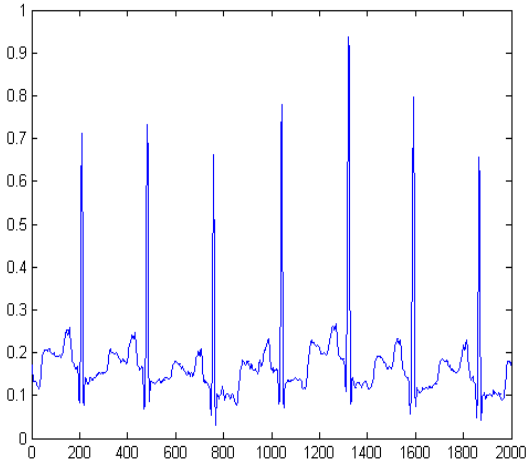



## The block representations of proposed classifier models

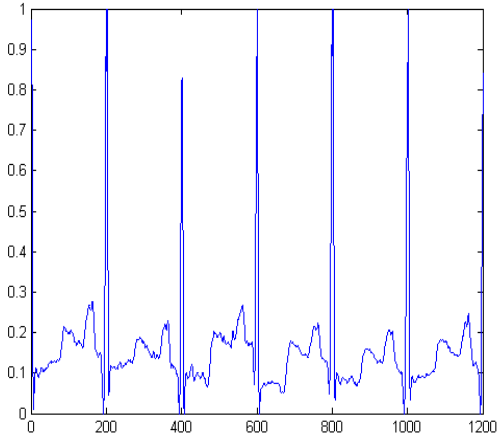




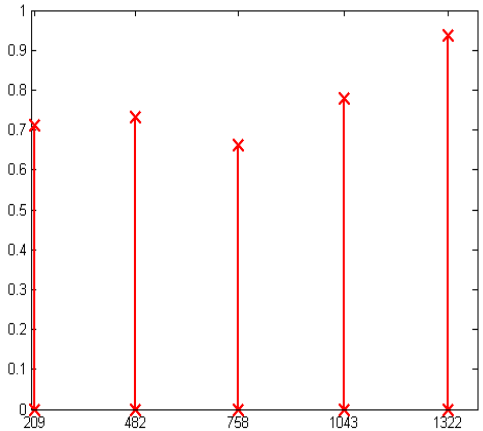

(a) Original signal




(b) Normalized and filtered ECG signal




(c) ECG signal that RR intervals were arranged




(d) The sample point of R beat in (b) signal as 200 samples

Figure: Filtering and QRS detection results belong to normal sinus rhythm

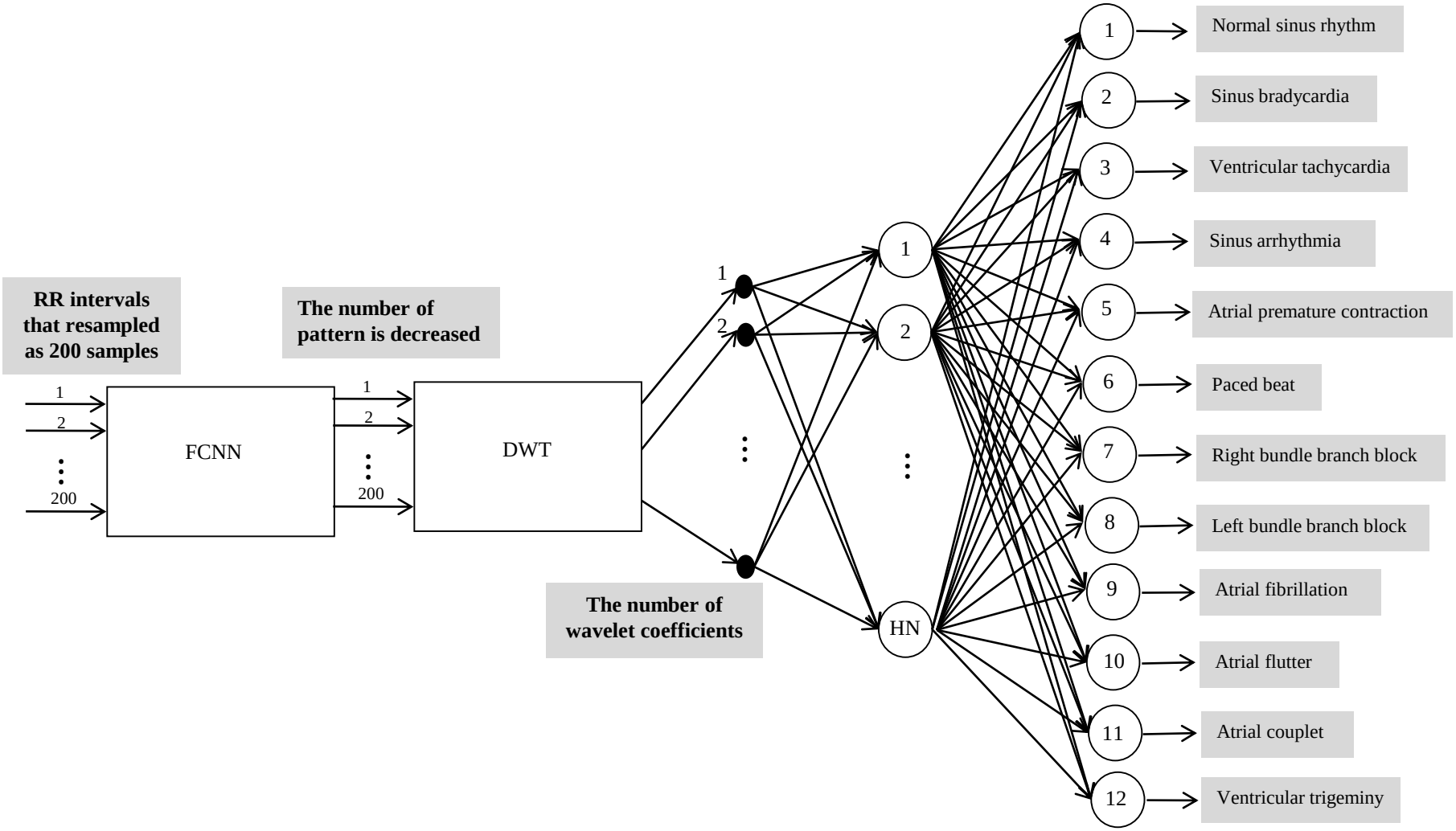



Figure: The hybrid FCWNN structure (HN: The number of hidden nodes)

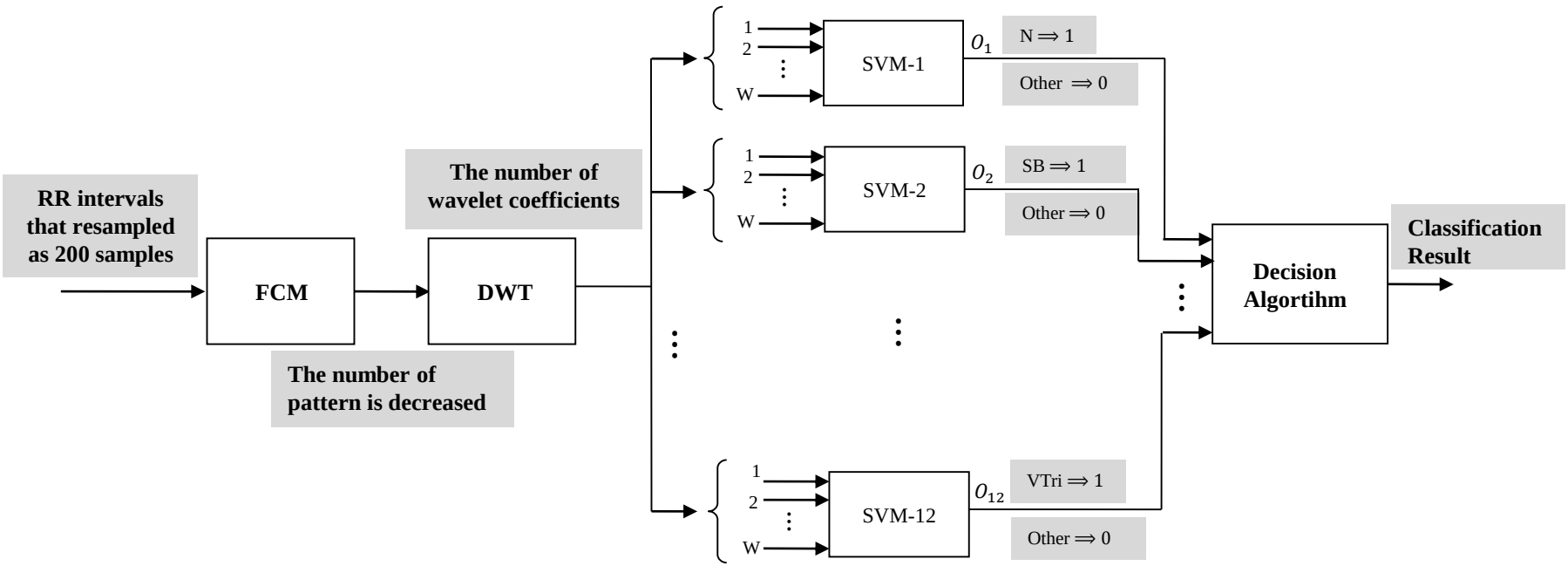



Figure: The FCWSVM structure that designed according to one against all principle (W:The number of wavelet coefficients)

# MLAs for Arrhythmias

**Table:** Comparison FCWNN and FCWSVM structures with other studies in literature

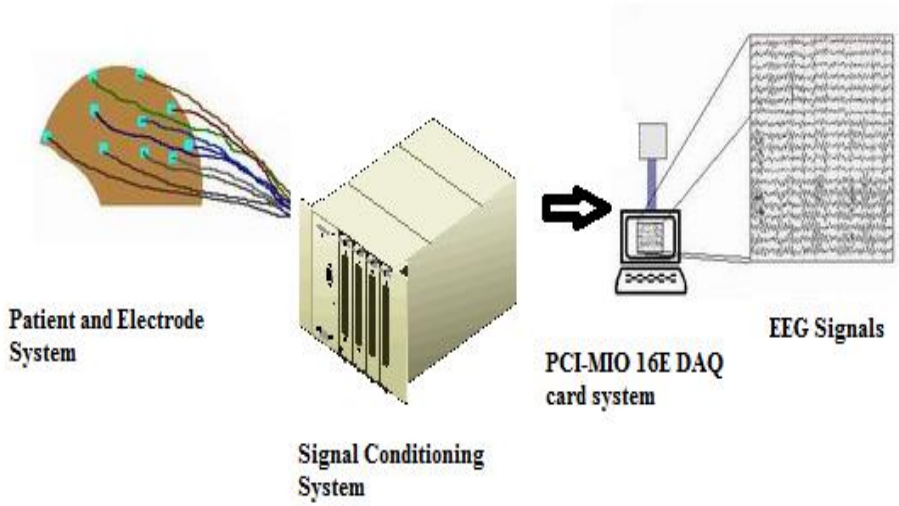
| Author | Feature Extraction | Classifier architecture | The number of classes | Accuracy rate (%) |
|---|---|---|---|---|
| Y.P. Meau et al. [1] | DWT | Extended Kalman Filter based MLP | 5 | 93.28 |
| S.N. Yu, K.T. Chou[2] | ICA | Neural network | 8 | 98.37 |
| S.N. Yu, K.T. Chou[3] | ICA | Neural network | 8 | 98.71 |
| S.N. Yu, Y.H. Chen[4] | DWT | Probabilistic neural network | 6 | 99.48 |
| S. Osowski et al. [5] | - | Weighted voting | 7 | 98.63 |
| H. Hosseini et al.[6] | - | Two-stage ANN classifier | 6 | 90 |
| E.D. Übeyli [7] | Eigen vector | SVM | 4 | 98.3 |
| F. Melgani, Yakoub Bazi [8] | - | PSO and SVM | 6 | 92.3 |
| B. Doğan, M. Korürek [9] | - | Kernelized FCM and hybrid Ant colony. | 6 | 96.26 |
| C.P. Shen et al. [10] | Wavelet-based features | Modified SVM | 12 | 98.92 |
| R. Ceylan, Y. Özbay, B. Karlık [11] | - | FCNN | 10 | 100* |
| Y.Özbay, R.Ceylan, B.Karlik [12] | DWT | FCWNN | 10 | 100* |
| Y.Özbay, R.Ceylan, B.Karlik [13] | - | FCNN | 5 | 100* |
| İ. Güler and E.D.Übeyli[14] | DWT | Combined neural network | | 96.94 |
| Y.Özbay, R.Ceylan, B.Karlik | DWT | FCWNN | 12 | 99.62 |
| Y.Özbay, R.Ceylan, B.Karlik | DWT | FCWSVM | 12 | 98.86 |

(ICA: Independent Component Analysis, DWT: Discrete Wavelet Transform, MLP: Multilayer perceptron)
(*: This accuracy rate was obtained in training. There is not test accuracy rate for all of the classes in this studies)

3. Application:

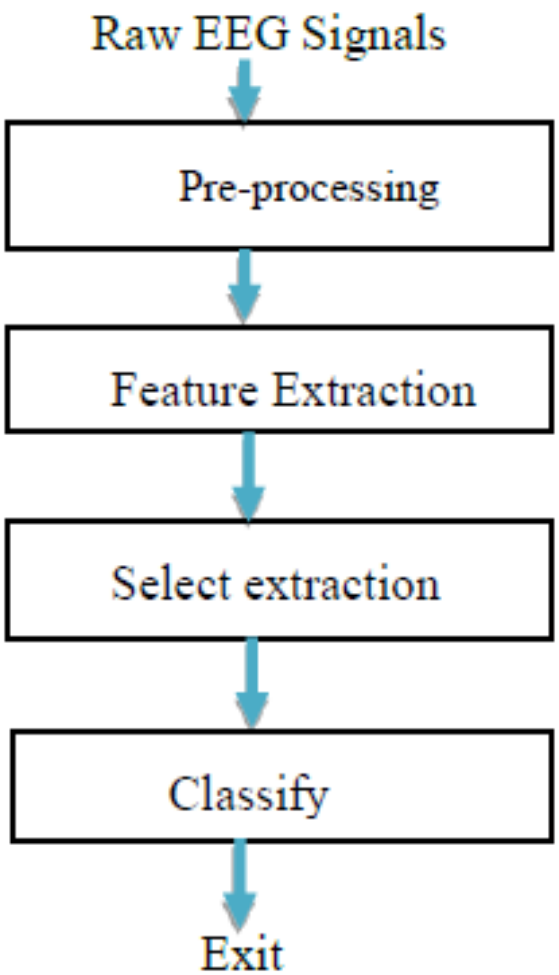
# Machine Learning Algorithms for Recognition of Epileptic Seizures in EEG

The aim of this study is diagnose epileptic seizure by using machine learning algorithms with EEG data.
EEG data are extracted by discrete wavelet transform (DWT) and AR models.
EEG is non-stationary signal.
Data are admitted to neurology department of Medical Faculty Hospital of Dicle University.
400 people who 200 of them are epilepsy and others healthy.
Time frequency methods are DWT and AR method.




The length of the signal which will be modeled level of stationary of the data.

# MLAs for EEG




**1. DWT**

- advantage model for non-stationary signals
- Optimum time-frequency resolution

**2. AR Method**

- Parametric method for spectral analysis
- AR method such as selection of the optimum estimation method
- Selection of the model order
- Wavelet transform vector size is 400x129
- Auto regressive extraction input vector size is 400x15.

**RESULTS:**

- k-fold cross-validation for Artificial Neural Networks (ANN), Naive Bayesian, k-Nearest Neighbor (k-NN), Support Vector Machines (SVM) and k-Means.
- Wavelet transform method is achieved with k-NN,
- k-NN is effective algorithms,
- Wavelet transform is better than the AR method for EEG signals,
- Recognition of the epileptic seizure by using k-NN and ANN are faster and have better accuracy than literature studies,
- k-means algorithm has been observed to give the lowest performing.

**Table:** Training data set accuracy rates of used classifiers for two feature extractions

| Classifiers | Wavelet | AR Model |
|---|---|---|
| ANN | %99.75 | %99.50 |
| SVM | %99.5 | %99.50 |
| Naïve Bayes | %99.5 | %98.00 |
| k-Means | %58.5 | %96.50 |
| k-NN | %100 | %99.75 |

4. Application:

# Machine Learning Algorithms for Classification of Biomedical Sounds

**Biomedical Sounds:**

- Human body organs such as lungs, heart etc. produce different kinds of sounds during their activities.
- The presence and absence of these sounds, or being different from usual may be indicative of variety problems related to the organs or systems .
- For example, wheezing in lung sounds may be sign of asthma disease. Cardiac murmurs may indicate a problem with the heart or circulatory system.
- Therefore, sounds that occur in the body are used to diagnose and treat various medical disorders and are called biomedical sounds.
- Physicians listen to heart, stomach, lungs, blood vessels etc. by placing the stethoscope on the skin of body and then evaluate state of organs by interpreting them. But; The response of stethoscope and external noise .
- The proper diagnosis also requires significant training and experience of the medical personnel.
- Stethoscope may be unreliable in noisy environments such as ambulance, a busy emergency room etc.

**Biomedical Sounds:**

During the last two decades, much research has been carried out on computer-based biomedical sound analysis. These studies generally examined under three main groups.

- Hardware/Equipment weighted studies to record the audio signals and to create a database
- Filtration studies to distinguish sounds from a variety of noise
- The analysis and classification of sound signals

In literature, numerous and different research has been carried out on analysis, processing and classification of biomedical sounds. Various signal processing and machine learning methods are used for these purposes.

**Mostly Used Analysis Methods**

Frequency Analyses Methods: Fourier Transforms, Parametric Methods (AR, ARMA), Time-Frequency Analysis Methods: Wavelet Transforms

**Mostly Used Machine Learning Methods**

For the classification of these sounds, usually machine learning algorithms such as ANNs, k-NN, and SVM are used.

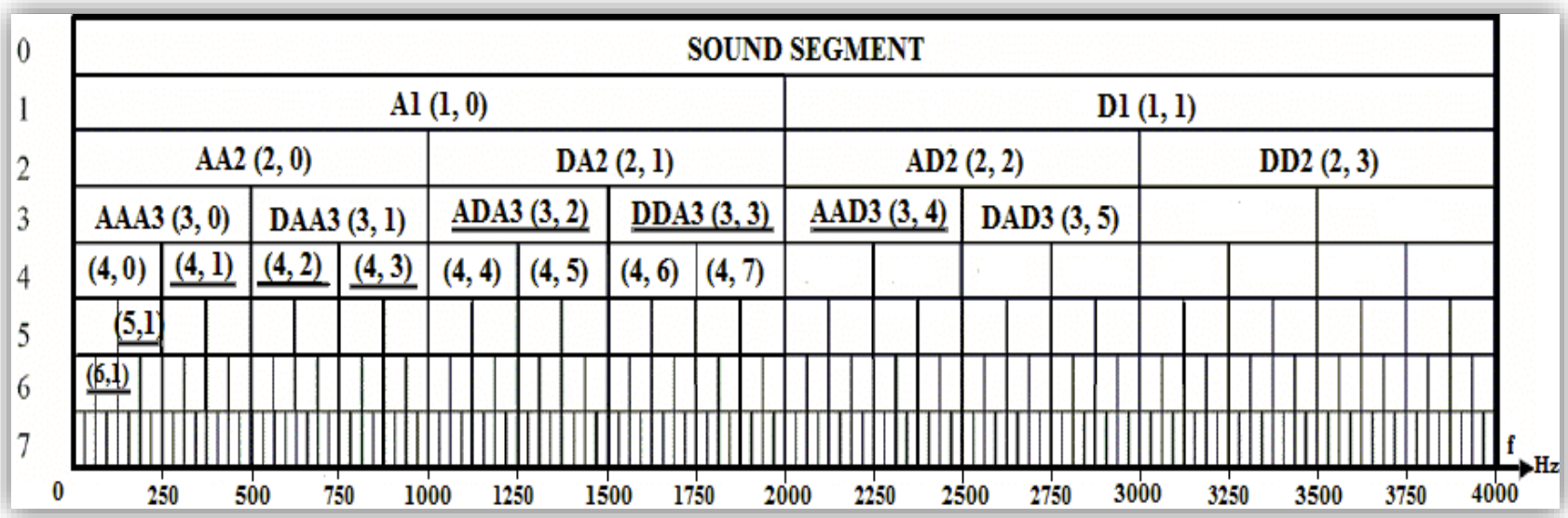
**Data:**

The lung sounds data belongs to 11 people. While 6 people have asthma disease, 5 people haven't any lung disorders. Lung sounds are 8000 Hz sampling frequency.

**Pre-processing of Data:**

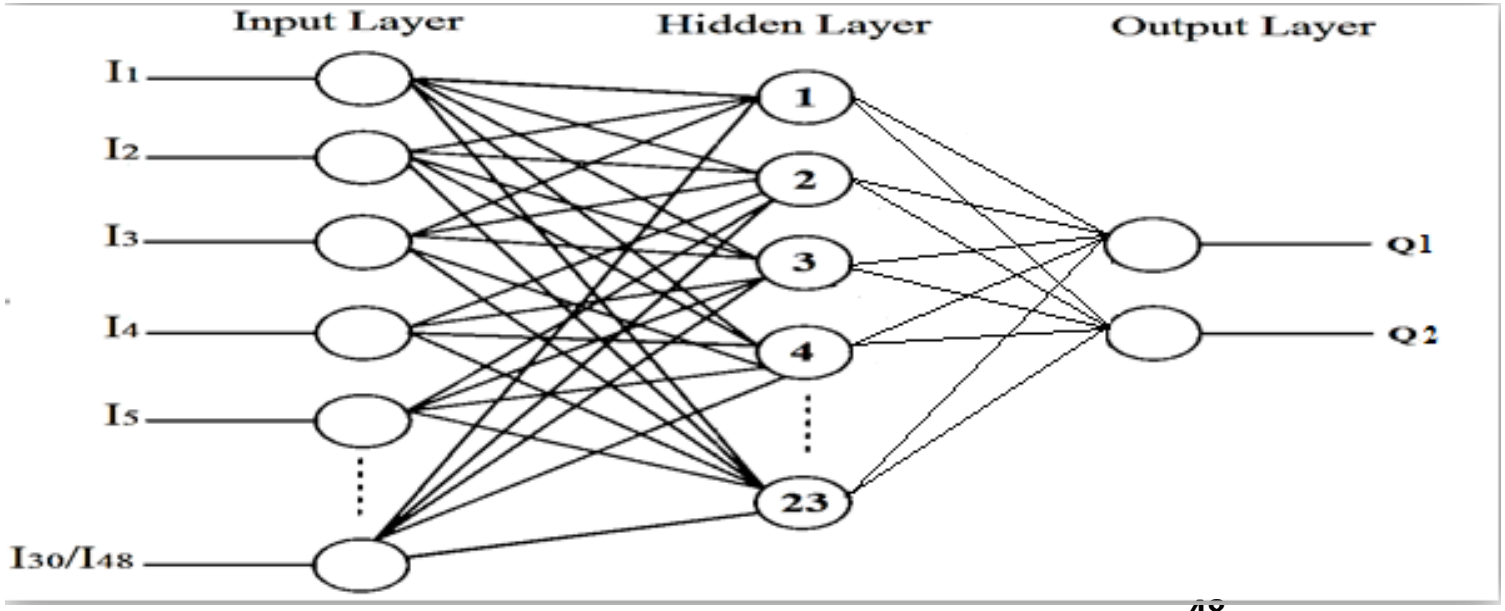
Because of mix of heart, muscle and other sounds , recorded lung sounds are not completely distinguishable. Hence we use filters to minimize these unwanted sounds. Lung sound signals usually lies between 100-2000 Hz.

In wavelet packet transform, both lower and higher frequency bands are decomposed into two sub-bands. Thereby wavelet packet gives a balanced binary tree structure. Sub-bands were selected from the WPT tree to represent sound segment.



In our study, ANN with back propagation algorithm was used to classify lung sounds into two class namely normal and asthma. Statistical features were used as inputs into network.

The numbers of neurons in the input layer are 30 and 48 for feature vectors obtained using DWT and WPT respectively. 23 neurons which gave the best performance is used in hidden layer. The number of neurons in the output layer is 2 for both DWT and WPT.



# Analysis and Classification of Heart Sounds

- **Data:**
  - 9 different types of heart sound were analyzed and classified. Heart sounds are 44100 Hz sampling frequency.
  - Duration of sounds is nearly 14-15 seconds and each of these sounds has 17 heart beat (period).
- **Used Heart Sounds:**
  - Opening snap(OPS)
  - Aortic stenosis (AST)
  - Mid-systolic click + Late systolic murmur (MCC+LSM)
  - Normal heart sounds (First - S1 and Second - S2)
  - Third heart sound (S3)
  - Fourth heart sound (S4)
  - Ventricular septal defect (VSD)
  - Patent ductus arteriosus (PDA)
  - Atrial septal defect (ASD)

# MLAs for Heart Sounds

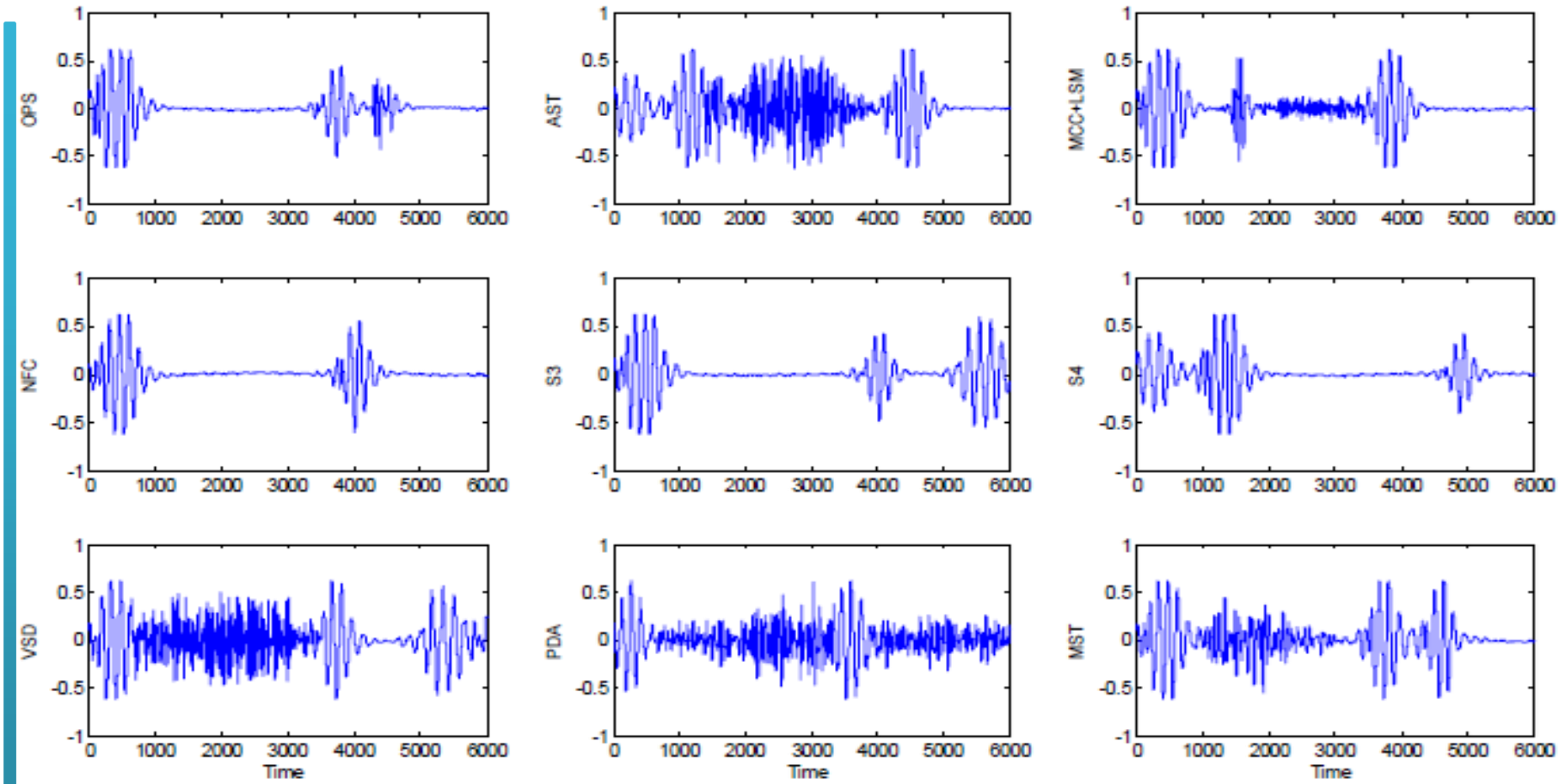



Figure: One period of nine different heart sounds

## Pre-processing of Data:

- Heart sound signals is usually found in the 20-600 Hz frequency band.
- Normal heart sounds are found in the frequency range of 20-200Hz. Heart murmurs are usually scattered throughout the 30-600 / 700Hz frequency range.
- Heart sounds were filtered using a band-pass filter Butterworth, with 3rd order and cut-off frequencies set at 20Hz and 600Hz.
- After filtration, the heart sounds are segmented into a small duration of a complete one(1) cycle of heart beat.

## Feature Extraction of Heart Sounds

- Fast Fourier Transform (FFT) based Welch, Autoregressive (AR)-Burg and Autoregressive Moving Average (ARMA) methods were used to compute power spectrum densities of heart sounds.
- The power spectrum of signals gives the distribution of the signal power among various frequencies.
- Power spectrum densities were considered feature vectors of heart sounds data.
- However, the resulting feature vector size (size of the power spectral density) is 513. The size is too large and must be reduced for successful classification.
- Principal component analysis (PCA) and linear discriminant analysis (LDA) are performed and used to reduce size of the feature vectors.

**Classification of Heart Sounds:**

- Support Vector Machines (SVM) and k-nearest neighbor (k-NN) classifier was used to classify heart sounds into nine class.
- Distance metric was selected Euclidean and k value was selected 3 for K-NN.
- Kernel function of SVM is determined as linear function.
- Classification process was carried out with 15 cross validation.

| Feature Extraction Methods | Classification Methods | |
|---|---|---|
| | K-NN (CC%) | SVM (CC%) |
| Welch-PCA(10) | 98,00 | 98,67 |
| Welch-LDA(8) | 99,33 | 99,33 |
| Burg-PCA(8) | 98,67 | 99,33 |
| Burg-LDA(8) | 100 | 100 |
| ARMA- PCA(9) | 88,06 | 93,39 |
| ARMA-LDA(8) | 92,06 | 95,39 |

CC: Classification Accuracy

Thank you for your attention!
Questions & Answers

- For further information:

Prof. Dr. Bekir KARLIK

Selcuk University, Konya, Turkey

bkarlik@selcuk.edu.tr