# 4. Classification of physiological signals

## 1. Introduction :

The diagnosis of various diseases on the basis of physiological signals is based on a set of data that can be classified into two main families:

- The first family includes data on the patient and his or her medical record, in particular age and sex.
- The second family represents data derived from physiological signals and is known as "features". These features include statistical, temporal and/or frequency attributes.
- Artificial intelligence is a fundamental branch of computer science whose aim is to simulate the behaviour of the human brain.

There are two main types of neural network.

a. Feed-back networks
In this type of network, each cell is linked to all the others and even has feedback on itself. Signals can be propagated forwards or backwards.

b. Feed-ForWord networks
We also call it a layered network, because the output of any layer cannot be propagated back to previous layers.

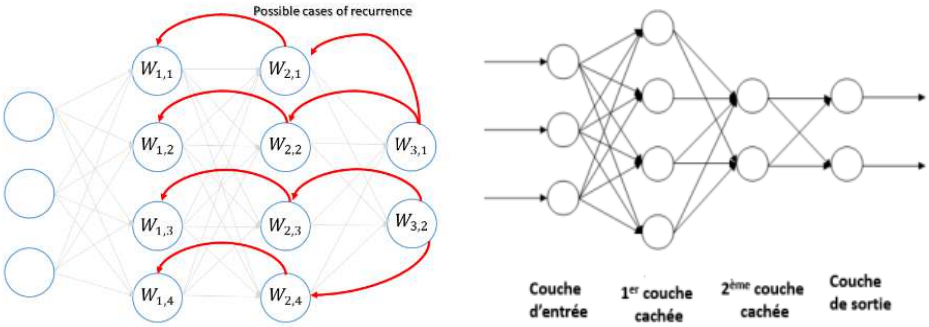



Figure 1: (a) Feed-back networks (b) Feed-ForWord networks

Most learning algorithms for formal neural networks are optimisation algorithms: they use non-linear optimisation methods to minimise a cost function which is a measure of the difference between the actual responses of the network and its desired responses. This optimisation is carried out iteratively, by modifying the weights according to the cost function.

It can be divided into three families:

1. Supervised learning
Supervised learning of a neural network involves modifying the weights as long as the response corresponding to each input is not close enough to the desired response.
So we have the right answer and we use it to measure the error rate in the network response, then we modify the connections according to this error.

2. Unsupervised learning

   Unsupervised learning of neural networks consists, as in the case of supervised learning, of modifying the weights of the neural connections. In this case, however, the examples in the training database are data alone, and it is not possible to modify the network weights according to an error in the desired responses, since no response is known a priori.

3. Reinforcement learning

   The reinforcement learning algorithm is qualitative, as opposed to supervised or unsupervised learning. It is not necessary to have the desired outputs for the training examples, only an "overall" assessment of the network's behaviour for each of the examples processed. This algorithm can be applied to any network structure.

   The only requirement is to have stochastic (binary) output neurons. The response of the neural network is thus a function of the inputs and also of the output neurons. This introduces a degree of randomness into the system's behaviour. If the response provided by the system is considered to be good, the algorithm tends to favour the appearance of this behaviour by reducing the randomness. If the overall system response is considered to be poor, the aim is to avoid the subsequent appearance of this behaviour.

In both cases (supervised and unsupervised learning), the only way to improve the system's responses is to adjust the weights of the connections.

The learning of the multi-layer perceptron is supervised, and consists of adapting the weights of the neurons so that the network is capable of performing a given transformation.

The aim of classification is to identify the classes to which objects belong, based on descriptive traits (attributes, characteristics, etc.).

The attributes presented in the matrix (base2) are used to build the PMC (multi-layer perceptron) that we have chosen as the classifier. It operates in two phases:

The learning phase: during this phase, the PMC learns the different configurations of the vectors of characteristic parameters of the biomedical signals it wishes to classify:

- Propagation, which consists of presenting an input configuration to the network, then propagating it to the output configuration via the hidden layers.
- After the propagation process, back-propagation consists of minimising the quadratic error between the calculated outputs and the desired outputs. The error is considered to be a function of the synaptic weights.

The test phase: known as the generation phase, which consists of presenting an unknown vector of parameters at the input of the network and observing the response it provides.

In this phase, we test the network, and it decides whether the signal presented is for a normal case or presents a pathology.

$$recognition\ rate = \frac{number\ of\ cases\ recognised}{total\ number\ of\ cases}$$

*Allocation of results*: each signal is associated with a Y value which takes either 1 for the presence of a pathology or 0 otherwise.

To test our algorithm in a relevant way, we judged the quality of the result obtained compared to the estimates often used in the literature.

*Sensibility*: it corresponds to the True Positive Fraction.

$$Sensibility = \frac{TP}{TP + FN}$$

The specificity: it corresponds to the False Positive Fraction.

$$specificiy = \frac{TN}{TN + FP}$$

With

TP (true positive: VP): The lesion is correctly detected by the test.

FP (false positive: FP): When the test result is declared positive but it was not.

FN (false negative: FN): When the test result is declared negative but in fact it is positive.

TN (true negative:VN): The lesion is correctly declared as negative.

| | Pathologic cases | Healthy Case | Quantifier |
|---|---|---|---|
| Positive Test | $VP$ | $FP$ | $Vpp = \frac{VP}{VP + FP}$ |
| Negative Test | $FN$ | $VN$ | $Vpn = \frac{VN}{VN + FN}$ |
| Quantifier | $Se = \frac{VP}{VP + FN}$ | $Sp = \frac{VN}{VN + FP}$ | |

Table 1: Parameters for judging the quality of the results obtained

$Vpp$ is the positive predictive value, if the test is positive, the probability that the patient is actually sick.

$Vpn$ is the negative predictive value, if the test is negative, the probability that the patient is actually healthy.

In this PW, in order to classify the attributes (featurs) presented in the 'base' matrix, we will use two algorithms: K-Nearest Neighbor (KNN), and Support Vector Machines (SVM).

1. **K-Nearest Neighbor (KNN)**

The purpose of the K nearest neighbors (KNN) method is to classify target points (unknown class) according to their distances with respect to points constituting a learning sample (i.e. whose class is known a priori).

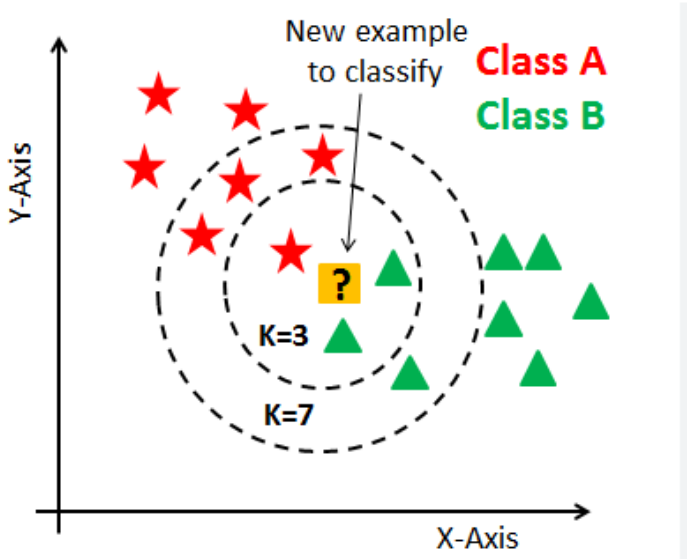



**Figure 2:** Operating principle of the KNN classifier.

2. **Support Vector Machines (SVM)**

Support vector machines or wide-margin separators are a set of supervised learning techniques intended to solve discrimination and regression problems at a hyperplane (feature space) in which the data is separated into several classes whose boundary is farthest away from the data points (or "maximum margin").

SVMs are a generalization of linear classifiers.

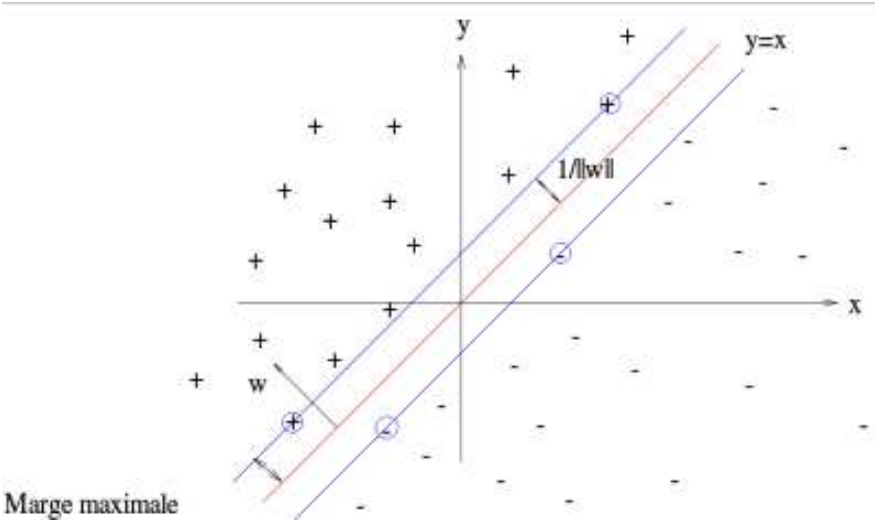



**Figure 3**: Working principle of the SVM classifier.

**Application :**

1- Develop the algorithm of the decision tree classifier (AD).

2- Apply the AD classifier on the database 'base1'.

3- Compare the results obtained using the three classifiers KNN, SVM, and AD