Hidden Neural Network for Complex Pattern Recognition: A Comparison Study with Multi-Neural Network Based Approach

Lilia Lazli[1*], Mounir Boukadoum[2]

[1]LRI/ERIA Laboratory, Badji Mokhar University, Algeria

[2]University of Quebec at Montreal (UQAM), Canada

[*1]l_lazli@yahoo.fr, [2]Boukadoum-mounir@uqam.ca

Abstract- **When the feature space undergoes changes, owing to different operating and environmental conditions, multi-aspect classification is almost a necessity in order to maintain the performance of the pattern recognition system and improve robustness and reliability in decision making. This is an important issue being investigated in ANN research, in many cases, the problems can be solved more effectively by combining one or two other techniques rather than implementing ANN exclusively. New learning methods, especially multiple classifier systems, are now actively studied and applied in pattern recognition. So, the main goal of this paper is to propose two hybrid models and compare your performance in complex pattern recognition problem: speech recognition and biomedical diagnosis.**

This paper compare, the performance obtained with (1) Multi-network RBF/LVQ structure, we use involves Learning Vector Quantization (LVQ) as a competitive decision processor and Radial Basis Function (RBF) as a classifier. (2) Hybrid HMM/MLP model using a Multi Layer-Perceptron (MLP) to estimate the Hidden Markov Models (HMM) emission probabilities. For pre-classification, the k-means clustering algorithm is proposed to obtain optimum information for the biomedical and speech training data for the proposed hybrid models.

Keywords- Medical Diagnosis; Speech Recognition; K-means Clustering; Hybrid RBF/LVQ Model; Hybrid HMM/MLP Model

I. INTRODUCTION

Indeed, since about twenty years ago, the otoneurology functional exploration possesses a tool to analyze objectively the state of the nervous conduction of additive pathway. The PERA's classification is a first step in the development of a diagnosis tool assisting the medical expert. The classification of these signals presents some problems, because of the difficulty to distinguish one class of signal from the others. The results can be different for different test sessions for the same patient. In [9], the authors present the HMM as stochastic model and apply them in the biomedical area in order to build a Computer Aided Medical Diagnosis (CAMD) tool. In the pattern recognition domain, HMM techniques hold an important place. There are two reasons why the HMM has occurred. First, the models are very thick in mathematical structure and hence can form the theoretical basis for use in a wide range of applications. Second, the models when applied properly, work very well in practice for several important applications.

The used biomedical basis used by authors in [9] is composed of 3 categories to be recognized: N, E and R classes. Each category is modeled by a discrete HMM with KM clustering. For the phase of generalization, the authors used a global DB. The average of recognition rate for the biomedical basis is 84.48% for the totality of the DB.

In many pattern classification problems the availability of multiple looks at an object can substantially improve robustness and reliability in decision making. The use of several aspects is motivated by the difficulty in distinguishing between different classes from a single view at an object. It occurs frequently that returns from two different objects at certain orientations are so similar that they may be easily confused. Consequently, a more reliable decision about the presence and type of an object can be made based upon observations of the received signals at multiple aspect angles. This allows for more information to accumulate the size, shape, composition and orientation of the objects, which in turn yields more accurate discrimination. Moreover, when the feature space undergoes changes, owing to different operating and environmental conditions, multi-aspect classification is almost a necessity in order to maintain the performance of the pattern recognition system [1, 8].

In this paper, we propose two original hybrid approaches on classification of electrical signals (speech and biomedical signals). The classification of these signals presents some problems, because of the difficulty to distinguish one class of signal from the others. The results can be different for different test sessions for the same pattern (speaker or patient).

First, we have developed a serial multi-neural network approach that involves both Learning Vector Quantization (LVQ) [2, 18] and Radial Basis Function (RBF) [3, 19] Artificial Neural Networks (ANN). These two models of ANN are particularly adapted for classification tasks and represent a very active research topic for many researchers who are trying to solve problems for which classical solutions have been limited, it is also admitted that a flat neural structure does not represent the more appropriated way to approach "intelligent behavior". The approach we propose uses a Multi-Neural Network (MNN)

DOI : 10.5963/LSMR0306003

architecture [8], so the association of two neural models improves the global order of the non-linear approximation capability of such global neural operator, compared to each single neural structure (in our application RBF and LVQ based approach) constituting the MNN system. This technique allows to fill in the gap induced by the RBF network, and thus, to refine the classification.

Second, we present the Hidden Markov Models (HMM) and apply them to complex pattern recognition problem. We attempt to illustrate some applications of the theory of HMM to real problems to match complex patterns problems as those related to biomedical diagnosis or those linked to social behavior modeling (speech recognition). In the pattern recognition domain, and particularly in speech recognition, HMM techniques hold an important place [2, 20, 21]. There are two reasons why the HMM exists. First, the models are very rich in mathematical structure and hence can form the theoretical basis for use in a wide range of applications. Second, the models, when applied properly, work very well in practice for several important applications. However, standard HMM require the assumption that adjacent feature vectors are statistically independent and identically distributed [1, 4-6]. These assumptions can be relaxed by introducing ANN in the HMM framework. Among these, the hybrid approach using the Multi-Layer Perceptron (MLP) ANN type to estimate HMM emission probabilities has recently been shown to be particularly efficient for many pattern systems [5, 6, 11, 12, 22-26]. We then propose a second hybrid system, a hybrid HMM/MLP model for Arabic speech recognition and biomedical diagnosis which makes it possible to join the discriminating capacities, resistance to the noise of MLP and the flexibilities of HMM in order to obtain better performances than traditional HMM.

In this paper first, we give a brief survey of classification methods in pattern recognition, especially to overview statistical and ANN methodologies. We then discuss the strengths and weaknesses of these methods, identify the needs of improved performance in pattern recognition, and suggest some research directions of pattern classification that can help meet these needs. Next, we will focus on the speech classification and biomedical diagnosis, so in related sections, we will clarify the principle of our hybrid HMM/MLP and RBF/LVQ systems, in Section III, we present first the speech and biomedical Data Bases (DBs) uses in our application. The acoustic pre-processor used for both the DBs experiments is presented after and the proposed clustering procedure is resumed. Next, the RBF/LVQ multi-neural network based approach and hybrid HMM/MLP proposed architecture are presented respectively in Section IV and V. In Section VI, we present the obtained classification results for the two hybrid systems. A comparison study with the classical approaches' (RBF, LVQ and HMM) has been made. In Section VII, we present the comparison study with a related work. Finally, in Section VIII, we conclude and give the prospects that follow our work.

II. RELATED WORK

To take full advantage of large sample data, the pattern recognition community turned attention to learning-based classification methods, especially ANN from the late 1980s and the 1990s. Due to the close connection between ANN and statistical pattern recognition, statistical classification methods are also considered seriously from then. Meanwhile, the research activities in pattern recognition and machine learning communities are becoming close to each other [27, 28].

In the following, we briefly review the methods that have been successfully applied to pattern recognition.

A. Statistical Methods

Statistical classification methods are rooted in the Bayes decision rule. In the case of 0-1 loss, the input pattern is classified to the class of Maximum A Posteriori (MAP) probability, which is computed by the Bayes formula from the a priori probability and the conditional probability density. Statistical classifiers are divided into parametric ones and non-parametric ones depending on the probability density estimation approach. Parametric classifiers assume for each class a known form of Density Function (DF), usually a Gaussian function, with unknown parameters estimated on training samples by Maximum Likelihood (ML). Non-parametric classifiers approximate arbitrary DF by interpolating the local densities of training samples (Parzen window), or estimate the a posteriori probabilities directly from samples (k-Nearest Neighbor (k-NN)).

When assuming Gaussian density and equal a priori probabilities, the Bayesian discriminant function is equivalent to a Quadratic Discriminant Function (QDF), which is often taken as a standard classifier in benchmarking. When further assuming that the Gaussian density functions of all classes share a common covariance matrix, the QDF is reduced to a Linear Discriminant Function (LDF). If more restrictively, the conditional DF is spherical Gaussian with equal variance, the discriminant function is reduced to the Euclidean distance from class mean, which was often taken in early feature matching methods. The QDF does not necessarily outperform the LDF because it has as many parameters as square of feature dimensionality, and so, is sensitive to the training sample size.

Under the umbrella of statistical pattern recognition are also feature selection and transformation methods. Feature transformation can reduce the dimensionality of feature space and often improve the classification accuracy. Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA) are two linear subspace learning methods that have been popularly used.

HMM are learnable finite stochastic automates. Nowadays, they are considered as a specific form of dynamic Bayesian

DOI : 10.5963/LSMR0306003

networks based on the theory of Bayes [29]. A HMM consists of two stochastic processes. The first stochastic process is a Markov chain that is characterized by states and transition probabilities. The states of the chain are externally not visible, therefore “hidden”. The second stochastic process produces emissions observable at each moment, depending on a state-dependent probability distribution.

HMM are a dominant technique for sequence analysis and they owe their success to the existence of many efficient and reliable algorithms. Mathematical and algorithmic basics of Expectation Maximization (EM) algorithm, specifically for HMM-Applications, are the Baum-Welch and the Viterbi algorithms [29, 30].

HMM are used in many areas in modern sciences or engineering applications, e.g. in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges. Other areas where the use of HMM and derivatives becomes more and more interesting are biosciences, bioinformatics and genetics [20, 21, 29-33].

Parameter optimization methods by error minimization will be reviewed in the context of ANN.

B. Artificial Neural Networks

ANN provide a method to characterize synthetic neurons to solve complex problems in the same manner as the human brain. For many years, especially since the middle of the last century, an interest in studying the brain’s mechanism and structure has been increasing. This growing research interest has led to the development of new computational models, connectionist systems or ANN [34-37], based on the biological background, for solving complex problems like pattern recognition, and fast information processing and adaptation.

Feed-forward neural networks, including Single-Layer Perceptron (SLP), MLP [38, 39], RBF network [3, 19], Higher-Order Neural Network (HONNs) [41], etc., have been widely applied to pattern recognition. Usually, each output node of the network corresponds to a class, and the maximum output gives the decision of classification. The connecting weights are usually adjusted to minimize the square error between the outputs and target values on training samples (supervised learning).

The MLP is flexible to approximate nonlinear functions and capable of separating patterns of complicated distributions [38, 40]. This power makes it a popular tool for pattern classification. Many works in speech recognition have taken the MLP as a standard classifier. The generalization performance of MLP can be improved by weight decay, local connection (local receptive fields), weight sharing, structure selection and stopping by cross-validation, etc.

The RBF network has one hidden layer of Gaussian functions, which are combined linearly by the output nodes. In early stage, the parameters of RBF networks were usually estimated in two phases: Gaussian parameter estimation by clustering and weight learning by error minimization. Since the clustering procedure does not consider the separability of patterns, the Gaussian parameters learned in this way do not lead to good classification performance.

A substantial improvement is to adjust all the parameters simultaneously by error minimization [9]. This makes the RBF network competitive with the MLP in classification accuracy.

Some unsupervised learning methods have also been applied to pattern recognition, among them are competitive learning for Vector Quantization (VQ, can be used for learning prototypes for each class) and auto-association network. The LVQ algorithm of Kohonen [2] learns class prototypes with the aim of separating the samples of different classes. LVQ is a supervised learning method and can give higher classification accuracy than VQ. We view VQ and LVQ as neural-like methods because like neural networks, the parameters (prototypes) are adjusted in online mode (stochastic gradient descent, iteratively on training samples).

The Discriminative Learning Quadratic Discriminant Function (DLQDF) [35] can also be viewed as a neural-like classifier. The DLQDF inherits the structure and initial parameters from the MQDF, but the parameters are optimized on training samples by minimizing the classification error by stochastic gradient descent.

This is an important issue being investigated in ANN research [1, 6, 24, 25], with the development of research and applications, ANN have been integrated or fused with other methods of soft computing and signal processing [42, 43]. The fusion is to combine or cascade different computing methods with ANN to improve system performance over an individual technique. In many cases, the problems can be solved more effectively by combining one or two other techniques rather than implementing ANN exclusively. In this way, the fused methods complement each other to enhance the ability of data interpretation and modeling and to avoid subjectivity in the operation of the training algorithm with ANN individually. New learning methods, especially multiple classifier systems, are now actively studied and applied in pattern recognition.

So, we proposed two hybrid models and compare your performance in complex pattern recognition problem: speech recognition and biomedical diagnosis. First, we have developed a serial multi-neural network approach that involves both LVQ and RBF ANN, these two models of ANN are particularly adapted for classification tasks. Next, a hybrid HMM/MLP system was developed combining HMM and MLP in which MLP targets and weights are iteratively re-estimated to guarantee the increase of the posterior probabilities of the correct model.

DOI : 10.5963/LSMR0306003

In Sections III, IV and V, we will clarify the principle of our hybrid HMM/MLP and RBF/LVQ systems.

III. DATABASES CONSTRUCTION AND CLUSTERING

This section describes the used DataBases (DBs) and clustering approach.

A. Speech Databases

Three speech DBs have been used in this work:

1) The first one is referred to as DB1, and the isolated digits task has 13 words in the vocabulary: 1, 2, 3, 4, 5, 6, 7, 8, 9, zero, oh, yes, no. They are spoken by 30 speakers, producing a total of 3900 utterances (each word should be marked 10 times). The digits database has about 30,000 frames of training data. This first corpus consists of isolated digits collected over the microphone.

2) The second DB, referred to as DB2, contained about 50 speakers saying their last name, first name, the city of birth and the city of residence. Each word should be marked 10 times. The used training set in the following experiments consists of 2000 sounds.

3) The third DB, referred to as DB3, contained the 13 control words (i.e. View/new, save/save as/save all) so that each speaker pronounces each control word 10 times. The used training set in the following experiments consists of 3900 sounds said by 30 speakers.

For each DBs, 12 speakers were used for training (a non-overlapping subset of these is used for cross-validation used to adapt the learning rate of the MLP), while the remaining 8 speakers were used for testing.

For feature extraction, we used the Perceptual Linear Predictive (PLP) analysis [14, 15] with Log-RASTA filter for reducing the effect of linear spectral distortion.

B. Biomedical Database

For task of biomedical DB diagnosis, the object of this survey is the classification of an electric signal coming from a medical test [7], experimental device is described in Figure 1. The used signals are called Potentials Evoked Auditory (PEA), examples of PEA signals are illustrated in Figure 2. Indeed, the exploration functional otoneurology possesses a technique permitting the objective survey of the nervous conduction along the auditory ways. The classification of the PEA is a first step in the development of a help tool to the diagnosis. The main difficulty of this classification resides in the resemblance of signals corresponding to different pathologies, but also in the disparity of the signals within a same class. The results of the medical test can be indeed different for two different measures for the same patient.

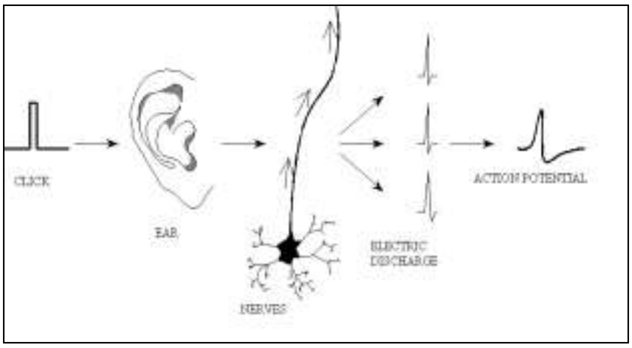


Fig. 1 PEA generation experiment

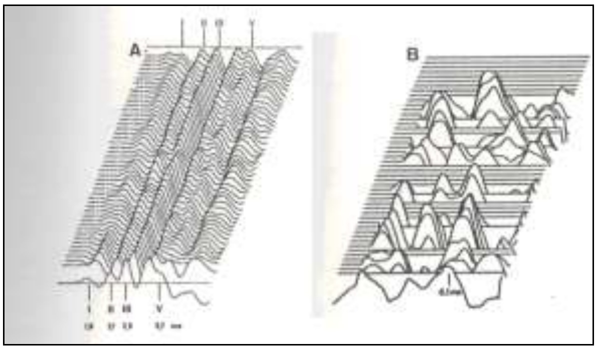


Fig. 2 (A) PEA signal for normal patient, (B) Patient with auditory disorder

The PEA signals are defined in a data base containing the files of 11185 patients. We have chosen 3 categories of patients according to the type of their trouble. The categories of patients are:

1) Normal (N): the patients of this category have a normal audition (normal class).

2) Endocochlear (E): these patients suffer from disorders that touch the part of the ear situated before the cochlea (Endocochlear class).

3) Retrocochlear (R): these patients suffer from disorders that touch the part of the ear situated to the level of the cochlea or after the cochlea. (retrocochlear class).

We selected 213 signals (correspondents to patients). So that every process (signal) contains 128 parameters. 92 among the 213 signals belong to the N class, 83, to the class E and 38 to the R class. The training basis contains 24 signals, of which 11 correspondent to the R, 6 to the E class end 7 to the N class.

DOI : 10.5963/LSMR0306003

C. K-Means Clustering

The primary task of clustering is to split a set of patterns into clusters with respect to a suitable similarity measure so that the degree of similarity is strong for data within the same cluster and weak for data in different clusters.

Several approaches to find groups in a given DB have been developed, but we focus on the KM algorithm [16, 17] as it is one of the most used iterative partitional clustering algorithms and because it may also be used to initialise more expensive clustering algorithms (e.g., the Expectation-Maximization algorithm).

Almost all partitional clustering methods are based upon the idea of optimizing a function F referred to as clustering criterion which, hopefully, translates one's intuitive notions on cluster into a reasonable mathematical formula. The function value usually depends on the current partition of the DB {C1,…, CK}. That is:

$$F \; : \; P_K(\Omega) \rightarrow \Re \tag{1}$$

where $P_K(\Omega)$ is the set of all the partitions of the database $\Omega = \{w_1,\ldots, w_M\}$ in K non-empty clusters. Each w_i of the M instances of the database Ω is a N-dimensional vector. Concretely, the KM algorithm finds locally optimal solutions using as clustering criterion F the sum of the L2 distance between each element and its nearest cluster center (centroid). This criterion is sometimes referred to as square-error criterion. Therefore, it follows that:

$$F\left(\{C_1,\ldots,C_K\}\right) \; = \; \sum_{i=1}^{K} \; \sum_{j=1}^{K_i} \; \left\| \; w_{ij} \; - \; \overline{w}_i \right\| \tag{2}$$

where K is the number of clusters, K_i the number of objects of the cluster i, wij is the j-th object of the i-th cluster and is the centroid of the i-th cluster which is defined as:

$$\overline{w}_i \; = \; \frac{1}{K_i}\sum_{j=1}^{K_i} w_{ij}, \; i \; = \; 1,\ldots, \; K \tag{3}$$

IV. MULTI-NEURAL NETWORK BASED APPROACH

A MNN could be seen as a neural structure including a set of similar neural networks (homogneous MNN architecture) or a set of diffrent neural nets (heterogeneous MNN architecture). On the other hand, both two above mentioned (homogeneous and heterogeneous MNN) could be organized in different manners. From a general point of view, three topologis [7] could characterize the MNN`s organization:

- Parallel organization: in the case, ANN`s are not inter-connected. The MNN input is dispatched to all neural networks composing such structure.
- Serial organization: in this case, the output of a given ANN composing the structure is the input of the following ANN.
- Serial/parallel organization: which combines the two structures above mentioned connections.

The problem on which our efforts have been focused concerns the speech and biomedical classification, where signals to be classified could represent a high resemblance. In such class of problems, a very fine separation should be performed in the features space. So, the use of single neural structures could lead, on the one hand, to a large number of neurons, and on the other hand, to a long learning process. Especially, when the application deals with real-time execution constraints: in our case, and in particular the PEA signals classification is intended to be used as part of the process in a computer-aided medical diagnosis tool, and so, the execution time constraint should be taken into account.

The main difficulty in classification of used signals is related, on the one hand, to a large varity of such signals for a same classification result (for example, the variation panel of corresponding PEA signals could be very large), and on the other hand, to the close resemblance between such signals for two different classification results. The serial homogeneous MNN is equivalent to a single neural network structure with a greater number of layers with different neuron activation functions. So the use of homogeneous MNN with a serial organization is here out of real interest. In the parallel homogeneous MNN configuration, each neural net operates as some "expert" (learning a specefic characteristic of the feature space). So the interest of parallel homogeneous MNN appears when a decision stage, to process the results pointed out by the set of such "expert", is associated to such MNN structure becomes then a serial/parallel MNN, needing an optimization procedure to determine the number of neural nets to be used.

This work proposes an intermediary solution: a two-stage serial heterogeneous MNN structure combining a RBF based classifier (operating as the first processing stage) with a LVQ based decision-classification stage.

Figure 3 represents the proposed serial heterogeneous MNN based architecture. The RBF model we use is a weighted-RBF model but a standard one and so, it performs the feature space mapping associating a set of "categories" to a set of "areas" of

DOI : 10.5963/LSMR0306003

the feature space. The LVQ neural model belongs to theclass of competitive neural network structure. It includes one hidden layer, called competitive layer. Even if the LVQ model has essentially been used for the classification tasks, the competitiv nature of its learning strategy (based on winner takes all strategy), makes it usable as a decision-classification operator.

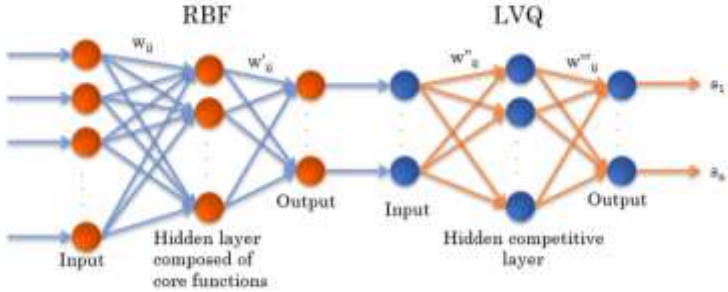


Fig. 3 Serial Multi-Neural Network based structure

On the other hand, the weighted nature of transfer functions between the input layer and the hidden one and between the hidden layer and the output one in this model allows non-linear approximation capability, making such neural net a function "approximation operator".

V. HYBRID HMM/MLP MODEL

The work at ICIS has provided us with further insight into the discriminant HMM, particularly on transition based models [10]. This new perspective has motivated us to further develop the original discriminant HMM theory [11], in which an MLP is trained to optimize the full posteriori probabilities of Markov models given the acoustic data via conditional transition probabilities, i.e., probabilities of the next state given the current state and the current acoustic vector. This approach uses posterior probabilities at both local and global levels and is more discriminant in nature. Using a recursive algorithm that is reminiscent of the EM algorithm for the estimation of data likelihoods.

The method is developped in the context of a statistical model for transition-based electrical signal recognition using MLP to generate probabilities for HMM. In the new approach, we use local conditional posterior probabilities of transitions to estimate global posterior probabilities of instance sequences given acoustic data (see Figure 4).

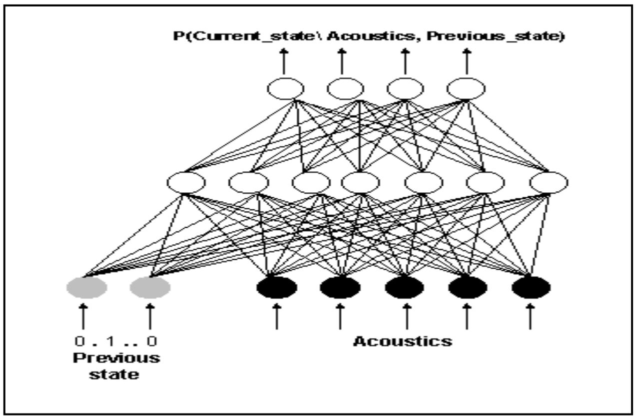


Fig. 4 An MLP that estimates local conditional transition probabilities

A. Estimating HMM Likelihoods with MLP

For statistical recognition systems, the role of the local estimator is to approximate probabilities or Probability Fensity Functions (PDF). Practically, given the basic HMM equations, we would like to estimate something like $p(x_n\backslash q_k)$, which is the value of the PDF of the observed data vector given the hypothesized HMM state. The MLP can be trained to produce the posterior probability $p(q_k\backslash x_n)$ of the HMM state give the acoustic data. This can be converted to emission PDF values using Bayes'rule.

Since the network outputs approximate Bayesian probabilities, $g_k(x_n,\Theta)$ is an estimate of:

$$p(q_k \backslash x_n) = \frac{p(x_n \backslash q_k)\, p(q_k)}{p(x_n)} \tag{4}$$

which implicitly contains the a priori class probability $p(q_k)$. It is thus possible to vary the class priors during classification without retraining, since these probabilities occur only as multiplicative terms in producing the network outputs. As a result,

DOI : 10.5963/LSMR0306003

class probabilities can be adjusted during use of a classifier to compensate for training data with class probabilities that are not representative of actual use or test conditions.

Thus, scaled likelihoods $p(x_n \backslash q_k)$ for use as emission probabilities in standard HMM can be obtained by dividing the network outputs $g_k(x_n)$ by the training set, which gives us an estimate of :

$$\frac{p(x_n \backslash q_k)}{p(x_n)} \tag{5}$$

During recognition, the scaling factor $p(x_n)$ is a constant for all classes and will not change the classification.

B. Global Posterior Probability Estimation

If X = {x_1, x_2, ..., x_N} is a squence of acoustic vectors and M_i a HMM, the optimal training and recognition criterion (actually minimizing the probability of errors) should be based on the posterior probabilities $P(M_i/X, \Theta)$. In standard HMM, using Bayes' rule, $P(M_i/X, \Theta)$ is usually expressed in terms of $P(X /M_i, \Theta)$ as:

$$P(M_i/X,\Theta) = \frac{P(X/M_i,\Theta)\ P(M_i/\Theta)}{P(X/\Theta)} \tag{6}$$

which separates the probability estimation process into language modeling and acoustic modeling in one particular way.

C. Motivations of HMM/ MLP Hybridation

MLP has several advantages that make them particularly attractive for pattern recognition [1, 5, 11], e.g. :

- They can provide discriminant learning between pattern units or HMM states that are represented by MLP output classes. That is, when trained for classification (using common cost functions such as MSE or relative entropy), the parameters of the MLP output classes are trained to minimize the discrimination between the correct output class and the rival ones.
- Because MLP can incorporate multiple constraints for classification, features do not need to be assumed independent. More generally, there is no need for strong assumptions about the statistical distributions of the input features (as is usually required in standard HMM).
- They have a very flexible architecture which easily accommodates contextual inputs and feedback, and both binary and continuous inputs.
- MLP are typically highly parallel and regular structures, which make them especially amenable to high-performance architectures and hardware implementations.

VI. CASE STUDY AND EXPERIMENTAL RESULTS

For biomedical DB, a codebook size of from M = 10 clusters have been used in recognition experiments. For the speech DBs, a codebook size of 320 clusters have been used in input layer of hybrid models because after feature extraction with log RASTA-PLP, the acoustic features were quantized into independent codebooks according to the k-means algorithm respectively:

- 128 clusters for the log-RASTA PLP vectors.
- 128 clusters for the first time derivative of cepstral vectors.
- 32 clusters for the first time derivative of energy.
- 32 clusters for the second time derivative of energy.

A. Comparison Study with Single RBF and LVQ Approaches

The structure of the RBF and LVQ for the respective single approaches is composed as follows (we give the proposed network parameters for the biomedical basis):

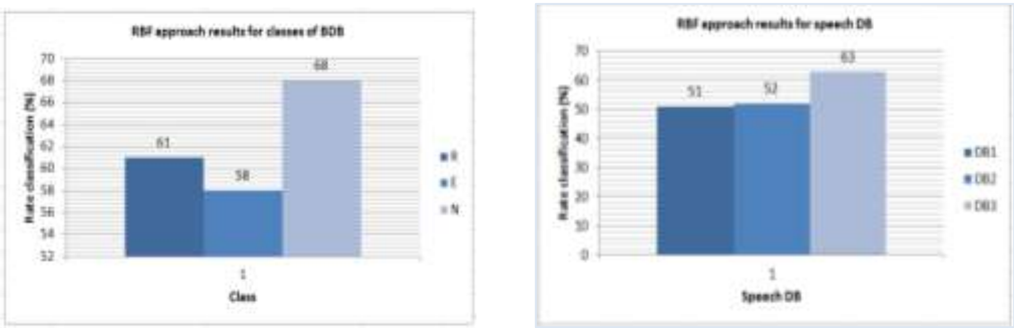
- The number of input neurons for RBF and LVQ corresponds to the number of components of the quantized vectors (10 inputs).
- The output layer of RBF and LVQ contains 3 neurons, corresponding to the 3 classes.
- For RBF, the number of neurons of the hidden layer (in this case, 22 neurons) has been determined to satisfy the following heuristic rule [13]: *Number of hidden neurons = (a number of entry neurons * number of output neurons)*$^{1/2}$

DOI : 10.5963/LSMR0306003

For LVQ, the number of hidden neurons (in this case, 10 neurons) has been determined by considering the number of subclasses we can count into the 3 classes.

For RBF, the learning DB contains 24 signals, 11 of which are R, 6 E and 7 N. For LVQ, the learning DB contains 20 signals, 6 of which are R, 7 E and 7 N. The obtained results are given in the Figure 5.

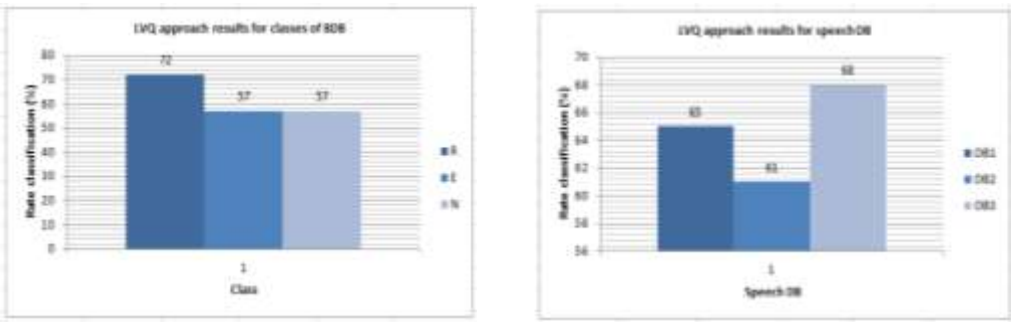
In the two cases, the learning DB has been successfully learnt. All of the learnt vectors are well classified in the generalization phase. The RBF network well classifies 62.3% of the full DB (including the learnt vectors), with a rate of correct classification of 61% for the R class, 58% for the E class and 68% for the N class. For the tree speech DBs, a rate of correct classification as follows: 51% for the DB1, 52% for the DB2, 63% for the DB3.



(a) (b)

Fig 5. Obtained results with RBF based approach for biomedical DB (a) and speech DBs (b)

The LVQ network well classifies 62% of the full medical DB (including the learnt vectors), with a rate of correct classification of 72% for the R class, 57% for the E class, 57% for the N class. For the tree speech DBs, a rate of correct classification as follows: 65% for the DB1, 61% for the DB2, 68% for the DB3. The obtained results are given in the Figure 6.



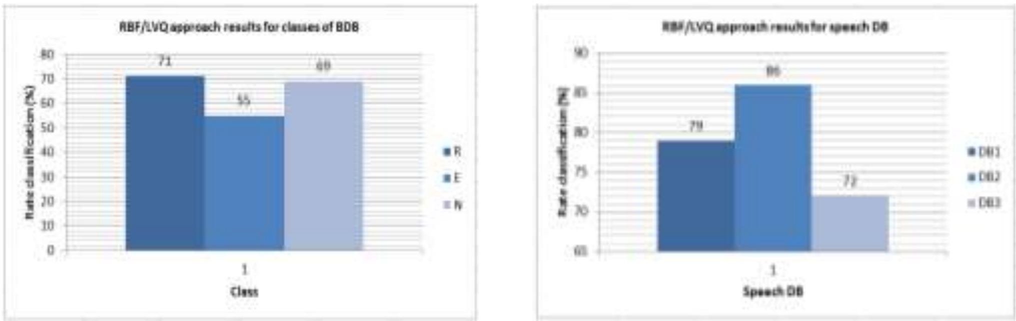
(a) (b)

Fig 6. Obtained results with LVQ based approach for biomedical DB (a) and speech DBs (b)

B. Results Relative to RBF/LVQ Based Multi-neural Network Approach

The present work uses the RBF/LVQ based serial heterogeneous MNN. The proposed networks parameters for the biomedical basis are as follow:

Concerning the RBF, the number of input neurons is 10 corresponding to the number of clusters of the quantized input vectors, and the output layer contains 3 neurons. The number of neurons of the hidden layer is 20 neurons.



(a) (b)

Fig 7. Obtained results with hybrid RBF/LVQ model for biomedical DB (a) and speech DBs (b)

DOI : 10.5963/LSMR0306003

For the LVQ, the number of input cells is equal to the number of output cells of the RBF. The output layer of the LVQ contains as many neurons as categories (3). The number of neurons in hidden layer is 8 neurons and has been determind by considering the number of subclasses we can count into the 3 classes. The learning DB contains 4 signals, 11 of which correspond to R disordrs, 6 to E disoders and 7 to N. for the generalization phase, we use the full DB, including the learning DB. Figure 7 gives the results of this experiment.

The learning DB has been successfully learnt. All of the learnt vectors are well classified in the generalization phase. We can see that this network well classifies 65% of the full DB (including the learnt vectors), with a rate of correct classification of 71% for the R class, 55% for the E class, 69% for the N class.

The behavior of the MNN concerning the R and N classes permits to obtain high rate of well classification of these classes. However, the classification rate of E signals is not satisfactory. Only about 55% of vectors are well classified.

For the tree speech DBs, a rate of correct classification is as follows: 79% for the DB1, 86% for the DB2, 72% for the DB3.

Comparingthe two single approaches, with our proposed MNN technique, we obtain:

– Better performance for the N class than the single LVQ ANN approach.

– Better rate for the R class than the single RBF ANN approach.

– Similar results, in the case of N and R class, to those respectively obtained for single RBF and single LVQ.

– Better rate for the all speech DBs than the single RBF or LVQ ANN based approach.

Therefore, the MNN structure combines the advantages of both LVQ and RBF ANN. Moreover, these globally better results of our MNN technique are achieved with low number of neurons in the ANN architecture, taken into account the difficulty of our problem.

C. Results Relative to Discrete HMM and Hybrid HMM/MLP Approach

Further assume that for each class in the vocabulary we have a training set of k occurrences (instances) of each class where each instance of the categories constitutes an observation sequence. In this work, the same HMM topology was used for all the experiments and the following HMM have been used:

1) Discrete HMM

For speech DBs, 10-state, strictly left-to-right, discrete HMM were used to model each basic unit (words). For the PEA signals, 5-state, discrete HMM were used.

For the tree speech DBs, a classification rate as follows: 87% for the DB1, 90% for the DB2 and 76% for the DB3. For the biomedical DB, the rate is 84%. Figure 8 gives the results of this experiment for biomedical DB and speech DBs.

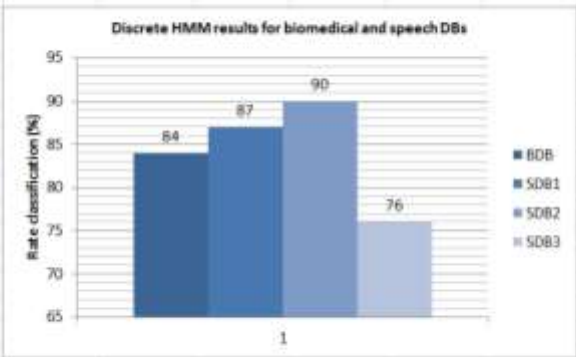


Fig. 8 Discrete HMM results for biomedical DB and speech DBs

2) Discrete HMM/MLP with Entries Provided by the KM Algorithm

10-state, strictly left-to-right, word HMM with emission probabilities computed from an MLP with 320 quantized acoustic vectors at the input of MLP for speech DBs, i.e., Each acoustic vector was represented by a binary vector composed of 4 fields respectively containing 128, 128, 32, and 32 bits. In each field, only one bit was "on" to represent the current associated cluster. A hidden layer of variable size (30 neurons), an output layer made up of more neurons than there are HMM states. For the PEA signals, a MLP with 10 neurons at the entry, 18 neurons for the hidden layer and 5 output neurons was trained. The three-layer MLP is trained by using stochastic gradient descent, and relative entropy as the error criterion.

For the tree speech DBs, a rate of correct classification as follows: 96% for the DB1, 94% for the DB2 and 80% for the DB3. For the biomedical DB, the rate is 94%. Figure 9 gives the results of this experiment for biomedical DB and speech DBs.

DOI : 10.5963/LSMR0306003

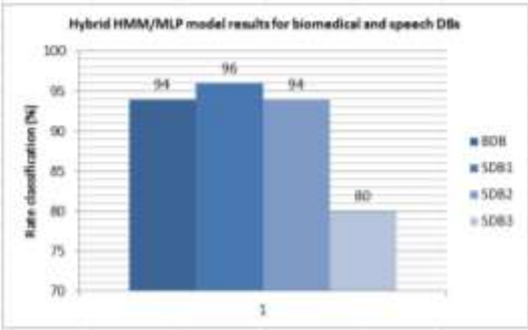


Fig. 9 Hybrid HMM/MLP results for biomedical DB and speech DBs

Compared to the obtained results with discrete HMM, the proposed hybrid HMM/MLP system offers better performance for all DB (biomedical and speech) than HMM results.

For the majority of cases, we can draw a preliminary conclusion from the results reported in Figure 5 to Figure 9 for the speech recognition and the PEA signals diagnosis: The hybrid discrete HMM/MLP approach using KM clustering always outperforms standard discrete HMM or RBF/LVQ ANN based approach.

VII. DISCUSSION

In [8], the proposed approach by the authors is based on MNN concept, combining a RBF based classifier operating as the first processing stage with a LVQ based decision classification stage. They used the RBF/LVQ based serial heterogeneous MNN, on the same biomedical basis to build a CAMD tool but the authors did not use the vector quantization phase and thus reduce the number of data in the input layer to the hybrid model. This related work has been considered as a reference for performance comparison with the obtained results by our system using a K-Means (KM) clustering.

The RBF/LVQ model based CAMD tool proposed in [8] well classifiers 62% of the full DB, including the learnt vectors, with a rate of correct classification of: 72% for the R category, 57% for the E and 57% for the N category.

We summarize in the next histogram the different results obtained by: 1) RBF/LVQ model without QV, 2) RBF/LVQ model with QV, 3) hybrid MM/MLP model with QV on the biomedical base and inserting those obtained in the related work reported in [8] in the same condition used RBF/LVQ model without QV (see Figure 10).

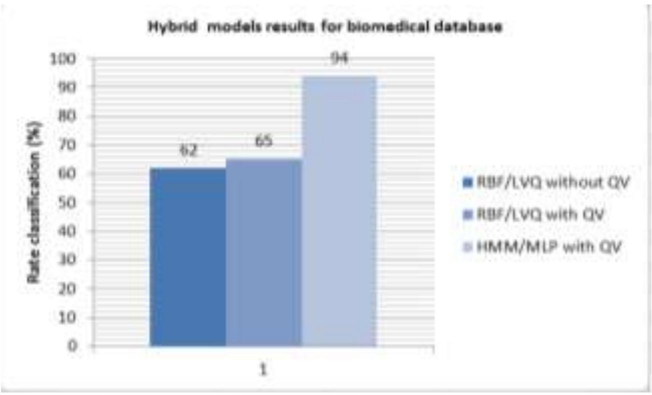


Fig. 10 Hybrid models results for biomedical DB

In comparison the two hybrid models are used in this present work:

- With the result obtained by the related work, we might deduce that the hybrid RBF/LVQ model with the k-means clustering, is more efficient than the same model without the use of k-means algorithm.
- With the proposed hybrid HMM/MLP model, a rate of correct classification for the biomedical DB was 94% for the totality of the basis. The obtained results are encouraging and show the feasibility of this hybridation based tool for CAMD.

VIII. CONCLUSION AND FUTURE WORK

This paper presented the test of two types of hybrid models in the framework of speech recognition and medical diagnosis. The first hybrid model is the MNN structure based on the KM clustering: the work use involves LVQ and RBF neural network. The first neural net (RBF) is used as a classifier, and the second one (LVQ) as a competitive decision processor. So, the

DOI : 10.5963/LSMR0306003

association of two RBF and LVQ neural networks improves the global order of the non linear approximation capability of such global neural operator, compared to each single neural structure constituting the MNN system.

For the second hybrid model, a discriminate training algorithm for hybrid HMM/MLP system based on the KM clustering is described. Our results on arabic speech and biomedical signals recognition tasks show an increase in the estimates of the posterior probilities of the correct class after training and significant decreases in error rates. These preliminary experiments have set a baseline performance for our hybrid HMM/MLP system, better recognition rates were observed. From the viewpoint of effectiveness of the models, it seems obvious that the hybrid HMM/MLP model is more powerful than discrete HMM or multi-network RBF/LVQ structure for the PEA signals diagnosis and speech DB recognition.

For the two types of signals recognition, the next work envisages improving the performance with a fusion scheme: cooperation of HMM with the RBF/LVQ multi-network structure in order to succeed to a hybrid model and compare the perfomance with the HMM/MLP model proposed in this paper.

REFERENCES

[1] L. Lazli, M.SELLAMI. “Connectionist Probability Estimators in HMM Speech Recognition using Fuzzy Logic,” MLDM 2003: the 3rd international conference on Machine Learning & Data Mining in pattern recognition, LNAI 2734, Springer-verlag, pp. 379-388, July 5-7, Leipzig, Germany, 2003.

[2] Blachnik, M., & Laaksonen, J. “Image classification by histogram features created with learning vector quantization,” Lecture Notes in Computer Science, 5163, 827-836, 2008.

[3] Buhmannm. D., “Radial Basis Functions,” Cambridge University Press, Cambridge, 2003.

[4] A. Hagen, A. Morris. “Comparison of HMM experts with MLP experts in the full combination multi-band approach to robust ASR,” To appear in International Conference on Spoken Language Processing, BEIJING, 2000.

[5] C. Ma, H. Kuo, H. Soltau, X. Cui, U. Chaudhari, L. Mangu, and C.-H. Lee. “A comparative study on system combination schemes for lvcsr, ” In Proceedings IEEE Int. Conf. ICASSP, pp. 4394-4397, 2010.

[6] E. Trentin and M. Gori, “A survey of hybrid ANN/HMM models for automatic speech recognition,” Neurocomputing, vol. 37, iss. 1-4, pp. 91-126, 2001.

[7] J-F. Motsh. “La dynamique temporelle du trons cérébral: Recueil, extraction et analyse optimale des potentiels évoqués auditifs du tronc cerebral,” Thesis, University of "Créteil", PARIS XII, 1987.

[8] A-S. Dujardin., Amarger, V., Madani, K. “Multi-neural networks approaches for biomedical applications: classification of brainstem auditory evoked potentials,” Published in: Neural Networks, 1999. IJCNN '99. IEEE International Joint Conference on Volume: 5, 3609-3613, 1999.

[9] L. Lazli, A-N. Chebira, K. Madani. “Hidden Markov Models for Complex Pattern Classification,” Ninth International Conference on Pattern Recognition and Information Processing, PRIP’07, BELARUS, 2007.

[10] N. MORGAN, H. BOURLARD, S. GREENBERG, H. HERMANSKY. “Stochastic perceptual auditory-event-based models for speech recognition,” In Proceedings Int. Conference on Spoken Language Processing, 1943-1946, Yokohama, Japan, 1994.

[11] H. Bourlard, N. MORGAN. “Connectionist Speech Recognition – A Hybrid Approach,” Kluwer Academic Publishers, 1994.

[12] S. RENALS, N. MORGAN, M. CHOEN, H.FRANCO, H.BOURLARD. “Connectionist probability estimation in the DECIPHER speech recognition system,” In Proceedings IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 601-604, California, 1992.

[13] M.I. JORDAN and al., “Hierarchical mixture of expert and the EM algorithm,” *Neural Computation*, vol.6, no.2, pp. 181-214, 1994.

[14] C.H.T. grace. “Combating Additive Noise and Spectral Distortion in Speech Recognition Systems with JAH-RASTA,” Master Thesis, *University of California at Berkely*, 1994.

[15] H. Hermansky. “RASTA Processing of speech,” IEEE Trans. On Speech and Audio Processing, vol.2, no.4, pp. 578-589, 1994.

[16] J. Ye, Z. Zhao, M. Wu, “discriminative K-Means for clustering,” Proceeding of the annual conference, 2007.

[17] A. Raftery and N. Dean, “variable selection for model based clustering”, Journal of American statistical association, 101(473): 168-178, 2006.

[18] Schneider, P., Hammer, B., Biehl, M. “Adaptive relevance matrices in learningvectorquantization,” Neural Computation, 21, 3532-3561, 2009.

[19] Martin D, “Radial Basis functions,” Book in Buhmann edition, 2003.

[20] Sha, F. & Saul, L., “Comparison of large margin training to other discriminative method for phonetic recognition by Hidden Markov Models,” Proc. ICASSP 2007, Honolulu, Hawaii, 2007.

[21] Cheng, C.-C.; Sha, F. & Saul, L.K., “Online Learning and Acoustic Feature Adaptation in Large Margin Hidden Markov Models,” In IEEE Journal of Selected Topics in Signal Processing, pp. 926-942, 2010.

[22] A. Abad, T. Pellegrini, I. Trancoso, and J. Neto, “Context dependent modelling approaches for hybrid speech recognizers,” in Interspeech, pp. 2950-2953, 2010.

[23] S. Espana-Boquera, M. J. Castro-Bleda, J. Gorbe-Moya. “Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models,” in IEEE journal of Transaction on Pattern analysis and machine intelligence, vol. 33 no. 4, pp. 767-779, 2011.

[24] S. Marukatat, T. Artières, B. Dorizzi, and P. Gallinari, “Sentence Recognition through Hybrid Neuro-Markovian modelling,” Proc. Int’l Conf. Document Analysis and Recognition, pp. 731-735, 2001.

[25] É. Caillault and C. Viard-Gaudin, “Mixed Discriminant Training of Hybrid ANN/HMM Systems for Online Handwritten Word

DOI : 10.5963/LSMR0306003

Recognition," Int'l J. Pattern Recognition and Artificial Intelligence, vol. 21, iss. 1, pp. 117-134, 2007.

[26] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A Novel Connectionist System for Unconstrained Handwriting Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, iss. 5, pp. 855-868, May 2009.

[27] Bishop, C.M., "Pattern Recognition and Machine Learning," Springer-Verlag New York, Inc.: Secaucus, NJ, USA, 2006.

[28] Mantzaris, D., Anastassopoulos, G., Iliadis L., et al., "A soft computing approach for osteoporosis risk factor estimation," in: IFIP International Federation for Information Processing, AIAI2010, Springer, 339: 120-127, 2010.

[29] Al-Ani, T. & Hamam, Y. "A low Complexity Simulated Annealing Approach For Training Hidden Markov Models," International Journal of Operational Research, vol. 8, iss. 4, pp. 483-510, 2010.

[30] Cheng, C.-C.; Sha, F. & Saul, L.K., "A fast online algorithm for large margin training of continuous-density hidden Markov models," In Proceedings of the Tenth Annual Conference of the International Speech Communication Association (Interspeech-09), pp. 668-671. Brighton, UK, 2009.

[31] Ince, H. T. & Weber, G.W. "Analysis of Bauspar System and Model Based Clustering with Hidden Markov Models," Term Project in MSc Program "Financial Mathematics – Life Insurance", Institute of Applied Mathematics METU, 2005.

[32] Kouemou, G. et al. "Radar Target Classification in Littoral Environment with HMMs Combined with a Track Based classifier," Radar Conference, Adelaide Australia, 2008.

[33] T.; Karmakar, C.K.; Khandoker, A.H.; Palaniswami, "Automatic Recognition of Obstructive Sleep apnoea Syndrome Using Power Spectral Analysis of Electrocardiogram and Hidden Markov Models," International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Sydney, NSW, 15-18 Dec., pp. 285-290, 2008.

[34] Hayati, M., and Mohebi, Z. "Application of artificial neural networks for temperature forecasting," World Acad. Sci., Eng., Technol. 28: 275-279, 2007.

[35] Maier, H., Jain, A., Dandy, G., and Sudheer, K. "Methods Used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future di rections," Environ. Modell. Software journal, Elsevier. Volume 25, Issue 8, Pages 891-909, 2010.

[36] Mantzaris, D., Anastassopoulos, G., and Adamopoulos, A. "Genetic algorithm pruning of probabilistic neural networks in medical disease estimation," Neural Networks, Elsevier, 24: 831-835. PMid: 21723704, 2011.

[37] Holger R. Maiera, Ashu Jainb, Graeme C. Dandya, K.P. Sudheerc, "Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions," Environmental Modelling & Software journal, Elsevier, vol. 25, iss. 8, pp. 891-909, 2010.

[38] Q. Zhu, B. Chen, N. Morgan, and A. Stolcke, "On Using MLP Features in LVCSR," in ICSLP, 2004.

[39] O. Vinyals and S. V. Ravuri, "Comparing multilayer perceptron to Deep Belief Net-work Tandem features for robust ASR," in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, pp. 4596-4599, may 2011.

[40] I.S. Isa, Z. Saad, S. Omar, M.K. Osman, K.A. Ahma, "Suitable MLP Network Activation Functions for Breast Cancer and Thyroid Disease Detectio", Second international conference on computational intelligence, modelling and simulation, IEEE, pp. 39-44, 2010.

[41] Baboo, S., and Shereef, I. "An efficient weather forecasting system using artificial neural network," Int. J. Environ. Sci. Dev. journal, Elsevier, 1: 321-326, 2010.

[42] Mantzaris, D., Anastassopoulos, G., Iliadis L., et al. "A soft computing approach for osteoporosis risk factor estimation," in: IFIP International Federation for Information Processing, AIAI2010, Springer, 339: 120-127, 2010.

[43] Sethuraman, J. "Soft computing approach for bond rating prediction," Artificial Neural Networks in Real-Life Appl., 202-219, 2006.

DOI : 10.5963/LSMR0306003