Design of Effective Neural Network Ensembles for Image Classification Purposes

Giorgio Giacinto and Fabio Roli

Dept. of Electrical and Electronic Eng., University of Cagliari, Italy

Piazza D'Armi, 09123, Cagliari, Italy

Phone: +39-070-6755874 Fax: +39-070-6755900 e-mail: {giacinto, roli}@diee.unica.it

Corresponding Author: Prof. Fabio Roli
Dept. of Electrical and Electronic Eng., University of Cagliari
Piazza D'Armi, 09123, Cagliari, Italy
Phone: +39-070-6755874 Fax: +39-070-6755900
e-mail roli@diee.unica.it

Abstract

In the field of pattern recognition, the combination of an ensemble of neural networks has been proposed as an approach to the development of high performance image classification systems. However, previous work clearly showed that such image classification systems are effective only if the neural networks forming them make different errors. Therefore, the fundamental need for methods aimed to design ensembles of "error-independent” networks is currently acknowledged. In this paper, an approach to the automatic design of effective neural network ensembles is proposed. Given an initial large set of neural networks, our approach is aimed to select the subset formed by the most error-independent nets. Reported results on the classification of multisensor remote-sensing images show that this approach allows one to design effective neural network ensembles.

Keywords: Neural Network Ensembles, Combination of Classifiers, Image Classification, Remote Sensing.

1. Introduction

Recently, in the field of pattern recognition, the combination of an ensemble of neural networks has been proposed as an approach to the development of high performance image classification systems [1,4,5,7,8,11,29]. Typically, the network outputs are combined by voting rules, belief functions, statistical techniques, Dempster-Shafer evidence theory, and other integration schemes [29]. Voting rules, for example, the majority rule, are used if the network outputs are regarded as simple classification "labels" [12]. Statistical combination methods are adopted to combine network outputs regarded as class posterior probabilities. As an example, a simple average or some other linear combination have been proposed [29]. If the network outputs are interpreted as fuzzy membership values, belief values or evidence values, then belief functions and Dempster-Shafer techniques are used [29]. Finally, it is also possible to combine the networks by a "meta" neural network, which takes the net outputs as input features [9].

The experimental results reported in the literature showed that the image classification accuracy provided by the combination of an ensemble of neural networks can outperform the accuracy of the best single net [1,4,5,7,8,11,29]. In addition, other advantages are provided by neural network ensembles in the context of image classification applications. For example, the combination of neural networks can be used as a "data fusion" mechanism where different nets process data from different imaging sensors. In the field of optical character recognition, different networks can take as inputs samples characterised by different feature vectors (e.g., shape and grey-level intensity features) [29].

However, the above also showed that neural network ensembles are effective only if the nets forming them make different errors. Hansen and Salamon showed that neural networks combined by the "majority" rule can provide increases in classification accuracy, only if the nets make independent errors [6]. Tumer and Ghosh pointed out that increases in accuracy depend on error-independence far more than on the particular combination method [27]. The review paper by Sharkey also stressed the fundamental role of error "diversity" in determining the effectiveness of a neural network ensemble [25]. Accordingly, most of the combination methods described in the literature assume that ensembles are formed by neural networks making independent classification errors.

Unfortunately, the reported experimental results pointed out that the creation of error-independent networks is not a trivial task, in the sense that, though different in terms of their weights, architectures, and other parameters, nets can exhibit the same pattern of errors, basically because of the so-called problem of network “symmetries” [6]. Therefore, the fundamental need for methods aimed to design ensembles of error-independent networks is currently acknowledged. However, in the pattern recognition field, most of the work focused on the development of combination methods. Few papers have been published in the neural networks literature addressing the problem of the creation of ensembles of error-independent nets [24]. An overview is given in Section 2.

In this paper, an approach to the automatic design of effective neural network ensembles is proposed (Section 3). Instead of attempting to design an ensemble of independent networks directly, a large set of independent but also error correlated nets is initially created. Given such a set, our approach is aimed to select the subset formed by the most error-independent nets. The extent to which our hypotheses can be deemed realistic, and the effectiveness of the proposed design approach are discussed in Section 3.3. Experimental results and comparisons are reported in Section 4. Conclusions are drawn in Section 5.

2. Related Work

In the neural network field, several methods for the creation of ensembles of neural networks making different errors have been investigated [25]. Such methods basically lie on “varying” the parameters related to the design and to the training of neural networks. In particular, the main methods in the literature can be included in one of the following categories:

- *varying the initial random weights:* an ensemble of nets can be created by varying the initial random weights, from which each network is trained;
- *varying the network architecture:* the nets forming the ensemble exhibit different architectures (basically, the net architectures are different in number of hidden neurones);
- *varying the network type:* different net types (e.g., multilayer perceptrons, radial basis functions neural networks, and probabilistic neural networks) can be used to create the ensemble members;

- *varying the training data:* an ensemble of nets can be created by training each network with a different learning set. This can be done in a number of different ways. For example, "sampling" the training data to obtain different learning sets, using learning sets extracted from different data "sources" (e.g., data from different imaging sensors), or by different pre-processing phases (e.g., sets formed by samples characterised by different "features").

Partridge experimentally compared the capabilities of the above methods to create error-independent networks [15]. He found the following ordering: varying the net type, varying the training data, varying the net architecture, and varying the initial random weights. He thus concluded that varying the net type and the training data are the two best ways for creating ensembles of networks making different errors. Such conclusions are basically shared by other researchers [24].

However, it has been noted that neural network ensembles could be created using a combination of two or more of the above methods (e.g., varying the net type plus varying the training data) [25]. Raviv and Intrator proposed a method that combines "sampling" and "pre-processing" [18]. The potentialities of such "hybrid" methods are easy to see, even if a clear assessment of their performances is beyond the current state of the art.

We think that the large set of hybrid methods that could be developed by combining the above methods points out very clearly the need for systematic approaches to the creation of neural network ensembles. In other words, in order to exploit effectively the currently available "tool box", the design of neural network ensembles must be based on engineering approaches. In our opinion, such approaches should give clear guidelines on the use of methods to generate ensemble "candidate" members and on the ways to select the most error-independent networks. Differently, as stated by Wolpert [28], the creation of effective neural network ensembles is running the risk of remaining a "black art".

Recently, researchers have started to investigate the problem of the engineering design of neural network ensembles. It seems to us that the proposed approaches can be classified into two main design strategies:

- the "direct" strategy;
- the "overproduce and choose" strategy.

The first design strategy is aimed to generate an ensemble of error-independent nets directly. Opitz and Shavlik presented and algorithm called ADDEMUP, which uses genetic algorithms to search directly for an ensemble of independent neural networks [14]. Rosen described a method that allows to train an ensemble of networks by backpropagation, not only to reproduce a desired output, but also to have their errors linearly decorrelated with the other networks [21]. Individual networks so trained are then linearly combined. Some methods developed in the context of “adaptive sampling and boosting” can also be regarded as direct design methods [3].

On the other hand, the "overproduce and choose" strategy is based on the creation of an initial large set of nets and the subsequent choice of the subset of the most error-independent nets. Partridge and Yates described a design method based on such a strategy [16]. They introduced some very interesting "error diversity" measures, which can be used to choose a subset of independent networks. However, they did not propose a systematic method to choose such a set. Only an experimental investigation of three heuristic techniques is described. In addition, the problem of optimality of such choice techniques is not discussed. Sharkey and Sharkey also described a design method that follows the "overproduce and choose" strategy [26]. Their choice algorithm is basically guided by a heuristic based on the evaluation of error correlation between pairs of nets.

3. Design of Effective Neural Network Ensembles

3.1 Basic concepts

As pointed out in the previous sections, the main determinant of the effectiveness of a neural network ensemble is the extent to which the nets forming the ensemble exhibit an error diversity, in the sense that they make different errors. Therefore, our approach to the design of neural network ensembles is aimed to create a set of nets exhibiting the highest possible degree of error diversity.

Among the two main design strategies recently described in the neural networks literature, our approach follows the so called “overproduce and choose” strategy (see Section 2). The rationale behind this choice is that we think that the “direct” generation of error-independent nets is a very difficult problem that is beyond the current state of the neural computing theory. This opinion is shared by other researchers in the neural networks field [16,26]. In addition, the "overproduce and choose" strategy allows to exploit effectively the methods for the creation of ensemble members

(Section 2). All the available methods can be used in the overproduction phase to create a large set of “candidate” members. The choice phase is then aimed to select the most error-independent nets as ensemble members. Finally, it is worth noting that the "overproduce and choose" strategy is used successfully in other fields. For example, in the field of software engineering the development of programs for safety-critical applications is based on so-called N-version programming [13]. The basic idea is to produce N versions of a program, such that the versions fail independently, and then to combine the versions by a majority vote to produce a more reliable program.

With regard to the overproduction phase, we exploited the conclusions of Partridge concerning the most effective parameters in creating neural networks making different errors (see Section 2). Accordingly, different net types, for example, multilayer perceptrons, radial basis functions neural networks, and probabilistic neural networks, are used in the overproduction phase to generate the initial ensemble of candidate members.

Let E be the ensemble of N neural networks created by the overproduction phase:

$$E=\{n_1, n_2, \ldots, n_N\} \qquad (1)$$

The subsequent choice phase is aimed to select a subset of nets E* that can be effectively combined. It is easy to see that the optimal subset E* could be obtained by “exhaustive enumeration”, that is, by assessing the classification accuracies provided by all possible subsets of set E, and then choosing the subset exhibiting the best performance. In particular, for each subset E*, the accuracy provided by the combination of the nets forming it could be estimated on a “validation” set. Unfortunately, the number of possible subsets is equal to $\sum_{i=1}^{N}\binom{N}{i}$ and the size N of set E is large by definition of the overproduction phase. Therefore, such a “force brute” approach is not feasible.

Our approach to the choice phase is based on some hypotheses concerning set E (equations 2-4), which allow us to select a subset of nets that can be effectively combined, without any need for an exhaustive enumeration of all possible subsets. (As will be shown in the next section, a number of subsets at most equal to N is evaluated by our approach).

First of all, let us assume that set E is formed by the following union of M subsets, E_i:

$$E = \bigcup_{i=1}^{M} E_i \qquad (2)$$

where E_i meet the following assumption:

$$i,j \quad i \neq j \quad E_i \bigcap E_j = \qquad (3)$$

and the nets forming the above subsets satisfy the following conditions:

$$E_i, E_j, n_l, n_m, n_n, \; i \neq j, \; n_l, n_m \quad E_i, \; n_n \quad E_j \quad \text{prob}(n_l \text{ fails}, n_m \text{ fails}) > \text{prob}(n_l \text{ fails}, n_n \text{ fails}) \qquad (4)$$

In the above equation, the terms prob(n_l fails, n_m fails) and prob(n_l fails, n_n fails) are the compound error probabilities of related net pairs. Such error probabilities can be estimated by the number of coincident errors made by pairs of nets on a validation set. In particular, equation 4 states that the compound error probability between any two nets belonging to the same subset is higher than that between any two nets belonging to different subsets. In other words, we are assuming that set E is formed by clusters of error correlated nets, while nets belonging to different clusters should be more independent. It is easy to see that such a condition provides a useful “guide” for the choice of an ensemble of nets that can be combined effectively. In fact, according to equation 4, error-independent nets can be extracted from the subsets E_i.

Therefore, according to the hypotheses of equations 2-4, we defined a choice phase made up of the following steps:

- identification of subsets E_i;
- extraction of nets from the above subsets in order to create an ensemble E* formed by the most error-independent nets.

In the first step, it can easily be seen that the M subsets forming set E can be identified by a “clustering” algorithm grouping the neural nets according to compound error probability (see equation 4). The developed clustering algorithm is described in the next section.

After the subsets E_i, i=1...M, have been identified, one net is taken from each subset to create an ensemble $E^*=\{n^*_1, n^*_2, \ldots, n^*_M\}$ formed by the most error-independent nets. This extraction step is described in detail in the next section.

It is worth noting that the more error correlated the nets belonging to the same subset, and the higher the degree of error diversity exhibited by the nets belonging to different subsets, the more effective the above choice phase. In Section 3.3, we discuss in detail the extent to which our hypotheses can be deemed realistic and the proposed design approach effective.

3.2 The proposed approach

As described in the previous section, our design approach is made up of the overproduction and the choice phases. The overproduction phase basically exploits the conclusions of Partridge [15]. Therefore, in the following, we only give further details on the choice phase.

As previously described, the choice phase is made up of the following steps:

· Unsupervised learning for identifying subsets E_i, i=1...M

· Creation of the final ensemble E* by selection of nets from subsets E_i

Unsupervised learning for subset identification

This step of our choice phase is implemented by a clustering algorithm that basically groups the neural networks belonging to set E according to the compound error probability in agreement with equation 4. In order to explain better such "clustering of neural nets", it is worth noting the analogy with the well known problem of "data clustering" [10]. In our task, the nets belonging to set E play the role of the "data", while the subsets E_i represent the data "clusters". Analogously, compound error probability between two nets plays the role of distance measure used in data clustering. In particular, in order to perform such clustering of neural networks, it can easily be seen that two "distance" measures are necessary: one distance measure between two nets and another between two clusters of nets. We defined the first measure on the basis of compound error probability:

$$n_s, n_t \quad E \quad d(n_s, n_t) = 1 - \text{prob}(n_s \text{ fails}, n_t \text{ fails}) \qquad (5)$$

According to equation 5, the "further" the two nets, the fewer the coincident errors. Therefore, the above distance measure is aimed to group nets that make a large number of coincident errors, and to assign error-independent nets to different clusters.

The "distance" between any two clusters E_i and E_j was defined as the maximum "distance" between two nets belonging to such clusters:

$$E_i, E_j \quad i \neq j \quad d(E_i, E_j) = \max_{n_s \; E_i, n_t \; E_j} \{d(n_s, n_t)\} \qquad (6)$$

The rationale behind equation 6 can be seen on observing that two clusters containing two nets that make different errors must never be merged, not even if other nets belonging to such clusters are highly correlated, since the subset E* is then formed by extracting the most independent nets from each cluster. It is also worth noting that the same kind of distance measure is commonly used for data clustering purposes [10]. Finally, it is easy to see that equation 6 can also be used to measure the distance between a net and a previously formed cluster.

In our experiments, a hierarchical agglomerative clustering algorithm using the two above distance measures has been adopted [10]. Such an algorithm starts assigning to an individual cluster each of the N nets forming set E. Two or more of these trivial clusters are then merged, thus forming a second partition. The process is repeated to form a sequence of nested clusters. The distance measures are computed with respect to a validation set, in order to avoid "overfitting" problems. Further details on hierarchical agglomerative clustering can be found in [10]. It is worth noting that different clustering algorithms could also be adopted, as long as distance measures based on the compound error probability are used.

*Creation of the ensemble E**

At each iteration of the clustering algorithm, a candidate ensemble E* is created by taking one net from each of the M clusters formed. (It is easy to see that the size of the candidate ensembles varies during the clustering process, since a different number of clusters is formed at each iteration). In particular, for each cluster, the net that exhibits the maximum average distance from all other clusters is chosen. The distance between one net and one cluster is computed according to equation 6. For each candidate ensemble E*, the M nets are then combined by majority voting and the classification accuracy is computed on a validation data set. Finally, the performances of all the ensembles created during the clustering algorithm are compared and the one with the highest performance is chosen. It is worth noting that, in the worst case, a number of candidate ensembles equal to N is created during the clustering process. Therefore, our approach exhibits limited computational complexity.

3.3 Discussion

3.3.1 The hypotheses made

Our design approach is based on the assumption that the ensemble E created in the overproduction phase can be regarded as the union of disjoint clusters of nets (equations 2 and 3). In addition, the compound error probability between any two nets belonging to the same cluster should be higher than that between any two nets belonging to different clusters (equation 4). From an application viewpoint, we are therefore assuming that the overproduction phase generates an ensemble E formed by clusters of nets characterised by high within-cluster error correlations and low between-cluster error correlations. Such an assumption can be deemed very realistic because, as pointed out by many researchers [2,6], the methods used in the overproduction phase basically create neural networks related to different "local minima" of the error function used in network training processes. Therefore, we can assume that set E is formed by clusters of nets related to such local minima. With regard to the condition expressed by equation 4, according to the experiments reported in the literature, the nets belonging to a given local minimum are surely error correlated. On the other hand, as pointed out by Hansen and Salamon and other researchers [2,6], different local minima of the error function correspond to different ways of forming generalisations about the training set patterns, and consequently nets related to different minima can be regarded error-independent.

3.3.2 The effectiveness of the proposed approach

In order to discuss the effectiveness of our design approach, let us define the following "optimality criterion" for the solution provided by the choice phase:

$$S \quad E,\ S \neq E^* \quad p(\text{ensemble}(S)\text{ fails}) \geq p(\text{ensemble}(E^*)\text{ fails}) \qquad (7)$$

where the terms p(ensemble(S) fails) and p(ensemble(E*) fails) are the error probabilities of the ensembles based respectively on the combinations of the nets forming the ensembles S and E*. E* is the ensemble generated by our design approach, while the term S is any other subset of set E.

Equation 7 states that any subset of set E different from subset E* provides a higher error probability. Accordingly, in order to show that it is optimal, we should prove that our design approach provides the ensemble E* that satisfies equation 7.

It is easy to see that our choice phase is more likely to provide the optimal ensemble E* the more error correlated the nets belonging to the same subset and the higher the degree of error diversity exhibited by the nets belonging to different subsets. Such a claim can be seen on analysing the following "limit" case:

- nets belonging to a given cluster are completely error correlated (i.e., they make exactly the same errors);
- nets belonging to different clusters are totally independent (i.e., they make no coincident errors).

It is easy to see that, under the above limit conditions, the optimal ensemble E* satisfying equation 7 can be created by randomly choosing one net from each cluster. On the other hand, it is also quite clear that, in such a limit case, our choice phase can provide such an optimal ensemble. In fact, it is trivial for our clustering algorithm to identify the correct clusters, and our "extraction" step can obviously create the optimal ensemble E*.

Therefore, we can say that our design approach is more likely to provide the optimal solution of the design problem, the more the clusters formed by error correlated nets, and the more error-independent nets belonging to different clusters. In other words, the performances of our design approach depend on "compactness" and "separation" of the clusters in terms of error correlation among nets. A formal proof of the optimality of our approach under this kind of assumptions is currently being developed.

4. Experimental Results

4.1 The data set

The data set used for our experiments consists of a set of multisensor remote-sensing images related to an agricultural area near the village of Feltwell (UK). The images, each 250 x 350 pixels, were acquired by two imaging sensors installed on an airplane: a multi-band optical sensor (an Airborne Thematic Mapper sensor) and a multi-channel radar sensor (a Synthetic Aperture Radar).

For our experiments, six bands of the optical sensors and nine channels of the radar sensor were selected. Therefore, we used a set of fifteen images. Images related to the different sensors were spatially registered by using one of the radar images as a reference and by scaling and registering the optical images with reference to it. Figure 1 shows the "registered" images related to a band of the optical sensor (Fig. 1(a)) and a channel of the radar sensor (Fig. 1(b)). As the image classification process was carried out on a "pixel basis", each pixel was characterised by a fifteen-element "feature vector" containing the brightness values in the six optical bands and over the nine radar channels considered. For our experiments, we selected 10944 pixels belonging to five agricultural classes (i.e., sugar beets, stubble, bare soil, potatoes, carrots) and randomly subdivided them into a training set (5124 pixels), a validation set (582 pixels), and a test set (5238 pixels). We used a small validation set to simulate real cases where validation data are difficult to obtain. Validation data are extracted from the training sets. Consequently, large reductions of training sets are necessary to obtain large validation sets.

It is worth noting that other remote-sensing researchers commonly use the selected data set, and it can be considered a "benchmark" for the considered application (i.e., the land-cover classification using multisensor remote-sensing images). More details about this data set can be found in [20,22,23].

4.2 Experimentation Planning

Our experiments were mainly aimed to:

- evaluate the effectiveness of the proposed design approach;
- compare our approach with other design approaches proposed in the literature.

Concerning the first aim, we performed different overproduction phases, thus creating different initial ensembles E. According to Section 3.1, such sets were created using different neural network types, namely, MultiLayer Perceptrons (MLPs), Radial Basis Functions (RBF) neural networks, and Probabilistic Neural Networks (PNNs). For the MLP and the RBF networks, ensembles were created by varying the network architecture and initial random weights. In the following, for the sake of brevity, we report the results related to three initial sets E , here referred to as sets E^1, E^2, and E^3, generated by overproduction phases:

- set E^1 was made up of fifty MLPs. Five architectures with one or two hidden layers and various numbers of neurons per layer were used in the overproduction phase. For each architecture, ten training processes with different initial random weights were performed. All the networks had fifteen input units and five output units corresponding to the number of input features and data classes, respectively (Section 4.1);
- set E^2 was made up of nineteen MLPs and one PNN. Two different architectures with two hidden layers and different numbers of neurons per layer were used for the creation of such MLPSs (15-7-7-5 and 15-30-15-5). For the PNN, according to the experience gained in our previous work [22], an a priori fixed value of the smoothing parameter equal to 0.1 was selected;
- set E^3 was made up of the same MLPs belonging to set E^2, three RBF neural networks, and the same PNN of set E^2. The three RBF networks were created with three different trials of the k-means clustering algorithm used to define the network architecture.

With regard to the second aim of our experimentation, we compared our design approach with another approach based on the "overproduce and choose" strategy, namely, the one proposed by Partridge and Yates [16]. Accordingly, only the comparisons of the choice phases were carried out. In particular, we considered the two types of choice phases proposed by Partridge and Yates. One type is the so called “choose the best”. Such a method assumes a given size of the final ensemble E*. Then, selects the networks with the highest classification accuracies from set E to form set E*. The other choice method proposed by Partridge and Yates is the so-called “choose from subspaces”. For each network type, it chooses the net exhibiting the highest accuracy. (It is worth noting that the term “subspace” is therefore referred to the subset of nets related to a given network type).

4.3 Results and Comparisons

Experimentation with set E^1

The main aim of this experiment was to evaluate the effectiveness of our approach in the design of neural network ensembles made up of an unique type of net. It is worth noting that this is a difficult design task, since nets of the same type are poorly independent according to the Partridge results [15]. Our approach selected an ensemble E* made up of seven MLPs belonging to three different architectures. It should be noted that this is not an obvious result, as one should expect a set E* to be formed by taking one net for each of the five architectures considered. Table 1 shows the performances of the combination of nets selected by our design approach. For comparison purposes, the performances of the combination of nets forming the initial set E^1 and those provided by the ensembles created by the Partridge and Yates design methods are also reported. All the considered ensembles were combined by the majority rule. It is worth noting that a size of set E* of five was chosen for the "choose the best" method. The performances were assessed in terms of percentage of classification accuracy, rejection percentage, and difference between accuracy and rejection. A pattern was rejected when a majority of nets assigning it to the same data class is not present.

All values reported in Table 1 are referred to the test set. Table 1 shows that the performances of all three design methods are quite similar. Our method is slightly better, but the difference is small. This result can be justified by the fact that the initial set E^1 also provides similar performances. This means that set E^1 does not contain error-independent networks that can be selected by a design method to improve performances.

Experimentation with set E^2

This experiment was aimed to evaluate to what extent our design method can exploit error-independence to improve the performances of a set of "weak" neural networks. We therefore selected an initial set of nineteen MLPs, whose performances were not good (ranging from 80.41% to 85.05%). However, such MLPs were based on two different architectures and different initial random weights for a good degree of error uncorrelation. In addition, in order to create set E^2, we added one PNN that can be expected to be "independent" of the MLPs. Our design approach extracted from set E^2 an ensemble E* made up of two MLPs, characterized by two different architectures, and the PNN. Table 2 shows the performances of the combinations by the majority rule of the networks forming the considered ensembles. All the reported values are referred to the test set. A size of set E* of three was chosen for the "choose the best" method. The results show that our design approach was able to choose more independent networks than other approaches. This can be explained by the fact that nets that can be combined effectively can be chosen from a set of weak networks, only by a detailed analysis of error uncorrelation. This kind of analysis is not carried out by other methods.

Experimentation with set E^3

The aim of this experiment is basically similar to the previous one. Our design approach extracted a set E* formed by one MLP, one RBF neural network, and the PNN. Table 3 shows the performances of the combinations of networks forming the considered ensembles. All values are referred to the test set. A size of set E* of three was chosen for the "choose the best" method. It is easy to see that since our design approach clearly outperformed the others, similar conclusions as in the previous experiment can be drawn.

5. Conclusions

In this paper, an approach to the automatic design of neural network ensembles has been proposed. To the best of our knowledge, in the pattern recognition field, no previous work has clearly addressed such a problem. As pointed out in Section 2, some work has been carried out by neural network researchers. The experimental results reported in this paper showed the effectiveness of the proposed design approach. We also compared our design approach with another recently proposed in the neural network literature. The comparison showed that with our approach it is possible to select the ensemble made up of the most error independent networks. In addition, as compared with other approaches described in the literature, our approach provides a systematic method to choose such a set. Finally, it is worth noting that the assumptions in our approach can be deemed realistic for practical applications and, under specific hypotheses, it can provide the optimal solution for the design task at hand.

Acknowledgements

This work was supported by the Italian Space Agency, within the framework of the project "Metodologie innovative di integrazione, gestione, analisi di dati da sensori spaziali per l'osservazione della idrosfera, dei fenomeni di precipitazione e del suolo".

References

[1] Battiti, R. and Colla A.M., Democracy in neural Nets: Voting Schemes for Classification, Neural Networks 7 (1994) 691-707.

[2] Bishop, C.M., Neural Networks for Pattern Recognition (Oxford Univ. Press, NY, 1996).

[3] Freund, Y. and Shapire R.E, A decision-theoretic generalisation of on-line learning and an application to boosting, Journal of Computer and System Sciences 55 (1997) 119-139.

[4] Giacinto, G. and Roli F., Ensembles of Neural Networks for Soft Classification of Remote Sensing Images, Proc. of the European Symposium on Intelligent Techniques, Bari, Italy, (1997) 166-170.

[5] Giacinto, G., Roli F., and L.Bruzzone, Combination of Neural and Statistical Algorithms for Supervised Classification of Remote-Sensing Images, Pattern Recognition Letters, vol. 21, no. 5, May 2000, pp. 385-397

[6] Hansen, L. K. and Salamon P., Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 993-1001.

[7] Ho, T.K., Hull J.J. and Srihari S.N., Decision Combination in Multiple Classifier Systems, IEEE Trans. on Pattern Analysis and Machine Intelligence 18 (1994) 66-75.

[8] Huang, Y.S. and Suen C.Y., A method of combining multiple experts for the recognition of unconstrained handwritten numerals, IEEE Trans. on Pattern Analysis and Machine Intelligence 17 (1995) 90-94.

[9] Huang, Y.S., Liu K. and Suen C. Y., The combination of multiple classifiers by a neural network approach, Int. Journal of Pattern Recognition and Artificial Intelligence 9 (1995) 579-597.

[10] Jain, A.K.and Dubes R.C., Algorithms for clustering data (Prentice Hall, 1988).

[11] Kittler, J. , Hatef M., Duin R.P.W. and Matas J., On Combining Classifiers, IEEE Trans. on Pattern Analysis and Machine Intelligence 20 (1998) 226-239.

[12] Lam, L. and Suen C.Y., Application of Majority Voting to Pattern Recognition: An Analysis of Its Behavior and Performance, IEEE Trans. on Systems, Man and Cybernetics-Part A 27 (1997) 553-568.

[13] Littlewood, B.and Miller D.R., Conceptual modelling of coincident failures in multiversion software, IEEE Trans. On Software Engineering 15 (1989) 1569-1614.

[14] Opitz, D.W. and Shavlik J.W., Actively searching for an effective neural network ensemble, Connection Science 8 (1996) 337-353.

[15] Partridge, D., Network generalization differences quantified, Neural Networks 9 (1996) 263-271.

[16] Partridge, D. and Yates W.B., Engineering multiversion neural-net systems, Neural Computation 8 (1996) 869-893.

[17] Proc. Final Workshop MAESTRO-1 European SAR Campaign, ESTEC, Noordwijk, The Netherlands (European Space Agency Pub., 1992).

[18] Raviv, Y. and Intrator N., Bootstraping with noise: an effective regularization technique, Connection Science 8 (1996) 355-372.

[19] Richards, J.A and Jia X., Remote Sensing Digital Image Analysis - An Introduction (3rd ed. Springer Verlag, 1999).

[20] Roli F., Multisensor image recognition by neural networks with understandable behaviour International Journal of Pattern Recognition and Artificial Intelligence 10 (1996) 887-917.

[21] Rosen, B.E., Ensemble learning using decorrelated neural networks, Connection Science 8 (1996) 373-383.

[22] Serpico. S.B., Bruzzone L. and Roli F., An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images, Pattern Recognition Letters 17 (1996) 1331-1341.

[23] Serpico, S.B., and Roli F., Classification of multi-sensor remote-sensing images by structured neural networks, IEEE Trans. Geoscience Remote Sensing 33 (1995) 562-578.

[24] Sharkey, A.J.C. (Ed.), Special Issue: Combining Artificial Neural Nets: Ensemble Approaches, Connection Science 8 (1996).

[25] Sharkey, A.J.C., On Combining Artificial Neural Nets, Connection Science 8 (1996) 299-314.

[26] Sharkey, A.J.C. and Sharkey N.E., Combining Diverse Neural Nets, The Knowledge Engineering Review 12 (1997) 231-247.

[27] Tumer, K. and Ghosh J., Error correlation and error reduction in ensemble classifiers, Connection Science 8 (1996) 385-404.

[28] Wolpert, D.H., Stacked generalisation, Neural Networks 5 (1992) 241-259.

[29] Xu, L., Krzyzak A., and Suen C.Y., Methods for combining multiple classifiers and their applications to handwriting recognition, IEEE Trans. on Systems, Man, and Cyb. 22 (1992) 418-435.

TABLE CAPTIONS

Table 1 - The performances of the combination of the nets selected by our design approach are reported with reference to set E^1. For comparison purposes, the performances of the combination of nets forming the initial set E^1 and the ones provided by the ensembles selected by the Partridge and Yates design methods are also reported ("choose the best" and "choose from subspaces"). All considered ensembles were combined by the majority rule.

Table 2 - The performances of the combination of the nets selected by our design approach are reported with reference to set E^2. For comparison purposes, the performances of the combination of nets forming the initial set E^2 and the ones provided by the ensembles selected by the Partridge and Yates design methods are also reported ("choose the best" and "choose from subspaces"). All considered ensembles were combined by the majority rule.

Table 3 - The performances of the combination of the nets selected by our design approach are reported with reference to set E^3. For comparison purposes, the performances of the combination of nets forming the initial set E^3 and the ones provided by the ensembles selected by the Partridge and Yates design methods are also reported ("choose the best" and "choose from subspaces"). All considered ensembles were combined by the majority rule.

TABLE 1

Ensemble created by	%Accuracy	%Rejection	%(Accuracy-Rejection)
Our design method	90.52	0.82	89.70
Choose the best	90.10	0.49	89.60
Choose from subspaces	89.98	0.49	89.48
E^1	89.83	1.20	88.63

TABLE 2

Ensemble created by	%Accuracy	%Rejection	%(Accuracy-Rejection)
Our design method	91.31	2.20	89.11
Choose the best	88.78	1.65	87.13
Choose from subspaces	89.35	1.57	87.78
E^2	87.87	2.37	85.50

TABLE 3

Ensemble created by	%Accuracy	%Rejection	%(Accuracy-Rejection)
Our design method	94.83	4.71	90.11
Choose the best	88.78	1.65	87.13
Choose from subspaces	89.35	1.57	87.78
E^3	90.46	3.05	87.41

FIGURE CAPTIONS

Figure 1. Multisensor images used for experiments: (a) image acquired by the Airborne Thematic Mapper sensor in “channel 9” [29] and (b) image acquired by the radar sensor (SAR) in “band L”, using HV polarization [28].

Giorgio Giacinto and Fabio Roli

Design of Effective Neural Network Ensembles for Image Classification Purposes

Fig. 1a

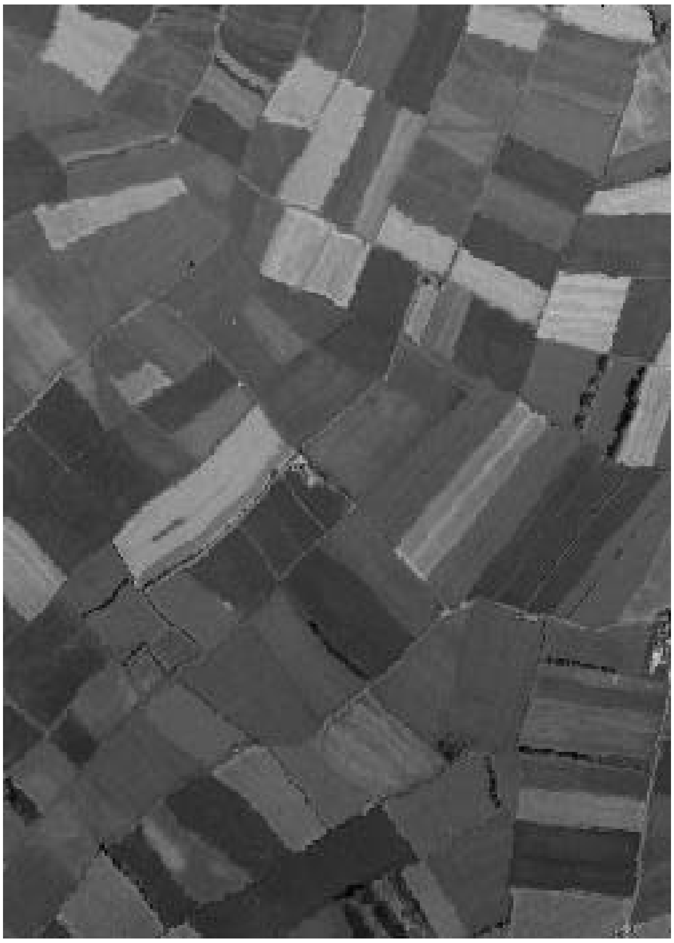

Image and Vision Computing

Giorgio Giacinto and Fabio Roli

Design of Effective Neural Network Ensembles for Image Classification Purposes

Fig. 1b