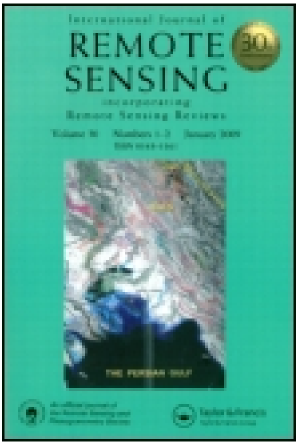


International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Neural network classification and novelty detection

M. F. Augusteijn & B. A. Folkert
Published online: 25 Nov 2010.

INT. J. REMOTE SENSING, 2002, VOL. 23, NO. 14, 2891–2902

Taylor & Francis
Taylor & Francis Group

Neural network classification and novelty detection

M. F. AUGUSTEIJN* and B. A. FOLKERT

Computer Science Department, University of Colorado, Colorado Springs, Colorado Springs, CO 80933, USA

(*Received 11 February 1999; in final form 25 January 2001*)

Abstract. Novel data are data belonging to classes not included in the training set of the classifier. Neural network classifiers tend to put these data into the category of patterns that most resemble the novel ones, rather than label them as novel or unknown. This work investigates the ability of the back-propagation neural network architecture to detect novel patterns and concludes that this method is unsuitable for this task. It also explores the applicability of a different neural network architecture, the probabilistic neural network, and finds that this method shows superior performance as an overall classifier when compared to back-propagation and, in addition, is able to identify novel patterns.

1. Introduction

The application of a neural network to classify various ground covers in remotely sensed, multispectral imagery has become an important alternative to classification by means of a maximum-likelihood classifier. The back-propagation neural network architecture (Rumelhart *et al.* 1986) is most frequently used for this purpose. Paola and Schowengerdt have performed an extensive comparison between back-propagation and maximum-likelihood classification applied to a ground cover identification task using six-band Thematic Mapper (TM) imagery (Paola and Schowengerdt 1995). They conclude that the back-propagation network, being non-parametric, appears more robust to training site selection and class definition, and that this classifier more easily accommodates heterogeneous categories such as 'urban residential' in which several ground cover signatures are mixed. They state that the maximum likelihood algorithm, on the other hand, is sensitive to the purity of the class signatures and performs poorly if these signatures are not pure. They also conclude that the fuzzy output of the neural network is related to class likelihood and may be at least as good an indicator of significant class mixing as maximum likelihood class density and *a posteriori* probability. The relationship between the output values of a back-propagation neural network and the *a posteriori* probabilities of the classes has previously been investigated (Richard and Lippmann 1991). Conclusions are based on experimental evidence, and this relationship has not been established in a theoretical manner.

*e-mail: mfa@cs.uccs.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160110055804

A problem related to the classification of mixed pixels and the estimation of *a posteriori* class probability is the recognition of novel patterns. Both back-propagation neural networks and maximum likelihood classifiers rely on supervised classification and make use of a training set of patterns with known categorization. Thus, the various ground cover categories must be identified in advance, and the classifier can only associate patterns with these known categories. However, small classes may be missed or new classes may unexpectedly appear. For example, it has been observed that rocks covered with bird droppings have significantly different spectral response patterns compared to clean rocks, but a classification scheme may not include a 'bird droppings' category. Ground cover identification quite generally is plagued with the problem that it may not be possible to assemble a complete training set of all possible ground covers present in some imagery. Novelty detection is the ability of a classifier to flag patterns significantly different from those in the training set as 'unknown' or 'novel'.

Novelty detection has been studied in a variety of applications. Parra *et al.*, for example, described the use of information preserving nonlinear maps to determine when a pattern is significantly different from those used for training and applied this method to motor fault detection. Their task consisted of noting early irregularities in electrical motors by monitoring the electrical current. The failure detector was trained with data supplied by a healthy motor and should indicate if the motor is going to fail (Parra *et al.* 1996). Roberts and Tarassenko grew a Gaussian mixture model to form a representation of a training set of 'normal' system states. They used a threshold to discriminate 'normal' from 'novel' patterns. Their application was medical signal processing; in particular, the detection of epileptic seizures (Roberts and Tarassenko 1994). Washburne *et al.* used a probabilistic neural network (PNN) to identify ground covers in remotely sensed imagery that were significantly different from a sample set stored in the PNN. They mentioned the problem of discriminating between land mines and various types of rocks in infrared imagery (Washburne *et al.* 1993).

Back-propagation, like most neural network architectures, will not automatically flag novel patterns as unknown. Instead, the network will attempt to classify each pattern into the closest matching category. This property, known as generalization, is one of the most important attributes of neural network classification and accounts for the ability of the network to deal appropriately with mixed pixels. It may be hypothesized though that if output values are indicative of *a posteriori* probabilities of class membership then these values could possibly be used to identify novel patterns. The idea is that a threshold may be set on the network's output values, and if no value exceeds this threshold then the pattern would be considered unknown. It should be noted, however, that previous work concerning the interpretation of output values as *a posteriori* probabilities has not addressed the situation of a network being presented with a pattern of a category not included in the training set (Richard and Lippmann 1991).

The purpose of our investigation was to determine if the output values of a back-propagation network could indeed flag the occurrence of novel patterns in ground cover classification of multispectral imagery. In addition, we also explored the capability of another network architecture, the PNN (Specht 1990). This architecture is somewhat related to the maximum likelihood classifier and calculates quasi class membership probabilities based on a training set of patterns with known categorization. Both neural network classifiers are applied to six-band (the thermal band is excluded) TM imagery of the United States Air Force Academy (USAFA) grounds located near Colorado Springs.

2. The neural networks and novelty detection

The standard back-propagation architecture is used in this study configured with a single hidden layer and a tangent-hyperbolic activation function for both hidden and output layers. This kind of architecture is considered a global network; the activation function of each unit serves as a separator partitioning pattern space into an infinite sub-space causing excitatory output (output value near +1) and another infinite sub-space causing inhibitory output (output value near –1) of this unit. The output layer has a unit for each category in the training set. Output units are trained to produce a value close to 1 when a pattern belonging to its corresponding class is presented and a value close to –1 in all other cases. During the test phase and in the absence of novelty detection, the unit with the highest output value is assumed to indicate the class of the input pattern.

Novelty detection may be implemented in this architecture in one of two ways. One scenario is to only consider the output of the unit showing the highest activation. If this activation is below a preset threshold (as selected by the user) then the pattern may be considered to belong to a novel category. The other scenario takes the entire output pattern into account. The distance between this output pattern and each one of the target patterns (the patterns used as targets during training) is calculated, and if the smallest of these distances is above a preset threshold then the input pattern is again considered to belong to a novel category. Both scenarios are explored in this study. Since the back-propagation network is well known and has been used in many studies in remote sensing, there is no need to repeat the formulas used to calculate the classification.

The PNN or Parzen classifier (Specht 1990) does not use an iterative training algorithm, but is constructed given a set of training patterns. It consists of an input layer, a pattern layer and a summation or output layer, as shown in figure 1. In its simplest form, this architecture contains as many pattern units as there are patterns in the training set, and the connections to these units are weighted with the feature values of these training patterns. This may lead to a large network. Alternatively,

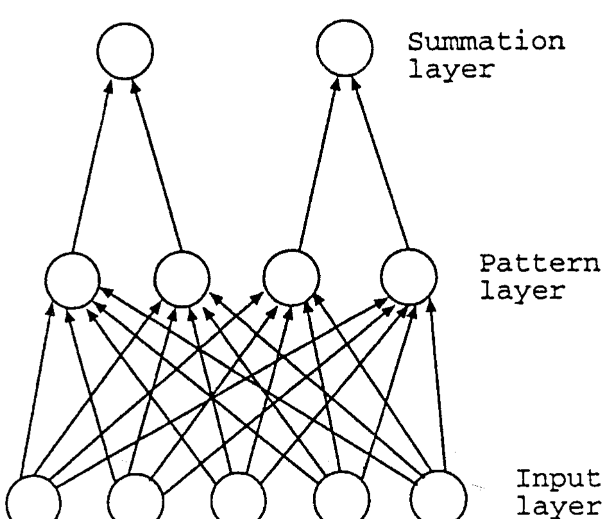


Figure 1. The architecture of the probabilistic neural network.

patterns belonging to the same category may first be clustered, and the cluster centre values may then be employed as connection weights.

Each pattern unit uses a Gaussian activation function of the form:

$$f(\bar{x}) \sim \exp\left(-\frac{(\bar{x}-\bar{x}_i)^2}{2\sigma^2}\right) \tag{1}$$

The symbol $\bar{x}$ denotes the pattern to be classified, $\bar{x}_i$ is the unit's weight vector (the training pattern or cluster centroid associated with that unit) and σ is a smoothing parameter set by the user. The summation layer contains as many units as there are categories in the classification task. The pattern and summation layers are not completely connected as shown in figure 1; only those pattern units belonging to the same category are connected to the summation unit corresponding to that category. These connections all carry a weight value 1; therefore, each unit in the summation layer simply adds the output values of the pattern units of its class, and this sum becomes its output value. When presented with a test pattern, each summation unit will calculate a quasi-probability of the pattern to belong to the associated category. These summation units may also incorporate a *prior* probability and a cost of misclassification associated with the category (Specht 1990). The classification of a test pattern is then determined by the summation unit showing the highest activation: it is assumed to belong to the category associated with that unit.

Pattern classification by means of a PNN is somewhat reminiscent to maximum likelihood classification in the sense that both use a mixture of Gaussians to estimate a probability measure of class membership for a given pattern and select the class for which this measure is highest. There are obvious differences. The PNN does not estimate Gaussian widths from the covariance matrix of the training data; it uses a constant smoothing parameter instead. The value of this parameter determines the amount of generalization present in the network. With little smoothing, the activation of a pattern unit drops off steeply when moving away from its centre, while a more gradual decrease in activation is seen when this parameter is given a larger value. The PNN architecture has this feature in common with the Radial Basis Function network, which, in its simplest form, also uses a constant width for its Gaussian activation functions (Moody and Darken 1989). The PNN has the important advantage of being simple to construct while avoiding the lengthy training sequences generally required for a back-propagation network. It is also easy to use for the novice practitioner of neural network technology since its execution requires the estimation of a single parameter value, and it has been shown that network performance does not crucially depend on this parameter value (Specht 1990).

In comparing the two neural network architectures, the PNN and back-propagation, it will be evident that the classification mechanism is quite different. While back-propagation is a global classifier, the PNN uses local support. The activation of each summation unit in the PNN architecture is determined by a mixture of Gaussians and will therefore only be significantly different from zero in a finite sub-space of the pattern space. Outside its sub-space, the summation unit will not measurably contribute to the classification. The detection of a novel pattern is also rather straightforward in this architecture: if the highest output value calculated by the summation layer is below a preset value then the pattern is assumed to belong to a class not represented in the PNN.

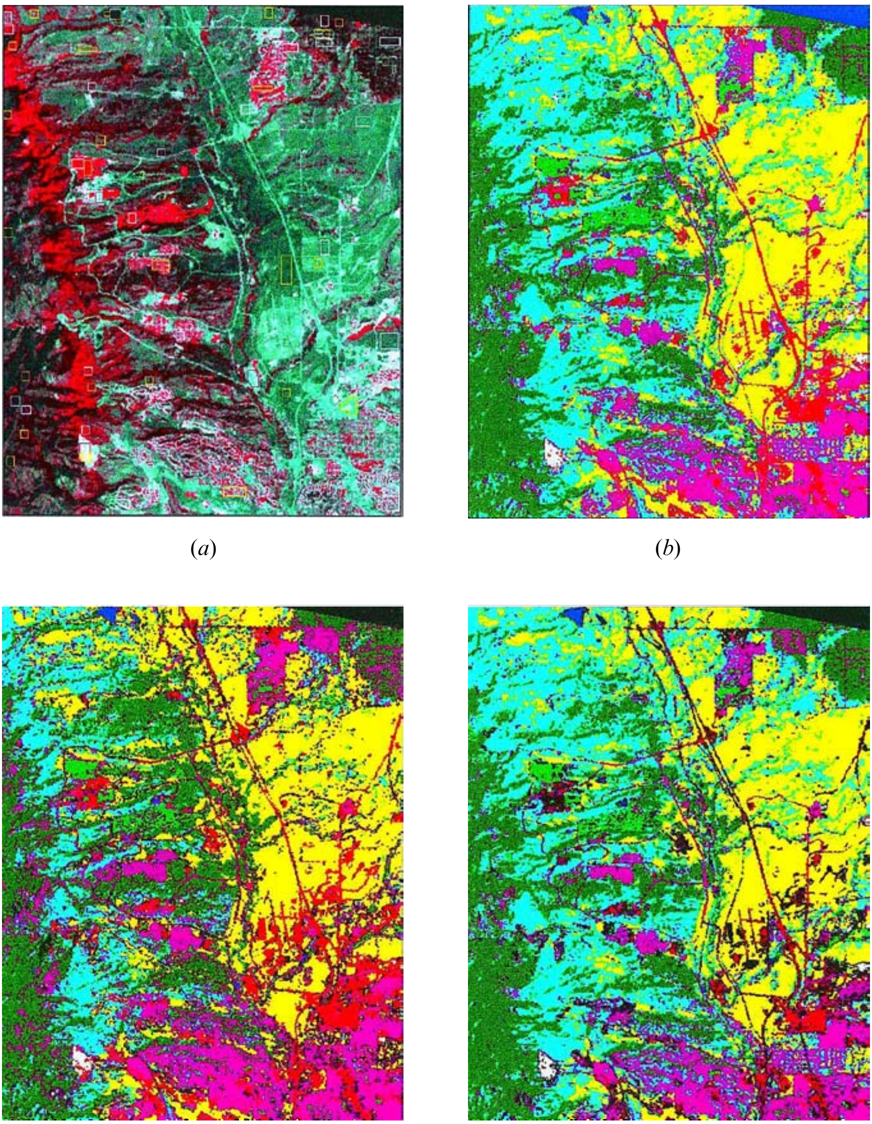
3. Neural network classification of remotely sensed data

A multispectral TM image of the United States Air Force Academy and surrounding grounds, acquired in the spring of 1993, provided the test data for the classification experiments. Six spectral bands were used; the thermal band was excluded. This image has also been used in an earlier classification study, and some of the results reported here may be compared with the ones obtained in this previous study (Augusteijn and Warrender 1999). A false colour image showing the near-infrared band mapped to red, the red band mapped to green and the green band mapped to blue is displayed in figure 2(*a*). Test and training sites are superimposed on this image. Some of its characteristic features are the golf course near the centre of the image, the relatively large areas of maintained grass of the cadet parade grounds and athletic fields north of the golf course, and the speckled housing areas on the Academy grounds as well as in the city of Colorado Springs (lower right corner) and the Black Forest area (upper right corner). The left edge of the image shows the foothills of the Rocky Mountains covered by Pike Forest. Palmer Lake is just visible near the top of the image and some smaller lakes can be found near the golf course area. The major highway (I-25) is surrounded by meadow shown as greenish areas in this spring image. The black region in the top right corner denotes a region for which no image data were available. The image covered roughly 203 km^2 and was registered as 539 by 419 pixels at 30 m resolution.

The land cover classification of this image employed eight categories: water, forest, scrub, meadow, rock, urban, lawn and housing. Scrub areas consist primarily of scrub oak, meadows are covered by wild grass, possibly with low shrubs as well, lawn is maintained grass, urban refers to large buildings and roads, while housing consists of residential areas covered by a mixture of buildings, roads, lawns, trees and shrubs.

Both types of neural networks were trained and tested using the same pattern sets acquired from the training and test sites. Ground truthing of these sites was performed by means of site visits. Each pattern consisted of six components representing reflectance values of a pixel in each one of the six bands scaled to a range from 0.0 to 1.0. The dataset obtained from these sites contained 1600 training patterns (200 per category) and 2439 test patterns. An equal number of training patterns was chosen for each class so that all classes were learned equally well. As there were some classes with very little available data (in particular, the water and rock class), some pixels were used more than once for these classes to get the required amount of data. All pixels available in the test sites were used to establish performance for the classifiers. Since the test data were not equally distributed over the categories, certain classes were more heavily weighted in the overall test results, but these were the classes more abundantly appearing in the image.

The code for the back-propagation architecture was developed at the University of Colorado, Colorado Springs. The network was configured with six input units, corresponding to the six pattern features and eight output units, one for each of the eight categories. Many training runs were performed to find appropriate parameter values; it was found that a network configured with 11 hidden units, combined with a learning rate of 0.045 and without a momentum term gave the best results. When the error threshold was set at 0.005 training of this network took from 30 to 60 minutes on a 100 MHz Pentium PC. The network's training and test performance is shown in the first row of table 1. The entire image was then classified by associating each pixel with the closest category based on the network's output values. The



(a) (b) (c) (d)

Figure 2. United States Air Force Academy grounds, north of Colorado Springs, CO, USA. (a) False colour image of the scene with training (white) and test (yellow) polygons, (b) classified map obtained with the PNN architecture without novel pattern identification, (c) classified map obtained with the back-propagation architecture including novel patterns, (d) classified map obtained with the PNN architecture including novel patterns. The colour palette for the classification maps is as follows: dark blue: water; dark green: forest; cyan: scrub; yellow: meadow; white: rock; red: urban; magenta: housin; and light green: lawn. The black areas denote novel patterns.

resulting classification map (not shown) looked plausible, but displayed several housing areas in the Pike Forest. This region is not inhabited, but consists of a mixture of bare soil, rock, trees and other vegetation. The spectral signatures of this class and the housing category are likely to be similar. The map also showed the

Table 1. Training and test performance of the neural network architectures.

Network architecture	Training performance (%)	Test performance (%)
Back-propagation	97.8	90.2
PNN without clustering	100	90.1
PNN with Kohonen clustering	93.5	91.2
PNN with Kohonen and LVQ clustering	96.0	97.0

region without image data classified as water. Since water has the lowest reflectance of all categories represented by this dataset, it is logical that a region showing no reflection is included in the water class.

The code for the PNN architecture was also developed at the University of Colorado, Colorado Springs. The first PNN was configured with 1600 pattern units, one for each training sample. Extensive experimentation with the smoothing parameter showed that smaller values resulted in better performance than larger ones. Eventually, a value $\delta = 0.035$ was chosen for all experiments. Training accuracy of this network was 100%, as expected, and the test set accuracy was comparable to that of back-propagation, as shown in table 1. However, this network is rather large, and pattern clustering was expected to lead to better performance. Clustering of the training data was performed by means of another neural network architecture: the Kohonen network (Kohonen 1982). This network provides unsupervised partitioning of the data into a preset number of clusters. Because of the unsupervised nature of this algorithm, it could only be applied to each category separately since each cluster must consist of patterns belonging to the same category. The number of clusters selected for each class was 50; thus, a PNN architecture was now configured with 400 cluster units, replacing the 1600 pattern units used before. Training performance of this network decreased while test performance improved slightly, as shown in table 1.

It was soon realized that the set of clusters obtained in this manner had a severe problem. Because the clusters of the various classes were generated separately, the sets of clusters tended to overlap. There was no overall partitioning of pattern space into regions containing patterns belonging to different classes. This could only be accomplished by means of a supervised clustering algorithm using the complete set of training samples. Learning Vector Quantization (LVQ) is appropriate for this task, but this architecture only trains well when properly initialized (Fausett 1994). It was decided to retain the number of 400 clusters and use the cluster centres resulting from the Kohonen network as initial values for LVQ. In this manner, LVQ was able to fine-tune these clusters and to separate the ones belonging to different categories resulting in a better, overall partitioning of the pattern space. When these cluster centres were used as the weights for the pattern units in the PNN both training and test performance increased considerably, as shown in table 1. This network was also used to classify the entire image; the resulting classification map is shown in figure 2(*b*). The PNN placed fewer houses in Pike forest than the back-propagation architecture, and its classification may therefore be considered more accurate.

4. Neural network novelty detection

As mentioned before, there are two methods by which back-propagation, at least in principle, can be made to identify patterns of categories not included in its training set. One method uses a threshold on the highest output value and declares a pattern as novel if this value remains below the threshold; the other method calculates the distance between the output pattern and all target patterns and proclaims a pattern as novel if the minimum distance is found to exceed a preset threshold. Both approaches were investigated and were found to lead to similar results. However, the back-propagation network is a global classifier; positive and negative output values are learned simultaneously and may therefore be considered equal in importance. Especially when the network is presented with data significantly different from all training samples, it seems advantageous to take the entire output pattern into account, not just the highest output value. Therefore, the Euclidean distance between the output pattern and target patterns was selected as the preferred criterion to identify novel patterns. This criterion is used in the following experiments.

An appropriate threshold value must now be chosen. This threshold determines the maximum distance between the pattern's output vector and the target vector of the nearest category that will allow the pattern to be classified in that category. Therefore, the threshold value serves as an additional criterion for classification: as a smaller value is selected, fewer patterns will be included into known categories and more will be labelled as novel. Effectively, the threshold value will decrease the generalization capability of the back-propagation network. Obviously, a very small threshold value will render the network incapable of any useful generalization. This observation also shows how an appropriate value may be chosen: the effect of the threshold value on test set classification must be established. The test patterns are all acquired from sites that have been visited, and ideally, none of them should be classified as novel. However, the threshold will cause some of these test patterns, previously classified correctly, to be excluded from their category because the distance between their output values and their category's target is above the threshold. A second threshold may now be set on the percentage of test patterns that is allowed to change from their correct category to this new novel designation. This percentage, in effect, constitutes a decrease in performance of the classifier and should therefore be kept small. This second threshold will be called the allowed test-change percentage: values varying from 1 to 5% were used in this study.

Once this percentage is set, the value of the minimum distance threshold can be experimentally determined from the test data. This distance threshold is then applied to the image, and all pixels now identified as novel are changed to black in the classification map. Table 2 shows the number of patterns classified as novel in the test data and the percentage of novel patterns in the image for several allowed test-change percentages. The classification map produced by the back-propagation network with an allowed test-change of 1% is shown as figure 2(*c*). It is observed that the non-imaged portion in the upper right corner is indeed identified as novel. Other novel pixels are primarily found along the border between two categories where the signatures of the image data are likely to be mixed. Some larger regions classified as novel are found in the Pike Forest. Comparing these regions with the imagery, it is seen that they are relatively dark compared to the surrounding area. Pike Forest contains many steep rock formations, and it is believed that these dark patches are caused by shadows cast by these rocks rather than by any novel ground cover.

A slightly modified procedure was followed when using the PNN architecture as a classifier. Here, the output values may be treated as quasi-probabilities, and a threshold is set on the largest value in the output of the network (the value indicating class membership). Only if this value exceeds the threshold will the pattern be

Table 2. Number of patterns classified as novel in the test set and percentage of novel patterns detected in the imagery depending on the decrease in test performance for the two network architectures.

Network architecture	Allowed test-change (decrease in test performance) (%)	Number of novel test patterns detected	Novel patterns in image (%)
Back-propagation	0.25	5	2.2
Back-propagation	0.50	11	4.7
Back-propagation	1.00	22	6.6
Back-propagation	3.00	65	13.3
Back-propagation	5.00	110	17.2
PNN	0.25	5	6.8
PNN	0.50	9	8.5
PNN	1.00	23	15.6
PNN	3.00	71	25.9
PNN	5.00	119	30.2

included into the category indicated by the associated output unit. The magnitude of this threshold is again determined by the allowed test-change percentage, measured as before. This threshold is then applied to the image data. Table 2 also contains the results for the PNN classification. It is seen that the PNN marks many more image pixels as novel than back-propagation for the same test-change percentage. A very low test-change threshold (0.25%) was used to obtain the classification map produced by the PNN (figure 2(*d*)). It is seen that the non-imaged area is again marked as novel as are many boundaries between two regions. Although the PNN also labelled some regions in the Pike Forest as novel, the area of these regions is less when compared with the map produced by back-propagation (figure 2(*c*)). The most significant difference displayed by the PNN map consists of the many novel patches found throughout the image. Close observations reveal that these patches are predominantly found in the housing class. This class is expected to produce the most heterogeneous spectral signatures. It is plausible to assume that the PNN is more capable than back-propagation to detect those regions within a heterogeneous class that show spectral signatures most different from those included in the training set.

When comparing the two classification maps of figure 2(*c*) and (*d*), it may be concluded that both architectures are capable to mark spectral signatures sufficiently different from those in the training set as 'unknown' or 'novel', although the PNN seems to be doing a better job. However, neither architecture found a large region of truly novel ground cover, most likely because such a region did not exist in the available data. In order to investigate how the two neural network classifiers might deal with an unknown ground cover, one ground cover was deliberately omitted from the training set. Maps produced by these partially trained classifiers can then be compared to the (complete) classification displayed in figure 2(*b*). A more thorough comparison of the capabilities of the two classifiers can be obtained in this manner.

The first class selected for omission was the water category. This class was chosen because its signature is significantly different from any of the other ground cover signatures. Water was found to be easily classified in the imagery and is only found in some small, well-defined regions. First, a back-propagation network was trained with the data of the remaining seven classes, and this network was subsequently used to classify the image while imposing different 'test change thresholds'. The results were surprising. At a threshold of 0.5% almost all novel pixels were found at the border of two categories of the imaged part of the data. No novel pixels were found in the water region; Palmer Lake as well as the two ponds near the golf course were classified as 'forest'. This classification did not change when the test-change threshold was increased; even at a threshold of 5% the network still insisted it found 'forest' covering the water regions. The PNN, on the other hand, when trained without the water class, did mark all water regions as novel at all thresholds used in the experiments.

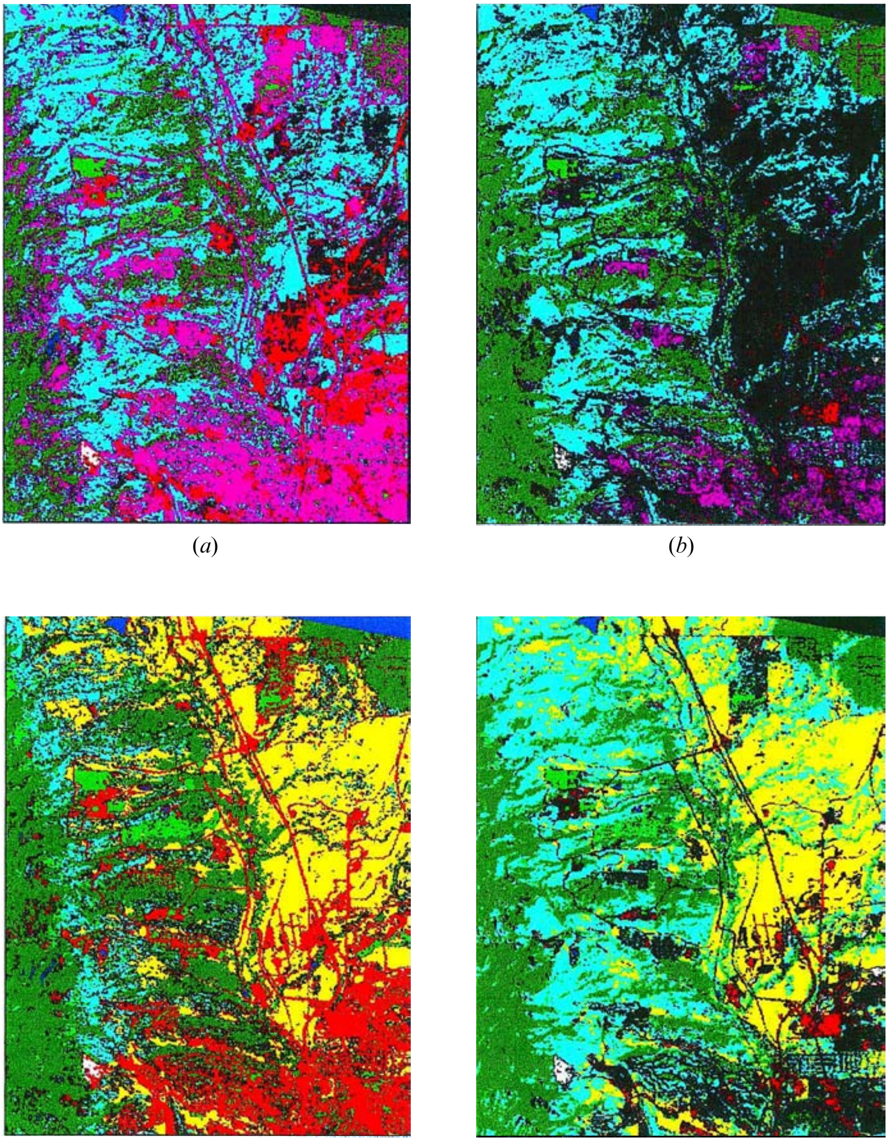
The next class to be omitted from the training data was meadow. This is a large class, but it was found to be accurately labelled in the test data. A back-propagation network trained with seven categories excluding the meadow patterns was used to classify the imagery with a test-change threshold of 5%. The resulting map is shown in figure 3(*a*). Even though some of the meadow region is now classified as novel, most of the meadow is put into the 'scrub' class and some is now labelled as 'housing'. This misclassification is somewhat plausible since some scrub is found mixed with meadow, still back-propagation seems to over-generalize in its abundant detection of scrub. The PNN again presents us with a much more accurate picture, as shown in figure 3(*b*), which displays the classification map obtained with a 3% threshold. It is seen from this figure that almost all former meadow regions have been blacked out. It should be noted, however, that the omission of this large class did cause some confusion among the remaining classes. This is particularly evident in regions where urban patterns border meadow; many of the urban patterns are also classified as novel. This must in part be attributed to the usage of an, at least for the PNN, large threshold value. A lower threshold was indeed found to alleviate this problem but also reduced the effectiveness of identifying meadow patterns as novel.

The last class to be removed was housing. This class is also well represented in the image data and was found to be the most difficult class to identify because of its heterogeneous signature. The results produced by a back-propagation network with a threshold value of 5% are shown in figure 3(*c*). This time, it is seen that back-propagation moved many of the housing patterns to the 'urban' category, presumably the most similar remaining class. The PNN, on the other hand, showed good results when applying a small threshold value. A map produced by the PNN with a threshold value of 0.25% is shown as figure 3(*d*). It should be noticed that even though most of the housing is now flagged as novel some urban areas are also no longer recognized. This again is indicative of the similarity between housing and urban regions; the largest buildings, however, are still correctly labelled as urban in figure 3(*d*).

5. Conclusion

This research investigated the ability of a neural network classifier to recognize new ground covers; that is, ground covers not represented in the training set. Two very different architectures were used, the back-propagation network, which is the most well known and most often applied architecture, and the less well known PNN. The networks were also compared with respect to their ability to classify known ground covers.

The PNN is the overall winner of this comparison. It showed better performance on a test set of patterns obtained from sites ground truthed by means of site visits. It produced an overall classification map that was believed to be more accurate than the one generated by back-propagation because it placed less housing in the Pike Forest which is known to be uninhabited. The PNN was also significantly more



(a) (b)

(c) (d)

Figure 3. United States Air Force Academy grounds, north of Colorado Springs, CO, USA. (a) Classification by means of the back-propagation architecture using meadow as a novel class; (b) similar to (a) using housing as a novel class; (c) classification by means of the PNN architecture using meadow as a novel class; (d) similar to (c) using housing as a novel class. The colour palette is the same as used in figure 2.

successful in marking new ground covers as novel after a known ground cover had been deliberately omitted from the training set. Back-propagation, on the other hand, almost without exception attempted to place new ground covers into the known category most similar to the new one.

The inability of back-propagation to recognize the limits of what it knows and doesn't know is closely related to the manner in which it learns a training set.

Back-propagation learns to separate categories by placing hyper-planes in pattern space between patterns identified as belonging to different classes. Regions without patterns do not receive these hyper-planes and are inadvertently attached to regions that do contain patterns, most likely to those regions with patterns most similar. In this manner, back-propagation is able to extend its classification to regions away from its training patterns. This is sometimes advantageous, but it prevents back-propagation to correctly mark patterns considerable different from its training set as novel. The PNN, on the other hand, is a local classifier; each pattern unit shows output values significantly different from zero only for a finite (generally small) region of pattern space surrounding its training samples. This architecture will display low values for all output units when presented with patterns considerably different from those in the training set. It is therefore well-suited to detect novel patterns.

An additional advantage of the PNN is that it does not require the lengthy training sequences necessary when training back-propagation. It was found, though, that the performance of the PNN is crucially dependent on the generation of useful cluster centres prior to construction of the network. Unsupervised clustering algorithms cannot be used on the complete pattern set and applying clustering to the patterns of each category separately did not work well. The final solution was to combine the unsupervised Kohonen method followed by supervised LVQ clustering. Other supervised methods may also have worked, but determining a good strategy for finding the pattern clusters definitely was the main challenge in getting the PNN architecture to generate high performance. This promising architecture still seems to be under-utilized in the classification of remotely sensed imagery.

References

Augusteijn, M. F., and Warrender, C. E., 1999, Fusion of image classifications using Bayesian techniques with Markov random fields. *International Journal of Remote Sensing*, **20**, 1987–2002.

Fausett, L., 1994, *Fundamentals of Neural Networks, Architectures, Algorithms, and Applications* (Englewood Cliffs, NJ: Prentice Hall).

Kohonen, T., 1982, Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43**, 59–69.

Moody, J., and Darken, C., 1989, Fast learning in a network of locally-tuned processing units. *Neural Computation*, **1**, 281–294.

Paola, J. D., and Schowengerdt, R. A., 1995, A detailed comparison of back-propagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 981–996.

Parra, L., Deco, G., and Miesbach, S., 1996, Statistical independence and novelty detection with information preserving nonlinear maps. *Neural Computation*, **8**, 260–269.

Richard, M. D., and Lippmann, R. R., 1991, Neural network classifiers estimate Bayesian *a posteriori* probabilities. *Neural Computation*, **3**, 461–483.

Roberts, S., and Tarassenko, L., 1994, A probabilistic resource allocating network for novelty detection. *Neural Computation*, **6**, 270–284.

Rumelhart, D. E., Hinton, G. E., and Williams R. J., 1986, Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1: Foundations,* edited by D. E. Rumelhart and J. L. McClelland (Cambridge, MA: MIT Press), pp. 318–362.

Specht, D. F., 1990, Probabilistic neural networks. *Neural Networks*, **3**, 109–118.

Washburne, T. P., Specht, D. F., and Drake, R. M., 1993, Identification of unknown categories with probabilistic neural networks. *Proceedings of the 1993 IEEE International Conference on Neural Networks, San Francisco, California, 28 March–1 April, 1993*, Volume 1 (Piscataway, NJ: IEEE Service Center), pp. 434–437.