Mon. Not. R. Astron. Soc. **353,** 211–220 (2004) doi:10.1111/j.1365-2966.2004.08056.x

Selection of quasar candidates from combined radio and optical surveys using neural networks

R. Carballo,[1⋆] A. S. Cofiño[1] and J. I. González-Serrano[2]

[1]*Departamento de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria ETS Ingenieros de Caminos, Canales y Puertos, Avenida de los Castros s/n, 39005 Santander, Spain*

[2]*Instituto de Física de Cantabria (CSIC-UC), Avenida de los Castros s/n, 39005 Santander, Spain*

Accepted 2004 May 23. Received 2004 May 20; in original form 2003 October 30

ABSTRACT

The application of supervised artificial neural networks (ANNs) for quasar selection from combined radio and optical surveys with photometric and morphological data is investigated, using the list of candidates and their classification from the work of White et al. Seven input parameters and one output, evaluated to 1 for quasars and 0 for non-quasars during the training, were used, with architectures 7 : 1 and 7 : 2 : 1. Both models were trained on samples of ∼800 sources and yielded similar performance on independent test samples, with reliability as large as 87 per cent at 80 per cent completeness (or 90 to 80 per cent for completeness from 70 to 90 per cent). For comparison, the quasar fraction from the original candidate list was 56 per cent. The accuracy is similar to that found by White et al. using supervised learning with oblique decision trees and training samples of similar size. In view of the large degree of overlapping between quasars and non-quasars in the parameter space, this performance probably approaches the maximum value achievable with this data base. Predictions of the probabilities for the 98 candidates without spectroscopic classification in White et al. are presented and compared with the results from their work. The values obtained for the two ANN models and the decision trees are found to be in good agreement. This is the first analysis of the performance of ANNs for the selection of quasars. Our work shows that ANNs provide a promising technique for the selection of specific object types in astronomical data bases.

Key words: methods: data analysis – methods: statistical – quasars: general.

1 INTRODUCTION

In the recent years large astronomical data bases based on surveys at different wavelengths have been made publicly available to the astronomical community. A full exploitation of these data bases will only be possible with the help of artificial intelligence (AI hereinafter) tools, which will allow the selection, classification and even the definition of particular object types within the data bases.

Artificial Neural Networks (ANNs hereinafter) are one of these tools. ANNs have been applied in astronomy for mainly the following problems: classification of stellar spectra (e.g. Bailer-Jones, Irwin & von Hippel 1998), morphological star/galaxy separation (e.g. Bertin & Arnouts 1996), morphological and spectral classification of galaxies (Folkes, Lahav & Maddox 1996; Lahav et al. 1996; Firth, Lahav & Somerville 2003) and, more recently, estimation of photometric redshifts of galaxies (Firth et al. 2003). A summary of the most relevant applications of ANNs in astronomy can be found in Tagliaferri et al. (2003).

⋆E-mail:carballor@unican.es

In this work we investigate the application of ANNs in a new domain, i.e. the effective selection of quasar candidates. Although ultimately an optical spectrum will be required to confirm the quasar classification and determine its redshift, an optimized selection of quasar candidates allows one to obtain large quasar samples with a reduction of telescope time. Large quasar samples are necessary to address important questions in cosmology, such as the comparison of the space distribution of quasars with that predicted by theory.

The test is based on a combined radio and optical survey including photometric and morphological data: the list of quasar candidates in White et al. (2000), drawn from the cross-correlation of the Very Large Array FIRST Survey and the Automatic Plate Measuring Machine (APM) catalogue of the POSS-I photographic *E* and *O* plates (McMahon & Irwin 1992). White et al. obtained the spectroscopic classification of 1130 of the candidates, 636 (56 per cent) being confirmed as quasars. These quasars form the FIRST Bright Quasar Survey of the North Galactic Cap (FBQS-2 hereinafter). From their results the authors explored for the first time the viability of AI tools for obtaining an a priori selection of the best quasar candidates from radio and optical photometric and morphological data. The tool used was supervised learning with the oblique decision

tree classifier OC1 (Murthy, Kasif & Salzberg 1994). The decision tree had a single output per object and the desired outputs or *targets* were set, during the learning (training), to 1 for quasars and to 0 for non-quasars. The authors found that a decision tree classifier, trained on data sets of about 800 objects, allowed one to obtain an efficient selection of the quasars, producing samples with reliability as high as 80 per cent at 90 per cent completeness. The work by White et al. opened the idea of the application of AI techniques for the effective selection of quasar candidates from combined radio and optical photometric surveys. Following this idea, we analysed – with the same sample of quasar candidates – a different classifier, i.e. supervised learning with ANNs, taking advantage of the wealth of training algorithms included in the MATLAB Neural Network Toolbox (http://www.mathworks.com/).

The layout of the rest of the paper is as follows. Section 2 starts with a brief description of the list of quasar candidates from White et al., including the selection criteria and the available parameters from FIRST and APM. Then the performance of the decision tree classifier reported by White et al. is summarized. In Section 3 we describe the techniques applied for the training and testing of the ANNs, we explore the performance of the model, basically in terms of the reliability and completeness of the samples of the best candidates, and finally we use the model to predict the quasar probabilities for the 98 FBQS-2 candidates without any spectrum. Through the discussion, our results are compared with those obtained in White et al. The main conclusions are presented in Section 4.

2 QUASAR SELECTION FROM THE FBQS-2 CANDIDATES VIA DECISION TREES

The candidates for FBQS-2 (White et al. 2000) were obtained from the correlation of the VLA FIRST radio survey (down to $S_{1.4\,\mathrm{GHz}} \sim 1$ mJy) with stellar sources on POSS-I(APM) with $E \leqslant 17.8$ and a blue colour ($O - E \leqslant 2$). The spectroscopic classification of 1130 of the 1238 candidates yielded 636 quasars, 96 narrow-line AGNs, 68 BL Lac objects, 190 H II galaxies, 52 passive galaxies and 88 stars. The selection efficiency for quasars was therefore $636/1130 = 56$ per cent ($704/1130 = 62$ per cent combining quasars and BL Lac). White et al. classified any object with broad emission lines as a quasar, i.e. they did not use the conventional cut at $M_B = -23$ to exclude lower-luminosity objects. Fifty of the 636 quasars fall into this low-luminosity category. The redshift range for the whole quasar sample was $z \sim 0$ to $z \sim 3.5$

White et al. present diagrams showing that the fraction of quasars varied with optical magnitude, optical colour, radio flux and radio-optical position separation. Based on these results, the authors suggested that AI methods could be used to assign the candidates an a priori probability of being quasars, $p(\mathrm{Q})$, before taking the spectra. They analysed the performance of the oblique decision tree classifier OC1, improved by using 10 trees instead of a single one, with a weighted voting scheme. The sample of quasars with spectroscopic classification was divided into five sets. Setting aside the first set, the remaining four were used for the training and the first one for the test. Repeating the procedure for each of the sets, the authors could use all the objects for the training and all the objects for the test.

The performance of any classifier can be quantified through two important parameters, which are the efficiency and the completeness of the subsamples built from the classifier as a function of the threshold used, $p_C(\mathrm{Q})$. For this case, the efficiency (or reliability) is the number of candidates above $p_C(\mathrm{Q})$ that are quasars divided by the total number of candidates above this threshold. The completeness is the number of quasars that are included above $p_C(\mathrm{Q})$ divided by the total number of quasars. By decreasing the threshold $p_C(\mathrm{Q})$ above which candidates are accepted, the completeness of the sample is increased, but probably at the cost of efficiency. For the extreme case of $p(\mathrm{Q}) \geqslant 0$ the FBQS-2 sample would be complete, i.e. it would include all the quasars among the candidates, but the reliability would drop to 56 per cent. White et al. found the voting decision trees to be a successful classifier, allowing one to construct subsamples of candidates ~87 per cent reliable at completeness 70–80 per cent or still very reliable, ~ 80 per cent, at 90 per cent completeness. The authors used seven input parameters: E, $O - E$, $\log_{10} S_p$ (where S_p is the FIRST peak flux density), S_i/S_p (where S_i is the FIRST integrated flux density), the radio-optical separation, and the point spread functions PSF(E) and PSF(O).

3 QUASAR SELECTION FROM THE FBQS-2 CANDIDATES VIA ANNS

3.1 Fitting and testing technique

An ANN is a computational tool that provides a non-linear parametrized mapping between a set of input parameters and one or more outputs. The type of ANN we used is the *multi-layer perceptron* (hereinafter MLP; Bishop 1995; Bailer-Jones, Gupta & Singh 2001). A particular ANN architecture may be denoted as $N_{in} : N_1 : N_2 : \ldots : N_{out}$, where N_{in} is the number of input parameters, N_1 is the number of nodes in the first hidden layer and so forth, and N_{out} is the number of nodes in the output layer. The nodes are connected, each connection carrying a *weight* and each node carrying a *bias*. For our work we assume that every node is connected to every node in the previous layer and every node in the next layer only. Each node receives the output of all the nodes in the previous layer and produces its own output, which then feeds the nodes in the next layer. At a node in layer s the following calculation is obtained:

$$z = \sum_{j=1}^{m} x_j^{s-1} w_{jk}^{s} + b_k, \tag{1}$$

where x_j^{s-1} are the inputs from the previous layer (with m nodes), and w_{jk} and b_k are, respectively, the weights and the bias of the node. Then the signal output of the node is $x_k^s = g(z)$, where g is the non-linear activation or transfer function. In order to obtain the correct mapping, a set of representative input–output data is used for the training, a process in which the weights and biases are optimized to minimize the error in the outputs.

We used the same set of seven input parameters adopted by White et al. (2000), normalizing each of them to the range $[-1, 1]$. In order to have homogeneous data sets we did not include the 28 candidates missed in APM E or O or both, and for which White et al. used the APS magnitudes (Minnesota Automated Plate Scanner POSS-I catalogue, Pennington et al. 1993). The number of FBQS-2 candidates with spectral classification and homogeneous data was reduced to 1112 for this reason. The distribution is as follows: 627 quasars, 94 narrow-line AGNs, 67 BL Lacs, 187 H II galaxies, 51 passive galaxies and 86 stars.

The network output consisted of a single node, with the target values set during the training as 1 for quasars and 0 for non-quasars. For the output node we used a logarithmic sigmoid activation function of the form $g(z) = 1/[1 + \exp(-z)]$, in the range [0,1]. The actual output of the ANN is then an estimate of the probability that the source is a quasar, and it is denoted as $p(\mathrm{Q})$ (Richard & Lippmann 1991). The transfer function for the hidden

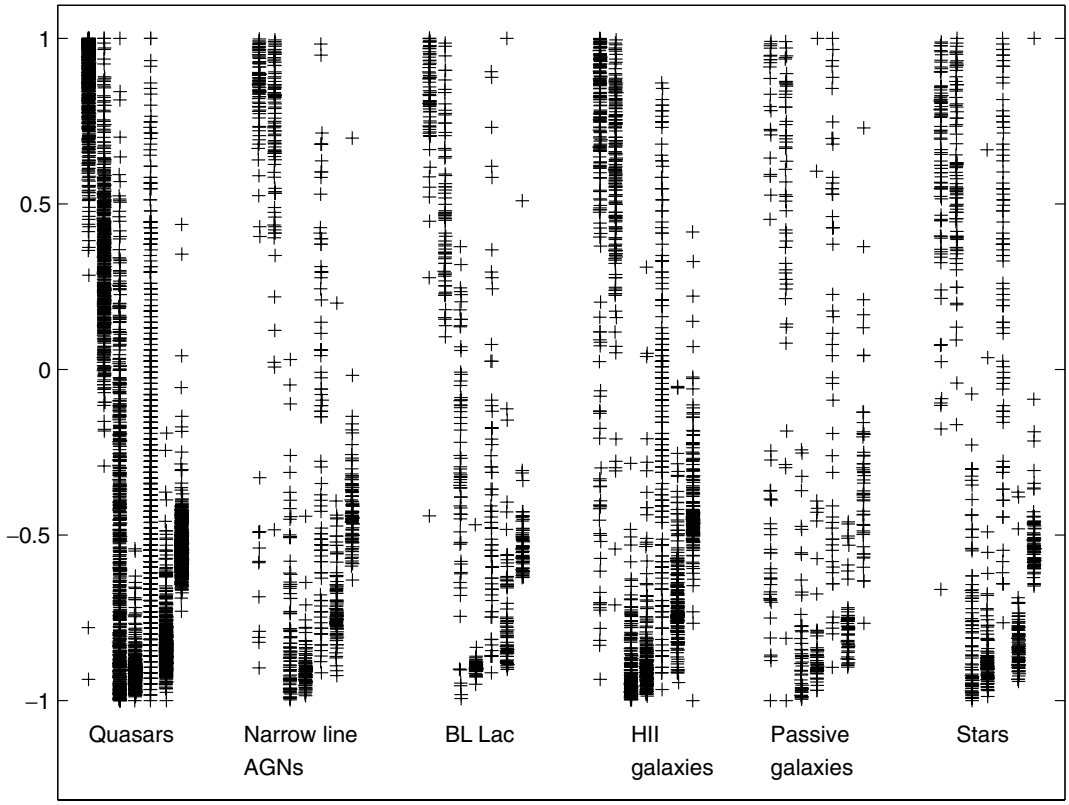

Figure 1. Normalized parameters E, $O - E$, $\log_{10} S_p$, S_i/S_p, radio–optical separation, PSF(E) and PSF(O), ordered from left to right, for each class.

nodes was $g(z) = \tanh(z) = 2/[1 + \exp(-2z)] - 1$, in the range $[-1, 1]$.

Fig. 1 shows the normalized input parameters for the six classes. Although the parameter space covered varies between classes, there is a large degree of overlapping between quasars and non-quasars, which will certainly limit the performance of the classification. In fact, the category of non-quasars has an increased intrinsic scatter due to the presence of classes with different physical nature and covering different regions of the parameter space.

The error function we used was the mean of the squared errors, of the form

$$mse = \frac{1}{N}\sum_{i=1}^{N}(p_i - t_i)^2, \tag{2}$$

wherep_i and t_i are, respectively, output (probability of being a quasar) and target value for the ith object. The sum of the squared errors has been widely used as the minimizing error function for classification with the MLP (Richard & Lippmann 1991; Bishop 1995; Lahav et al. 1996; Bailer-Jones et al. 2001; Ball et al. 2004). Although on theoretical grounds there are more appropriate error functions for classification, such as *cross-entropy* (which assumes the expected noise distribution for discrete variables), the sum-of-squares error has proven to yield the same performance as cross-entropy for MLP classification on real-world problems with large data bases (Richard & Lippmann 1991). In addition, the sum-of-squares error has the advantage that the determination of the network parameters represents a linear optimization problem, in particular, the powerful Levenberg–Marquardt algorithm for parameter optimization is applicable specifically to a sum-of-squares error function (Bishop 1995). Based on these results, we used the *mse* error function and applied the Levenberg–Marquardt algorithm, which is the default optimization technique used for *batch-training* (weights and biases updated after all the input vectors are presented to the network) in the MATLAB Neural Network Toolbox. The Levenberg–Marquardt algorithm is the fastest method for training moderate-sized neural networks (Hagan & Menhaj 1994).

Regardless of the optimization algorithm employed, one of the main problems in the training process is that of 'overfitting', i.e. the ANN tends to memorize the outputs, instead of modelling the general intrinsic relationships in the data. In order to reduce this problem we used training with validation error. With this method, the training that is being carried out in the *training set* is automatically stopped when the error obtained running the trained network in another set, the *validation set*, does not decrease for a given number of iterations. We adopted for this parameter, known as maxfail in MATLAB, a value of 20, instead of the default value of five iterations. An additional independent set, the *test set*, is needed to evaluate the ANN performance.

Following the procedure adopted by White et al., we divided the initial sample of candidates into four subsets or folds (they used five) of approximately similar size. In contrast to White et al., who selected the folds randomly, we chose them to have similar fractions of the different object types as the total sample. Setting aside each subset, the remaining three were used for the training and validation, and the subset itself was used for the test. The size of the test fold, i.e. about 275 objects (1/4 of the candidates), was selected to insure the inclusion of about a dozen objects of the classes with fewer members, like passive galaxies and BL Lacs. The three subsets used for training and validation were first combined and then randomly divided into two groups: one forming the training set, with 2/3 of the candidates, and the other forming the validation set,

with the remaining 1/3, each of them with similar proportions of object types as the total sample. Repeating the procedure for each of the four folds, we obtained four different classifiers, with the advantage of having used all the objects for the training/validation and all the objects for the test, and therefore having optimized the statistics.

The ANN was run 10×10 times per fold. The first factor accounts for the obliged repetition in an algorithm that includes randomization (for example in the seeds for the initial weights of the ANN) to avoid poor local minima. The second one arises from the use of 10 different splittings to separate the training and the validation sets. In order to choose the best ANN of the 100 runs, we first selected the splitting of the training and validation sets that gave the lowest value of the mean squared error, $\overline{mse}$, averaged over the 10 fits. Adopting $mse = 0.5 \times mse_{\rm train} + 0.5 \times mse_{\rm valid}$ for each run,

$$\overline{mse} = 0.5 \times \overline{mse}_{\rm train} + 0.5 \times \overline{mse}_{\rm valid}. \quad (3)$$

We then checked if the relation

$$\overline{|mse_{\rm train} - mse_{\rm valid}|/mse_{\rm train}} < 0.15 \quad (4)$$

was satisfied for the splitting, to ensure that the errors in the validation and training were not only small, but also roughly similar (within 15 per cent on average). In the case that this condition failed, the splitting with the next minimum value $\overline{mse}$ was checked for condition (4) and so forth. Once the splitting was selected we chose amongst the 10 ANNs the one with the minimum value of $mse = 0.5 \times mse_{\rm train} + 0.5 \times mse_{\rm valid}$ and satisfying

$$|mse_{\rm train} - mse_{\rm valid}|/mse_{\rm train} < 0.15. \quad (5)$$

For a few cases two or more fits had the same minimum; in these cases we took the first fit in the running order. In the end we had a final ANN for each of the four test sets. Running each ANN for its corresponding test set we were able to obtain the values p(Q) for the 1112 candidates.

3.2 Results

We used two different ANN architectures. The first one, denoted as 7 : 1, does not include hidden layers and it is also known as a *logistic discrimination model*. The second architecture includes a hidden layer with two nodes, and it is denoted as 7 : 2 : 1. As we shall see, the performance of the classifier does not improve with the inclusion of a hidden layer (increasing the free parameters of the ANN from eight to 19); therefore, more complex architectures were not explored. At the end of this subsection we present the quasar probabilities obtained from the fitted ANNs for the list of FBQS-2 candidates without optical spectroscopy.

3.2.1 Logistic discrimination model

Fig. 2 shows $mse_{\rm train}$ and $mse_{\rm valid}$ for the 400 networks run (100 networks per test set $\times$ four test sets). We recall that each group of 100 networks is divided in 10 blocks, each of them corresponding to a different splitting of the validation-training sets, and each block is made of 10 fits. Some of the networks or whole splittings (blocks) produce peaks in $mse_{\rm train}$, in $mse_{\rm valid}$ or in both. The splittings showing peaks tend to have a higher $\overline{mse}$ than the splittings lacking them; hence our choice of the minimum $\overline{mse}$ to select the best splitting. For 54 per cent of the networks the validation set stopped the training. The number of iterations in cases of validation stop is very small (below 12), with an average of ~ 4 compared with the average of ~ 23 found for networks that stopped for other reasons.

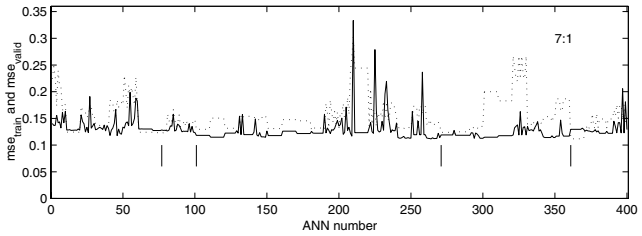

Figure 2. $mse_{\rm train}$ (continuous line) and $mse_{\rm valid}$ (dotted line) for the 400 ANNs run. The vertical lines mark the best ANN for each of the test sets.

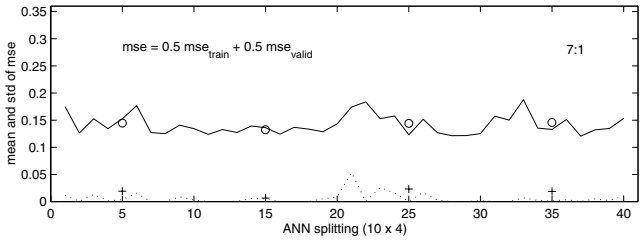

Figure 3. Mean (continuous line) and standard deviation (dotted line) of mse ($= 0.5 \times mse_{\rm train} + 0.5 \times mse_{\rm valid}$) over the 10 runs with similar *test*, *training* and *validation* sets. The first parameter is denoted as $\overline{mse}$ in the text. The meaning of the circles and crosses is explained in the text.

Fig. 3 shows $\overline{mse}$ and the standard deviation of mse over the ten runs for each of the 40 different splittings (10 splittings per test set $\times$ four test sets). The scatter ranges from 0 to about 0.05, with an average for the 40 splittings of 0.006. The influence of the initial values on the performance of the selected network is negligible, since the standard deviation of mse due to different initiations is clearly much lower than the mean values. The same occurs for the separation of the training and validation sets; the circles in Fig. 2 show the average of $\overline{mse}$ over the 10 different splittings per test set, and the averages are significantly larger than their standard deviations, symbolized as crosses. Finally, the figure also shows that the four mean values (one per test set) are very similar, with the standard deviation being much lower than the average (this average and standard deviation are not shown in the figure). The last result demonstrates that the performance of the network does not depend strongly on the particular selected *test set* either. The four best ANNs (one per test set) are marked with vertical lines in Fig. 2; Table 1 summarizes some of their parameters.

So far we have discussed the mean squared errors obtained in the training process. Column four of Table 1 gives the mean squared errors obtained for the *test sets*. The latter mse values generally show good agreement with those obtained for the training and validation sets. However, a more interesting parameter for the purpose of assessing the performance of the network is the normalized error

Table 1. Parameters of the four selected ANNs for the logistic model.

$mse_{\rm train}$	$mse_{\rm valid}$	$N_{\rm iter}$	$mse_{\rm test}$	$\overline{E}$
0.128	0.122	3[a]	0.119	0.49
0.118	0.130	18	0.133	0.54
0.119	0.124	23	0.154	0.63
0.129	0.112	11[a]	0.122	0.49

[a]The validation set results stopped the training.

function (Bishop 1995), of the form

$$\overline{E} = \sum_{i=1}^{N}(p_i - t_i)^2 \Bigg/ \sum_{i=1}^{N}(\overline{t} - t_i)^2, \quad (6)$$

where $\overline{t}$ is the mean of the target data over the test set. This error function equals unity when the model is as good a predictor of the target data as the simple model $p = \overline{t}$, and equals zero if the model predicts the data values exactly. The value we found is around 0.55. Although the model is not good enough for classification, the results are powerful for our aim of selecting the best candidates. In fact, compared with the model that takes $p = \overline{t}$, which would give $mse_{\rm test} \sim 0.246$, the *mse* obtained with the ANNs is reduced by about a factor two. In the next paragraphs we present the results of the model in terms of completeness and efficiency of the subsamples of *the best candidates* that can be drawn from the ANN model.

Fig. 4 shows the distribution of p(Q) for the 1112 candidates and the logistic discrimination ANN. The model gives probabilities above 0.5 for most of the quasars, although there is a large number of them with probabilities below this value. Narrow-line AGNs, H II galaxies, passive galaxies and stars tend to give low probabilities, and the model therefore provides a good means of rejecting objects of these types. The number of BL Lac objects is small, and their distribution of p(Q) is rather flat, and even slightly increased at high probabilities, therefore the current model is not able to reject these sources. The last problem was also found by White et al. using the decision tree classifier.

The efficiency and completeness of the sample as a function of the quasar probability threshold $p_{\rm C}$(Q) are shown in Fig. 5. The logistic model allows one to obtain a high reliability at a high completeness: for completenesses of 70, 80 and 90 per cent the corresponding reliabilities are 90, 87 and 81 per cent. Fig. 6 shows the fraction of candidates that are quasars as a function of p(Q). The fraction is slightly above p(Q) (about 0.1 in the range from p(Q) 0.3 to 0.8), i.e. the likelihood that a candidate turns out to be a quasar is slightly larger than the probability given by the model.

Fig. 4 shows that the majority of the high-p(Q) candidates that are not quasars are BL Lac objects. Taking $p(\mathrm{Q}) > 0.75$, there are 353 quasars, 24 BL Lacs, two narrow-line AGNs, three H II galaxies, a passive galaxy and three stars. An inspection of the input

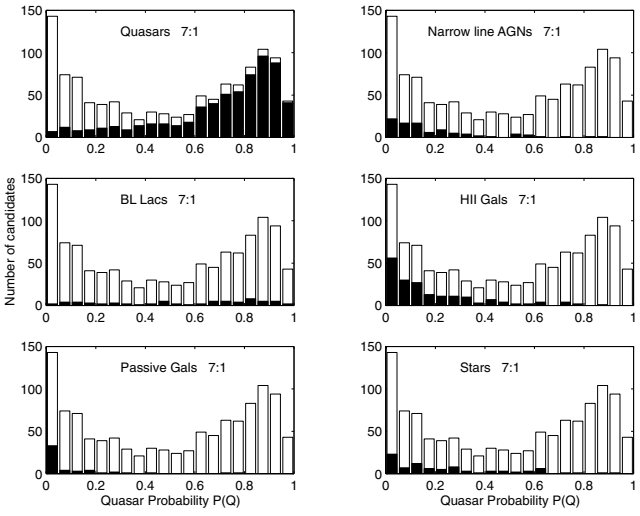

Figure 4. Distribution of p(Q) for the logistic discrimination model. The shaded distributions correspond to the objects of the indicated types (quasars, narrow-line AGNs, BL Lacs, etc.).

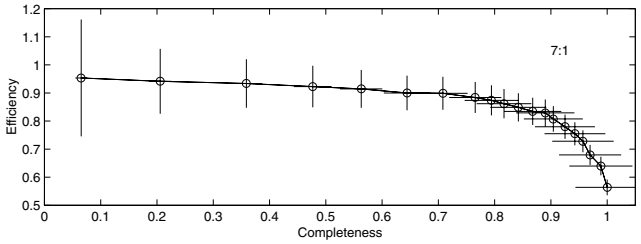

Figure 5. Efficiency versus completeness of the sample for $p(\mathrm{Q}) > 0, 0.05, 0.1, 0.15, 0.2, \ldots, 0.95$ (right to left) and the logistic model. Poissonian errors were assumed.

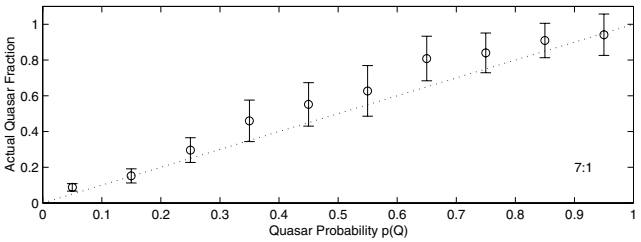

Figure 6. Fraction of candidates that are quasars as a function of p(Q) for the logistic model. Poissonian errors were assumed.

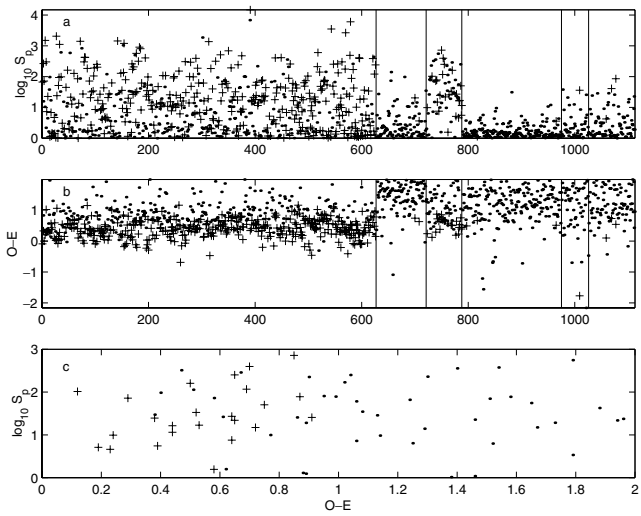

Figure 7. (a) and (b): $\mathrm{Log}_{10}(S_{\rm p})$ and $O - E$ for the classes of quasars, narrow-line AGNs, BL Lacs, H II galaxies, passive galaxies and stars (separated by vertical lines). (c): $\mathrm{Log}_{10}(S_{\rm p})$ versus $O - E$ for the BL Lacs. Crosses correspond to $p(\mathrm{Q}) > 0.75$ for the 7 : 1 model.

parameters for the non-quasars revealed as the most outstanding result that the *whole population* of BL Lacs has radio fluxes higher than those found for the remaining non-quasar classes, and similar to those typically found in quasars – see Fig. 7(a). About 36 per cent of the BL Lacs have $p(\mathrm{Q}) > 0.75$ and Figs 7(b) and (c) show that these correspond to the cases with bluer $O - E$ colours. The efficiency of quasar selection using the cut at $p(\mathrm{Q}) = 0.75$ is 91 per cent (353/386) and increases to 98 per cent considering quasar or BL Lac selection (377/386). The corresponding completeness would be 56 per cent (353/627) for quasars and 54 per cent (377/694) for quasars or BL Lacs. The completeness decreases in the latter case since *only* blue BL Lacs are confused with quasars.

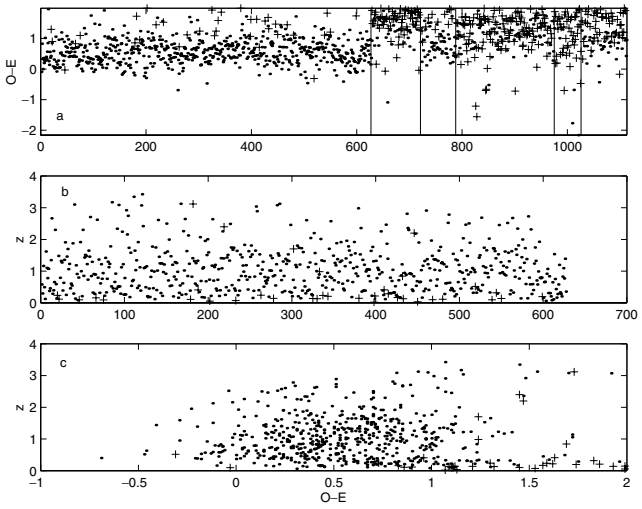

Figure 8. (a): $O - E$ for the classes of quasars, narrow-line AGNs, BL Lacs, H II galaxies, passive galaxies and stars (separated by vertical lines). (b) and (c): z and z versus $O - E$ for the quasars. Crosses correspond to $p(\mathrm{Q}) < 0.2$ for the 7 : 1 model.

At the other extreme, there are 36 quasars with probabilities $p(\mathrm{Q}) < 0.2$, and their most significant differences with respect to the remaining quasars are their redder $O - E$ colours and lower redshifts, with twenty-five of them at $z < 0.25$ – see Figs 8(a), (b) and (c). The misclassified quasars also differ, although to a lower extent, in their larger integrated-to-peak radio flux ratio, larger optical–radio separation and wider PSF. The low probabilities found for the low-z quasars should not be regarded as a limitation of the classifier, since at low redshifts the host galaxy is expected to be slightly resolved and to have a noticeable contribution to the total 'galaxy + quasar' emission. This contribution, imperceptible at higher redshifts, is the most likely explanation for the differences in the input parameters between low-z quasars and the remaining quasars.

If only the quasars with $z \geqslant 0.25$ are considered, the fraction of them with $p(\mathrm{Q}) < 0.2$ drops from 6 per cent (36/627) to 2 per cent (11/558). As for the probability cut $p(\mathrm{Q}) > 0.75$, the efficiency remains at 91 per cent and the completeness increases from 56 to 62 per cent. White et al. also found, using the decision tree classifier, that the great majority of quasars with low probabilities were at low redshift (out of 30 quasars with $p(\mathrm{Q}) < 0.2$, 24 had $z < 0.25$).

3.2.2 ANN with a hidden layer

In this subsection we present the results for the ANN model 7 : 2 : 1. Fig. 9 shows mse_{train} and mse_{valid} for the 400 networks. Three main differences with respect to the model without a hidden layer are revealed: (i) the distribution of mse_{train} is more noisy but there

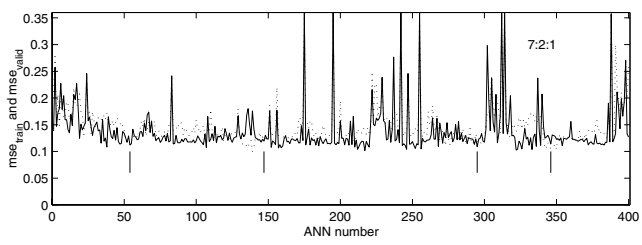

Figure 9. Similar to Fig. 1, but for the 7 : 2 : 1 architecture.

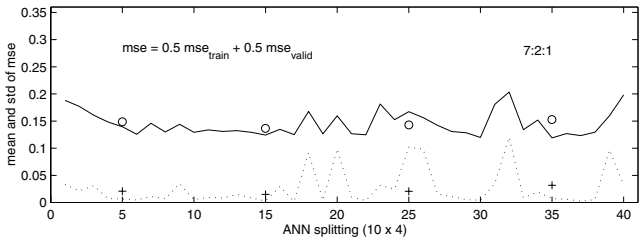

Figure 10. Similar to Fig. 2, but for the 7 : 2 : 1 architecture.

is a better agreement between mse_{train} and mse_{valid}, (ii) validation stopping dominates (96 per cent of the cases) over the remaining reasons for training stopping, and (iii) the number of iterations in cases of validation stopping reaches higher values, with an average of ~ 33 iterations. Fig. 10 shows the average and standard deviation of *mse* over the ten runs for each of the 10 splittings and for each of the four test sets. The large variations of *mse* are clearly evident from this figure: the standard deviation of *mse* ranges from 0.002 to 0.12, with an average for the 40 splittings of 0.027, i.e. 4.5 times larger than for the 7 : 1 architecture. However, for most of the training-validation-test configurations the scatter of *mse* is still much lower than the mean value. Considering the averages per test set, the mean values for $\overline{mse}$ are also significantly larger than their standard deviations (denoted with circles and crosses, respectively) and the same occurs considering the average over the four test sets. As occurred for the 7 : 1 model, the performance of the selected network does not depend strongly on changes of the initiation values, splitting for training-validation or choice of test set.

The relevant parameters of the four selected ANNs are summarized in Table 2. The *mse* values for the test sets generally show a good agreement with the values obtained for the train and validation sets. Both mse_{test} and the normalized error function $\overline{E}$ are on average similar to those obtained for the 7 : 1 architecture.

Fig. 11 shows the distribution of $p(\mathrm{Q})$ for the 1112 candidates and the 7 : 2 : 1 model. The distribution is more peaked towards the extreme values of the probabilities than in the logistic model. In this respect, the 7 : 2 : 1 model gives a better agreement with the results from OC1 than the logistic model. As occurred for the logistic model and OC1, all the non-quasar classes except for the BL Lacs tend to give low probabilities.

The efficiency and completeness of the sample as a function of the quasar probability threshold are very similar to the values found for the logistic model. For completenesses of 70, 80 and 90 per cent, the corresponding reliabilities are 88, 87 and 81 per cent, respectively. Again there is a very good agreement between $p(\mathrm{Q})$ and the likelihood that a candidate with $p(\mathrm{Q})$ turns out to be a quasar (measured by the fraction of candidates at this $p(\mathrm{Q})$ that are quasars), except for $p(\mathrm{Q})$ around 0.55, where the likelihood is increased by an amount around 0.1.

Table 2. Parameters of the four selected ANNs for the model 7 : 2 : 1.

mse_{train}	mse_{valid}	N_{iter}	mse_{test}	$\overline{E}$
0.112	0.126	14[a]	0.110	0.45
0.119	0.121	8	0.127	0.52
0.109	0.118	16[a]	0.146	0.60
0.114	0.106	27[a]	0.127	0.51

[a]The validation set results stopped the training.

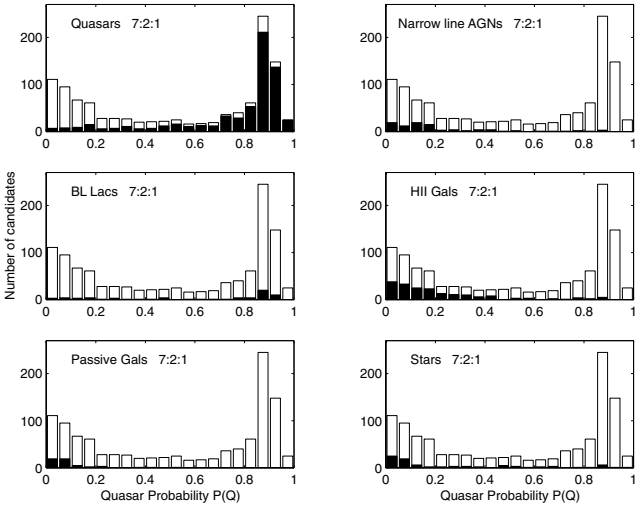

Figure 11. Similar to Fig. 3, but for the 7 : 2 : 1 architecture.

Taking $p_C(Q) = 0.85$, there are 372 quasars, 30 BL Lacs, three narrow-line AGNs, five H II galaxies and eight stars above this cut. A result similar to the one obtained in Fig. 7 for the logistic model is found for the 7 : 2 : 1 architecture: the majority of the high-$p(Q)$ non-quasars are blue BL Lac objects. The efficiency of quasar selection for this threshold is 89 per cent and increases to 96 per cent considering quasar or BL Lac selection. The corresponding completenesses would be 59 per cent for quasars and 58 per cent for quasars or BL Lacs.

Regarding the limit of low probabilities, we find 39 quasars with $p(Q) < 0.2$, 22 of them with redshifts below 0.25. We find similar results to those presented in Fig. 8 for the logistic model: most of the misclassified quasars have lower redshifts and redder $O - E$ colours than the remaining quasars, as well as higher integrated-to-peak radio flux ratios and wider PSFs, and these results are indicative of an appreciable contribution from the host galaxy to the total 'galaxy + quasar' emission. If only the quasars with $z \geqslant 0.25$ are considered, the fraction of them with $p(Q) < 0.2$ decreases from 6 per cent to 3 per cent. As for the probability cut $p(Q) > 0.85$, the efficiency would remain at 89 per cent and the completeness would increase from 59 per cent to 65 per cent.

Table 3 presents a summary of the performance of the two ANN models. Both use similar training set sizes and achieve similar efficiencies for completeness in the range from 70 to 90 per cent. The main difference is that the distribution of $p(Q)$ is more peaked towards the extreme values (0 and 1) for the model with a hidden layer than for the logistic one.

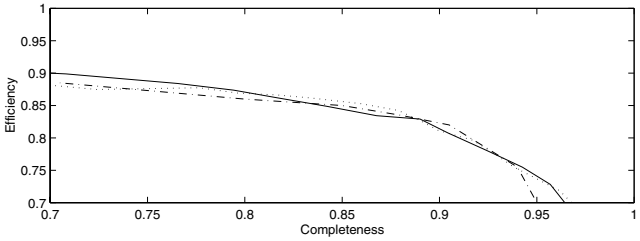

Figure 12. Efficiency versus completeness for the 7 : 1 ANN (continuous line), the 7 : 2 : 1 ANN (dotted line) and the OC1 decision tree (dash–dotted line).

Fig. 12 shows that the distribution of efficiency versus completeness for the two ANN models and the oblique decision tree OC1 are very similar. The agreement obtained for the three different classifiers favours the interpretation that the found accuracy – ∼87 per cent at 80 per cent completeness – is more limited by the data structure itself (i.e. the large degree of overlapping between quasars and non-quasars in the input parameter space) than by the complexity of the algorithms. The ANN and the decision tree classifiers both point to BL Lacs (blue BL Lacs for ANNs) and low-z quasars as the object types that most severely limit the accuracy of quasar selection, the former producing intruders (false alarms), and the latter misclassifications.

Owens, Griffiths & Ratnatunga (1996) apply the decision tree OC1 for the morphological classification of galaxies taken from the ESO-LV catalogue (Lauberts & Valentijn 1989), and present a comparison of their results with those obtained by Storrie-Lombardi et al. (1992) for the same sample using ANNs. The classification *into six classes* has an overall efficiency around 63 per cent for the two methods, with a difference between them lower than 3 per cent. Owens et al. (1996) attribute the similarity found to limitations in the classification accuracy intrinsic to the data base (errors in the assumed classification for some of the galaxies and a poorly defined separation between classes). Our work and that of Owens et al. (1996) show examples of classification from astronomical data bases in which OC1 and ANNs give similar performances, probably at the limit set by the data base itself, whose attributes do not provide enough information for a more accurate classification.

Table 3. Efficiency and completeness of quasar selection from the sample of FBQS-2 candidates using ANNs.

Size of spectroscopically identified sample 1112
Training+validation set size ∼ 840
Total number of quasars 627

ANN 7 : 1			ANN 7 : 2 : 1		
Completeness/Efficiency	70 per cent/90 per cent 80 per cent/87 per cent 90 per cent/81 per cent		Completeness/Efficiency	70 per cent/88 per cent 80 per cent/87 per cent 90 per cent/81 per cent	
$p(Q) < 0.2$	Quasars	36	$p(Q) < 0.2$	Quasars	39
$p(Q) > 0.75$	Candidates	386	$p(Q) > 0.85$	Candidates	418
	Quasars	353		Quasars	372
	BL Lacs	24		BL Lacs	30
	Efficiency for quasars	91 per cent		Efficiency for quasars	89 per cent
	Efficiency for quasars + BL Lacs	98 per cent		Efficiency for quasars + BL Lacs	96 per cent
	Completeness for quasars	56 per cent		Completeness for quasars	59 per cent
	Completeness for quasars + BL Lacs	54 per cent		Completeness for quasars + BL Lacs	58 per cent

3.2.3 Predictions for the FBQS-2 candidates without spectroscopic classification

The ANN models 7 : 1 and 7 : 2 : 1 were used to estimate the probabilities p(Q) for the 98 FBQS-2 candidates without spectral classification in White et al. We adopted four classifiers per model, corresponding to the four selected ANNs (parameters described in Tables 1 and 2). Fig. 13(a) shows the probabilities obtained with the 7 : 1 model – plotted with a different line type for each ANN – and using OC1 (White et al.). There is a good agreement between the probabilities predicted with the four ANNs and between them and the values from OC1. Similar results are found for the 7 : 2 : 1 model – Fig. 13(b). The probabilities obtained for the ANN models and OC1 are listed in Table 4. For the ANN models we give the mean and standard deviation of p(Q) over the four selected ANNs.

Fig. 14 shows $\overline{p}$(Q) for the 7 : 2 : 1 model versus $\overline{p}$(Q) for the 7 : 1 model. The agreement between the two predictions is generally very good, with the largest discrepancies occurring at intermediate probabilities, where the candidates' parameters fit neither the quasar class nor the non-quasar class. The average difference between $\overline{p}$(Q) for the two models (7 : 2 : 1 − 7 : 1) is only 0.04, and the standard deviation of the difference is 0.06.

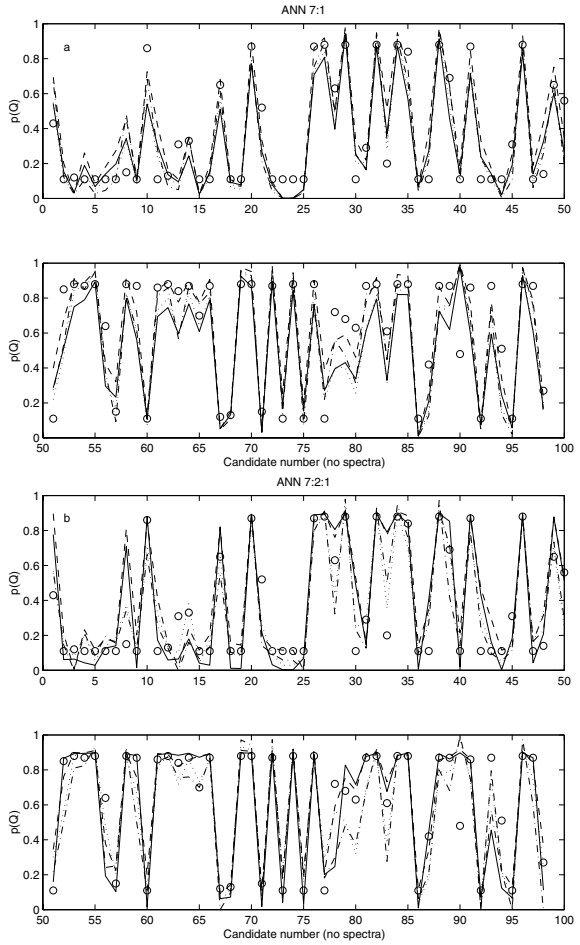

Figure 13. p(Q) distribution for the 98 FBQS-2 candidates without spectroscopic classification in White et al. (2000) using the ANN architectures 7 : 1 (a) and 7 : 2 : 1 (b). The four line types correspond to the four selected ANNs. Circles correspond to the p(Q) values obtained by White et al. with OC1.

Fig. 15(a) shows p(Q) for OC1 versus $\overline{p}$(Q) for the two ANN models. Fig. 15(b) is a similar plot in which the abscissa corresponds to the average of $\overline{p}$(Q) for the two ANN models. The mean and standard deviation of the differences are (0.04, 0.14) for OC1 − 7 : 1, (0.005, 0.13) for OC1 − 7 : 2 : 1 and (0.02, 0.13) for OC1 minus the average of the two ANN models. Figs 14 and 15, and the standard deviation values, show that the agreement in $\overline{p}$(Q) between the two ANN models is better than the agreement between any of them (or their average) and OC1.

The 98 sources in Table 4 were sought for associations in the NASA Extragalactic Database (NED). Five of them are confirmed extragalactic sources with spectroscopic redshift. FBQS J09 1309.2+41 3635 and FBQS J12 5018.1+36 4914 are classified as Ultraluminous Infrared Galaxies with redshifts 0.22 and 0.279 in Stanford, Stern & de Breuck (2000). The authors state that the majority of the ULIRGs in their sample are star-forming galaxies, and this interpretation is consistent with the low quasar probability we found, of ∼0.07 and ∼0.105, respectively (0.31 and 0.11 with OC1). FBQS J12 0354.7+37 1137, with $z = 0.401$, has broad emission lines (Appenzeller et al. 1998), and therefore corresponds to the quasar classification in our study, and it has in fact a quasar probability ∼0.94 (0.88 for OC1). A similar case is FBQS J12 5142.2+24 0435, with $z = 0.188$ and broad emission lines (Chen et al. 2002), and also a high quasar probability ∼0.82 (0.86 for OC1). FBQS J15 3411.3+26 2124 has $z = 0.1294$ and spectral type 'possibly Seyfert' (Keel, de Grijp & Miley 1988). We measure for the source p(Q) ∼ 0.04 (OC1 gives 0.11), which favours a spectral type 2 Seyfert galaxy, with narrow emission lines. In addition, four objects in Table 4 have spectroscopic classification in the recent Sloan Digital Sky Survey Data Release 2 (SDSS DR2). FBQS J08 3522.7+42 4258, FBQS J15 3402.2+42 5249 and FBQS J16 4733.9+36 4055 are quasars at $z = 0.805$, $z = 0.649$ and $z = 1.566$, with p(Q) ∼ 0.71, ∼ 0.88 and ∼0.73 (0.86, 0.88 and 0.87 with OC1). FBQS J07 4342.2+32 1543 is a star with p(Q) ∼ 0.48 (0.15 for OC1). Summarizing these results, the inspection of NED and SDSS DR2 shows that the five FBQS candidates classified as quasars have in fact rather high quasar probabilities – ∼0.94, ∼0.82, ∼0.71, ∼0.88 and ∼0.73 – whereas the two ULIRGs and the star have p(Q) values ∼0.07, ∼0.105 and ∼0.48, reinforcing the high efficiency of the ANN models.

4 CONCLUSIONS

In this work we have analysed the performance of neural networks for the selection of quasar candidates from combined radio and optical surveys with photometric and morphological data. Our work was based on the candidate list leading to FBQS-2 (White et al. 2000), and the input parameters used were radio flux, integrated-to-peak flux ratio, photometry and point spread function in the red and blue bands, and radio–optical position separation.

Two ANN architectures were investigated: a logistic model (7 : 1) and a model with a hidden layer with two nodes (7 : 2 : 1), and both yielded similarly good performances, allowing one to obtain subsamples of quasar candidates from FBQS-2 with efficiencies as large as 87 per cent at 80 per cent completeness. For comparison, the quasar fraction from the original candidate list was 56 per cent. More complex architectures were not explored, since the inclusion of the hidden layer – increasing the free parameters from 8 to 19 – did not improve the performance of the network. The efficiencies we find for completeness in the range 70 to 90 per cent are 90–80 per cent, similar to those found by White et al. using the oblique decision tree classifier OC1 and a similar sample size for the

Table 4. Quasar probabilities for the 98 candidates without spectroscopic classification in White et al. (2000).

Name	ANN 7 : 1		ANN 7 : 2 : 1		OC1	Name	ANN 7 : 1		ANN 7 : 2 : 1		OC1
FBQS J	$\overline{p}$(Q)	σ	$\overline{p}$(Q)	σ	p(Q)	FBQS J	$\overline{p}$(Q)	σ	$\overline{p}$(Q)	σ	p(Q)
07 1505.4+34 0501	0.61	0.08	0.70	0.17	0.43	12 2251.3+33 1640	0.25	0.07	0.39	0.15	0.56
07 1650.6+35 0520	0.15	0.04	0.14	0.06	0.11	12 2407.3+37 5332	0.30	0.07	0.26	0.08	0.11
07 1903.2+34 2550	0.03	0.01	0.04	0.04	0.12	12 2520.4+29 2420	0.55	0.09	0.67	0.17	0.85
07 3018.1+22 4502	0.17	0.07	0.17	0.09	0.11	12 2856.6+35 5635	0.85	0.07	0.87	0.04	0.88
07 3237.9+34 2952	0.06	0.02	0.07	0.04	0.11	12 3659.5+42 3641	0.85	0.04	0.87	0.03	0.87
07 3317.3+22 3725	0.12	0.06	0.16	0.03	0.11	12 3757.9+22 3430	0.94	0.03	0.91	0.02	0.88
07 3833.5+36 0957	0.18	0.07	0.16	0.02	0.11	12 4327.8+23 2811	0.36	0.06	0.32	0.13	0.64
07 4342.2+32 1543	0.40	0.08	0.56	0.24	0.15	12 4444.5+22 3305	0.20	0.10	0.19	0.06	0.15
08 2711.2+22 3323	0.10	0.02	0.12	0.08	0.11	12 4840.4+24 1240	0.84	0.05	0.90	0.02	0.88
08 3522.7+42 4258	0.63	0.09	0.78	0.10	0.86	12 4958.8+24 5233	0.62	0.07	0.70	0.16	0.87
08 5552.7+38 4325	0.31	0.08	0.29	0.10	0.11	12 5018.1+36 4914	0.10	0.02	0.11	0.07	0.11
08 5624.8+34 5024	0.12	0.04	0.12	0.04	0.13	12 5142.2+24 0435	0.77	0.07	0.86	0.03	0.86
09 1309.2+41 3635	0.09	0.03	0.05	0.04	0.31	12 5256.9+25 2503	0.82	0.06	0.88	0.01	0.88
09 1833.8+31 5620	0.30	0.06	0.24	0.10	0.33	12 5444.7+42 5305	0.63	0.10	0.79	0.07	0.84
09 1845.7+23 3833	0.02	0.00	0.07	0.04	0.11	13 1823.4+26 2623	0.84	0.06	0.84	0.06	0.87
09 3456.7+26 3054	0.15	0.03	0.14	0.08	0.11	13 1848.3+25 2815	0.70	0.09	0.80	0.09	0.70
10 1355.2+30 0546	0.59	0.07	0.71	0.13	0.65	13 2324.1+25 1809	0.84	0.05	0.87	0.02	0.87
10 2802.9+30 4743	0.08	0.02	0.09	0.06	0.11	13 4531.0+25 5504	0.05	0.00	0.05	0.04	0.12
10 2857.6+34 4054	0.07	0.01	0.08	0.06	0.11	13 4540.0+28 0123	0.13	0.02	0.11	0.03	0.13
10 3346.3+23 3220	0.81	0.05	0.88	0.03	0.87	14 0819.3+29 4950	0.95	0.02	0.93	0.03	0.88
10 3818.1+42 4442	0.28	0.05	0.20	0.05	0.52	14 1257.7+23 2618	0.92	0.04	0.91	0.02	0.88
10 5330.9+33 1342	0.07	0.02	0.09	0.04	0.11	14 3655.7+23 4928	0.03	0.00	0.06	0.04	0.15
10 5653.3+33 1945	0.00	0.00	0.05	0.05	0.11	14 4053.9+27 0642	0.94	0.04	0.93	0.03	0.87
11 0113.8+32 3155	0.00	0.00	0.05	0.05	0.11	14 4755.7+38 2813	0.19	0.05	0.16	0.08	0.11
11 2242.8+41 4355	0.04	0.01	0.04	0.04	0.11	14 5007.2+31 5050	0.90	0.05	0.91	0.01	0.88
11 3020.4+42 2204	0.76	0.07	0.84	0.04	0.87	15 0228.5+35 4455	0.13	0.05	0.14	0.08	0.11
11 3124.2+26 1951	0.87	0.05	0.90	0.01	0.88	15 0428.0+26 2419	0.81	0.07	0.88	0.03	0.88
11 3324.7+32 3449	0.45	0.05	0.58	0.24	0.63	15 0435.8+33 5728	0.27	0.08	0.22	0.05	0.11
11 3442.0+41 1330	0.96	0.02	0.93	0.04	0.88	15 0555.4+42 4415	0.48	0.09	0.43	0.18	0.72
11 3609.0+36 0641	0.26	0.06	0.37	0.16	0.11	15 1314.9+34 2111	0.47	0.08	0.64	0.17	0.68
11 3639.1+37 2651	0.19	0.03	0.18	0.07	0.29	15 1627.3+30 5220	0.34	0.09	0.52	0.21	0.63
11 3707.7+29 0324	0.92	0.03	0.91	0.01	0.88	15 1913.4+25 2134	0.71	0.10	0.79	0.11	0.87
11 3921.2+35 0748	0.37	0.10	0.61	0.21	0.20	15 2049.1+37 5219	0.85	0.06	0.90	0.02	0.88
11 4048.0+33 2908	0.91	0.03	0.91	0.02	0.88	15 2158.4+38 1814	0.39	0.08	0.52	0.22	0.61
11 4111.1+30 0442	0.65	0.08	0.80	0.08	0.84	15 2547.2+42 5210	0.87	0.05	0.88	0.00	0.88
11 5244.4+31 1123	0.06	0.01	0.08	0.06	0.11	15 3402.2+42 5249	0.87	0.04	0.88	0.01	0.88
11 5943.8+30 3348	0.28	0.07	0.33	0.06	0.11	15 3411.3+26 2124	0.01	0.00	0.06	0.05	0.11
12 0354.7+37 1137	0.95	0.03	0.93	0.03	0.88	15 3420.2+41 3007	0.19	0.06	0.29	0.12	0.42
12 0908.4+26 5131	0.57	0.09	0.67	0.18	0.69	15 3521.6+33 1826	0.76	0.07	0.83	0.05	0.87
12 1147.1+24 0736	0.13	0.03	0.13	0.08	0.11	15 3818.6+41 0548	0.73	0.07	0.80	0.09	0.87
12 1232.3+42 5821	0.76	0.06	0.83	0.06	0.87	15 4007.6+25 2836	0.99	0.01	0.95	0.05	0.48
12 1355.3+36 5255	0.26	0.05	0.36	0.11	0.11	15 4049.2+39 0351	0.67	0.08	0.80	0.07	0.86
12 1529.6+39 1200	0.14	0.03	0.17	0.08	0.11	15 5537.5+22 1327	0.06	0.02	0.04	0.04	0.11
12 1727.8+29 0449	0.02	0.01	0.05	0.05	0.11	15 5723.9+42 0825	0.67	0.09	0.64	0.16	0.87
12 1902.5+22 2416	0.18	0.06	0.15	0.03	0.31	16 0531.1+24 3147	0.22	0.09	0.20	0.05	0.51
12 2004.3+31 1148	0.89	0.04	0.89	0.01	0.88	16 2237.8+23 5943	0.05	0.02	0.07	0.05	0.11
12 2034.6+36 3357	0.12	0.06	0.13	0.06	0.11	16 3718.8+27 2607	0.96	0.02	0.93	0.03	0.88
12 2208.1+24 0012	0.33	0.08	0.27	0.06	0.14	16 4733.9+36 4055	0.69	0.09	0.76	0.12	0.87
12 2221.3+37 2335	0.66	0.06	0.80	0.09	0.65	17 0753.9+27 2418	0.18	0.04	0.18	0.16	0.27

training. The lack of a clean separation between quasars and non-quasars in the parameter space certainly limits the accuracy of the classification, and the agreement in the performances obtained favours in fact the interpretation that the three classifiers approach the maximum value achievable with this data base. Although neither of the two artificial intelligence tools provides a secure quasar classification (say efficiency larger than 95 per cent for a reasonable completeness), they are powerful for prioritizing targets for observation.

We report the probabilities obtained with the two ANN models for the 98 FBQS-2 candidates without spectroscopic classification in White et al. Our results are compared with those found by White et al. using OC1. The three models are found to be in agreement, with a better match between the two ANN models (standard deviation of the difference in probabilities $\sim$0.06) than between them and OC1 (standard deviation $\sim$0.13).

To our knowledge, this is the first work exploring the performance of ANNs for the selection of quasar samples. Our study demonstrates

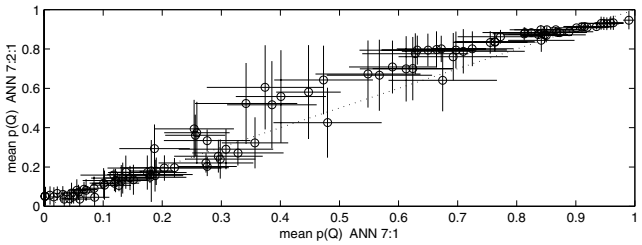

Figure 14. Comparison between $\overline{p}$(Q) for the 7 : 1 and the 7 : 2 : 1 ANNs, for the FBQS-2 candidates without spectroscopic classification in White et al. (2000).

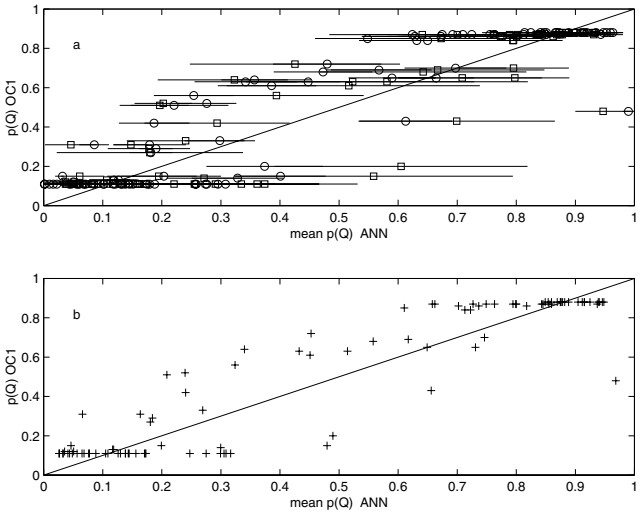

Figure 15. (a) p(Q) OC1 versus $\overline{p}$(Q) for the 7 : 1 (circles) and 7 : 2 : 1 (squares) ANN models. (b) p(Q) OC1 versus the average of $\overline{p}$(Q) for the 7 : 1 and the 7 : 2 : 1 ANN models.

the ability of ANNs for automated classification in astronomical data bases.

ACKNOWLEDGMENTS

We thank the anonymous referee for useful comments on the manuscript. RC and JIGS acknowledge financial support from DGES project PB98-0409-c02-02 and from the Spanish Ministerio de Ciencia y Tecnología under project AYA 2002-03326. This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the US Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are University of Chicago, Fermilab, Institute for Advanced Study, Japan Participation Group, Johns Hopkins University, Los Alamos National Laboratory, Max-Planck-Institute for Astronomy (MPIA), Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, United States Naval Observatory and University of Washington. MATLAB is a trademark of The MathWorks, Inc.

REFERENCES

Appenzeller I. et al., 1998, ApJS, 117, 319
Bailer-Jones C. A. L., Irwin M., von Hippel T., 1998, MNRAS, 298, 361
Bailer-Jones C. A. L., Gupta R., Singh H. P., 2001, in Gupta R., Singh H. P., Bailer-Jones C. A. L., eds, Automated Data Analysis in Astronomy. Narosa Publishing House, New Delhi, p. 51
Ball N. M., Loveday J., Fukugita M., Nakamura O., Okamura S., Brinkmann J., Brunner R. J., 2004, MNRAS, 348, 1038
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bishop C. M., 1995, Neural Networks for Pattern Recognition. Oxford Univ. Press, Oxford
Chen Y., He X.-T., Wu J.-W., Li Q.-K., Green R. F., Voges W., 2002, AJ, 123, 578
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195
Folkes S. R., Lahav O., Maddox S. J., 1996, MNRAS, 283, 651
Hagan M. T., Menhaj M., 1994, IEEE Trans. Neural Networks, 5, 989
Keel W. C., de Grijp M. H. K., Miley G. K., 1988, A&A, 203, 250
Lahav O., Naim A., Sodré L., Jr., Storrie-Lombardi M. C., 1996, MNRAS, 283, 207
Lauberts A., Valentijn E. A., 1989, The Surface Photometry Catalogue of the ESO-Uppsala Galaxies. European Southern Observatory, Munich
McMahon R. G., Irwin M. J., 1992, in MacGillivray H. T., Thomson E. B., eds, Digitised Optical Sky Surveys. Kluwer, Dordrecht, p. 417
Murthy S. K., Kasif S., Salzberg S., 1994, J. Artif. Intell. Res., 2, 1
Owens E. A., Griffiths R. E., Ratnatunga K. U., 1996, MNRAS, 281, 153
Pennington R. L., Humphreys R. M., Odewahn S. D., Zumach W., Thurmes P. M., 1993, PASP, 107, 279
Richard M. D., Lippmann R. P., 1991, Neural Comput., 3, 461
Stanford S. A., Stern D., de Breuck C., 2000, ApJS, 131, 185
Storrie-Lombardi M. C., Lahav O., Sodré L. Jr., Storrie-Lombardi L. J., 1992, MNRAS, 259, 8
Tagliaferri R. et al., 2003, Neural Networks, 16, 297
White R. L. et al., 2000, ApJS, 126, 133

This paper has been typeset from a TEX/LATEX file prepared by the author.