The Astrophysical Journal Supplement Series, 152:201–209, 2004 June

AUTOMATED CLASSIFICATION OF 2000 BRIGHT *IRAS* SOURCES

Ranjan Gupta
IUCAA, Post Bag 4, Ganeshkhind, Pune 411 007, India; rag@iucaa.ernet.in

Harinder P. Singh
Department of Physics and Astrophysics, University of Delhi, Delhi-110007, India; hpsingh@physics.du.ac.in

and

K. Volk and S. Kwok
Department of Physics and Astronomy, University of Calgary, 500 University Drive N W, Calgary, AL T2N 1N4, Canada;
volk@iras.ucalgary.ca, kwok@iras.ucalgary.ca
Received 2001 October 17; accepted 2004 March 11

ABSTRACT

An artificial neural network (ANN) scheme has been employed that uses a supervised back-propagation algorithm to classify 2000 bright sources from the Calgary database of *Infrared Astronomical Satellite* (*IRAS*) spectra in the region 8–23 μm. The database has been classified into 17 predefined classes based on the spectral morphology. We have been able to classify over 80% of the sources correctly in the first instance. The speed and robustness of the scheme will allow us to classify the whole of the Low Resolution Spectrometer database, containing more than 50,000 sources, in the near future.

Subject headings: infrared: galaxies — methods: data analysis

On-line material: machine-readable table

1. INTRODUCTION

The *Infrared Astronomical Satellite* (*IRAS*) Low Resolution Spectrometer (LRS) recorded spectra of some 50,000 sources in blue (8–15 μm) with $\lambda/\Delta\lambda \sim 40$ and in red (13–23 μm) with a resolution $\sim$20. A total of 5425 objects with better quality spectra were included in the Atlas of Low-Resolution *IRAS* Spectra (1986, hereafter the Atlas). Volk & Cohen (1989a) published spectra of 356 *IRAS* point sources with $F_\nu(12\ \mu\text{m}) > 40$ Jy that were not included in the Atlas. These brighter sources were classified into nine classes based upon the spectral morphology. Sixty percent of the sources have silicate emission and red-continuum spectra associated with H II region sources. No emission-line sources formed part of the set of 356 spectra. This sample was also used to test the classification scheme of *IRAS* sources based on broadband colors. Classifiable spectra were found for 338 of the sources in the sample. The remaining 18 sources had either extremely noisy or incomplete spectra. It was found that some class of sources overlapped on the color-color diagrams and, therefore, the nature of some of the *IRAS* sources could not be determined from the *IRAS* photometry.

Volk et al. (1991) published an additional 486 spectra belonging to sources with 12 μm fluxes between 20 and 40 Jy that were also not in the Atlas. Classifiable spectra were found for 424 sources. The spectra were classified into nine groups as in Volk & Cohen (1989a) that describe the astrophysical nature of these sources. Kwok et al. (1997) processed 11,224 spectra (including sources in the Atlas), corresponding to a flux limit of 7 Jy at 12 μm. These spectra were also classified by eye and put into nine classes based on the presence of emission and absorption features and on the shape of the continuum. They identified optical counterparts of these *IRAS* sources in the existing optical and infrared catalogs and listed the optical spectral types if they were known.

It is evident that large databases like the one referred to above require automated schemes for any analysis. Artificial neural networks (ANNs) have been employed extensively in several branches of astronomy for automated data analysis (Lahav & Storrie-Lombardi 1994). ANNs have been used previously by the IUCAA group in three distinct areas of stellar astronomy. They have been applied to classify digitized optical and ultraviolet spectra (Gulati et al. 1994a, 1994b; Singh et al. 1998); to compare a set of observed spectra of F and G dwarfs with a library of synthetic spectra (Gulati et al. 1997a; Gupta et al. 2001); and to determine the reddening properties of hot stars from the low-dispersion ultraviolet spectra (Gulati et al. 1997b).

We have attempted to classify the 2000 brightest sources from the Atlas into 17 classes by means of ANNs. In the next section we describe features of the 17 new classes. In § 3, we present details of the ANN scheme. Results are discussed in § 4, and important conclusions of the study are presented in the last section. Further, an Appendix has been added to explain the general ANN architecture for the benefit of the readers.

2. SPECTRAL CLASSES

The set of 2000 spectra were classified into the 17 classes by eye by one of us (K. Volk). These classifications are assumed to from the reference set of "correct" classifications. Such by-eye classification has an element of subjectivity in it, and is more apt to have problems when the spectra are noisy. This set of 2000 sources are the brightest sources in the Atlas, and they have the least problems with noise or spectral peculiarities. This makes the by-eye classification as accurate as possible despite the element of subjectivity.

The groups used for spectral classification are an extension of the nine classes described in Volk & Cohen (1989a). The current scheme introduces more classes, for three reasons:

1. The "H II region sources" group shows a variety of features, and so separate classes have been made for the more obvious feature types so that these classes are, one hopes, more uniform.

TABLE 1
A Short Description of the 17 ANN Training Classes

Source Type	Description	ANN Class
Stellar	Stellar photospheric spectra; B type to early M type	2
	Lower temp. stellar continuum spectra with no features: mid to late M type, little circumstellar dust	3
Carbon stars	Strong 11.3 μm emission	1
	Lower temp. continuum ($\sim$250 K), e.g., AFGL 3068 and no strong features	10
	Weaker 11.3 μm emission	15
Oxygen-rich AGB stars	Silicate absorption on an intermediate temp. continuum (400–200 K)	5
	Stronger 10 μm emission on a higher temp. continuum ($\geq$600 K)	6
	Weaker 10 μm emission on a higher temp. continuum ($\geq$600 K)	12
	10 μm features in transition from emission to absorption	13
Planetary nebulae (PN) and post-AGB sources (PPN)	Emission lines (PN)	0
	Silicate emission on a low temp. continuum (oxygen-rich PN and PPN)	8
	Lower temp. ($\sim$150 K) featureless (carbon-rich PN and PPN)	11
	21 μm emission feature (carbon-rich PPN)	14
ISM spectra: mostly H II regions; includes a few PN, PPN and galaxies	UIR or PAH features on a flat continuum (H II regions, galaxies, a few PN and PPN)	4
	Low temp. continuum ($\leq$100 K) without strong features	7
	10 μm absorption on a low temp. continuum ($\leq$100 K), includes a few PPN	9
	UIR or PAH features on a low temp. continuum ($\leq$100 K), includes a few PN	16

2. A few new classes have been created within the 10 μm silicate feature and 11.3 μm SiC feature groups to divide the objects with weaker features from those with stronger features.
3. Some small groups of unusual sources can be identified from the "usual/unknown" group of Volk & Cohen (1989a); hence, these were given their own classes here such as class 14 for 21 μm feature sources.

The division between the spectra with "stronger" and "weaker" features was made at LRS type 25 or 45 (for silicate emission sources and SiC emission feature sources respectively), since the original LRS classes are based upon the feature strength. Thus, the spectra with strong features, LRS types 25–29 or 45–49, are separated from the spectra with weaker features, LRS types 21–24 or 41–44. All objects with LRS classes near the dividing line were examined by eye to make sure that the LRS type accurately reflects the feature strength. In some cases this led to spectra being placed into the "stronger" or "weaker" feature groups even when the LRS type is not as expected. In the large majority of cases no problem with the LRS type was found.

A short summary of the 17 classes is given in Table 1, with representative spectra plotted in Figure 1. A detailed description of the different classes and their relation to the original LRS types now follows:

(0) Line emission sources.—This class is for spectra with strong emission lines (line peak to continuum ratio of 1 or larger) and no obvious dust emission or absorption features. These are nearly all planetary nebula sources. A few cases of H II region sources that may show the 12.8 μm [Ne II] line could have been included here, but they were left in the other "H II region" groups since the 12.8 μm emission line seen in LRS spectra is sometimes due to an instrumental problem rather than being an actual emission line, and since the possible [Ne II] lines were relatively weak. Comparatively few H II region spectra have strong enough emission lines as observed by the LRS instrument to be potentially included in this group. In the original LRS classification, these are generally classified as types 81–96.

The prototype spectrum in Figure 1 shows several emission lines: 9.0 μm [Ar III], 12.8 μm [Ne II], 15.5 μm [Ne V], and 18.7 μm [S III]. Other objects in this class tend to have fewer emission lines than this example.

(1) Sources with stronger 11.3 μm SiC *emission features*.—In this group one has evolved carbon stars undergoing mass loss and showing a strong 11.3 μm emission feature. In the original LRS classification these would be type 45–49 sources, implying the feature peak to estimated continuum ratio is larger than 1.648.

(2) Stellar continuum sources.—Here is the class for stars without circumstellar dust shells that radiate significantly in the 8–23 μm region, including Vega, Sirius, Aldebaran, and Antares. Most of these sources are of K type to early-M type. Only a few have spectral types earlier than K0. In the original LRS classification these would be type 16–19 sources.

(3) Featureless sources with cooler color temperatures.—This group is as in Volk & Cohen (1989a). The continuum color temperature is lower than expected for the photospheric

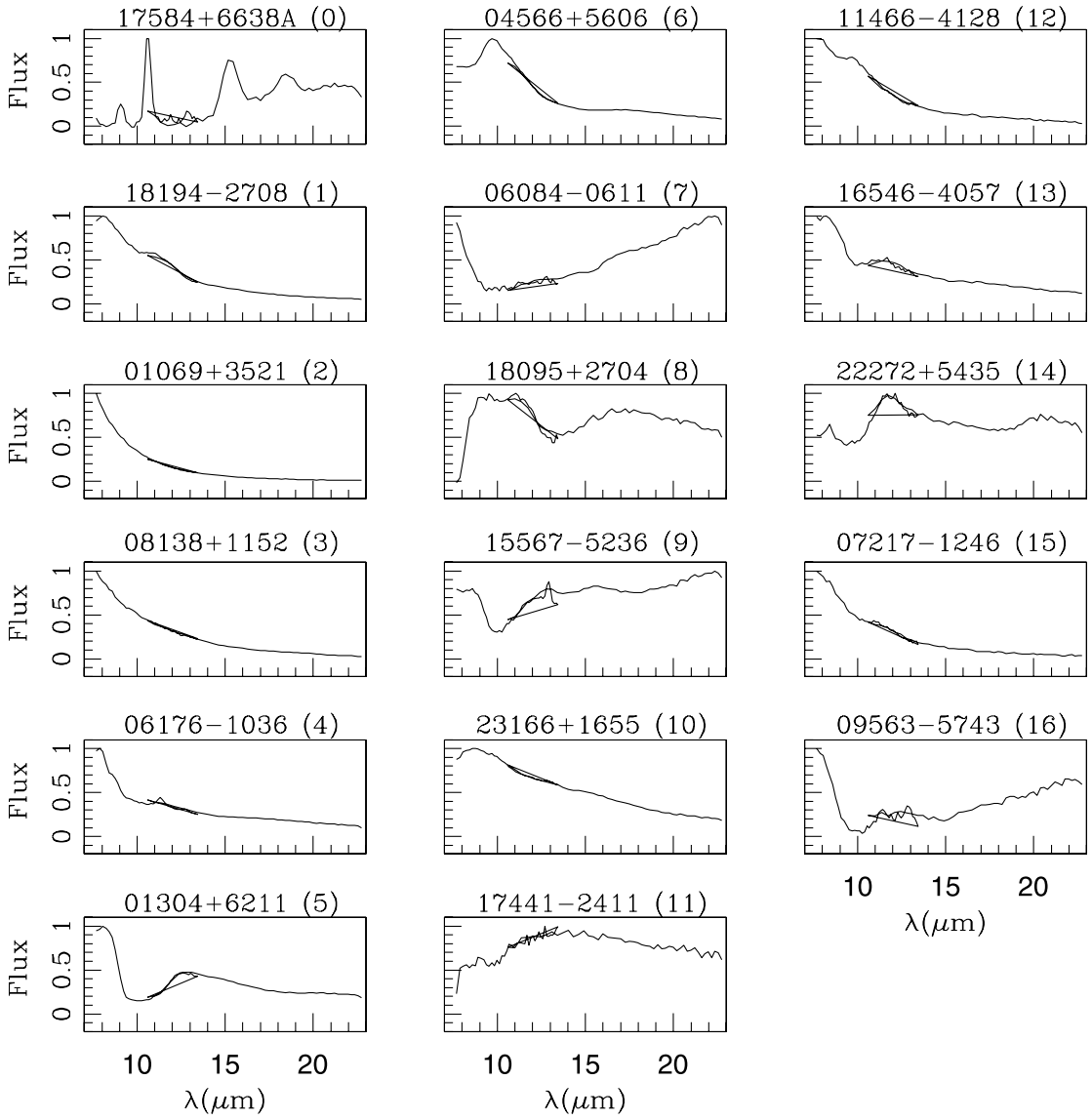

FIG. 1.—LRS spectra for the 17 training classes. There are 10 training spectra each for the 17 Catalog classes, but only one representative spectrum is shown. For each spectrum the *IRAS* name and the Catalog class (*in parenthesis*) are given above the plot. Please note that there exists a crossover region between the "blue" and the "red" spectra (as a result of the instrument's settings) ranging between 10 and 14 μm, and this is seen in these plots.

emission of any normal star, including late M-type stars. Aside from the unusually low color temperature for a stellar photosphere, there is no overt sign of any dust emission. Many of these sources have optical counterparts among known Mira or irregular variables of M type. In the original LRS classification, these would be type 12–15 sources. All featureless continuum sources with λF_λ ratios of less than 3 between 7.9 and 11 μm but which still have declining continua over the LRS wavelength range are placed in this class.

(4) Sources with UIR features at 7.6/8.5/11.3/12.5 μm.—This group includes sources with the UIR features, except for those with very steep ("red" according to the Atlas definition) continua, which form class 16. These would ideally be type 80 sources in the original LRS classification, but there was considerable confusion with other types of objects.

As long as the spectrum signal to noise ratio is reasonable, it is possible to objectively distinguish between UIR feature spectra and silicate absorption, as discussed in Volk & Cohen (1989b).

(5) Sources with 10 μm *silicate absorption*.—This class includes objects with silicate absorption features, but we attempt to exclude cases of compact H II regions with foreground silicate absorption, which have much lower continuum color temperature than the "normal" silicate absorption feature sources such as IRAS 01304+6211 (see Fig. 1). If the spectral continuum is rising from 7.6–23 μm and there is a silicate absorption feature at 10 μm, the spectrum is put in class 9. In the original LRS classification, class 5 sources correspond to types 31–39. An additional refinement is that sources that are in transition between silicate emission and silicate absorption at 10 μm are grouped in class 13.

(6) Sources with strong silicate emission at 10 μm.—Spectra with a falling continuum over the LRS wavelength range and having a silicate emission feature with a peak greater than 1.628 times the continuum are in this class. These would be types 25–29 in the original LRS classification. Similar spectra but with weaker features are put into class 12.

(7) Sources with low temperature ($\leq$100 K) *dust continuum emission*.—In this group are mostly H II region spectra that do not show very strong features owing to silicates or the UIR features. The continuum must rise over the LRS wavelength

range from 7.6–23 μm for a spectrum to be included in this class, corresponding roughly to a color temperature of less than 100 K, but generally the continuum shape cannot be described by a single color temperature. In the original LRS classification these would have been given some type in the 60s, 70s, or 80s. In a few unusual cases there can be other types of spectra mixed into this group. When there are accompanying strong features such as the 10 μm feature in emission or absorption the spectra are put into classes 8, 9, or 12.

(8) Sources with silicate emission and a low temperature dust continuum.—Here are spectra with continuum shape similar to that of class 7 but with a clear 10 μm silicate emission feature. In principle this corresponds to types 61–69 in the original LRS classification. Some of the sources are H II regions, but others are young planetary nebulae such as Hb 12, Vy 2-2, and SwSt-1; others are their immediate progenitors caught in the phase between the asymptotic giant branch phase and the planetary nebula phase (''proto–planetary nebulae'' or PPNs) such as HD 161796 and IRAS 18095+2704. The spectrum of 18095+2704 is shown in Figure 1; this has unusually strong features, and the continuum is not as steeply rising as is typically the case for spectra in this class.

(9) Sources with silicate emission and a low temperature dust continuum.—This is another class in which the continuum is required to rise over the LRS wavelength range from 7.6–23 μm. Here we include spectra showing the 10 and 18 μm silicate features in absorption. These are nearly all compact H II region sources with (presumably) foreground cold dust in the associate molecular cloud. In the original LRS classification these would be type 71–79 sources.

(10) Extreme carbon stars.—The prototype for this group is AFGL 3068 (IRAS 23166+1655), which is known to be a carbon star with an extremely optically thick dust shell. The spectra are selected for this group on the basis of having low color temperatures (of order 300 K) in the 10 μm region of the spectrum and of having no strong emission or absorption features. In some cases there is some type of weak feature around 11 μm, which might be due to the 11.3 μm SiC feature in emission or weak absorption. These objects are discussed in Volk et al. (1992).

(11) Low color temperature featureless spectra.—The group of spectra have low color temperature (about 150–200 K) featureless continua. They have noticeably lower color temperatures than the extreme carbon star spectra (see Fig. 1) but peak somewhere in the LRS wavelength range and decline thereafter, unlike the H II region spectra in classes 7, 9, and 16. A few of these sources are known to be either young planetary nebulae or PPNs, and they all seem to have carbon-based dust shells. These objects are discussed by Volk, Kwok & Woodsworth (1993). In the original LRS classification they were generally assigned type 05 or 50.

(12) Sources with weaker silicate emission features.—This group of spectra is the analog of class 6, but they have feature peak to continuum ratios of less than 1.628. In the original Atlas classification these would be type 21–24 spectra. In various cases the feature was not detected at all in the original LRS Atlas classification, and these were generally classified as type 15–18.

(13) Sources with intermediate optical depth silicate dust shells.—Prototype spectra for this class include IRAS 19192+0922, IRAS 17125−4814, IRAS 15119−6453, and IRAS 16546−4047 (see Fig. 1). All these spectra have continua that fall between 7.6 and 23 μm. The optical depth of the dust shells in these objects is such that the 10 μm feature is in transition between emission and absorption while the 18 μm feature is still in emission. In the original LRS classification these were often confused with SiC emission spectra, and so given erroneous types of 41–45.

(14) Source with the 21 μm *feature*.—A small number of sources are known to have a dust feature at 21 μm, one possible identification of which is TiC grains (von Helden et al. 2000). The feature was discovered in the LRS spectra. Less than 20 such sources are known, the brighter of which are used to define the group here. All sources in this group are carbon-rich PPNs. A few extreme carbon stars and the planetary nebula IC 418 appear to have very weak 21 μm features, but these require the *Infrared Space Observatory* to detect, and so are not included here.

(15) Sources with weaker 11.3 μm SiC *emission*.—All spectra from carbon stars with the 11.3 μm feature but with a feature-to-continuum ratio of less and 1.628 are put in this group, which is closely analogous to class 1. In the original LRS classification these would be of type 41 to 44.

(16) Cool continuum sources with strong UIR features.—Any spectrum that has a continuum that rises from 7.6–23 μm (allowing for the possibility of a strong 7.6 μm feature at the short-wavelength end of the spectrum) and strong UIR features at 7.6 and 11.3 μm is placed in this class. These are H II region spectra with UIR features. In the original LRS classification these would be of type 80 or 81. In some cases it is possible that there is an underlying silicate absorption component as well as the UIR features, causing a very large rise at the short-wavelength end of the LRS spectrum. However, these spectra lack any 18 μm absorption feature, which makes the presence of strong 10 μm silicate absorption less likely.

3. ARTIFICIAL NEURAL NETWORK SCHEME

In this analysis, we have used the multilayer back-propagation (MBPN) neural network scheme with supervised learning as described in Gulati et al. (1994a, 1994b). This scheme requires a set of spectra to be predefined as the training set and should include all the classes of the unknown set that the network is supposed to classify. The algorithm trains on this training set and subsequently in the test phase, classifies the unknown test set into the predefined classes.

A training set of 170 spectra having 10 sources for each training class was set aside for training purposes. Figure 1 shows a representative sample of each of these 17 classes. Each spectrum consists of 93 flux values in the range of 8–23 μm. We may mention that 148 numbers of the 170 training spectra were included in the 2000 test spectra for validity checks.

The ANN scheme involved using this set of 170 input spectra with 17 assigned classes (0–16) to train the ANN algorithm. We refer to these classes as the catalog classes. The ANN configuration used was 93-16-16-17, implying 93 data points for each of the 17 output spectral classes with two hidden layers of 16 nodes each (a detailed explanation on the ANN architecture is given in the Appendix). The training session involved an iterative procedure with the network weights getting modified at each iteration. During an iteration, the computed output and the desired output were compared and the resultant error was then utilized to modify the network weights for the next iteration using a back propagation algorithm. The learning or training was stopped once the error was minimized to a predefined level and the network weights were considered to be frozen (Gulati et al. 1994a). A learning curve for 10,000 iterations is given in Figure 2 (*left panel*).

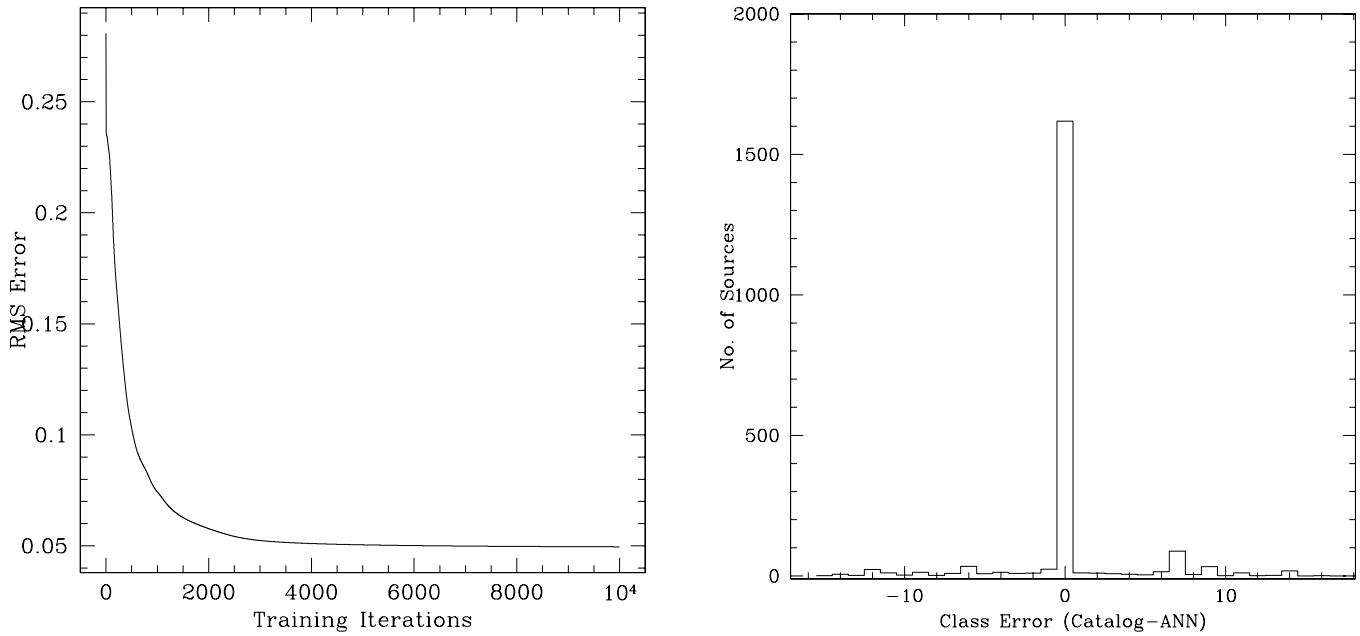

FIG. 2.—ANN learning curve for 10000 iterations (*left panel*). The rms error refers to the rms difference between the computed output and the desired output for each iteration. The right panel shows the histogram of classification accuracy for the 2000 test patterns with 17 training classes.

Subsequent to the above training session, the test session uses the frozen weights determined above to perform the classification of the 2000 test set of spectra into 17 catalog classes.

4. RESULTS AND DISCUSSION

A general picture of the result of our ANN classification scheme is given in Figure 2 (*right panel*) in the form of a histogram. A total of 1618 spectra have been classified correctly out of the total sample of 2000 source spectra indicating a success rate of 80% at the first instance. Table 2 provides a summary of the ANN classification of the 2000 test sources into the 17 groups or classes. Table 3 lists 382 of the misclassified sources with their catalog class and the corresponding ANN class. In the following, we describe some finer details of the classification accuracy for individual classes 0 to 16.

Out of a total 23 spectra for class 0, 19 were classified correctly. From the four that were misclassified, one each was classified as class 7, 8, 13, and 16. Table 3 lists the four spectra of class 0 that were misclassified. Source 22036+5306 has been put in class 16 by the ANN probably because of the strong feature at around 8 μm, typical of class 16.

Out of a total 136 spectra for class 1, 113 were classified correctly. From the 23 that were misclassified, 11 were misclassified into class 13, six were misclassified into class 15, three were misclassified into class 4, two were misclassified into class 6, and one was misclassified into class 14. The misclassified sources of class 1 are listed in Table 3. Out of these, 18 spectra seem to belong to the class assigned by the ANN and not the catalog class, specially those put into classes 13 and 15. ANN classification accuracy of class 2 has the best agreement with the catalog class. Out of 89 spectra of class 2,

TABLE 2

2000 TEST SOURCES CLASSIFIED BY ANN INTO 17 CATALOG CLASSES

	ANN CLASS																	
CATALOG CLASS	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	TOTAL
0	19	...	...	...	...	...	...	1	1	...	...	...	...	1	...	...	1	23
1	...	113	...	...	3	...	2	...	...	...	...	...	...	11	1	6	...	136
2	...	...	88	1	...	...	...	...	...	...	...	...	...	...	...	...	...	89
3	...	1	5	85	...	...	...	...	...	...	...	...	1	1	...	1	...	94
4	...	...	1	...	34	22	...	...	...	2	1	...	...	10	2	4	11	87
5	...	...	...	...	4	88	...	...	...	2	...	...	...	1	1	...	7	103
6	...	1	...	...	...	...	679	...	8	...	9	...	33	5	...	...	...	735
7	3	...	...	...	...	...	...	45	...	2	...	2	...	...	...	...	1	53
8	1	...	...	...	...	...	...	...	12	...	...	...	...	...	...	...	...	13
9	...	...	...	...	...	3	...	4	...	26	...	...	...	...	4	...	3	40
10	...	1	...	...	...	...	2	...	...	...	38	...	...	3	...	...	...	44
11	1	...	...	...	...	...	...	...	...	...	...	8	...	...	3	...	...	12
12	...	6	...	27	...	...	15	...	...	...	...	...	149	1	...	...	...	198
13	...	...	...	1	3	4	85	...	3	...	8	...	...	96	...	...	...	200
14	...	...	...	...	...	...	...	...	...	...	1	...	...	...	2	...	...	3
15	...	18	2	1	3	...	...	...	...	...	...	...	...	3	2	111	...	140
16	1	1	...	...	...	1	...	2	...	...	...	...	...	...	...	...	25	30

TABLE 3
LIST OF LRS SOURCES MISCLASSIFIED BY ANN

IRAS Designation	Catalog Class	ANN Class
04395+3601	0	13
17069−4149	0	7
21014−1133	0	8
22036+5306	0	16
03293+6038	1	15
04340+4623	1	13
15380−6545	3	12
17282−5102	3	2
17504−0234	3	2
19369+2823	3	1
20056+1834	3	13
22476+4047	3	2
14198−6115	4	5
14206−6151	4	16
15246−5612	4	16
15535−5328	4	5
16041−4912	4	5
16204−4717	4	5

NOTES.— Table 3 is available in its entirety in the electronic edition of the *Astrophysical Journal Supplement*. A portion is shown here for guidance regarding its form and content.

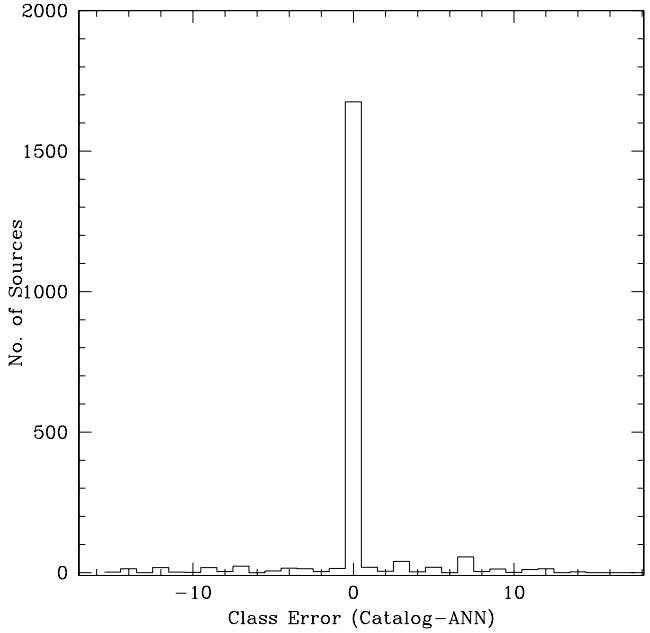

FIG. 3.—Histogram of classification accuracy for the 2000 test patterns with 16 training classes (class 12 trained as class 6).

88 spectra were classified correctly, while one spectrum was misclassified into class 3.

Of the 94 spectra for class 3, 85 were classified correctly. Of the nine that were misclassified, five were misclassified into class 2 and one each into classes 1, 12, 13 and 15. At least one source (20056+1834) has a spectrum that does not seem to belong to class 3 and has been put in class 13 by the ANN.

Of a total of 87 sources of class 4, 34 were correctly classified, while 53 were misclassified; 22 into class 5, 11 into class 16, 10 into class 13, four into class 15, two each into class 9 and 14, and one each into class 2 and 10. As mentioned in § 2, class 4 corresponds to type 80 sources in the original LRS classification in which there was a lot of confusion with the other types of sources. ANN put 22 sources into class 5, which has sources with 10 μm silicate absorption feature. We note that in the training sample of class 4, there are at least two sources (16367−4701 and 19327+3024) that might be resulting in contamination of this class with class 5 and hence less than efficient training for class 4.

Out of a total of 103 spectra for class 5, 88 were correctly classified. Seven spectra were misclassified into class 16, four into class 4, two into class 9, and one each into class 13 and 14. ANN has correctly picked up at least three spectra from the catalog that should be put in class 4 (Table 3).

Class 6 (strong 10 μm feature) has the largest sample of 735 sources. Out of these, 679 were classified correctly and 33 were misinterpreted as class 12 (weak 10 μm feature). Similarly, class 12 has 198 spectra, out of which 15 were misclassified as class 6, while 27 were misclassified into class 3. As discussed later, if one were to treat a weak and a strong 10 μm feature sources as belonging to a single class, percentage of correctly classified sources will be in excess of 90% for these classes.

For class 7, out of the 53 sources, 45 have been classified correctly. Three sources were incorrectly classified into class 0, while two sources each were incorrectly classified into classes 9 and 11. Class 8 has 13 sources, out of which only one has been wrongly classified into class 0.

Class 9 has 40 sources in all. Twenty-six have been correctly classified, four each have been misclassified into classes 7 and 14, and three each into classes 5 and 16. Class 10 has 44 sources in all, and 38 have been correctly classified, but three have been wrongly classified to class 13, two into class 6, and one into class 1. Class 11 has 12 sources, out of which eight were correctly classified. Three sources were incorrectly classified into class 14 and one into class 0.

As mentioned in § 2, in the original LRS classification, class 13 spectra were often confused with SiC feature sources. In the ANN classification also, out of the 200 class 13 sources, 104 were wrongly classified (85 into class 6), while 96 were correctly classified. The wrongly classified sources have been listed in Table 3 for the purpose of review by the human classifiers.

Class 14 has three sources, out of which one was misclassified into class 10. Class 15 had 140 sources, out of which 111 have been correctly classified. But 18 have been classified into class 1 and three each into classes 13 and 4; two each into class 2 and 4, and one into class 3. Finally, 25 sources out of a total of 30 of class 16 were classified correctly. Two sources were wrongly classified into class 7, while one each were classified into classes 0, 1, and 5.

If we assume classes 6 and 12 to be the same, (class 12 trained as class 6) and redo the classification exercise, we obtain the overall classification histogram as shown in Figure 3. A total of 1675 spectra have now been classified correctly, which is about 84% of the total test data set. This process also improved the classification of the class 6 sources, as shown in Figure 4.

We split off class 12 from class 6 because it improved the classification of silicate emission objects. With the full range of silicate emission objects in one class there was more dispersion of values and the ANN system had more chance of misclassification. However, it does not matter that sources are misclassified from class 6 to class 12 or class 12 to class 6 because the dividing line between these two groups is not sharp and there is a continuous range of silicate feature strengths, and some objects must fall at the boundary.

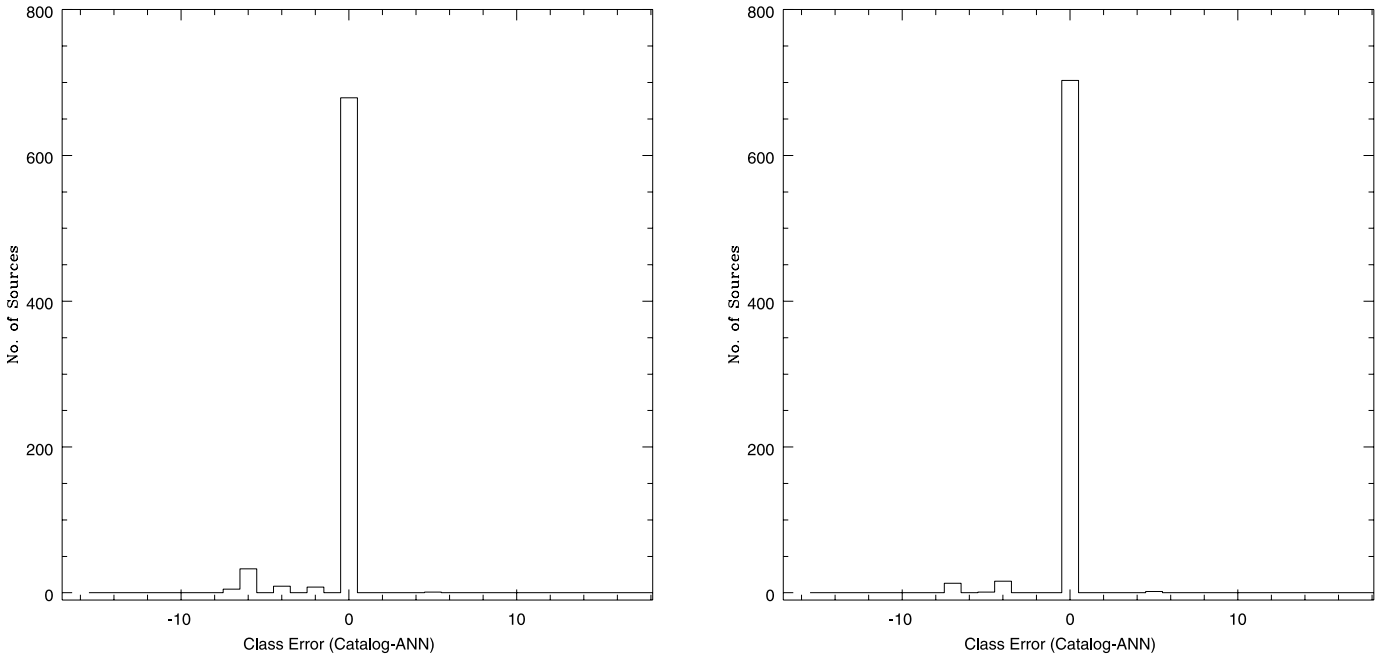

FIG. 4.—Histogram of classification accuracy for the 735 test patterns of class 6 with 17 training classes (*left panel*) and with combined training class 6 and class 12 (*right panel*).

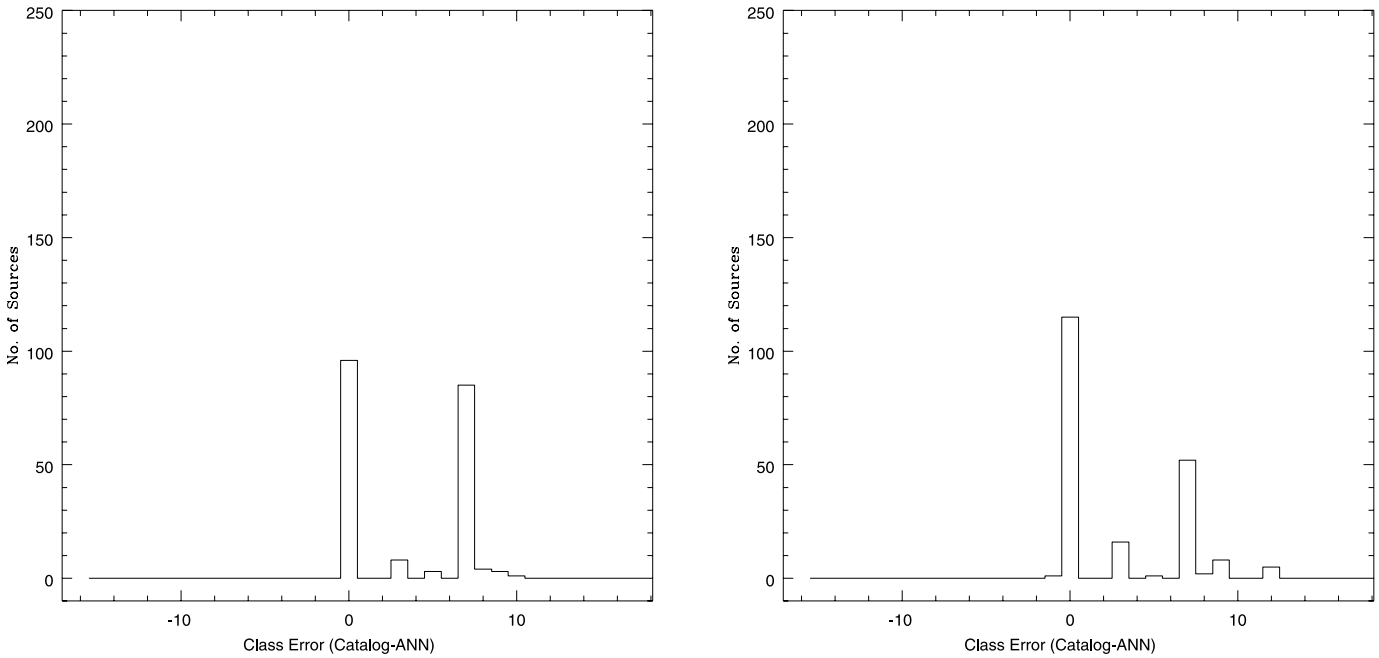

FIG. 5.—Histogram of classification accuracy for the 200 test patterns of class 13 with 17 training classes (*left panel*) and with combined training class 6 and class 12 (*right panel*).

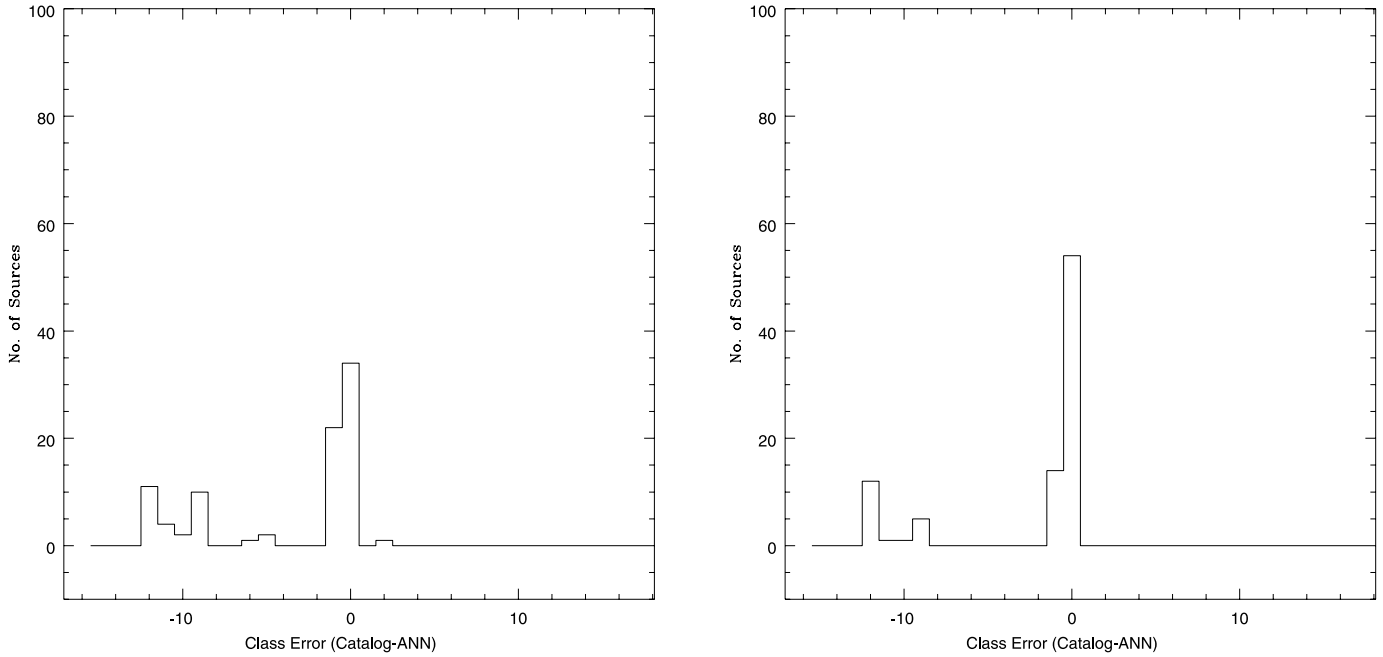

FIG. 6.—Histogram of classification accuracy for the 87 test patterns of class 4 with 17 training classes (*left panel*) and with combined training class 6 and class 12 (*right panel*).

Figure 5 shows the improvement in the class 13 classification, namely, earlier 85 sources of this class were wrongly classified as class 6, which has now reduced to 50 wrong classifications. This process also improved the classification of class 4, as seen in Figure 6. There were such improvements noticed in general, and Figure 3 is really a combined effect of all these.

5. CONCLUSIONS

We have demonstrated in this paper the application of the ANN scheme to a large database of 2000 *IRAS* spectra and have been able to correctly classify more than 80% of the data set. The misclassified spectra were looked in detail and most of them could have been wrongly cataloged or had features that would have been confusing for even a human classifier.

We stress here that the speed and robustness of this scheme can be very useful for classifying the whole of the LRS database containing over 50,000 sources.

R. G. and H. P. S. are grateful to S. K. and K. V. for kind hospitality while on a visit to Calgary. We thank the two anonymous referees for very useful comments.

APPENDIX

ARTIFICIAL NEURAL NETWORK ARCHITECTURE

In a feedforward neural network there are several inputs, a few hidden layers, and several outputs (Bailer-Jones et al. 2002). See Figure 7 for a block view of this architecture. Each node in the input layer holds a value, x_i. In our example application, the input vector, $(x_1, x_2, \ldots, x_i, \ldots)$, is the spectrum with 93 flux values, and the output vector, $(y_1, y_2, \ldots, y_l, \ldots)$, has 17 nodes. Each of the input nodes connects to every node in the next layer of nodes, the first "hidden" layer, and each of these connections has a weight, $w_{i,j}$, associated with it. A node in the hidden layer forms a weighted sum of its inputs, and passes this through a *nonlinear transfer function*, such that the output from the jth hidden node is

$$p_j = \tanh\left(\sum_i w_{i,j} x_i\right).$$

These values are passed to a second hidden layer, which performs a similar processing, the output from that layer being the vector $\boldsymbol{q}$:

$$q_k = \tanh\left(\sum_j w_{j,k} p_j\right).$$

The output layer then performs a simple sum of its inputs, so that the network output, y_l, is

$$y_l = \sum_k w_{k,l} q_k.$$

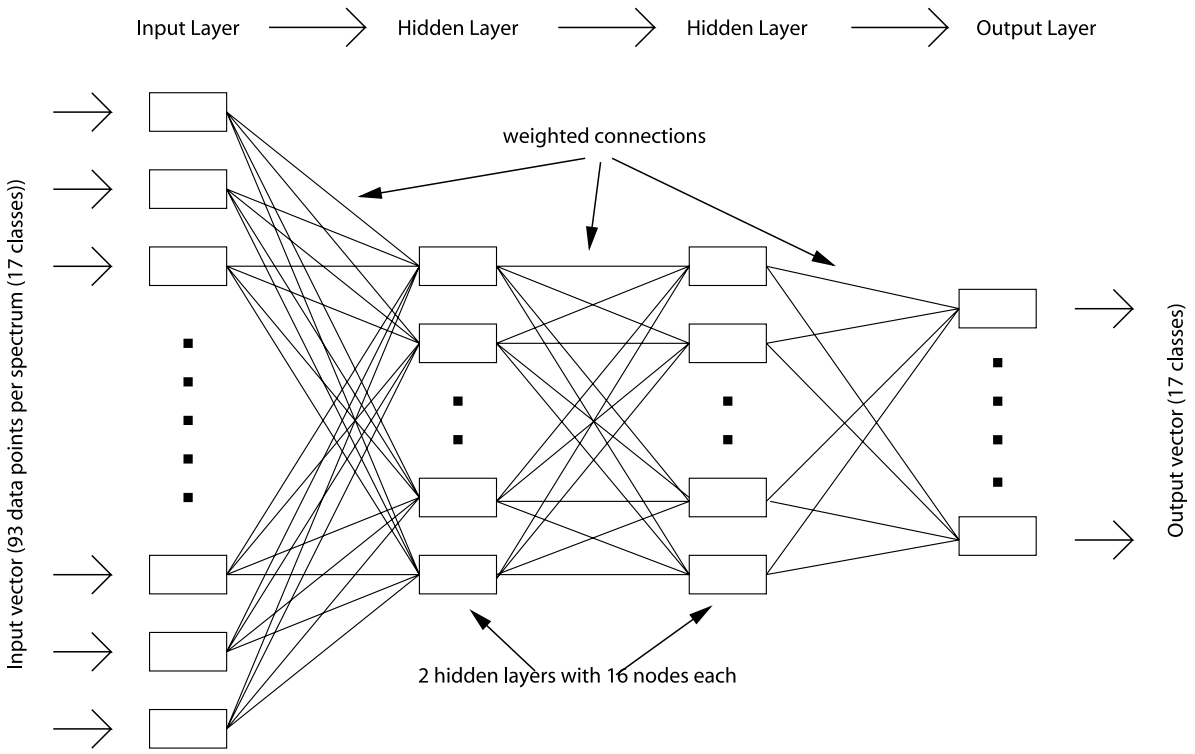

FIG. 7.—Feedforward artificial neural network architecture for the problem under investigation.

The tanh function in the hidden layers provides the nonlinear capability of the network. Other nonlinear functions are possible; the sigmoidal function $(1/(1 - \exp[-\sum wx]))$ is used here. Both functions map an infinite possible input range onto a finite output range, -1 to $+1$ in the case of tanh. This imitates the transfer function of neurons.

REFERENCES

Atlas of Low-Resolution *IRAS* Spectra. 1986, *IRAS* Science Team, prepared by F. M. Olnon & E. Raimond (A&AS, 65, 607) (Atlas)
Bailer-Jones, C. A. L., Gupta, R., & Singh, H. P. 2002, in Automated Data Analysis in Astronomy, ed. R. Gupta, H. P. Singh, & C. A. L. Bailer-Jones (New Delhi: Narosa Publ. House), 51
Gulati, R. K., Gupta, R., Gothoskar, P., & Khobragade, S. 1994a, ApJ, 426, 340
———. 1994b, Vistas Astron., 38, 293
Gulati, R. K., Gupta, R., & Rao, N. K. 1997a, A&A, 322, 933
Gulati, R. K., Gupta, R., & Singh, H. P. 1997b, PASP, 109, 843
Gupta, R., Gulati, R. K., & Singh, H. P. 2001, in Proc. 11th Cambridge Workshop, Cool Star, Stellar Systems, and the Sun—Challenges for the New Millennium, ed. R. J. Garcia-Lopez, R. Rebelo, & M. R. Zapatero Osorio, CD-ROM (San Francisco: ASP), 791
Kwok, S., Volk, K., & Bidelman, W. P. 1997, ApJ, 112, 557
Lahav, O., Storrie-Lombardi, M. C. 1994, Vistas Astron., 38, 249
Singh, H. P., Gulati, R. K., & Gupta, R. 1998, MNRAS, 295, 312
Volk, K., & Cohen, M. 1989a, AJ, 98, 931
———. 1989b, AJ, 98, 1563
Volk, K., Kwok, S., & Langill, P. P. 1992, ApJ, 391, 285
Volk, K., Kwok, S., Stencel, R. E., & Brugel, E. 1991, ApJS, 77, 607
Volk, K., Kwok, S., & Woodsworth, A. W. 1993, ApJ, 402, 292
von Helden, G., Tielens, A. G. G. M., van Heijnsbergen, D., Duncan, M. A., Hony, S., Waters, L. B. F. M., & Meijer, G. 2000, Science, 288, 313