Galaxy Zoo: reproducing galaxy morphologies via machine learning[★]

Manda Banerji,[1,2]† Ofer Lahav,[1] Chris J. Lintott,[3] Filipe B. Abdalla,[1]
Kevin Schawinski,[4,5]‡ Steven P. Bamford,[6] Dan Andreescu,[7] Phil Murray,[8]
M. Jordan Raddick,[9] Anze Slosar,[10] Alex Szalay,[9] Daniel Thomas[11]
and Jan Vandenberg[9]

[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT*
[2]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA*
[3]*Department of Physics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH*
[4]*Department of Physics, Yale University, New Haven, CT 06511, USA*
[5]*Yale Center for Astronomy & Astrophysics, Yale University, PO Box 208121, New Haven, CT 06520, USA*
[6]*Centre for Astronomy and Particle Theory, School of Physics & Astronomy, University of Nottingham,*
University Park, Nottingham NG7 2RD
[7]*LinkLab, 4506 Graystone Avenue, Bronx, NY 10471, USA*
[8]*Fingerprint Digital Media, 9 Victoria Close, Newtownards, Co. Down, Northern Ireland BT23 7GY*
[9]*Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA*
[10]*Berkeley Center for Cosmological Physics, Lawrence Berkeley National Laboratory & Physics Department,*
University of California, Berkeley, CA 94720, USA
[11]*Institute of Cosmology and Gravitation, University of Portsmouth, Mercantile House,*
Hampshire Terrace, Portsmouth, Hants PO1 2EG

ABSTRACT

We present morphological classifications obtained using machine learning for objects in the Sloan Digital Sky Survey DR6 that have been classified by Galaxy Zoo into three classes, namely early types, spirals and point sources/artefacts. An artificial neural network is trained on a subset of objects classified by the human eye, and we test whether the machine-learning algorithm can reproduce the human classifications for the rest of the sample. We find that the success of the neural network in matching the human classifications depends crucially on the set of input parameters chosen for the machine-learning algorithm. The colours and parameters associated with profile fitting are reasonable in separating the objects into three classes. However, these results are considerably improved when adding adaptive shape parameters as well as concentration and texture. The adaptive moments, concentration and texture parameters alone cannot distinguish between early type galaxies and the point sources/artefacts. Using a set of 12 parameters, the neural network is able to reproduce the human classifications to better than 90 per cent for all three morphological classes. We find that using a training set that is incomplete in magnitude does not degrade our results given our particular choice of the input parameters to the network. We conclude that it is promising to use machine-learning algorithms to perform morphological classification for the next generation of wide-field imaging surveys and that the Galaxy Zoo catalogue provides an invaluable training set for such purposes.

Key words: methods: data analysis – galaxies: general.

1 INTRODUCTION

Classification of galaxies has been a long-term goal in astronomy (e.g. van den Bergh 1998, and references therein). While classification by human eye is still common, there have been several attempts to use machine-learning techniques. For example, Lahav et al. (1995, 1996) showed that artificial neural networks can successfully reproduce visual classifications. In recent years, artificial

neural networks have gained prominence as a successful tool for calculating photometric redshifts (e.g. Firth et al. 2003; Collister & Lahav 2004), particularly with regard to the next generation of galaxy surveys (e.g. Abdalla et al. 2008; Banerji et al. 2008). However, artificial neural networks were first applied to astronomical data sets in order to classify stellar spectra (von Hippel et al. 1994; Bailer-Jones, Irwin & von Hippel 1998) and galaxy morphologies (Storrie-Lombardi et al. 1992; Naim et al. 1995; Folkes, Lahav & Maddox 1996). Astronomical data sets have grown considerably in size in the last decade owing largely to the advent of mosaic CCDs that can be used on large telescopes in order to image large areas of the sky down to very faint magnitudes. The Sloan Digital Sky Survey (SDSS; York et al. 2000) has led to the construction of a data set of around 230 million celestial objects. This data set has already been used for morphological classification using automated machine-learning techniques (Ball et al. 2004). Future generations of wide-field imaging surveys such as the Dark Energy Survey,[1] PanStarrs[2] and Large Synoptic Survey Telescope (LSST)[3] will reach new limits in terms of the size of astronomical data sets. Clearly, automated classification algorithms will prove invaluable for the analysis of such data sets, but these algorithms are yet to be applied on such scales.

The Galaxy Zoo project[4] launched in 2007 has led to morphological classification of nearly one million objects from the SDSS DR6 through visual inspection by more than 100 000 users (Lintott et al. 2008). This project has resulted in a remarkable data set that can be used for studies of the formation and subsequent evolution of galaxies in our Universe. In Lintott et al. (2008), the Galaxy Zoo classifications have been compared to those by professional astronomers showing that there is remarkable agreement between classifications by members of the general public and the professionals. The biases in the Galaxy Zoo classifications have been studied in detail by Bamford et al. (2009) who go on to correct for these biases and use the data set to study the relationship between galaxy morphology, colour, environment and stellar mass. This data set, however, also presents us with the unique opportunity to compare human classifications to those from automated machine-learning algorithms, on an unprecedented scale. If the neural network is shown to be as successful as humans in separating astronomical objects into different morphological classes, this could save considerable time and effort for future surveys while ensuring uniformity in the classifications.

In this paper, we explore the ability of artificial neural networks to classify astronomical objects from the SDSS into three morphological types – early types, spirals and point sources/artefacts. In Section 2, we describe the Galaxy Zoo catalogue. In Section 3, the artificial neural network method is presented. Section 4 details the different choices of input parameters that are used for classification. We present our results in Section 5 and draw some conclusions in Section 6.

2 THE GALAXY ZOO CATALOGUE

Galaxy Zoo is a web-based project that aimed to obtain morphological classifications for roughly a million objects in the SDSS by harnessing the power of the Internet and recruiting members of the public to perform these classifications by eye. The first part of this project is now complete and the morphological classifications subsequently obtained have been described in detail in Lintott et al. (2008), where these classifications have also been shown to be credible based on comparison with classifications by professional astronomers. The classifications have also been used in a number of interesting science papers, e.g. in the identification of a sample of blue early type galaxies in the nearby Universe (Schawinski et al. 2009) and to study the spin statistics of spiral galaxies (Land et al. 2008), and the power of this data set is proving enormous for studies of both galaxy formation and evolution (Bamford et al. 2009). Our goal is now to assess whether morphological classifications such as those from Galaxy Zoo can be reproduced for even larger data sets likely to become available with the next generation of galaxy surveys through the use of automated machine-learning algorithms such as artificial neural networks. Before we proceed, we caution the readers that the Galaxy Zoo catalogue is not represented by a simple selection function as is the case for both volume and flux-limited samples. It contains objects from both the Main Galaxy Sample and the Luminous Red Galaxy sample of the SDSS and as such over-represents the number of distant red galaxies in the Universe. This is not particularly important for the aims of this paper as we are simply attempting to reproduce the human classifications using machine learning. However, when using these classifications for scientific analysis, further cuts could be applied in order to remove this bias.

The Galaxy Zoo catalogue that we use in this paper is the combined weighted sample of Lintott et al. (2008). This contains morphological classifications for 893 212 objects into four morphological classes – ellipticals, spirals, mergers and point sources/artefacts. Note that the only stars in the Galaxy Zoo catalogue are bright stars with haloes or diffraction spikes which are thus interpreted as extended objects by the automatic star/galaxy separation criteria, and the only point sources included are those objects whose spectra are best fitted by galaxy templates. The classification by each user on each object is weighted such that users who tend to agree with the majority are given a higher weight than those who do not. The final morphology of the object is then a weighted mean of the classifications of all users who analysed it. Full details of the weighting scheme are provided in Lintott et al. (2008). The weighted catalogue of objects contains the SDSS object ID and three additional columns with the weighted fraction of vote of the galaxy being an elliptical, spiral or point source/artefact – i.e. classified as *don't know* by Galaxy Zoo users – between 0 and 1. Any star-forming irregular galaxies are also put into this *don't know* class by the Galaxy Zoo users. If the sum of the fractional votes in each of these three classes is less than 1, the remaining fraction of vote is assigned to the merger class. Note that this data set is affected by a luminosity, size and redshift-dependent classification bias as is the case for most morphologies derived from flux-limited data sets. Bamford et al. (2009) have derived corrections to remove this classification bias from the data. However, in this paper we work with the original catalogue of morphologies as the classification biases are not particularly important for the aims of this paper.

We match the Galaxy Zoo catalogue to the SDSS DR7 PhotoObjAll[5] catalogue in order to obtain input parameters for the neural network code. These input parameters are described in detail in Section 4. Before input into the neural network, we apply cuts to our sample and remove objects that are not detected in the g, r and i

---

[1] https://www.darkenergysurvey.org
[2] http://pan-starrs.ifa.hawaii.edu/public/
[3] http://www.lsst.org
[4] http://www.galaxyzoo.org

[5] Note that the SDSS object IDs correspond to objects in DR6 and DR7.

bands and those that have spurious values and large errors for some of the other parameters used in this study. Darg et al. (2010) have already discussed issues to do with merger classification within Galaxy Zoo and constructed a sample of ∼3000 merging pairs from the Galaxy Zoo data. In this paper, we classify objects as ellipticals, spirals or point sources/artefacts using our machine-learning code and note that the Darg et al. (2010) data set may be used in future for the classification of mergers although this has not been attempted in this paper. We therefore also remove the few well-classified mergers with a fraction of vote of being a merger greater than 0.8 from the sample as we are not attempting to classify the mergers in this work. This leads to a sample of ∼800 000 objects. Further cuts are then applied to define a *gold* sample where the fraction of vote for each object belonging to any one of three morphological classes – ellipticals, spirals and point sources/artefacts – is always greater than 0.8. This gold sample contains ∼315 000 objects and is essentially equivalent to the clean sample of Lintott et al. (2008). The neural network is run on the gold sample as well as the entire sample.

It is also the case that faint discy objects are more likely to be classified as ellipticals unless the spiral arms can be clearly seen. The elliptical sample therefore also probably contains a reasonable number of lenticular systems, and we therefore refer to this morphological class as early types throughout this paper. We therefore also consider a sample of objects with $r < 17$ that is defined as our bright sample and should suffer from a lower level of contamination in the early type class. This magnitude limit is the same as that imposed on objects that were used to determine user weights in Lintott et al. (2008). The bright sample contains ∼340 000 objects and has fewer 'well-classified' early types than the gold sample as discussed later.

3 ARTIFICIAL NEURAL NETWORKS

We use an artificial neural network code (Ripley 1981, 1988; Bishop 1995; Lahav et al. 1995; Naim et al. 1995; Collister & Lahav 2004) for classification in this paper. It has already been shown that on a de Vaucouleurs-type system, T, which spans values from −5 to 10, human expert classifiers agree to rms $\Delta_T = 1.8$ and that such agreement can be obtained by a neural network when trained on the classifications of one of the experts (Lahav et al. 1995; Naim et al. 1995). The neural network used in our study is made up of several layers, each consisting of a number of nodes. The first layer receives the input parameters described in detail in Section 4 and the last layer outputs the probabilities for the object belonging to the three morphological classes. All nodes in the hidden layers in between are interconnected and connections between nodes i and j have an associated weight, w_{ij}. A training set is used to minimize the cost function, E (equation 1), with respect to the free parameters w_{ij}:

$$E = \sum_k (T_{NN}(w_{ij}, p_k) - T_{eye,k})^2,$$ (1)

where T_{NN} is the neural network probability of the object belonging to a particular morphological type, p_k are the input parameters to the network and $T_{eye,k}$ are the fractional weighted votes in the training set in this case assigned by Galaxy Zoo users.

If the data are noisy or the network is very flexible, a validation set may be used in addition to the training set to prevent over-fitting. During the initial set-up, one has to specify the architecture of the neural network – the number of hidden layers and nodes in each hidden layer. We choose a neural network with two hidden layers with $2N$ nodes each, where N is the number of input parameters. The architecture of the network is therefore N:2N:2N:3. Note that

increasing the number of nodes further either by adding nodes to existing hidden layers or by adding more hidden layers to the network does not result in any substantial improvement to the classifications. The three nodes in the output layer give the probability of the galaxy being an early type, spiral and point source/artefact, respectively, between 0 and 1. Neural nets with this type of output are statistical Bayesian estimators and therefore the sum of all three outputs is roughly, although rarely, exactly equal to 1 [see e.g. appendix C of Lahav et al. (1996), and references therein]. Note that this differs from the Galaxy Zoo fractional votes which always add up to exactly 1 over all four morphological classes – early types, spirals, point sources/artefacts and mergers. As mentioned earlier, the mergers are not classified by the neural network in this paper.

4 INPUT PARAMETERS

When using an automated machine-learning algorithm, the choice of input parameters may be crucial in determining how well the network can perform morphological classifications. Ideally, one wishes to choose a set of parameters that show marked differences across the three morphological classes. In addition, it may be useful to define a set of parameters that is independent of the distance to the object. For example, colours may be used instead of magnitudes and ratios of radii could replace individual radius estimates. Fig. 1 illustrates the key role that the input parameters play in the morphological classification. The human eye sees an image and performs morphological classifications. The same image is also used to derive input parameters such as those that will be discussed in this section. The neural network uses these parameters as well as a training set based on the human classifications to derive its own morphological classifications, T_{NN}. These are then compared to the human classifications, T_{eye}, in order to assess the success of the machine-learning algorithm in reproducing what is seen by the human eye.

In this section, we consider two sets of input parameters based on these criteria but make no additional effort to fine-tune and optimize these parameters to perform this morphological classification. The first set of parameters listed in Table 1 has been used extensively in the literature for morphological classification in the SDSS (e.g. Ball et al. 2004). They include the $(g − r)$ and $(r − i)$ colours derived from the dereddened model magnitudes although these have not
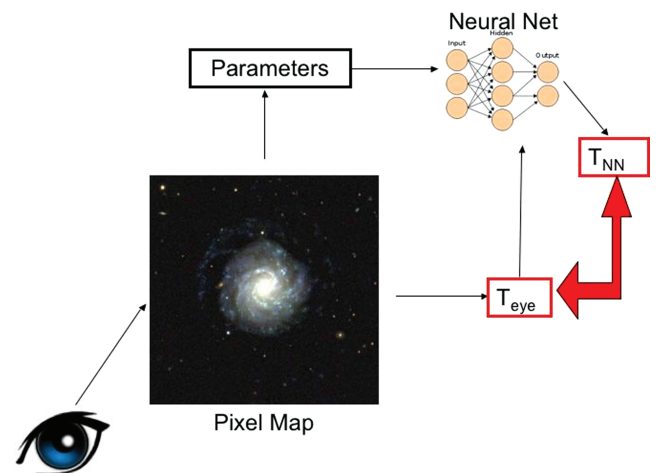
Figure 1. Schematic of how both the human eye and machine-learning algorithms such as artificial neural networks perform morphological classification and determine parameters such as those listed in Tables 1 and 2 from the galaxy images.

Table 1. First set of input parameters based on colours and profile fitting.

Name	Description
dered_g-dered_r	$(g - r)$ colour
dered_r-dered_i	$(r - i)$ colour
deVAB_i	de Vaucouleurs fit axial ratio
expAB_i	Exponential fit axial ratio
lnLexp_i	Exponential disc fit log likelihood
lnLdeV_i	de Vaucouleurs fit log likelihood
lnLstar_i	Star log likelihood

been *k*-corrected to the rest frame. We do not use the *k*-corrected colours as this would require a redshift to be measured for the object and therefore reduce the total number of objects in our sample. Furthermore, the vast majority of objects in future large-scale photometric surveys will not have secure spectroscopic redshifts and we aim to assess how effective machine learning is as a tool for morphological classification in such surveys. The other parameters considered are the axial ratios and log likelihoods associated with both a de Vaucouleurs and an exponential fit to the two-dimensional galaxy image. The de Vaucouleurs profile is commonly used to describe the variation in surface brightness of an elliptical galaxy as a function of radius whereas the exponential profile is used to describe the disc component of a spiral galaxy. In addition, the log likelihood of the object being well fitted by a point spread function (PSF), *lnLstar*, helps in distinguishing extended galaxies from more point-like sources. Note that the sample only contains those objects that are well fitted by a PSF that also have spectra that are best fitted by a galaxy template. Stars have already been removed from the sample. The $(g - r)$ and $(r - i)$ colours have been chosen as the images used in the Galaxy Zoo classifications were composites of images in these three bands. All other parameters correspond to the *i*-band images only.

The distribution of the $(g - r)$ colour, de Vaucouleurs fit axial ratio and log-likelihood fit to a de Vaucouleurs profile for the different morphological classes is illustrated in Fig. 2, where we plot the fraction of *gold* early types, spirals and point sources/artefacts as a function of these parameters. These histograms are constructed using only objects with a fraction of vote greater than 0.8 in each of the three classes from Galaxy Zoo, i.e. the *gold* sample. This threshold is arbitrary and certainly results in the loss of many true early types and spirals from our sample. However, choosing a high threshold also ensures that the samples we use to construct histograms do not suffer much from contamination. Throughout the rest of this paper, the Galaxy Zoo early types, spirals and point source/artefacts always refer to the *gold* sample with a fraction of vote greater than 0.8. Note that the fraction of vote is different from the classification probability although the two are highly correlated. Many objects with fractional votes less than 0.8 are also well-classified early types or spirals, but in this paper we only choose to consider those objects that are the most cleanly classified by members of the public for checking against the corresponding neural network classifications.

We can immediately see from Fig. 2 that different parameters allow us to distinguish between different morphological classes. As expected, early types are found to be redder than spirals whereas the point sources and artefacts have a wide range of colours. The axial ratio obtained from a de Vaucouleurs fit to the galaxy images is closer to unity for early type systems (typically ∼0.8) compared to spirals (typically ∼0.3) and has a bimodal distribution for the point sources and artefacts. The log likelihood associated with the

de Vaucouleurs fit is also larger for the early types than the spirals and largest for the point sources and artefacts.

The second set of input parameters described in Table 2 does not use the colours of galaxies or any parameters associated with profile fitting for morphological classification. Instead, a new set of shape and texture parameters are used as well as the concentration. The concentration is given by the ratios of radii containing 90 and 50 per cent of the Petrosian flux in a given band. *mRrCc* is the second moment of the object intensity in the CCD row and column directions measured using a scheme designed to have an optimal signal-to-noise ratio (Bernstein & Jarvis 2002):

$$mRrCc = \langle y^2 \rangle + \langle x^2 \rangle, \tag{2}$$

where *x* and *y* are image coordinates relative to the centre of the object in question and, for example,

$$\langle y^2 \rangle = \frac{\sum I(y, x) w(y, x) y^2}{\sum I(y, x) w(y, x)}. \tag{3}$$

Moments are measured using a radial Gaussian weight function, $w(y, x)$, iteratively adapted to the shape and size of the object. The ellipticity components, *mE1* and *mE2*, defined in equations (4) and (5) and a fourth-order moment (equation 6) are also specified:

$$mE1 = \frac{\langle y^2 \rangle - \langle x^2 \rangle}{mRrCc} \tag{4}$$

$$mE2 = 2\frac{\langle xy \rangle}{mRrCc} \tag{5}$$

$$mCr4 = \frac{\langle (y^2 + x^2)^2 \rangle}{\sigma^4}, \tag{6}$$

where σ is the size of the Gaussian weight. We use the ellipticity components to define the adaptive ellipticity, aE, for input into the neural network, where

$$aE = 1 - \sqrt{\frac{1 - \sqrt{mE1^2 + mE2^2}}{1 + \sqrt{mE1^2 + mE2^2}}}. \tag{7}$$

The final parameter in Table 2 is the texture or coarseness parameter described in Yamauchi et al. (2005). This essentially measures the ratio of the range of fluctuations in the surface brightness of the object to the full dynamic range of the surface brightness and is expected to vanish for a smooth profile but become non-zero if structures such as spiral arms appear.

The distribution of the concentration, adaptive ellipticity and texture parameters constructed using only the *gold* objects in the Galaxy Zoo catalogue with a fraction of vote greater than 0.8 are shown in Fig. 2. The concentration parameter is larger for early types compared to spirals. This is consistent with the previous studies by Shimasaku et al. (2001) and Strateva et al. (2001) who find the inverse concentration index to be larger for spirals than for ellipticals. The adaptive ellipticity is large for the spirals, small for the early types and slightly smaller still for the point sources and artefacts. The texture parameter, although roughly similar for the three morphological classes, is still slightly larger for spirals as compared to the early types and point sources suggesting that the latter have the smoothest surface brightness profiles.

We summarize the results of running the neural network with these different choices of input parameters in the next section.

5 RESULTS

The neural network code is run using three different sets of input parameters – (i) colours and profile-fitting parameters from Table 1,
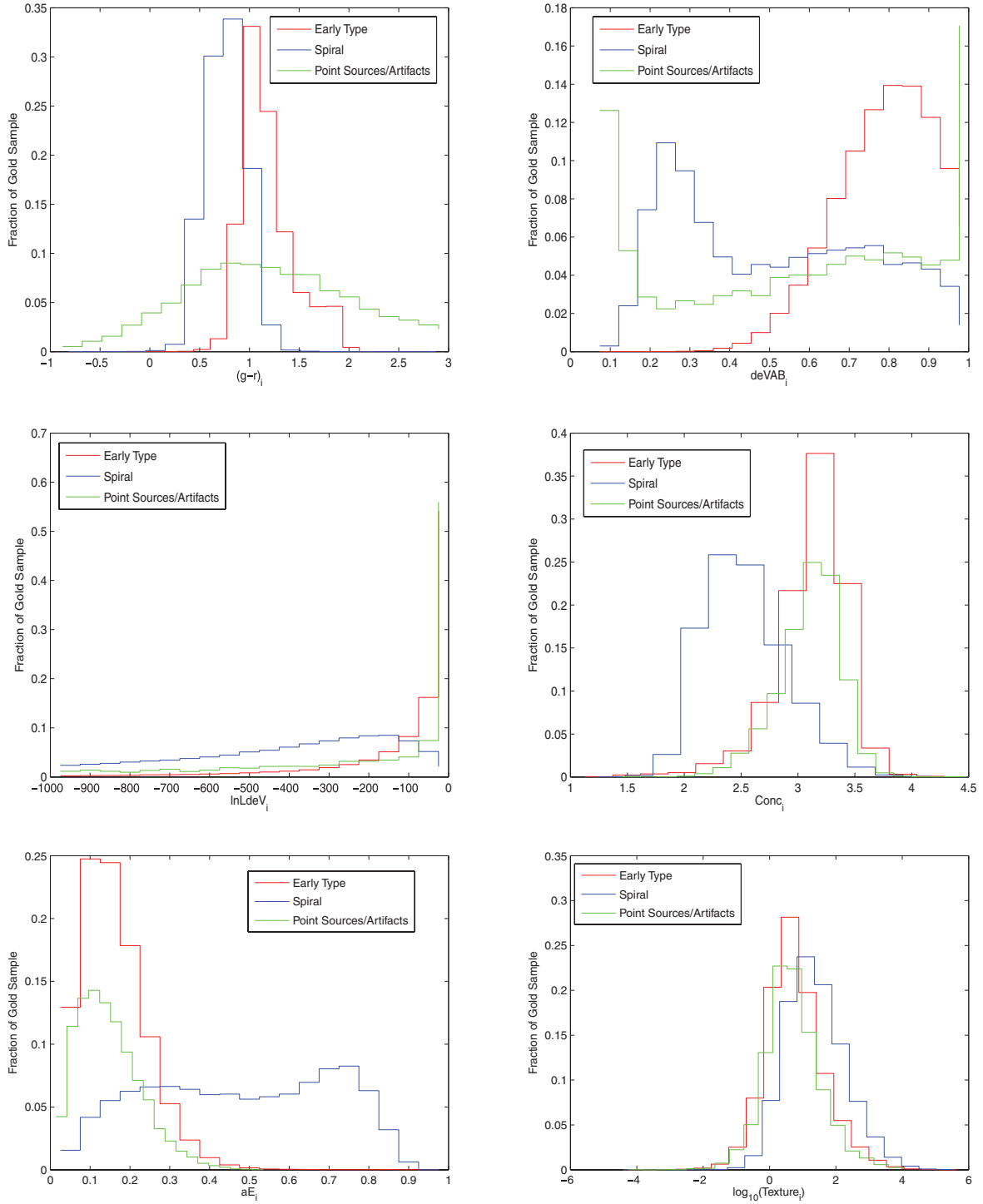
Figure 2. Histograms of some of the input parameters in Tables 1 and 2 for the gold sample of early types, spirals and point sources/artefacts.

Table 2. Second set of input parameters based on adaptive moments.

Name	Description
petroR90_i/petroR50_i	Concentration
mRrCc_i	Adaptive (+) shape measure
aE_i	Adaptive ellipticity
mCr4_i	Adaptive fourth moment
texture_i	Texture parameter

(ii) concentration, adaptive shape parameters and texture from Table 2 and (iii) a combined set of 12 parameters. We also define three samples on which the neural network is run. The first sample is the entire catalogue of ∼800 000 objects out of which 50 000 are used for training and 25 000 for validation. This is found to be a sufficiently large training set for these purposes, and no significant improvement in the classifications is seen on increasing the size of the training set further. The second sample defined as the gold sample contains only objects with a fraction of vote greater than

0.8 assigned to them in any one of the three morphological classes by Galaxy Zoo users. This is essentially the same as the clean sample in Lintott et al. (2008). This gold sample contains ∼315 000 objects, and once again 50 000 are used for training and 25 000 for validation. In this gold sample, ∼65 per cent are early types, ∼30 per cent are spirals and ∼5 per cent are point sources and artefacts. These fractions should not, however, be interpreted as the true ratio of different morphological types in the Universe as the fraction of early types is much higher in a flux-limited sample such as this. In addition, the choice of an arbitrary fractional vote threshold of 0.8 for morphological classifications in Galaxy Zoo means that many true spirals and ellipticals do not make our cut. We also do not attempt to remove any classification bias as was done in Bamford et al. (2009) as this is not particularly important for the aims of this paper. Finally, we consider a bright sample with $r < 17$ only. In this bright sample, ∼55 per cent are early types, ∼40 per cent are spirals and ∼5 per cent are point sources and artefacts and we can therefore see that the bias towards early types is somewhat reduced on removing the faint galaxies from the sample.

5.1 The entire sample

In this section, we summarize the results of running the neural network on the entire sample of objects using the three different sets of input parameters. In Figs 3–5, we plot the neural network probability of the object belonging to a morphological class versus the percentage of genuine objects in that class that are discarded on applying this probability threshold as well as percentage of contaminants that enter the sample. This allows us to determine the optimum probability threshold for the neural network that should be chosen for membership into each morphological class. This threshold is such that the number of contaminants equals the number of genuine objects in that class that are discarded. This is found to be 0.71 for early types, 0.50 for spirals and 0.24 for point sources/artefacts when using the parameters in Table 1; 0.68 for early types, 0.50 for spirals and 0.10 for point sources/artefacts when using the parameters in Table 2; and 0.73 for early types, 0.58 for spirals and 0.26 for point source/artefact when using the combined set of parameters.

For the Galaxy Zoo classifications, we consider the fractional vote threshold to be 0.8 for the galaxy belonging to a particular morphological class. In Tables 3–5, we summarize the results of running the neural net on the entire sample using the three different sets of input parameters. The tables give the percentage of early types, spirals and point sources/artefacts in Galaxy Zoo that are put into the different classes by the neural network after assuming the optimum neural network probabilities already mentioned in each of the three classes. Throughout this paper, the lowercase names – early type, spiral, point source/artefact – correspond to the Galaxy Zoo classifications and the uppercase names – EARLY TYPE, SPIRAL, POINT SOURCE/ARTEFACT – correspond to the neural network classifications. Note that as we are attempting to minimize both the number of contaminants and the number of genuine objects discarded in each of the three classes, the total number of objects is
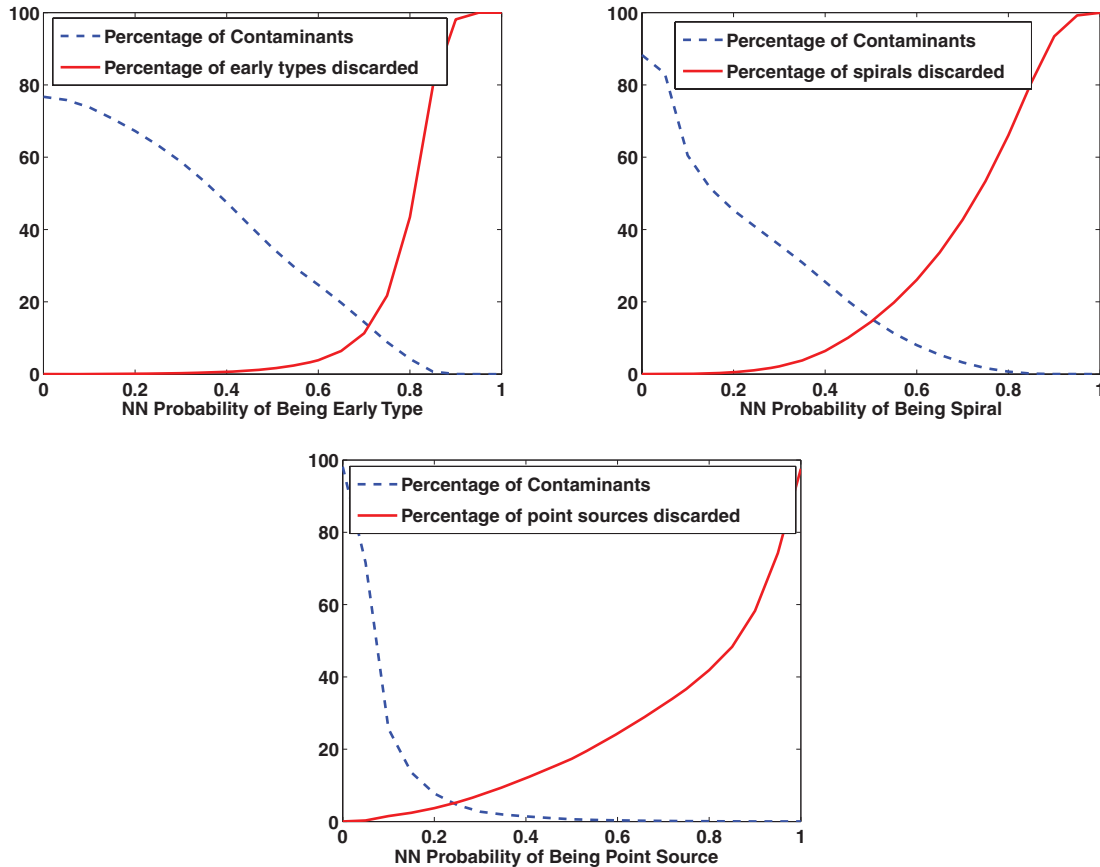
Figure 3. The neural network probability of a galaxy being an early type (top left), spiral (top right) and point source/artefact (bottom) versus the percentage of contaminants as well as the percentage of Galaxy Zoo objects in these classes that are discarded. These results are obtained using the seven input parameters in Table 1 based on colours and traditional profile fitting.
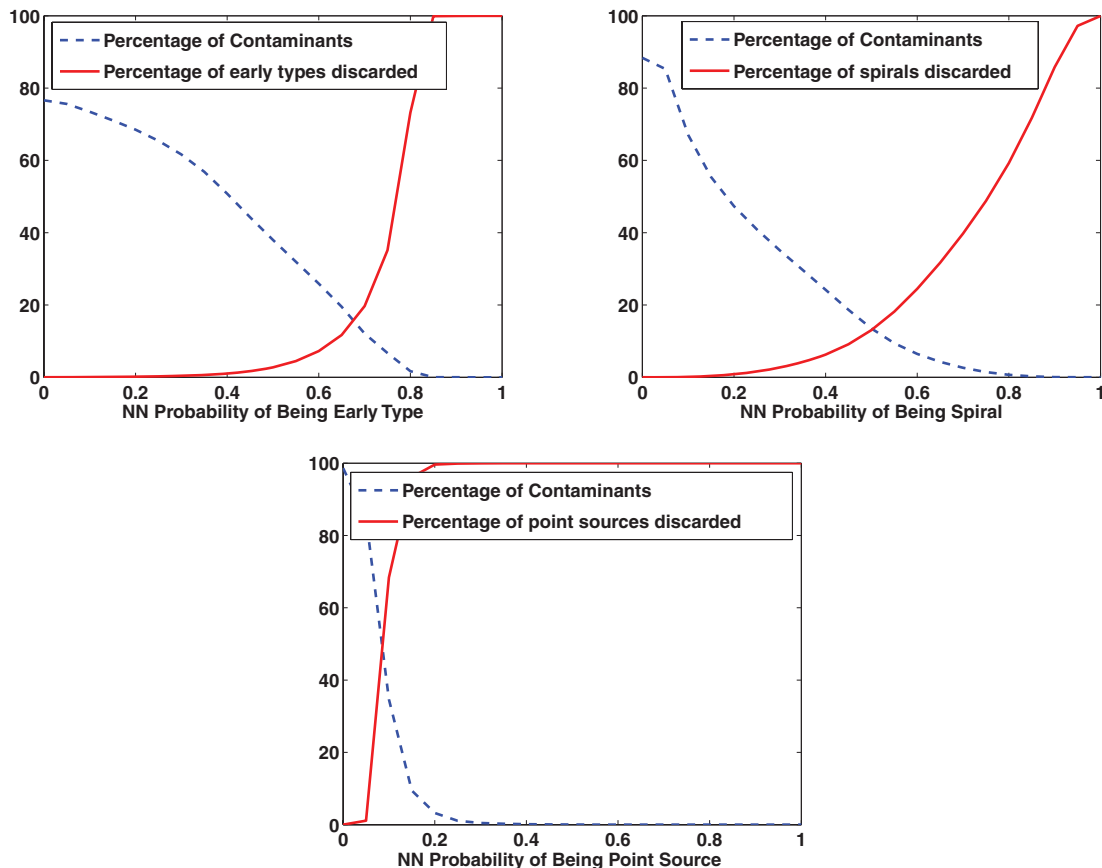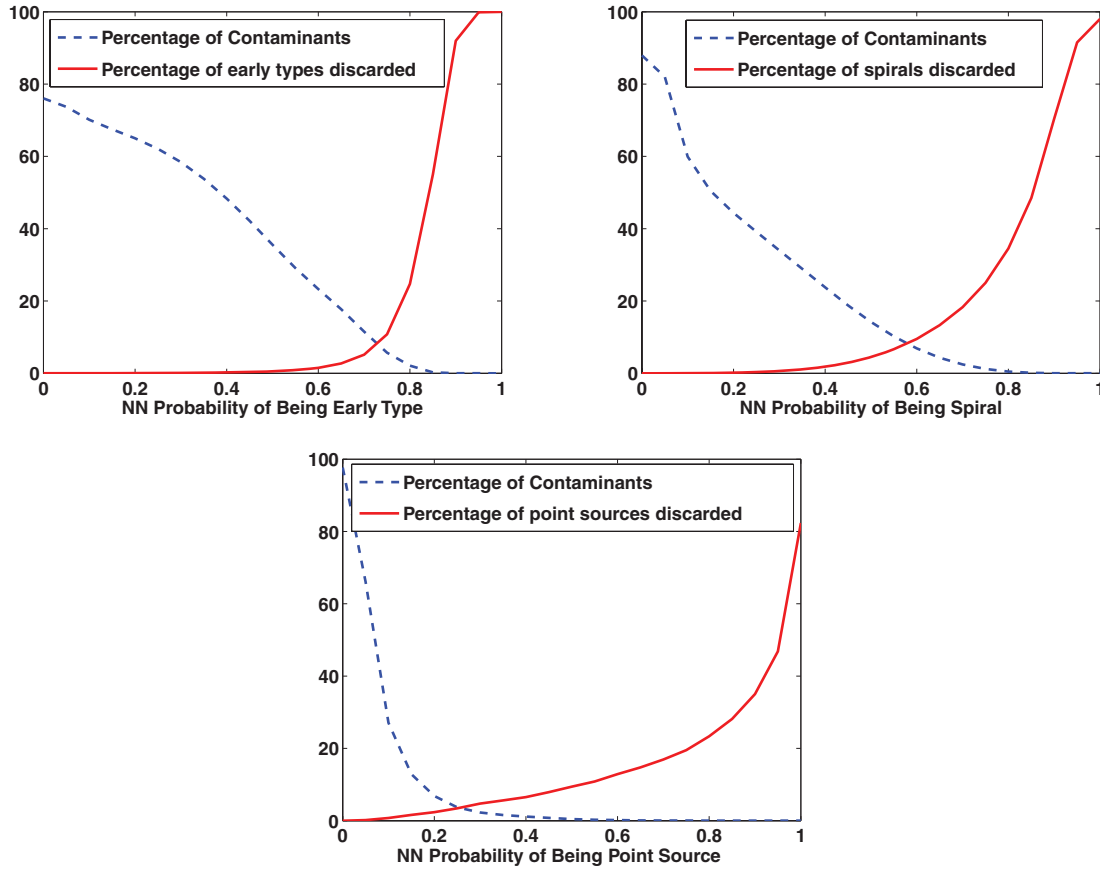
Figure 4. The neural network probability of a galaxy being an early type (top left), spiral (top right) and point source/artefact (bottom) versus the percentage of contaminants as well as the percentage of Galaxy Zoo objects in these classes that are discarded. These results are obtained using the five input parameters in Table 2 that use adaptive moments, concentration and texture.

not necessarily conserved. There may also be a significant number of poorly classified objects with a spread of likelihoods over two or more morphological classes. These poorly classified objects will have a probability of less than the chosen optimum threshold probability in all three morphological classes. Some of these objects that are poorly classified by the neural network may have been well classified by Galaxy Zoo which is why the sum of the columns in Table 3 is typically less than 100 per cent. Some objects may also be put into more than one class due to the assumed neural network threshold probabilities. For example, an object with a neural net probability of greater than 0.1 in the POINT SOURCE/ARTEFACT class could potentially also have a neural net probability of greater than 0.5 in the SPIRAL class and therefore be classified as both. This is why the sum of the columns in Table 4 is typically greater than 100 per cent. In the tables presented throughout this analysis, we are not summarizing the results of unique classifications for every single object but rather the best results that could be obtained for each subsample assuming neural network threshold probabilities that result in the same number of contaminants into the sample as genuine objects that are discarded. If we were to consider the best results for the entire sample as a whole, the chosen probability thresholds would probably be different.

Using the traditional colour and profile-fitting parameters, 87 per cent of early types, 86 per cent of spirals and 95 per cent of point sources/artefacts are correctly classified by the neural network. We note, however, that while colours are sensitive to the star formation history of a galaxy, the morphology essentially measures

the dynamic history. Although the two are correlated, they are not necessarily the same and therefore it is important to use colours in conjunction with other parameters when performing morphological classifications. Looking more closely at the Galaxy Zoo early types that were misclassified by the neural net as spirals as well as the spirals that are misclassified by the neural net as early types, there is some evidence that these galaxies may be red spirals or blue ellipticals. 45 Galaxy Zoo early types are classified as spirals with a probability of greater than 0.8 by the neural net. Similarly, 40 Galaxy Zoo spirals are classified as early types with a probability greater than 0.8. Out of these, 21 early types and nine spirals have SDSS spectra available. Using the criterion in Baldry et al. (2004) to isolate red and blue galaxies, we find that six out of the nine spirals are red and 10 out of the 21 early types are blue. In other words, the colour information used as an input to the neural network may be biasing the morphological classification. However, due to the small numbers of misclassified galaxies with SDSS spectra available, a definite statement cannot be made and we leave this to future work.

The adaptive shape parameters are very good for distinguishing between spirals and early types, and these parameters result in accurate classifications for 84 per cent of early types and 87 per cent of spirals. However, this set of parameters gives very poor results for point source/artefact and only 28 per cent are correctly classified by the neural network. This is because these parameters are very similar for point sources/artefacts and early types as can be seen in some of the histograms in Fig. 2. The spirals, on the other hand, have

Figure 5. The neural network probability of a galaxy being an early type (top left), spiral (top right) and point source/artefact (bottom) versus the percentage of contaminants as well as the percentage of Galaxy Zoo objects in these classes that are discarded. These results are obtained using the combined set of 12 input parameters from Tables 1 and 2.

Table 3. Summary of results for the entire sample when using input parameters specified in Table 1 – colours and traditional profile fitting.

			Galaxy Zoo	
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	87	0.3	0.3
N	Spiral	0.6	86	2.2
N	Point source/artefact	0.7	0.5	95

Table 4. Summary of results for the entire sample when using input parameters specified in Table 2 – adaptive moments.

			Galaxy Zoo	
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	84	0.5	85
N	Spiral	1	87	0.8
N	Point source/artefact	32	6.5	32

very different adaptive shape parameters. As there are many more early types in the training set compared to point sources/artefacts, the neural network cannot differentiate between the two and assigns most of the point sources/artefacts to be early types. Also, Fig. 4 shows that the optimum neural network probability for point source/artefact classification using this set of input parameters is as low as 0.1. This means that many objects will be classified both as a point source and a galaxy assuming this probability and for this reason, the sums of the columns in Table 4 are always greater than 100 per cent as there are objects in common between the classes. We also investigate whether the evidence for a bias due to colour in the morphological classifications is removed once the colours

Table 5. Summary of results for the entire sample when using input parameters specified in Tables 1 and 2.

| | | Galaxy Zoo | | |
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	92	0.07	0.6
N	Spiral	0.1	92	0.08
N	Point source/artefact	0.2	0.2	96

are removed as input parameters to the network. With the adaptive shape parameters as inputs, we find 208 Galaxy Zoo early types to be misclassified as spirals with a probability greater than 0.8 by the neural network and similarly 26 spirals are misclassified as early types with the same probability. Out of these, 116 early types and only eight spirals have SDSS spectra. Two out of the eight spirals are red but 46 of the 116 early types are blue, once again suggesting that there may still remain a bias, especially against blue ellipticals even when the colour parameters are removed as inputs to the neural net. This may suggest that at least some blue ellipticals have more structure than their red counterparts. However, we emphasize that the sample sizes here are too small to make any quantitative statement about the extent of this bias and we defer this to future work.

On adding the profile fitting and colour parameters to the adaptive shape parameters, the results are now considerably improved for all three classes. 92 per cent of early types, 92 per cent of spirals and 96 per cent of point sources/artefacts are correctly classified by the neural network. Lintott et al. (2008) have compared the Galaxy Zoo classifications to those by professional astronomers and found an agreement of better than 97 per cent between these samples. However, the samples used in Lintott et al. (2008) are the Morphologically Selected Ellipticals in SDSS (MOSES) sample of Schawinski et al. (2007) whose objective was to generate a very clean set of elliptical galaxies, and the detailed classifications of Fukugita et al. (2007) which are very sensitive to the E/Sa boundary. If the professional astronomers were set the same task as the Galaxy Zoo users – i.e. a clean division of spirals and ellipticals in the SDSS – the scatter between them and the Galaxy Zoo users may well be worse than this. We have shown that with a set of

12 intelligently chosen parameters, which are easily available but still by no means optimal for performing morphological classification, the neural network results agree to better than 90 per cent with those from Galaxy Zoo. It can therefore certainly be expected that by using a set of better tuned input parameters to the network, the machine-learning algorithm will be able to classify galaxies with comparable or less scatter than that produced by human classifications.

5.2 The gold sample

We now describe the results of running the neural network on our gold sample using the three different sets of input parameters. These results are summarized in Tables 6–8 where we consider the percentage of Galaxy Zoo early types, spirals and point sources/artefacts that have also been put into these classes by the neural network assuming a probability of greater than 0.8 now for the neural network classification. Once again it can be seen that the second set of input parameters involving just the concentration, texture and the adaptive shape parameters performs very poorly in classifying the point sources and artefacts. However, the combined set of input parameters leads to correct neural network classifications for 97 per cent of early types and 97 per cent of spirals, on par with the agreement between Galaxy Zoo and professional astronomers. The success for point sources/artefacts is somewhat lowered compared to using the entire sample. This is because when considering the entire sample we assume that all objects with a neural network point source/artefact probability of greater than ∼0.2 do indeed belong to the point sources/artefacts class, whereas in this section we require the probability to be greater than 0.8.

Table 6. Summary of results for the gold sample when using input parameters specified in Table 1 – colours and traditional profile-fitting.

| | | Galaxy Zoo | | |
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	93	0.6	0.9
N	Spiral	0.7	90	2.3
N	Point source/artefact	0.04	0.07	82

Table 7. Summary of results for the gold sample when using input parameters specified in Table 2 – adaptive moments.

| | | Galaxy Zoo | | |
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	92	0.8	92
N	Spiral	0.6	89	0.5
N	Point source/artefact	0	0	0

Table 8. Summary of results for the gold sample when using input parameters specified in Tables 1 and 2

| | | Galaxy Zoo | | |
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	97	0.2	1.3
N	Spiral	0.1	97	0.2
N	Point source/artefact	0.05	0.02	86

Table 9. Summary of results for the bright sample when using input parameters specified in Tables 1 and 2.

| | | Galaxy Zoo | | |
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	94	0.1	0.3
N	Spiral	0.1	92	0.1
N	Point source/artefact	0.2	0.2	98

Table 10. Summary of results for the entire sample when using input parameters specified in Tables 1 and 2 and only bright galaxies with $r < 17$ to train the network.

| | | Galaxy Zoo | | |
		Early type (per cent)	Spiral (per cent)	Point source/artefact (per cent)
A	Early type	91	0.05	0.8
N	Spiral	0.08	92	0.1
N	Point source/artefact	2	0.1	95

5.3 The bright sample

In this section, we run the neural network code using the combined set of 12 input parameters on our bright sample with $r < 17$. This allows us to perform two tests. First, we look at whether the bright galaxies in general have better classifications compared to the entire sample. This is done by comparing Tables 5 and 9. Secondly, we train our neural network on the bright sample and use this trained network to perform morphological classifications for the entire sample. This allows us to quantify the effects of magnitude incompleteness in the training set on the morphological classifications. The results are summarized in Tables 9 and 10. Once again the probability threshold required in the neural net output for classification is determined by requiring the percentage of contaminants to be equal to the percentage of genuine objects in that class that are discarded on applying the threshold. This optimum probability threshold is very similar for the bright sample to those shown in Fig. 5 but slightly higher in all three classes when the bright galaxies are used for training and used to classify all galaxies.

Comparing Tables 5 and 9, we can see that the results are slightly better when performing morphological classifications on the bright sample with $r < 17$ compared to the entire sample with $r < 17.77$, in both cases using a complete training set. This is to be expected as it is easier to distinguish between early types and spirals in a brighter sample. When the training is performed using an incomplete training set with $r < 17$ and all objects with $r < 17.77$ classified, the neural network still manages to perform these classifications with more than 90 per cent agreement with Galaxy Zoo users. The magnitude incompleteness in the training set does not seem to affect the classifications as most of the input parameters to the neural network considered in this study have been chosen to be distance independent and so their distribution does not really change from a shallow to a deeper sample. This is promising for using automated machine-learning algorithms to perform morphological classification for future deep surveys using the Galaxy Zoo classifications on the shallower SDSS as a training set.

6 CONCLUSIONS

In this study, we have used a machine-learning algorithm based on artificial neural networks to perform morphological classifications for almost one million objects from SDSS that were classified by eye as part of the Galaxy Zoo project. The neural network is trained on 75 000 objects and using a well-defined set of input parameters that are also distance independent, we are able to reproduce the human classifications for the rest of the objects to better than 90 per cent in three morphological classes – early types, spirals and point sources/artefacts. The Galaxy Zoo catalogue provides us with a training set of unprecedented size for automated morphological classifications via machine learning. Specifically, we draw the following conclusions.

(i) Using colours and profile-fitting parameters as inputs to the neural network, 87 per cent of early type classifications, 86 per cent of spiral classifications and 95 per cent of point source/artefact classifications agree with those obtained by the human eye. However, there is some evidence to suggest that a non-negligible fraction of red spirals and blue ellipticals are misclassified by the network when using this set of parameters.

(ii) When parameters that rely on an adaptive weighted scheme for fitting the galaxy images are used, the neural network is unable to distinguish between early types and point source/artefact as these parameters are very similar for the two classes.

(iii) A combination of the profile fitting and adaptive weighted fitting parameters results in better than 90 per cent agreement between classifications by humans and those by the neural network for all three morphological classes. This is approaching the success of Galaxy Zoo users in reproducing the classifications by professional astronomers and will certainly surpass this with a more finely tuned set of input parameters than we have considered in this paper.

(iv) The optimum neural network probability for a galaxy belonging to a particular morphological class is such that the percentage of contaminants is equal to the percentage of genuine objects in that class that are discarded on cutting the sample using this threshold. This optimum probability depends on both the input parameters and the morphological class of the object. We find that early types generally have a high optimum probability ($\sim$0.7) whereas the point sources and artefacts have a very low optimum probability ($\sim$0.2). Therefore, the same object could be put into more than one class by the neural network if the classifications were performed using the optimum threshold probabilities.

(v) For our gold sample, the early type and spiral classifications by the neural network match those by the human eye to better than 95 per cent.

(vi) Using a bright sample to train the neural network and performing morphological classifications for a deeper and fainter sample still results in better than 90 per cent agreement between the neural network and human classifications in all three morphological classes. This is because we have deliberately chosen our input parameters to be distance independent.

(vii) However, other sources of incompleteness in the training sets also need to be examined before the role of the Galaxy Zoo data in training morphological classifiers for future surveys can be fully understood.

The penultimate point in particular illustrates the power of the machine-learning algorithm in fully exploiting data from future wide-field imaging surveys such as the Dark Energy Survey, Pan-STARRS, HyperSuprime-Cam, LSST, Euclid, etc. Such surveys will obtain images for hundreds of millions of objects down to very deep magnitude limits. We have shown that by using the wealth of information made available through the Galaxy Zoo project as a training set, the machine-learning algorithm can quickly and accurately classify the vast numbers of objects that will make up future data sets into early types, spirals and point sources/artefacts. However, if the Galaxy Zoo catalogue is to be used as a training set for automated machine-learning classifications of mergers with the next generation of galaxy surveys, a more robust catalogue of visually classified mergers such as that of Darg et al. (2010) needs to be used. Also, it is worth emphasizing that the images obtained from the next generation of wide-field surveys will need to have the necessary pixel size and resolution required to derive photometric parameters such as those used as inputs to the neural network in this paper.

This paper has also examined the effect of magnitude incompleteness in the training set on automated morphological classifications. In the future, more work needs to be done on investigating other sources of incompleteness in the training set before these data can be used effectively to train machine-learning morphological classifiers for future surveys. For example, in this paper we have found some evidence that a non-negligible proportion of galaxies that are misclassified by the neural network are either red spirals or blue ellipticals. This misclassification is almost certainly due to the sparsity of such objects in the training set and needs to be addressed in future machine-learning papers that use these data. We have also only used objects with a fraction of vote greater than 0.8 from Galaxy Zoo to compare to our neural network classifications. In future work, we hope to address how excluding 'intermediate' objects from our comparison is likely to bias our results. Nevertheless, we conclude that morphological classification via machine learning looks promising in allowing for many more detailed studies of the processes involved in galaxy formation and evolution with the next generation of galaxy surveys.

REFERENCES

Abdalla F. B., Amara A., Capak P., Cypriano E. S., Lahav O., Rhodes J., 2008, MNRAS, 387, 969
Bailer-Jones C. A. L., Irwin M., von Hippel T., 1998, MNRAS, 298, 361
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Ball N. M., Loveday J., Fukugita M., Nakamura O., Okamura S., Brinkmann J., Brunner R. J., 2004, MNRAS, 348, 1038
Bamford S. P. et al. 2009, MNRAS, 393, 1324
Banerji M., Abdalla F. B., Lahav O., Lin H., 2008, MNRAS, 386, 1219
Bernstein G. M., Jarvis M., 2002, AJ, 123, 583
Bishop C. M., 1995, Neural Networks for Pattern Recognition. Oxford Univ. Press, New York
Collister A. A., Lahav O., 2004, PASP, 116, 345
Darg D. W. et al., 2010, MNRAS, 401, 1043
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195
Folkes S. R., Lahav O., Maddox S. J., 1996, MNRAS, 283, 651
Fukugita M. et al., 2007, AJ, 134, 579
Lahav O. et al., 1995, Sci, 267, 859
Lahav O., Naim A., Sodré L. Jr, Storrie-Lombardi M. C., 1996, MNRAS, 283, 207
Land K. et al., 2008, MNRAS, 388, 1686
Lintott C. J. et al., 2008, MNRAS, 389, 1179
Naim A., Lahav O., Sodre L. Jr, Storrie-Lombardi M. C., 1995, MNRAS, 275, 567
Ripley B. D., 1981, Spatial Statistics. Wiley, New York
Ripley B. D., 1988, Statistical Inference for Spatial Processes. Cambridge Univ. Press, Cambridge
Schawinski K., Thomas D., Sarzi M., Maraston C., Kaviraj S., Joo S.-J., Yi S. K., Silk J., 2007, MNRAS, 382, 1415
Schawinski K. et al., 2009, MNRAS, 396, 818

Shimasaku K. et al., 2001, AJ, 122, 1238

Storrie-Lombardi M. C., Lahav O., Sodre L. Jr, Storrie-Lombardi L. J., 1992, MNRAS, 259, 8P

Strateva I. et al., 2001, AJ, 122, 1861

van den Bergh S., 1998, Galaxy Morphology and Classification. Cambridge Univ. Press, Cambridge

von Hippel T., Storrie-Lombardi L. J., Storrie-Lombardi M. C., Irwin M. J., 1994, MNRAS, 269, 97

Yamauchi C. et al., 2005, AJ, 130, 1545

York D. G. et al., 2000, AJ, 120, 1579

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