J. Astrophys. Astr. (2022) 43:76
https://doi.org/10.1007/s12036-022-09871-2

BINARIES

Machine learning in astronomy

AJIT KEMBHAVI[1,*] and ROHAN PATTNAIK[2]

[1] Inter-University Centre for Astronomy and Astrophysics (IUCAA), Pune 411007, India.
[2]School of Physics and Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA.
*Corresponding author. E-mail: akk@iucaa.in

MS received 9 December 2021; accepted 3 June 2022

Abstract. Artificial intelligence techniques like machine learning and deep learning are being increasingly used in astronomy to address the vast quantities of data, which are now widely available. We briefly introduce some of these techniques and then describe their use through the examples of star-galaxy classification and the classification of low-mass X-ray binaries into binaries, which host a neutron star and those which host a black hole. This paper is based on a talk given by one of the authors and reviews previously published work and some new results.

Keywords. Low-mass X-ray binaries—star-galaxy classification—machine learning—classification.

1. Introduction

Abundant astronomical data is now freely available because of surveys like the the Sloan digital sky survey (SDSS, see Abdurro'uf *et al.* (2022) for the latest data release and references) and the more recent Subaru hyper suprime-cam survey (see Aihara *et al.* 2022 for the latest data release and references). Conventional data analysis techniques will seriously constrain the scientific projects which can be undertaken with such databases due to the sheer volume of the data. To tap the full potential of the data, it is necessary to use artificial intelligence techniques like machine learning (ML) and deep learning (DL), which have evolved rapidly over the past few decades (e.g., Baron 2019), making them very useful for a variety of applications. These developments and the availability of software platforms like TensorFlow 2 and Keras (TensorFlow 2 is a free and open-source software library for ML, DL, etc., developed by Google researchers, see https://www.tensorflow.org/guide/effective_tf2. Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow 2, developed by Chollet, François *et al.* see, https://keras.io) have enabled astronomers to use ML and DL for addressing the very large volumes of imaging, spectral and catalogue data that are now easily accessible to them.

Some examples of application of ML and DL to astronomy include photometric redshift estimation (D'Isanto & Polsterer 2018; Pasquet *et al.* 2019), gravitational lensing identification (Cheng *et al.* 2020), light curve classification (Lochner *et al.* 2016; Mahabal *et al.* 2019; Möller & de Boissière 2020), stellar spectrum classification and interpolation (Kuntzer *et al.* 2016; Sharma *et al.* 2020a, b), galaxy morphology classification (Dieleman *et al.* 2015; Abraham *et al.* 2018; Domínguez Sánchez *et al.* 2018; Barchi *et al.* 2020; Walmsley *et al.* 2020), and star-galaxy classification (Philip *et al.* 2002; Ball *et al.* 2006; Vasconcellos *et al.* 2010; Abraham *et al.* 2012; Soumagnac *et al.* 2015; Kim & Brunner 2017; Clarke *et al.* 2020).

In the following, we will briefly describe a few important ML and DL techniques and two illustrative applications to astronomy: star-galaxy separation and establishing the identity of the compact object in low-mass X-ray binaries on the basis of their X-ray energy spectra.

This article is part of the Special Issue on "Astrophysical Jets and Observational Facilities: A National Perspective".

Published online: 16 October 2022

2. Machine learning and deep learning algorithms

A number of algorithms are conventionally grouped together under ML. These are used to build computer programs which automatically improve with experience. While the algorithms can be used for classification and regression, we will be considering only the former in our examples. For a classification problem, the input to the program consists of a large number of training examples with each example having a number of measurable attributes, and belonging to one of several defined classes. The program learns from the training set to distinguish between the classes based on the attributes. After the learning is complete, the program is able to predict the class of previously unclassified objects on the basis of their attributes, i.e., the program is able to generalize the classification, learned using a finite training sample and to the examples beyond the training set. This process is known as supervised learning, because the input training set includes the known class of the objects in the sample. In unsupervised learning, the program can be asked to classify a training sample into a given number of classes on the basis of attributes with no prior classification being specified. Unsupervised learning is useful when possible novel classification schemes are to be investigated.

Machine learning includes algorithms like random forest (RF) and artificial neural networks (ANN) (see e.g., Mitchell 1997), while DL, which is also a part of ML, but is generally mentioned separately includes algorithms like convolutional neural networks (CNN), recurrent neural networks (RNN), generative adversarial networks (GNN), etc., (see e.g., Lecun *et al.* 2015; Goodfellow *et al.* 2016; Guo *et al.* 2016). We will briefly describe below about ANN, random forest and CNN, and two applications which will provide some insight into how the algorithms work.

2.1 *Artificial neural networks*

The basic unit of an ANN is a artificial neuron, which is historically loosely based on a biological neuron. An ANN with an input layer, a hidden layer and an output layer is shown in Figure 1. A neuron in the hidden layer receives inputs from all neurons in the preceding input layer and contributes to every output neuron. The output of the jth neuron is given by

$$o_j = \sigma(y_j), \quad y_j = \sum_{i=0}^{4} w_{ji} x_i,$$

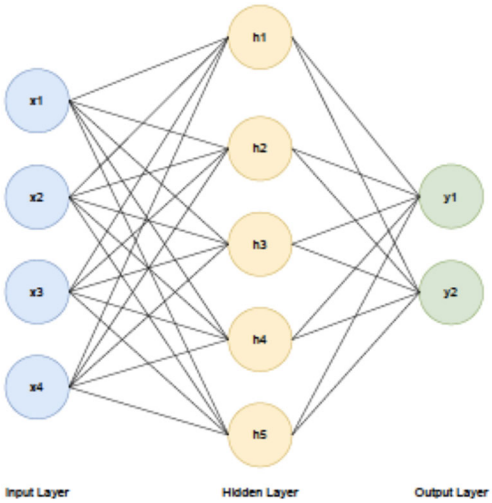

Figure 1. An ANN with four inputs, one hidden layer and two output classes.

where the w_{ji} is known as the weights and σ is a non-linear activation function. The usual forms of σ used are the sigmoid function

$$\sigma(y) = \frac{1}{1 + \exp(-y)}$$

or the rectified linear unit ReLU function $\sigma(y) = \max(0, y)$, which is preferred because it leads to faster training for many layered ANN. To avoid dead neurons, which have output zero and which sometimes arise when ReLU is used and a leaky ReLU is used, where $\sigma(y) = y$ for $y \geq 0$ and $\sigma(y) = 0.01y$ for $y < 0$. In general, there can be any number of input nodes, and depending on the complexity of the classification boundaries in the multi-dimensional space of the inputs, there can be several hidden layers. Each neuron in the ANN is connected to all the neurons in the preceding and succeeding layers, so the ANN is said to be fully connected.

For supervised learning, training the network consists of suitably adjusting the weights so that the desired known output is obtained for a given input. In the star-galaxy problem described below, a certain number of parameters are measured for objects which are known to be stars or galaxies. The parameters for a large number of such known objects are in turn input to a network, with desired output -1 for a star and 1 for a galaxy (say). A loss function is now defined which compares the output for a starting random choice of weights with the desired inputs for the training set. The loss function is then minimized over the multiple passes of the training data, using a technique known as back propagation (Mitchell 1997). Suitable definitions are used to

measure accuracy (or precision), which is the ratio of true positives to the sum of true and false positives over all classes, and the completeness (or recall), which is the ratio of true positives to the sum of true positives and false negatives over all classes. There are many variations possible on the basic structure of an ANN described here, and there are details of many matters and technicalities which we have not mentioned.

2.2 *Random forest*

Random forest (RF) is an ensemble technique which is used to boost the prediction made by an individual decision tree (Breiman 2001). A decision tree is one of the most intuitive, yet powerful ML algorithms (Breiman *et al.* 1984). A decision tree is made up of branches of nodes, where sets of if-this-then-that rules are applied to the features of the input data, and based on the result, leads down to one of the branches of the tree. The final layer of nodes, also known as leaf nodes, contain a predicted class label which is compared to the expected class for a particular input vector. Although the decision tree algorithm has proven to be very efficient (see e.g., Vasconcellos *et al.* 2011), a decision tree, if improperly trained, can at times over-fit the data (Mitchell 1997, Chapter 3). The idea behind RF is to combine the decisions of several such trees to improve upon the decision of a single over-trained tree. Taking a majority-vote over the decision of all the trees, helps in reducing the variance of the predictions (Breiman 2001). The probability of a source belonging to one class or the other is also calculated in a similar way, i.e., by dividing the number of trees that predicted the same class by the total number of trees. We illustrate the decision making process of a RF algorithm in Figure 2.

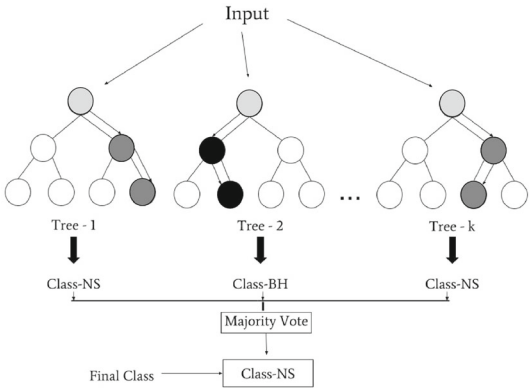

Figure 2. Illustration of the decision making procedure in a random forest algorithm.

2.3 *Deep learning: convolutional neural networks*

Conventional ML uses as inputs features extracted from the raw data, like images or spectra. As the data get complex, for example, when images of galaxies are to be classified, extracting a manageable number of features can be very difficult or even impossible. If all the pixels of an image were to be used directly as input to an ANN, then for CCD images with millions of pixels, the number of input nodes, hidden nodes and weights for all the connections would become so large as to make the network unmanageable. DL techniques are designed to get over these difficulties by using raw data to avoid the extraction of features to be used as inputs to the network. In fact, the network itself extracts features from the raw data by convolving it with a set of filters to provide a representation of the image at a more abstract level for the classification (Lecun *et al.* 2015). The classifier in the network uses downsized extracted features, so that number of weights remain manageable.

In convolutional neural networks (CNN), is the only DL technique that we will consider in this paper. A CNN has three types of layers, convolutional layers, pooling layers and fully connected layers. In a network, the convolutional layers alternate with the pooling layers and following a number of such pairs, there are the fully connected layers, which lead to the final classified output. In a convolutional layer, the input image is convolved with a set of kernels to generate feature maps. The kernels for extracting specific features like edges at various orientations and other motifs are not provided by the user, they are learned by the network for the input data being used. A convolutional layer is followed by a pooling layer, in which the dimensionality of a feature map is reduced by taking the maximum (or average) pixels over a 2×2 array, say, which strides over a feature map. After the final pooling layer, a set of fully connected layers is used to produce the output. These perform like a conventional ANN and the output provides the categories for classification with a probability associated with each class. The weights for all the layers are trained using back propagation, as in the case of an ANN. The need for a CNN and its operation can be better understood through the problem of star-galaxy classification discussed in Section 3.

3. Star-galaxy classification

Very compact galaxies or large galaxies at great distances, can resemble stars in their appearance in

astronomical images at optical or near-infrared wavelengths. So separating such galaxies from stars for galaxy surveys can be difficult on the basis of just their images. Stars have the appearance of the point spread function (PSF) for the image, which is determined by the earth's atmosphere to a large extent with contributions made by the telescope optics and structure, etc. In the ideal case, the PSF is a 2-dimensional circular Gaussian with a full-width at half-maximum (FWHM) of ∼1 arc-sec in good-seeing conditions, but is somewhat distorted in practice. Compact or distant galaxies which are significantly larger than the PSF can be easily distinguished as such. But the images of galaxies approaching the PSF in size are more difficult to distinguish: even though their shape is different in detail from the shape of the PSF, the differences are small and hard to discern. An expert astronomer would be needed to separate such galaxies from stars, but the task would take simply too long for large data sets, and therefore there is a very good case for the use of ML.

3.1 *Star-galaxy classification with ANN*

We will discuss here one of the early works, which uses ML for star-galaxy classification (Philip *et al.* 2002; a few other early results are cited in the reference). Philip *et al.* (2002) used R band images from the publicly available NOAO deep wide field survey (Jannuzi *et al.* 2000). The training set was constructed from a sub-image of the R band image NDWFSJ1426p3456, which had the best-seeing conditions for the data released in 2001 A total of 402 objects in the training set, in the magnitude range of 20–26, were visually classified as stars or galaxies independently by two of the authors, and the ∼2% cases where the classification turned out to be different, were resolved by a joint inspection. The final training set had 83 stars and 319 galaxies. The number of stars was smaller than the number of galaxies because of the high galactic latitude of the field and the faintness of the objects.

For each object, three parameters were measured: (1) an elongation measure, which is the ratio of the second order moments along the major- and minor-axes of the faintest isophote, (2) a standardized FWHM measure, which is the logarithm of the ratio of the FWHM of the object to the FWHM of the PSF for the image and (3) a gradient parameter which is the logarithm of the ratio of central peak count to the FWHM of the object, normalized to the standardized FWHM measure. A difference boosting neural network (DBNN, Philip *et al.* 2002) was used in the training with the three parameters as the inputs. The trained network was used to classify a test set consisting of a total of 154 stars and 558 galaxies from the two sub-images of the field, which had been previously visually classified as stars and galaxies as in the case of the training set. An overall accuracy of 98.1% classification was obtained, which was better than the 96.1% accuracy obtained for the same test set using SExtractor (Bertin & Arnouts 1996).

In the above project, the size of the training and test set was small because of the difficulty in visually identifying compact galaxies. The small sample can result in the overfitting of the network, which while providing good accuracy in the training, does not perform well in generalizing to a variety of images outside the training set. Moreover, it would be difficult to visually identify large samples of galaxies and to measure parameters for the training of galaxies which are faint and/or irregular, like the galaxies shown in Figure 3. It is therefore, necessary to use a technique based on DL, so that raw images of stars and galaxies can be used in the training. It is also necessary to have large samples of known galaxies and stars for the training of a DL based network.

3.2 *Star-galaxy classification with CNN*

We will now consider work by Chaini *et al.* (2022) on star-galaxy classification which satisfies the above requirements. The authors consider a more general star-quasar-galaxy separation with the aim being to identify these objects on the basis of their photometric properties alone, but we will only consider the star-galaxy sector. The training sample consists of spectroscopically identified stars and galaxies from SDSS data release 16 (DR16, Ahumada *et al.* 2020). The spectroscopic identification is secure and no visual identification is necessary. In the training only photometric data of the five SDSS passbands u, g, r, i, z are used. The sample is limited to compact objects, which are defined as those which have the ratio

$$c = \frac{\text{half light radius}}{\text{FWHM}} < 0.5.$$

Here, half light radius, or de Vaucouleurs radius, is the radius containing half of the total light of the object, and the average of the ratio in the five passbands is used. When only a faint subset is to be considered, the criterion for faintness is that the average of the five band magnitudes $\langle m \rangle > 20$. The training dataset consists of 80,000 objects each for the two classes, chosen randomly from larger sets, which satisfy the criteria.

The CNN used in this case is based on the inception network (Szegedy *et al.* 2014) and has five dense layers of size 1024, 512, 256, 128 and 64, respectively (a dense

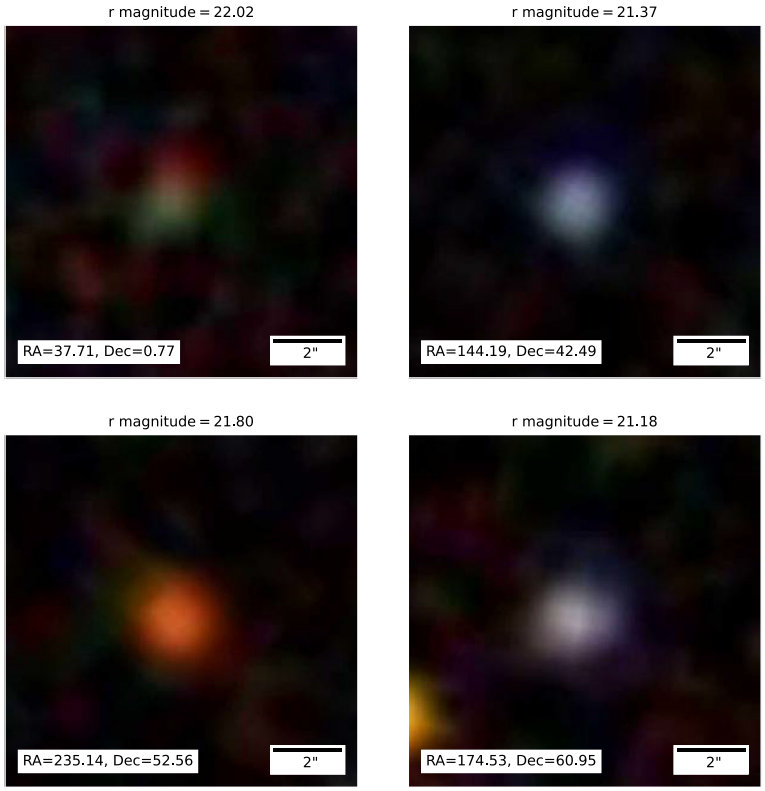

Figure 3. Four faint compact galaxy r, g, b band composite images from the SDDS.

layer has its neurons connected to every neuron in the preceding layer) with varying kernel sizes. Each layer is itself made up of four parallel convolutional layers, each activated by a leaky ReLU and averaging pooling layers are used. The loss function used is known as categorical-cross entropy and a total of 25,544,807 neurons are trained. The final layer contains a softmax function, which provides the output as probabilities for the two classes, star and galaxy.

When the training, validation and test samples are all drawn from the dataset described above, and compact objects are considered, the accuracy for star-galaxy separation reached with the CNN is 97.4%. If compact and faint objects with average of the five band magnitudes $\langle m \rangle > 20$ are considered, the accuracy drops marginally near to 95.2%. The reason for the drop of course is that the fainter objects have poorer signal-to-noise ratio, so the separation into stars and galaxies is more difficult.

Chaini *et al.* (2022) have also used photometric parameters for each object provided by the SDSS data processing pipeline to carry out the separation using an ANN. The parameters are the magnitudes in the five bands corrected for extinction, the half light radius, FWHM of the PSF and the extinction in each of the five bands, and the colours u–g, g–r, r–i and i–z, making a total of 24 parameters. The accuracy reached for the separation with the ANN is 97.9% and 96.0%, respectively, for compact objects, and compact and faint objects. Chaini *et al.* (2022) further consider an ensemble of the CNN and ANN that they call MargNet, which has a combined accuracy of 98.1% and 96.9% for the two cases. Using the ensemble of CNN and ANN is therefore the best option. The important point to note here is that the CNN works directly on star and galaxy images, and does not need any measured parameters and provides high accuracy. It is therefore useful even when a dependable pipeline for measuring photometric parameters is not available. Other examples of star-galaxy-quasar classification using CNN include Kim & Brunner (2017) and Clarke *et al.* (2020).

4. Low-mass X-ray binaries

Low-mass X-ray binaries (LMXBs) are binary systems where one of the components is a black hole (BH) or a neutron star (NS); the other component is a less massive star, usually on main sequence or an evolved star of mass $M < 1\ M_{\odot}$. Some LMXBs show long quiescent periods, which can last from a few months to decades, when the source is very faint. There are also short periods, lasting from days to months, when the source is in

outburst, with the flux increasing by several orders of magnitude (see e.g., McClintock & Remillard 2006).

The energy spectra of LMXB systems are described by two main components: (1) a thermal component which is usually described by a multi-color disc blackbody, thought to be produced by an accretion disc and (2) a hard component thought to be produced by a corona, which is a region of hot plasma around the compact object. This component is usually described by a thermal Comptonization model. The contribution of these components to the X-ray emission varies during an outburst, leading to modification of its spectral and timing properties. References to the details about LMXB energy spectra are provided in Pattnaik *et al.* (2021).

One of the important questions about LMXBs is whether the compact object in the binary is a NS or a BH. The nature of the compact object has a significant impact on the physical interpretation of the observations. With the large scale sky surveys and transient search programs, e.g., INTEGRAL/JEM-X (Lund *et al.* 2003), Swift/BAT transient monitor (Krimm *et al.* 2013), MAXI (Matsuoka *et al.* 2009), eROSITA (Merloni *et al.* 2012), the sample of LMXBs keeps increasing. Such newly detected transient sources are usually characterized by their fast variation (days) of luminosity by orders of magnitude. Early identification of the nature of the compact object is very important for the community to be able to plan observing campaigns (Middleton *et al.* 2017).

There are only a few methods for identifying the nature of the compact object. For example, coherent pulsations and the presence of thermonuclear bursts (for reviews see, Lewin *et al.* 1993; Cumming 2004; Galloway *et al.* 2008; Strohmayer *et al.* 2018), indicate that the compact object is a NS. Based on the mass function of the X-ray binary system, if the mass of the compact object is estimated to be greater than about 3 $M_{\odot}$, then the compact object can be taken to be a black hole. Apart from that, one can surmise the nature of the compact object by comparing its X-ray timing and spectral properties and X-ray-radio correlation with those of sources where the nature of the compact object is known.

One technique that is yet to be fully explored to classify LMXBs is the use of ML algorithms. ML has been used by Huppenkothen *et al.* (2017) to classify light curves of the unusual BH X-ray binary GRS 1915+105. It has also been used by Gopalan *et al.* (2015) to distinguish between different types of X-ray binaries. We describe below how ML can be applied to the X-ray energy spectra of LMXB to identify the nature of the compact object.

4.1 *Data*

We used the Rossi X-ray timing explorer (RXTE) mission (Bradt *et al.* 1993) data archive,[1] which provides more than 8500 observations of 33 NS systems and more than 6000 observations of 28 BH systems. We used data from the proportional counter array (PCA, Glasser *et al.* 1994) instrument aboard RXTE, which has an energy range of 2–60 keV to create the energy spectra. We selected a total of 61 sources, which are classified as BH or NS binaries, with classification well established (see e.g., Corral-Santana *et al.* 2016; Tetarenko *et al.* 2016, for BH). In the dataset, we have a fairly balanced representation of the two classes, with 8669 observations from 33 sources identified as neutron-star LMXBs (58%) and 6216 observations from 28 sources identified as black-hole LMXBs (42%). The number of observations per source varies greatly from source to source. A few sources have >1000 observations while some have <20 observations. Some details of the procedure followed to obtain the energy spectra for our analysis are described in Pattnaik *et al.* (2021).

For each observation, we used 43 channels in the energy range of 5–25 keV. The number of channels is kept fixed at 43 for all the observations since ML algorithms require each observation used in the training and testing to have the same size. The 43 count rate values are used directly as an input vector for the algorithm.

4.2 *Random forest for classifying LMXB*

We wish to determine the nature of the compact object on the basis of the energy spectrum. The object can be any one of two types, a black hole or a neutron star. The training set consists of 14,885 spectra of 61 X-ray binaries for which we know which of the two kinds the compact object is. This is a supervised binary classification problem of classifying an X-ray spectrum into two labeled classes. There are several ML algorithms that can be used for handling this type of binary classification problem. From those, we have to choose the one which provides the best accuracy, i.e., the highest percentage of correct classifications. We experimented with a number of algorithms including classification and regression trees (CART), more commonly known as decision trees (Breiman *et al.* 1984), random forest (RF) (Breiman 2001) which we briefly described in Section 2.2, XGBoost (XGB) (Chen & Guestrin 2016), logistic regression (LR) (Cox 1958), K-nearest neighbors (KNN) (Cover & Hart 2006) and support vector

[1] https://heasarc.gsfc.nasa.gov/cgi-bin/W3Browse/w3browse.pl.

machines (SVM) (Cortes & Vapnik 1995). These are all traditional ML algorithms that are usually known to show satisfactory performance even with a limited amount of data. They also have significantly lower execution times compared to DL methods (see e.g., Kotsiantis *et al.* 2007).

To establish the best algorithm, we compared their performance using accuracy as a metric. Here accuracy is defined as the ratio of the number of observations correctly classified to their class (neutron star or black hole), to the total number of observations. Using a k-fold cross-validation technique (Burman 1989), in which the set of 14,885 observations is split into training and test sets in many different ways, we find that RF provides the best accuracy of 91 ± 2% and use it in the subsequent analysis. We implement the RF algorithm using the `scikit-learn`[2] (Pedregosa *et al.* 2011) library of python.

4.3 *Methods and results*

We apply the RF algorithm with the best combination of hyper-parameters to the dataset described above. Hyper-parameters are a set of parameters defined prior to the training process that are used to tune the performance of the ML algorithm. Since the dataset contains 14,885 observations for 61 individual X-ray sources, each source has multiple observations taken at different times. The LMXB are variable in nature, so observations for the same source taken at different times typically sample a different physical spectral state, which correspond to different geometrical configurations in the source. We can therefore, assume that each observation for a given source is considered independent of the other observations. Traditionally, the data set of 14,885 spectra would be randomly split into a training set containing 90% (say) of the sources with the other 10% (say) forming the test set. However, this can lead to biases because in this method, some spectra of a given source can belong to the training set, while other spectra of the same source could belong to the test set. This can lead to overestimation of the accuracy reached and affect the predictive power of the nature of the compact source for the spectra of new LMXB. We therefore, chose not to use this method.

We find that optimal use of the data is made when we keep all observations from one source as the test data, while using all the remaining sources for training, and this experiment is repeated for each source. This provides the results for all the observations from each of the 61 sources. The size of the training and test sets vary in each run and each model uses one source less than the total number in the data. The final model is trained on the entire dataset. We show in Figure 4, the accuracy obtained for each source using this method. There are four sources that lie below the 50% average accuracy mark. The sigma-clipped average accuracy is 87 ± 13%, which gives a lower bound proxy on the performance of our final model.

While the RF algorithm provides good overall classification of low-mass X-ray binary sources into BH and NS types, there are a few sources for which the accuracy is low and most of the observations of those sources are misclassified. Four sources, XTE J1118+480 (BH), XTE J1748−288 (BH), IGR J00291+5934 (NS) and 1A 1246−588 (NS), have <50% accuracy, out of which the observations of XTE J1118+480 and XTE J1748−288 are consistently misclassified with overall accuracy percentage of ∼30%. Two factors that can influence the energy spectra are the signal-to-noise ratio (SNR) and the physical state of LMXB systems at the time of observation.

SNR is defined as the net count rate divided by the error in the net count rate for each spectrum. The SNR over the sample ranges from <4 to >5800. We find that for observations with SNR <100, the distribution of predicted probabilities peaks at 0.58. For SNR in the range of 100–1000 and for SNR > 1000, the distribution peaks are at 0.87 and 0.91, respectively. The performance of the classification model therefore, improves with the increase in SNR. Among the sources which were misclassified, only 1A1246−588 had an average SNR < 100. Therefore, there are reasons other than low SNR for the poor classification of sources. We find that the algorithm performs better for soft-state observations as compared to hard-state observations for individual sources (see Figures 8 and 9 of Pattnaik *et al.* 2021).

4.4 *Prediction for a sample of sources*

We have used the RF model trained on all 61 sources to predict the classification of 13 systems which have a total of 766 spectra, but where the nature of the compact object is still not established (Table 1). If >50% of the spectral observations of a source were predicted to belong to a particular class, then that class was assigned to the source. It is seen from the table that five sources have very few observations (<10) that meet our criteria for good data and it is difficult to make any comments on the predicted classes for these sources. The remaining eight sources all have >30 observations each and six of these sources are classified as BH LMXBs, while

[2] https://scikit-learn.org/stable/.

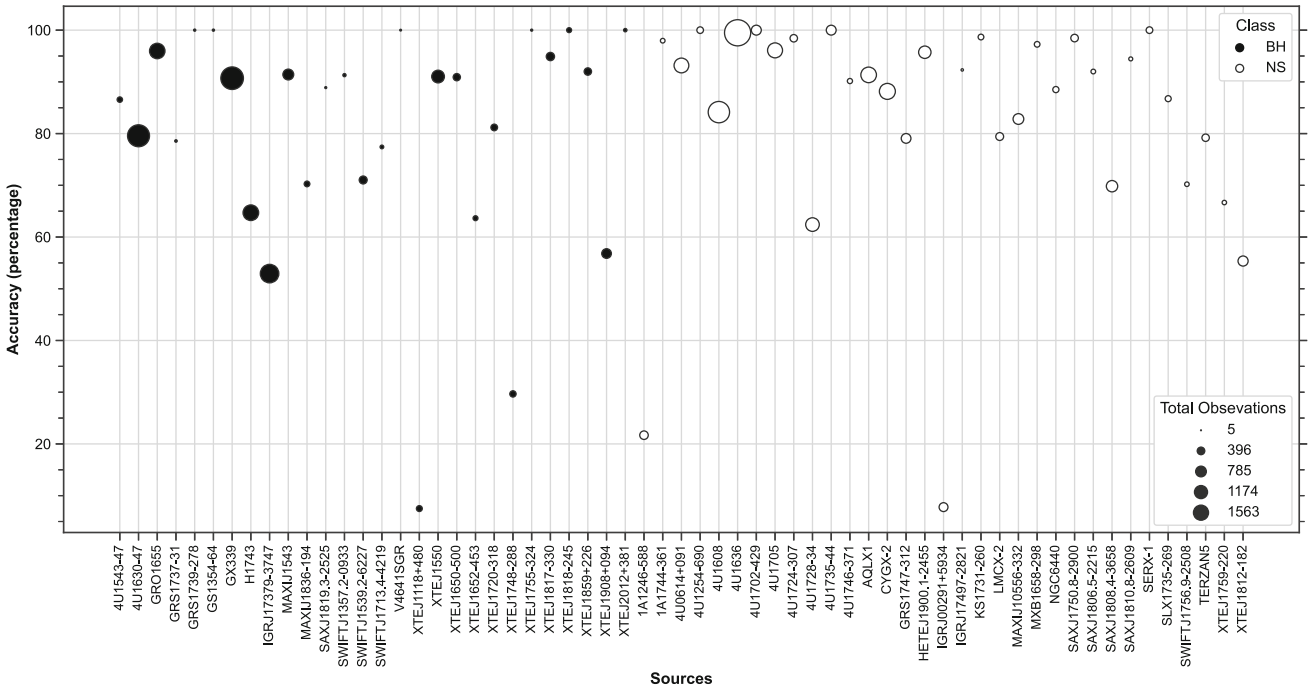

Figure 4. Plot showing individual source wise accuracy using the leave-one source out method of cross-validation. The filled circles are black hole binaries while open circles are neutron star binaries. The area of the points corresponds to the number of observations in each source. The figure is based on (Pattnaik *et al.* 2021).

Table 1. Classification results for sources in the prediction set. A class was assigned to a source if the majority of its observations were predicted to belong to that class. In cases where the ratio was 50-50 (XTE J1719-291) it is indicated that the source can belong to either class. Table taken from (Pattnaik *et al.* 2021).

Source name	Total obs.	Class (predicted)	Prediction (%)	Avg. SNR
4U1822−371	97	BH	55.67	67
4U1957+11	121	BH	72.73	22.38
IGRJ17285−2922	5	BH	60	10.03
IGRJ17494−3030	97	NS	54.64	25.84
SAXJ1711.6−3808	34	NS	94.12	34.35
SLX1746−331	65	BH	87.69	26.82
SWIFTJ1842.5−1124	49	BH	51.02	25.71
XTEJ1637−498	76	BH	65.79	8.41
XTEJ1719−291	2	NS/BH	50	2.82
XTEJ1727−476	4	BH	100	6.3
XTEJ1752−223	210	BH	67.14	56
XTEJ1856+053	5	BH	100	10.75
XTEJ1901+014	1	BH	100	1.1

two sources are classified as NS LMXBs. Six of these eight sources have prediction percentage >60%, while the remaining two sources have prediction in the range of 50–60% . All the 13 sources have an average SNR <100, which is the region where the algorithm has the worst performance.

4.5 *Discussion*

Our classification model is trained specifically on RXTE data and cannot be used directly to classify the energy spectra from other X-ray missions. It is in principle possible to train a classification model for different missions using data from the missions, but there may not be enough data in every case to train a ML algorithm. However, the concept of transfer learning could be employed to train an algorithm for another instrument with limited data using our pre-trained classification model for RXTE data.

Adding more information as input to the algorithm can also be explored as a means of improving the current level of accuracy reached for all the sources in our dataset. One way of doing that would be to combine the energy spectra with the power spectra of all the observations for each source.

There is now a considerable amount of data obtained with LAXPC detector on AstroSat (Yadav *et al.* 2021). It should be possible to apply ML and DL techniques to the data to have further useful information from it than has been done so far using conventional methods.

Acknowledgements

This paper is based on a talk given by one of the authors, Ajit Kembhavi at the 'Astrophysical jets and observational facilities: National perspective' meeting at ARIES, Nainital in April 2021, which described ML and DL techniques as well as work on star-galaxy classification by Chaini *et al.* (2022) and on the classification of LMXB by Pattnaik *et al.* (2021). The authors wish to thank an anonymous referee for suggestions which helped to significantly improve the manuscript.
The data underlying this paper are publicly available in the High Energy Astrophysics Science Archive Research Center (HEASARC) at https://heasarc.gsfc.nasa.gov/db-perl/W3Browse/w3browse.pl and the SDSS archives.

References

Abdurro'uf, Accetta K., Aerts C., *et al.* 2022, The Astrophysical Journal Supplement, 259, 35. https://doi.org/10.3847/1538-4365/ac4414
Abraham S., Philip N. S., Kembhavi A., Wadadekar Y. G., Sinha R. 2012, Monthly Notices of the Royal Astronomical Society, 419, 80
Abraham S., Aniyan A. K., Kembhavi A. K., Philip N. S., Vaghmare K. 2018, Monthly Notices of the Royal Astronomical Society, 477, 894
Ahumada R., Prieto C. A., Almeida A., *et al.* 2020, The Astrophysical Journal Supplement, 249, 3. https://doi.org/10.3847/1538-4365/ab929e
Aihara H., AlSayyad Y., Ando M., *et al.* 2022, Publications of the Astronomical Society of Japan, 74, 247. https://doi.org/10.1093/pasj/psab122
Ball N. M., Brunner R. J., Myers A. D., Tcheng D. 2006, The Astrophysical Journal, 650, 497
Barchi P. H., *et al.* 2020, Astronomy and Computing, 30, 100334
Baron D. 2019, Machine learning in astronomy: a practical overvie, 1904.07248
Bertin E., Arnouts S. 1996, Astronomy & Astrophysics, 117, 393. https://doi.org/10.1051/aas:1996164
Bradt H., Rothschild R., Swank J. 1993
Breiman L. 2001 Machine Learning, 45, 5
Breiman L., Friedman J., Olshen R., Stone C. 1984, Group, 37, 237
Burman P. 1989, Biometrika, 76, 503
Chaini S., Bagul A., Gondkar A., Sharma K., Vivek M., Kembhavi A. 2022, Photometrical identification of compact galaxies, stars and quasars using multiple neural networks, in preparation
Chen T., Guestrin C. 2016, in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, p. 785
Cheng T.-Y., Li N., Conselice C. J., Aragón-Salamanca A., Dye S., Metcalf R. B. 2020, Monthly Notices of the Royal Astronomical Society, 494, 3750
Clarke A. O., Scaife A. M. M., Greenhalgh R., Griguta V. 2020, Astronomy & Astrophysics, 639, A84
Corral-Santana J. M., Casares J., Muñoz-Darias T., *et al.* 2016, The Astrophysical Journal, 587, A61
Cortes C., Vapnik V. 1995, Machine learning, 20, 273
Cover T., Hart P. 2006, IEEE Trans. Inf. Theor., 13, 21
Cox D. R. 1958, Journal of the Royal Statistical Society: Series B (Methodological), 20, 215
Cumming A. 2004, Nuclear Physics B Proceedings Supplements, 132, 435
Dieleman S., Willett K. W., Dambre J. 2015, Monthly Notices of the Royal Astronomical Society, 450, 1441
D'Isanto A., Polsterer K. L., 2018, Astronomy & Astrophysics, 609, A111
Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L. 2018, Monthly Notices of the Royal Astronomical Society, 476, 3661
Galloway D. K., Muno M. P., Hartman J. M., Psaltis D., Chakrabarty D., 2008, The Astrophysical Journal Supplement Series, 179, 360
Glasser C. A., Odell C. E., Seufert S. E. 1994, IEEE Transactions on Nuclear Science, 41, 1343
Goodfellow I., Bengio Y., Courville A. 2016, Deep Learning, The MIT Press
Gopalan G., Vrtilek S. D., Bornn L. 2015, The Astrophysical Journal, 809, 40
Guo, Y., Liu, Y., Oerlemans, A., *et al.* 2016, Neurocomputing, 187, 27
Huppenkothen D., Heil L. M., Hogg D. W., Mueller A. 2017, Monthly Notices of the Royal Astronomical Society, 466, 2364
Jannuzi B. T., Dey A., Tiede G. P., Brown M. J. I., NDWFS Team 2000, AAS
Kim E. J., Brunner R. J., 2017, Monthly Notices of the Royal Astronomical Society, 464, 4463
Kotsiantis S. B., Zaharakis I., Pintelas P. 2007, Emerging artificial intelligence applications in computer engineering, 160, 3
Krimm H. A., *et al.* 2013, The Astrophysical Journal Supplement Series, 209, 14
Kuntzer T., Tewes M., Courbin F. 2016, Astronomy & Astrophysics, 591, A54
Lecun Y., Bengio Y., Hinton G. 2015, Nature, 521, 436. https://doi.org/10.1038/nature14539
Lewin W. H. G., van Paradijs J., Taam R. E. 1993, Space Science Reviews, 62, 223
Lochner M., McEwen J. D., Peiris H. V., Lahav O., Winter M. K. 2016, The Astrophysical Journal https://doi.org/10.3847/0067-0049/225/2/31
Lund N., *et al.* 2003, The Astrophysical Journal, 411, L231
Mahabal A., *et al.* 2019, Publications of the Astronomical Society of the Pacific, 131, 038002

Matsuoka M., *et al.* 2009, Publications of the Astronomical Society of Japan, 61, 999
McClintock J. E., Remillard R. A., 2006, Black hole binaries, 157
Merloni A., *et al.* 2012, 1209.3114
Middleton M. J., *et al.* 2017, New Astronomy, 79, 26
Mitchell T. 1997b, Machine Learning (New York: McGraw-Hill)
Möller A., de Boissière T. 2020, Monthly Notices of the Royal Astronomical Society, 491, 4277
Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D. 2019, Astronomy & Astrophysics, 621, A26
Pattnaik R., Sharma K., Alabarta K., *et al.* 2021, Monthly Notices of the Royal Astronomical Society, 501, 3457
Pedregosa F., *et al.* 2011, Journal of Machine Learning Research, 12, 2825
Philip N. S., Wadadekar Y., Kembhavi A., Joseph K. B., 2002, Astronomy & Astrophysics, 385, 1119
Sharma K., Kembhavi A., Sivarani T., Abraham S., Vaghmare K., 2020a, Monthly Notices of the Royal Astronomical Society, 491, 2280
Sharma K., *et al.* 2020b, Monthly Notices of the Royal Astronomical Society, 496, 5002
Soumagnac M. T., *et al.* 2015, Monthly Notices of the Royal Astronomical Society, 450, 666
Strohmayer T. E., *et al.* 2018, The Astronomer's Telegram, 11507
Szegedy C., Liu W., Jia Y., *et al.* 2014, 1409.4842
Tetarenko B., Sivakoff G., Heinke C., Gladstone J. 2016 The Astrophysical Journal Supplement Series, 222, 15
Vasconcellos E. C., de Carvalho R. R., Gal R. R., *et al.* 2010, Astro-Phys. https://doi.org/10.1088/0004-6256/141/6/189
Vasconcellos E., De Carvalho R., Gal R. 2011, The Astronomical Journal, 141, 189
Walmsley M., *et al.* 2020, Monthly Notices of the Royal Astronomical Society, https://doi.org/10.1093/mnras/stz2816,491, 1554
Yadav J. S., Agrawal P. S., Misra R., Roy J., Pahari M. R. K. 2021, Journal of Astrophysics and Astronomy, 496, 5002