# A Classifier to Detect Elusive Astronomical Objects through Photometry

D. Bhavana<sup>1</sup>, S. Vig<sup>1\*</sup>, S.K. Ghosh<sup>2</sup>, and Rama Krishna Sai S. Gorthi<sup>3</sup><sup>†</sup>

Indian Institute of Space science and Technology, Thiruvananthapuram 695547, India

- <sup>2</sup> Tata Institute of Fundamental Research, Mumbai 400 005, India

Accepted XXX. Received YYY; in original form ZZZ

The application of machine learning principles in the photometric search of elusive astronomical objects has been a less-explored frontier of research. Here we have used three methods: the Neural Network and two variants of k-Nearest Neighbour, to identify brown dwarf candidates using the photometric colours of known brown dwarfs. We initially check the efficiencies of these three classification techniques, both individually and collectively, on known objects. This is followed by their application to three regions in the sky, namely Hercules  $(2^{\circ} \times 2^{\circ})$ , Serpens  $(9^{\circ} \times 4^{\circ})$  and Lyra  $(2^{\circ} \times 2^{\circ})$ . Testing these algorithms on sets of objects that include known brown dwarfs show a high level of completeness. This includes the Hercules and Serpens regions where brown dwarfs have been detected. We use these methods to search and identify brown dwarf candidates towards the Lyra region. We infer that the collective method of classification, also known as ensemble classifier, is highly efficient in the identification of brown dwarf candidates.

**Key words:** methods: statistical – (stars): brown dwarfs – infrared: stars – techniques: miscellaneous – techniques: photometry

# INTRODUCTION

Classification of astronomical objects has always posed a problem to researchers. Classification usually depends on characteristic spectral features of a set of objects that can be observed through photometry at certain wavelengths in the absence of spectroscopic data. Such photometrybased classification schemes have been traditionally implemented by applying colour and magnitude cuts to the data (van der Veen & Habing 1988; Allen et al. 2004). The advent of large all-sky photometric surveys, each identifying millions of new objects, necessitates automated techniques for classification. It is clear that increasing the number of features used for classification improves accuracy. The need for handling multi-dimensional feature spaces (wherein different classes can be clearly distinguished) has led to the introduction of machine learning algorithms for this purpose (Ho & Agrawala 1968).

Among the earliest machine-learning methods used for classification in astronomy are the k-Nearest Neighbour (k-NN) method and the Neural Network (NeuN) algorithm. These computational techniques for recognising pat-

\* Email: sarita@iist.ac.in

† E-mail: rkg@iittp.ac.in

terns have been used for astronomical classification since 1970s. A statistical nearest-neighbour test was employed by Bogart & Wagoner (1973) to study clustering of galaxies and QSOs. Heydon-Dumbleton et al. (1989) used an automatic classification procedure for star-galaxy classification using NeuN. Thereafter, Odewahn et al. (1992) used the perceptron and backpropagation neural network algorithms to create accurate classifiers for separating star and galaxy images. Similar networks were also used for morphological classification of galaxies by Storrie-Lombardi et al. (1992). Nowadays, more sophisticated techniques are used for classification, such as the Support Vector Machines (SVMs) (Krakowski et al. 2016; Kurcz et al. 2016), Random Forest algorithm (Nakoneczny et al. 2018) and deep learning (González et al. 2018). In the current work, we attempt to create an ensemble classifier by applying simple machine learning techniques like k-NN and NeuN for identification of objects like brown dwarfs, which are rarely observed despite being theoretically predicted to be in abundance (Mužić et al. 2017).

Brown dwarfs are objects with mass so low that they cannot sustain hydrogen fusion in their cores. They are capable of fusing deuterium, and the minimum mass for deuterium-burning is defined as the lower mass limit for a brown dwarf (adopted by the International Astronomical

<sup>&</sup>lt;sup>3</sup>Indian Institute of Technology, Tirupati 517506, India

Union in 2002, Spiegel et al. 2011). Brown dwarfs fall under 3 spectral types, L, T and Y. Some of the hottest brown dwarfs discovered have been found to be of late M-type as well, but since the photometric properties of M-type brown dwarfs are very similar to the M-type main-sequence stars, we have omitted them from the brown dwarf category in this study. Brown dwarfs are extremely faint and cool; T and Y dwarf temperatures can range from 300 K to 1300 K, and visible magnitudes of even the hotter L-dwarfs fall above 20 mag (Costa et al. 2005). Hence, they are very difficult to detect. Their emission peaks in the infrared with a distinctive spectral energy distribution arising from strong molecular absorption features (Warren et al. 2007; Stephens et al. 2009). Foraging the immediate solar neighbourhood for such cold objects is one of the goals of the all-sky mission performed by the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010). WISE is an Earth-orbiting NASA mission that surveyed the entire sky simultaneously at wavelengths of 3.4, 4.6, 12, and 22  $\mu$ m, hereafter referred to as bands W1, W2, W3, and W4, respectively. These bands have been selected so as to uniquely identify brown dwarfs based on their spectral features (Kirkpatrick et al. 2011).

Most of the brown dwarf searches till date have used generic colour cuts to identify candidates, which are then confirmed or rejected on the basis of additional spectroscopy and proper motion studies (Cushing et al. 2011; Kirkpatrick et al. 2011; Tinney et al. 2012; Thompson et al. 2013). However, Luhman (2012) cautions against the use of theoretical magnitudes and colours for this purpose. Instead, he recommends the use of photometry of known low-mass objects, to guide the identification of candidates. But this has been attempted by very few as of now. For example, Marengo & Sanchez (2009) present a statistical method for the photometric search of rare astronomical sources using the k-NN method. Here k-NN acts as a non-parametric classifier, by deciding the class of a new set of data based on its distance from a class of brown dwarf templates, where the distance is defined in multi-dimensional colour and magnitude space. We have included the above technique, hereby termed as k-NN threshold distance, in our analysis. In addition, we have applied a modified version of k-NN distance method, namely k-NN classification, whereby additional classes (i.e. different sets of background objects which are not brown dwarfs) are considered for improvement in the analysis. The third method used in the present work is NeuN, a parametric classifier. Although this method has been used for astronomical classification problems as mentioned earlier, it has not been applied to the specific problem of brown dwarf classification problem till now.

In this work, we specifically aim to (i) pose the identification of brown dwarfs as a classification problem rather than a (colour) threshold based identification, (ii) identify satisfactory training data for brown dwarfs as well as background classes, and (iii) examine a suitable approach for identifying brown dwarf candidates using their infrared photometry by applying well-known methods like k-NN classifier and NeuN. In addition, we propose an ensemble classifier for the same. The ensemble classifier will identify the final brown dwarf candidates on the basis of a majority vote, with each individual classifier being given equal weightage. Simultaneously, we also use different training sets and compare their efficiencies to determine an optimum set.

The organisation of this paper is as follows. Section 2 of this paper describes the machine learning methods, while Sect. 3 provides details of the data used and features employed to form the training sets. In Sect. 4, the outcome of the classification techniques are determined through efficiencies of few optimum training sets. These training sets are applied to certain regions in sky in Sect. 5 and inferences drawn in Sect. 6. A short summary of results is provided in Sect. 7.

#### 2 CLASSIFICATION SCHEMES

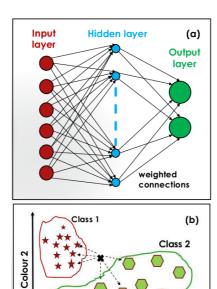
The data used for the photometric classification include the colours of objects as derived from near and mid-infrared photometric bands. The colours are designated as features in machine-learning parlance, and these are described in Sect. 3.1. Objects are classified based on their photometric similarity to known objects of a given class. These latter objects are referred to as templates. Below, we briefly introduce the classification schemes.

#### 2.1 Neural Networks

Neural networks are a family of machine learning algorithms developed for solving difficult pattern classification problems. Inspired by biological neural networks (Beale & Jackson 1990), they consist of individual processing units, called neurons or nodes. A network is constructed using layers of neurons. The first and last layer are the input and the output layer respectively, while the layers in between are designated as hidden layers. The first mathematical models of neurons, called perceptrons (Rosenblatt 1958), were capable of identifying only linearly separable patterns. A network of such neurons formulated in a feed forward architecture with input, output and one or more hidden layers, called a feed forward neural network (or a multilayer perceptron), was found to possess far greater classifying power. The layers are connected through one-directional weighted pathways between nodes in different layers. The input features for classification correspond to the nodes in the input layer, while the number of nodes in the the output layer is decided by the number of output classes. The number of hidden neurons and layers are set according to the complexity of the classification problem, which decides the non-linearity of the network. A schematic of this method is shown in Fig. 1(a).

Object classification was identified as one of the areas in astronomy where NeuN methods were likely to make an impact (Miller 1993). What sets NeuN apart from the more conventional rule-based classifiers is their ability to learn from examples (Cho et al. 1991). This learnt information is stored in the network in the form of weights along the connecting pathways between individual neurons. This allows the network to generalise, making them capable of classifying patterns which may not be included in the initial training set.

We create a simple feed-forward NeuN with the help of the MATLAB neural network toolbox. This employs supervised learning, with one hidden layer of neurons having 100 neurons. For training the network, we provide it with a training data set of example inputs and their corresponding desired outputs. Here, the inputs are a set of six colours



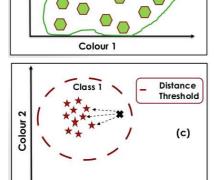


Figure 1. Schematic of the three classification techniques used in this study: (a) Neural networks, (b) k-NN Classification for k = 3, and (c) k-NN Threshold Distances for k = 3.

Colour 1

(these are discussed in detail in Section 3.1) for each object and the output is set to 0 or 1 depending on the class of the object. The network weights, which are initially set to random values, eventually store learnt patterns after training. During learning, the network weights are repeatedly adjusted in order to minimize the error between the obtained and expected output. This process, termed as backpropagation (Rumelhart et al. 1986), is repeated until either the entire training set is correctly classified, or the network is unable to minimize the error term further.

# 2.2 k-Nearest Neighbour Approach to Classification (k-NN-C)

In k-NN classification, the class of an object is decided based on its distance to a specific class of templates, where distances are defined in a multi-dimensional colour space. Usually, applications of the Nearest Neighbour methods consider the k-nearest templates (of any class) in a parameter space and decide the class based on the majority vote of these templates (Popescu et al. 2018; Miettinen 2018; Wallace et al. 2019; Akras et al. 2019). However, in the present case we select k templates from each class and estimate an average

distance to each class. We then determine the class to which the object has the shortest distance, and classify the object as belonging to that class. This approach can be considered as a modification of the threshold distance method used by Marengo & Sanchez (2009). While these authors estimate the distance to a single class, we consider a training sample consisting of multiple classes: the brown dwarf and background. This approach is helpful for the current classification problem since the templates of one class are clustered in a small region and those of the other class (i.e. the background objects) are sparsely distributed. The same 6 colours used by the NeuN classifier as input parameters are again used here to create the multi-dimensional colour-colour space. A schematic of this method can be found in Fig. 1(b).

An averaged Euclidean metric is preferred for multidimensional spaces as the distance does not increase with increase in the number of dimensions. While calculating the distance between the  $i^{th}$  test object and the  $j^{th}$  template in the training set for each colour  $d_l(i, j)$ , their photometric uncertainties (denoted by  $\sigma_i$  and  $\sigma_j$ ) are also taken into account. The k-NN distance of the i<sup>th</sup> test sample to each class is the weighted average of the Euclidean distances D(i, j) to the nearest k templates of that class, where the weights w(i,j) are introduced to reduce the influence of isolated templates that happen to be much farther away than the nearest neighbors. A Gaussian kernel is very effective for this task (Marengo & Sanchez 2009) which is given by Eqn. 3. In addition to using uncertainties of the input colours in the k-NN distance equation, they also incorporate a sparseness factor  $\sigma_s$ , in order to account for lack of observed templates of a particular class in a given location in the colour-colour space. This factor is a measure of how far apart the templates are with respect to each input colour.

$$\sigma_s(j) = \frac{1}{k} \sqrt{\sum_{t=1}^k d_l(t,j)^2} \tag{1}$$

$$\sigma_l(i,j) = \sqrt{(\sigma_i^2 + \sigma_j^2)} + \sigma_s(j)$$
 (2)

$$w(i,j) = exp\left[-\left(\frac{D(i,j)}{k \times \sqrt{\sum_{l=1}^{N} \sigma_l(i,j)^2}}\right)^2\right]$$
(3)

The k-NN distance of the  $i^{\rm th}$  test sample to a given class is given by the equation:

$$D_{kNN}(i) = \frac{\sum_{j=1}^{k} D(i,j) \times w(i,j)}{\sum_{i=1}^{k} w(i,j)}.$$
 (4)

The final classification is based on the minimum k-NN distance of the  $i^{\rm th}$  test sample to each class.

# 2.3 k-NN Threshold Distances (k-NN-TD)

In this method, the k-NN distance of each test object is calculated from a training sample consisting entirely of brown dwarfs, and objects whose k-NN distance is within a certain defined threshold, are classified as brown dwarfs. This can be visualised through the schematic given in Fig. 1(c). This

method is identical to the one used by Marengo & Sanchez (2009) and the same formula, as given in Eqn. (4) is used for calculating the k-NN distance. This application requires templates only for the search class, relying on the assumption that the templates are an accurate representation of the class, and that the selected features (colours) chosen in the analysis are sufficient to provide an effective discrimination.

The threshold distance, denoted by  $D_{th}$ , and the number of neighbors, k, are optimized for maximum completeness and rejection efficiency using the bootstrap method (Hastie et al. 2001). The completeness ( $\mathscr{C}$ ) and rejection ( $\mathscr{R}$ ) efficiencies are defined as follows:

Completeness, 
$$\mathscr{C} = \frac{\text{No. of true positives}}{\text{No. of expected brown dwarfs}}$$
 (5)

Rejection Efficiency, 
$$\mathcal{R} = \frac{\text{No. of true rejections}}{\text{Total no. of background objects}}$$

Samples of the template and background classes are created and then tested for  $\mathscr C$  and  $\mathscr R$  for different values of k and  $D_{th}$ . In order to standardize this approach, we select the k and  $D_{th}$  values which give the highest product of  $\mathscr C$  and

#### 2.4 Ensemble Classifier

We use the parametric (NeuN) as well as the two non-parametric (k-NN) techniques to create an ensemble classifier. Such a classifier will identify brown dwarfs by factoring in the outputs from all the individual classifiers and making the final decision on the basis of a majority vote. In other words, if either two or all three methods classify an object as a brown dwarf candidate, then the ensemble method would label this object as a brown dwarf candidate and vice-versa. In this method, each individual classifier is given equal weightage. This helps to reduce the misclassification due to the shortcomings of any one classifier. We attempt to quantify the performance of the ensemble classifier by calculating  $\mathscr C$  and  $\mathscr R$  in each case.

# 3 FEATURES AND DATA FOR CLASSIFICATION

The dearth of work done in this type of automated brown dwarf classification problem prompted us to create training datasets from scratch and we have decided on a set of features to be used for classification, after going through a wide variety of pre-existing literature. These are described in this section.

# 3.1 Colours used as input features

The data used for photometric classification includes the brightness of objects as measured in WISE and the Two Micron All-Sky Survey (Skrutskie et al. 2006, 2MASS). The WISE bands W1, W2 and W3 have been incorporated in the analysis. W4 has been excluded as the angular resolution in this band is lower than the other bands by a factor of  $\sim 2$ . In addition, the photometric uncertainties of

**Table 1.** Photometric colours used as features for brown dwarf classification, using WISE and 2MASS filters.

Colour	Characteristic
W1-W2	Methane absorption in W1
W2-W3	Methane absorption in lower bands
J-H	H <sub>2</sub> O absorption in J
J-W1	H <sub>2</sub> O absorption in J
J-W2	H <sub>2</sub> O absorption in J
J-K <sub>s</sub>	Presence of methane

many objects are not available in W4. Moreover, we find that its inclusion did not show any significant improvement in the results. In 2MASS, all the three bands: i.e. J (1.25  $\mu$ m), H (1.65  $\mu$ m), K<sub>s</sub> (2.17  $\mu$ m) bands have been considered. With these bands, it is possible to identify the near-infrared spectral features associated with H<sub>2</sub>O in the atmospheres of the L and T dwarfs (Stephens & Leggett 2004). We have selected six colours based on the spectral characteristics of brown dwarfs in the WISE and 2MASS filter combinations (Kirkpatrick et al. 2011; Marengo & Sanchez 2009; Faherty et al. 2016; Zhang et al. 2018). The colours used in this work are: W1-W2, W2-W3, J-H, J-W1, J-W2, J-K<sub>s</sub>, and the spectral characteristics of brown dwarfs that they describe are listed in Table 1.

# 3.2 Training Samples

The training sets are constructed using different combinations of brown dwarfs and background objects. The selection of templates used for the classification is described below.

- (i) Brown dwarfs The brown dwarfs used as templates for this work have been taken based on availability of infrared data in all the required bands. For this we have used brown dwarfs from the following research papers: Kirkpatrick et al. (2011), Thompson et al. (2013), and Best et al. (2018).
- (ii) Background objects In this classification problem, background objects include other sources such as stars in different evolutionary stages or galaxies. For this work, we have attempted to incorporate a variety of background object templates and only those have been included in the training sets, which were felt to have an effect on the classification on brown dwarfs and, hence, were likely contaminants. The objects were taken from existing literature and include NLS1 galaxies (Chen et al. 2017a), Ap and Am stars (Chen et al. 2017b), Young Stellar Objects (YSOs; Su et al. 2014; Fischer et al. 2016), Red Giants (Anders, F. et al. 2017), K-type stars (Pecaut & Mamajek 2016), and Mdwarfs (Best et al. 2018). Along with all these known background objects, some objects have been taken randomly from the WISE All-Sky point source catalog (Wright et al. 2010) by applying magnitude cuts so that they are brighter than the known brown dwarf templates in each band.

Apart from the above objects, some artificial samples are created for both classes with the characteristics of the known templates. This is carried out in a man-

ner similar to the bootstrap implementation described by Marengo & Sanchez (2009).

# 4 OPTIMAL TRAINING SETS AND PERFORMANCE OF CLASSIFICATION METHODS

It is well known that for any classification problem, diverse training data sets are of great importance. But, when it comes to brown dwarfs, few attempts have been made in this regard. Here, we have experimented with different catalogs of known brown dwarfs as well as various other stars and galaxies to create the brown dwarf and background classes, respectively. Distinct training sets were generated by combining specific template groups.

For each set,  $\mathscr C$  and  $\mathscr R$  are employed as the validation metrics. Each set is randomly divided into three parts, in the ratio 70:15:15, and labelled as training, validation and test samples, respectively. The validation set is used to tune the parameters of a classifier, for example, to tune the weights in a neural network. In the k-NN-TD method, the validation set was used to fix the distance threshold value which gave the maximum  $\mathscr{C}$ -  $\mathscr{R}$  efficiency product. The test set is then used to assess the performance of the final tuned classifier by the estimation of  $\mathscr{C}$  and  $\mathscr{R}$ . We carried out a rigorous 5-fold cross-validation for the verification of these efficiencies. In this method, each dataset is divided randomly into 5 parts. One part is taken as the test set whereas the other 4 are combined to form a training set. This testing is repeated for all the 5 parts, and the average  $\mathscr{C}$  and  $\mathscr{R}$  evaluated. We find that the efficiencies computed by both these methods are very similar (match within  $\sim 90\%$ ).

Based on the values of  $\mathscr{C}$  and  $\mathscr{R}$ , the composition of the sets were modified to select the best training sets for this particular problem. These are described below.

#### 4.1 2-Class Training Sets

The 2-class training sets include the classes of (i) brown dwarfs, and (ii) background objects. Three distinct groups were constructed for this purpose, using the combinations of objects described below. All the classification methods were applied on the training sets and the results are presented.

The k-NN-TD method is also discussed here although it employs only one class of brown dwarf templates as its training set. The threshold distance effectively demarcates the feature space into two classes, with all the objects falling within the threshold considered as brown dwarf candidates and the rest taken as other (background) objects. Thus, it is equivalent to a 2-class classification technique in this study, and we have compared its performance with the rest of the 2-class methods. We note that for all the 2-class training sets described below, only the brown dwarf template class in each group is used by the k-NN-TD classification method.

(i) Set A - This set comprises of the brown dwarfs from the sample of Kirkpatrick et al. (2011) and Thompson et al. (2013), in addition to the simulated objects created from these samples. The background class includes random WISE background objects and the corresponding simulated objects alongwith NLS1 galaxies. Note that the NLS1 galaxy sample

also includes point-like sources, with colours similar to those of brown dwarfs in the bands under consideration. The resultant efficiency values, on application of the classification techniques, are shown in Fig. 2. We find that the efficiencies are 100%. This is due to the relatively small variety of objects in the training samples, and a marked difference between the characteristic colours of the two classes, i.e., brown dwarfs and the WISE background.

(ii) Set B - The composition of this set is similar to Set A, the difference being the addition of YSO templates to the background class (Fischer et al. 2016). The inclusion of YSOs lowers the  $\mathscr R$  of the k-NN-C method to  $\sim 98\%$ , which in turn lowers the  $\mathscr R$  of the ensemble classifier by  $\sim 0.3\%$  (see Fig. 2). This could be because of the fact that YSOs have colours similar to those of brown dwarfs in the infrared bands with few being misclassified. We note that  $\mathscr C$  values remain unaffected.

(iii) Set C - This set includes the brown dwarfs from the samples of Kirkpatrick et al. (2011), Thompson et al. (2013) and Best et al. (2018). The background class comprises of all known objects (NLS1 galaxies, Ap and Am stars, YSOs, Red Giants, K-type stars, and M-dwarfs). The resultant efficiencies display a lowering of the  $\mathscr{C}$  and  $\mathscr{R}$  values with respect to sets A and B, see Fig. 2. The ensemble classifier performs marginally better than the others. The NeuN and k-NN-C methods give similar values of  $\mathscr{C}$  (98%) and  $\mathscr{R}$  (95%). k-NN-TD returns a better  $\mathcal{R}$  (98%) but a relatively lower  $\mathscr{C}$  (95%). It would appear that an increase in the variety of the samples has affected the efficiencies. Such diversification of training set is expected to help improve the  $\mathcal{R}$  of the classifiers during testing on a real scenario, as more nonbrown dwarf templates start finding representation in the background class.

The composition of the different training sets are summarised in Table 2.

# 4.2 3-Class Classification

The spectral categories of brown dwarfs (L, T, Y) can be utilized to make finer constraints in the feature space, for their identification. The L category of objects cannot be strictly considered as a brown dwarf class as it can also include lowmass stars. Therefore, in order to better evaluate the effects of L-dwarfs in the training sets on the classification of both types of objects (brown dwarfs and background), the NeuN and k-NN-C classifiers have been built with 3 categories: (i) T- & Y-dwarfs, (ii) L-dwarfs, and (iii) background objects. Each test object is classified into one of the three classes based on the features. The training set created for the 3class classification contains 2698 objects. There are 335 objects in the first category which has T- and Y-dwarfs taken from Kirkpatrick et al. (2011) and Thompson et al. (2013), along with T-dwarfs from Best et al. (2018). There are 1087 objects taken from the above samples in the second category (i.e. L-dwarfs). There are 1275 background objects comprising of random WISE background, NLS1 galaxies, YSOs from the Fischer et al. (2016) catalog and M-dwarfs. This composition is listed in Table 4. For this training set, only k-NN-C and NeuN methods are applicable as they cater to multiple classes.

The resultant efficiencies of this classification are dis-

Table 2. Composition of the 2-class Training Sets.

Set Brown Dwarfs			Background Objects				
	Composition	No. of objects	Composition	No of objects	Total no. of objects		
A	Kirkpatrick, Thompson, Simulated	1430	WISE background objects (Real&Simulated), NLS1 galaxies	1200	2630		
В	Kirkpatrick, Thompson, Simulated	1430	WISE background objects (Real&Simulated), NLS1 galaxies, Fischer YSOs	1275	2705		
C	Kirkpatrick, Thompson, Best T-dwarfs	344	NLS1 galaxies, Fischer YSOs, Ap & Am stars, Red Giants, K-type stars	325	669		

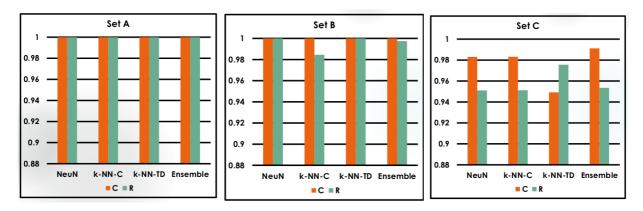


Figure 2. Completeness and Rejection Efficiencies of the 2-class training sets A, B and C, described in Sec 4.

 ${\bf Table~3.~Efficiencies~of~2-Class~Training~Sets,~described~in~Sec~4..}$ 

Completeness						Rejection	Efficiency	
Set	NeuN (%)	k-NN-C (%)	k-NN-TD (%)	Ensemble (%)	NeuN (%)	k-NN-C (%)	k-NN-TD (%)	Ensemble (%)
A	100	100	100	100	100	100	100	100
В С	100 98.3	100 98.3	100 94.9	100 99.1	$100 \\ 95.1$	$98.5 \\ 95.1$	100 97.6	99.7 95.35

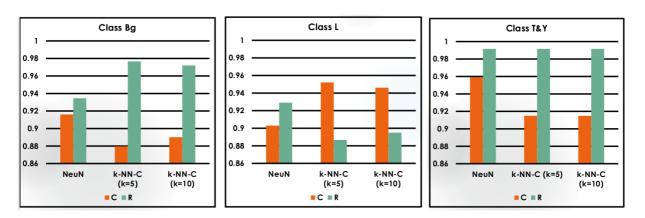


Figure 3. Completeness and Rejection Efficiencies of the 3-Class Classification (Background denoted by Bg, L-dwarfs and T, Y-dwarfs), described in Sec 4.2.

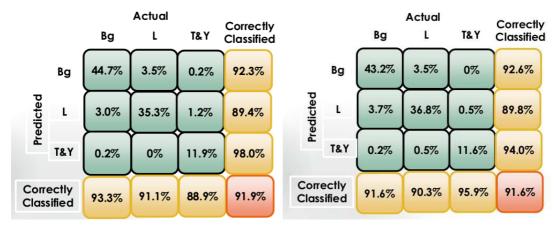


Figure 4. Confusion matrix for the 3-Class Classification (Background denoted by Bg, L-dwarfs and T, Y-dwarfs) by NeuN classifier for (left) validation and (right) test sets.

Table 4. 3 Class Classification-Training Set Composition

Set	Composition	No. of objects
3A	Best Dwarfs(1000 M,L,T), Kirkpatrick+Thompson, WISE background (Real), NLS1 galaxies, CM YSOs	2698 (Bg-1275, L-1087, T-335)

played in Table 5 and Fig. 3. We observe that  $\mathscr{R}$  values for the T- and Y-dwarfs are high (> 97%), but the  $\mathscr{C}$  values are relatively lower ( $\mathscr{R} \geq 88\%$ ). This can be attributed to the misclassification of early T-dwarfs as L-dwarfs. The L-dwarfs have lower  $\mathscr{R}$  values ( $\sim 91\%$ ) as compared to the T- and Y-dwarfs but the  $\mathscr{C}$  values are comparable ( $\geq 90\%$ ). The background class exhibits a trend similar to the T- and Y-dwarfs: low  $\mathscr{C}$  values(> 87%) but high  $\mathscr{C}$  values (> 87%).

The confusion matrix for the 3-class NeuN classifier is illustrated in Fig. 4. A confusion matrix is used to describe the performance of a classifier with multiple classes on a set of test data for which the true values are known. The rows of the matrix correspond to the class predicted by the NeuN classifier for a test object while the columns correspond to the actual class of the object. The cells along the diagonal represent the fraction of test objects correctly classified by the network. From the figure, we see that a relatively larger fraction of L-dwarfs (3.5%) are being misclassified as background when compared to T- & Y-dwarfs ( $\sim 0.5\%$  or lower). The fraction of correctly classified test objects of that class, given by the last cell at the end of each row, is lower for the L-dwarfs when compared to the T- and Y-dwarfs, as expected. All these results imply that a small fraction of the L-dwarfs are closer to the background templates, than brown dwarfs. The lowermost cell towards the right end gives the  $\mathscr{C}$  of the network for the given sample. We find that the  $\mathscr{C} > 91.5\%$  for NeuN. It is to be noted here that the ensemble classifier cannot be applied for the 3-class training sets as there are only two methods, and hence a majority vote would not be possible in some cases.

#### 4.3 Performance of Classification Methods

We summarise the performance of the classification techniques here. We find that all the three classification methods perform well on the training sets considered, with  $\mathscr{C} \geq 87\%$  and  $\mathscr{R} \geq 88\%$ . In the 2-class classification, the ensemble classifier performs as well as NeuN for the sets A and B, and marginally outperforms the other classifiers for set C. In the classification methods using 3-class training sets, it would appear that both NeuN and k-NN-C perform equally well. Thus, we see that the methods outlined above perform unequivocally well on the diverse training sets that have been created.

#### 5 TESTING ON REGIONS IN SPACE

Having analysed the performance of the classification methods to known objects, we apply them to objects in three different regions of the sky. Two of them, Hercules and Serpens, have been selected due to the relatively larger number density of known brown dwarfs in these regions. The third arbitrary region was selected based on the attribute that it should not have any known brown dwarf. For this, we considered a region towards the Lyra. The details of these three regions are given below:

- (i) A part of the constellation Serpens: This comprises a region of size  $4^{\circ} \times 9^{\circ}$ , centred on RA (J2000)=246.500°, Dec (J2000)=4.500°.
- (ii) The constellation Hercules: This comprises a region of size  $2^{\circ} \times 2^{\circ}$ , centred on RA (J2000)=268.000°, Dec (J2000)=17.000°.
- (iii) The constellation Lyra: This comprises a region of size  $2^{\circ}\times2^{\circ}$  centred on RA (J2000)=275.410°, Dec (J2000)=32.450°.

We include all sources for analysis in a given region whose infrared photometry in the requisite bands are available in the AllWISE Source Catalog (Cutri & et al. 2014). The catalog also lists the association of sources with 2MASS Point and Extended Source Catalog. The WISE photometric magnitudes, their uncertainties and the associated 2MASS

Table 5. Efficiencies of 3-Class Training Set.

			Completeness		Rejection Efficiency			
Set	Output	NeuN	k-NN-C	k-NN-C	NeuN	k-NN-C	k-NN-C	
	Class	(%)	k=5 (%)	k=10 (%)	(%)	k=5 (%)	k=10 (%)	
Validation	Bg	93.3	87.0	87.0	92.9	97.6	96.6	
	L	91.1	94.3	94.3	93.1	90.7	90.3	
	T and Y	88.9	95.8	95.8	99.7	97.5	98.3	
Test	Bg	91.6	88.0	89.0	93.5	97.7	97.2	
	L	90.3	95.2	94.6	92.9	88.7	89.5	
	T and Y	95.9	91.5	91.5	99.2	99.2	99.2	

photometry and uncertainities have been used for the classification. We restrict the search to point sources by taking only those sources with 2MASS extended source flag = 0, which indicates that the morphology of the source is consistent with a point source.

The Hercules region has 3 known T-dwarfs, one identified by Thompson et al. (2013) and two by Best et al. (2018). The Serpens region has 5 T-dwarfs, all of them identified by Best et al. (2018), and two of them were identified by Thompson et al. (2013). Therefore, if a classifier is able to identify the known brown dwarfs in these regions, it would further validate the performance of the classification techniques. For these regions, it is not possible to calculate  $\mathcal{R}$  as we do not know the exact number of brown dwarfs and background objects that are present. Therefore, in order to quantify how effectively the classifier rejects background objects, we define a new parameter called the rejection ratio (RR) in the following way.

Rejection Ratio, 
$$RR = \frac{\text{No. of objects rejected}}{\text{Total no. of objects in the region}}$$
 (7)

# 5.1 Serpens

The Serpens region contains 21,022 objects with photometry in all the requisite bands. This region contains 5 T-dwarfs. The classifier techniques developed till now are applied to the region. We find that, among the 2-class methods, NeuN performs exceptionally well and identifies all the five known brown dwarfs with RR of 99.9% and above, see Table 6. In the k-NN-C method, all the training sets, except Set A for k=10, identify all known dwarfs with a reasonably high RR. The performance of the k-NN-TD method, is quite poor with sets A and B unable to identify any known dwarf, whereas set C identifies all known dwarfs, albeit with a lower RR than the other methods ( $RR \sim 94\%$ ). The ensemble classifier also works well with sets B and C identifying all known dwarfs and set A identifying 4 out of 5. The RR values for the ensemble classifier are better than the k-NN methods, but not as high as the NeuN classifier. Among the training sets, only Set C identifies all 5 known dwarfs in all the methods. The other two sets work well in all classifiers except k-NN-TD. The classification methods using 3-class training sets also work extremely well (Table 7), identifying all 5 known dwarfs with a RR around 99.8%.

#### 5.2 Hercules

The selected region in Hercules contains 2,611 objects with photometry in all the requisite bands. The region is known to contain 3 known brown dwarfs. Application of the classification methods towards the objects in this region show that among the 2-class methods, NeuN classifier (Table 6) again produces the best results with two out of three training sets identifying all known brown dwarfs, all with RRvalues above 99.5%. This is followed by k-NN-C and k-NN-TD methods. In this region, the ensemble classifier performs better than the k-NN methods, but not as well as the NeuN classifier. Among the training sets, only Set C identifies all 3 known dwarfs and has a RR value above 99%. Training set A performs well in the NeuN (identifying all 3 known dwarfs with RR=99.53%) but is unable to identify the known dwarfs by the other two methods. The classification techniques using 3-class training sets perform well (Table 7), identifying all 3 known dwarfs with a RR value  $\sim 99.8\%$ .

### 5.3 Lyra

The selected region in the constellation Lyra has 4,620 objects with photometry in all infrared wavebands considered here. No known dwarf is found in this region. In this region, the NeuN and the ensemble classifiers (Table 6) work well giving RR above 99% for all training sets. The k-NN-C method does not fare so well in comparison, with only sets B and C having a RR of above 99%. The same follows for k-NN-TD method, with sets A and B having a RR of above 99%. But both the k-NN methods have RR > 96% for all sets. The 3-class classification techniques (Table 7) give a high RR, but classify an abnormally large number of objects as L-dwarfs, which may be due to the inclusion of early L-dwarfs from Best et al. (2018) in the 3-class training set.

#### 5.4 Search for counterparts

# 5.4.1 SIMBAD counterparts

The objects identified as brown dwarf candidates by the NeuN and ensemble classifiers for each region are given in Tables 8, 9, 10. We carried out a search in the SIMBAD astronomical database to see if these sources matched with any known objects identified previously. The search was applied based on the positional coordinates of the source and a search radius of 3.5". This corresponds to half the beam

Table 6. Results on the three regions in sky for 2-class training sets.

Tı	raining Sets		Set A			Set B			Set C	
Region	Method	$N^1$	KR <sup>2</sup>	RR <sup>3</sup> (%)	$N^1$	KR <sup>2</sup>	RR <sup>3</sup> (%)	$N^1$	$KR^2$	RR <sup>3</sup> (%)
	NeuN	16	5/5	99.9	7	5/5	99.9	20	5/5	99.9
	k-NN-C (k=5)	351	5/5	98.3	68	5/5	99.7	510	5/5	97.6
Serpens	k-NN-C (k=10)	562	4/5	97.3	78	5/5	99.6	123	5/5	99.4
	k-NN-TD ( $k$ =5)	45	0/5	99.8	1	0/5	100	1230	5/5	94.2
	k-NN-TD (k=10)	25	0/5	99.9	1	0/5	100	1294	5/5	93.8
	Ensemble	47	4/5	99.8	8	5/5	99.9	27	5/5	99.9
	NeuN	4	3/3	99.9	3	2/3	99.9	5	3/3	99.8
	k-NN-C (k=5)	22	2/3	99.2	5	2/3	99.8	72	3/3	97.2
Hercules	k-NN-C (k=10)	37	2/3	98.6	7	2/3	99.7	19	3/3	99.3
	k-NN-TD (k=5)	4	1/3	99.9	1	1/3	99.9	69	3/3	97.4
	k-NN-TD (k=10)	2	1/3	99.9	1	1/3	99.9	35	3/3	98.7
	Ensemble	3	1/3	99.9	1	1/3	99.9	7	3/3	99.7
	NeuN	7	-	99.8	1	-	99.9	2	-	99.9
	k-NN-C (k=5)	81	-	98.3	22	-	99.5	139	-	97.0
Lyra	k-NN-C (k=10)	119	-	97.4	23	-	99.5	46	-	99.0
	k-NN-TD (k=5)	9	-	99.8	0	-	100	149	-	96.8
	k-NN-TD (k=10)	6	-	99.9	2	-	99.9	162	-	96.5
	Ensemble	9	-	99.8	1	-	99.9	9	-	99.8

<sup>1</sup>N - Number of brown dwarfs

Table 7. Results on the three regions in sky for 3-class training sets.

Region	Method	No of objects in class Bg	No of objects in class L	No of objects in class T&Y	KR <sup>2</sup>	RR <sup>3</sup> (%)
Serpens	NeuN	18354	2629	39	5/5	99.8
	k-NN-C (k=5)	15827	5172	23	5/5	99.9
	k-NN-C (k=10)	2218	389	4	5/5	99.9
Hercules	NeuN	2395	208	8	3/3	99.7
	k-NN-C (k=5)	2196	411	4	3/3	99.9
	k-NN-C (k=10)	2218	389	4	3/3	99.9
Lyra	NeuN	3713	541	6	-	99.9
	k-NN-C (k=5)	3183	1071	6	-	97.0
	k-NN-C (k=10)	3300	955	5	-	99.0

 $^{2}$ KR - Ratio of known brown dwarfs  $^{3}$ RR - Rejection Ratio

of the W3 WISE band, which has the largest beams among the bands considered.

In the Serpens region, both the NeuN and ensemble classifiers together identify 59 objects, out of which 9 objects have counterparts in the SIMBAD database. We were pleasantly surprised to see one of them as a L-type brown dwarf , 2MASS J16192830+0050118. This was not included in the previous 5 known dwarfs based on the catalogs considered, as its corresponding WISE data was not linked to this particular object. The other 8 objects are M-type variable stars and galaxies. In the Hercules region, out of the 6 candidates, only one has a previously identified counterpart (an M-type variable star). The classifiers identify 15 brown dwarf candidates in the Lyra region, of which four have SIMBAD counterparts. Two of these are variable stars

and two are radio sources. While the association provides a certain estimate of the type of source, one caveat is that due to the WISE resolution, the search diameter considered for the positional search is not small, i.e. 7''.

#### 5.4.2 Gaia Counterparts

A search was also carried out using the Gaia database using the DR2 catalog (Gaia Collaboration et al. 2018) for the objects identified as brown dwarf candidates by the NeuN and ensemble classifiers. Again, a search radius of 3.5" was used and the nearest positional association was considered. For every source which had a corresponding Gaia identification, the parallaxes, proper motions and photometric magnitudes were obtained. The distance to each source was derived from

 $<sup>^2</sup>$ KR - Ratio of known brown dwarfs  $^3$ RR - Rejection Ratio

the parallax. Using the distance and apparent G magnitude, we estimated absolute G magnitude of each source. This was then was compared with the absolute magnitudes of known L and T brown dwarfs (Best et al. 2018) to estimate the spectral type of each source. The associated Gaia sources and their properties are listed in Table 11.

Three objects in Hercules, sixteen in Serpens and two in Lyra were found to have Gaia associations. Of these, only one object is of spectral type L or later and that is the known L-type brown dwarf 2MASS J16192830+0050118 which was identified earlier from the SIMBAD database. It is also the only object which is within a distance of 100 parsecs from the Earth. The majority of sources do not have Gaia counterparts. This is expected as the cooler brown dwarfs are likely to be faint at optical wavelengths.

# 6 INFERENCES

The NeuN classifier emerges as the best technique out of the individual methods, performing well in training set efficiency calculations as well as on tests in regions on the sky. The k-NN-C method with 3-class training set also holds promise with  $\mathcal{R}$  for T-dwarfs as high as 99%, and  $\mathcal{C}$  values in the range 90-95%. NeuN gives much better results than k-NN-C with 2-class training set while testing on specific regions in the sky. This method also identified a brown dwarf in the Serpens region which was not part of the initial training sets or the dwarfs identified by WISE. The k-NN-C using 2-class training set gives results comparable to NeuN while comparing training set efficiencies but does not hold up as much while testing on known test samples or new regions. The k-NN-TD method fares less well than the previously discussed methods in the training set efficiency calculations and the Hercules test region.

Thus, if the classification is to be implemented using a single method, the NeuN classifier would be an appropriate choice. But a better option would be to use the ensemble classifier, which performs reasonably well in all scenarios, and where the decision would not be dependent on a single classifier alone. Of all the training sets used (in 2-class classifications), training set C performs best on sources from given regions on the sky. It shows maximum efficiency in the Hercules and Serpens regions, and high rejection ratios in the Lyra region. But the set has low training set efficiency in cross-validation (one amongst the lowest in both k-NN methods). Training Set B was found to be effective in rejecting background sources but its performance in the Hercules region was poor compared to training set C. Training set A has high efficiencies and managed to identify the new brown dwarf in the Serpens region, but it does not hold up as well in the other aspects. Also, it has the least sample variety among the 3 training sets. The fact that Set C, which has the most sample variety, is also the best performer seems to indicate that there is a strong correlation between performance and background sample variety. Set B has intermediate sample variety and its performance is also intermediate that of the other 2 sets. Since astronomical objects are found in a vast variety, inclusion of different types of background objects helps machine-learning algorithms perform better.

A number of objects have been identified in the three regions using the classification techniques described in the

work. We note that the SIMBAD associations indicate that we are also selecting objects which are not brown dwarfs, viz. M-type stars, few galaxies and carbon stars. In one case, a radio source is also identified as a counterpart. Thus, one needs to probe in detail in order to confirm the associations of the identified candidates. However, the fact that known brown dwarfs have been identified by these methods provides strong support to the fact that the methods are effective and one can verify the nature of the other brown dwarf 'candidates' through follow-up spectroscopic studies. It is worth noting that the number of brown-dwarf candidates identified by these classifiers is much lower than those identified by traditional techniques of colour-magnitude restrictions which, in turn, saves time and resources required for the final verification.

We have used WISE and 2MASS data for this study as they were easily available for numerous different objects. A more robust classification calls for additional information in the form of identification of the background objects, eg. YSOs in a given star-forming region. Alternate examples include cross-identifications of background stars and galaxies across catalogs, or investigation of regions away from Galactic plane where the number density of stars is lower than the Galactic plane. The amount and nature of extinction towards each object is also expected to play a crucial role in ascribing a background source as a brown dwarf. This additional information provided to the classifiers can improve the classification and reliability of the methods. Lastly, we note that these classification techniques can be used to identify any group of elusive astronomical objects, by changing the input colours and training sets. Hence, this approach can serve as a base for classification of other astronomical objects as well.

#### 7 SUMMARY

- NeuN and k-NN methods have been used for classifying astronomical objects based on their photometric colours. Although the methods are general and can be applied to select any specific kind of astronomical objects, we have applied it to the specific case of brown dwarfs.
- In this study, apart from NeuN, we have used two different k-NN methods: k-NN-C and k-NN-TD, for classifying brown dwarfs, using six colours from WISE and 2MASS. We also propose an ensemble classifier which identifies brown dwarf candidates on the basis of a majority vote from the above three methods.
- A number of training sets have been constructed for testing the performance of the classifiers. This includes the 2-class and 3-class training sets.
- ullet In addition to the different techniques, we create different training sets by combining templates from various known brown dwarf and background object catalogs. The efficiencies for the sets, and for different methods, are then calculated by using  $\mathscr C$  and  $\mathscr R$  as the validation metrics.
- All the methods perform well on the training sets considered, with  $\mathscr{C} \geq 87\%$  and  $\mathscr{R} \geq 88\%$ . In the 2-class classification, both NeuN and the ensemble classifier emerge as the best methods. Both NeuN and k-NN-C perform equally well in the 3-class clssification methods.
  - We apply the methods and optimal training sets to

Table 8. Brown dwarf candidates identified by NeuN and Ensemble Classifier in Hercules, the alphabets in brackets indicate the training set.

WISE ID	$lpha_{ m J2000} \  m (deg)$	$\delta_{ m J2000} \  m (deg)$	Technique	SIMBAD Association	Category
J175510.28+180320.2	268.7928	18.0555	NeuN(A,B,C); Ensemble(A,B,C)	2MASS J17551062+1803203	Brown Dwarf
J175032.93+175904.2	267.6378	17.9844	NeuN(A,C); Ensemble(C)	2MASS J17503293+1759042	Brown Dwarf
J175454.47+164919.6	268.7273	16.8218	NeuN(A,B,C); Ensemble(C)	2MASS J17545447+1649196	Brown Dwarf
J174846.58+165936.8	267.1941	16.9935	NeuN(C); $Ensemble(A,C)$	No object found	
J175527.28 + 162322.9	268.8637	16.3897	Ensemble(A)	No object found	
J175338.61+164013.8	268.4109	16.6705	NeuN(A); Ensemble(C)	NSV 9827	Semi-regular pulsating star
J174811.68+174109.6	267.0487	17.6860	NeuN(C)	No object found	
J174827.01+175113.4	267.1125	17.8537	Ensemble(C)	No object found	
J174854.84 + 164232.2	267.2285	16.7089	Ensemble(C)	No object found	

Table 9. Brown dwarf candidates identified by NeuN and Ensemble Classifier in Lyra, the alphabets in brackets indicate the training set.

WISE ID	$lpha_{ m J2000} \  m (deg)$	$\delta_{ m J2000} \  m (deg)$	Technique	SIMBAD Association	Category
J181811.58+325942.6	274.5483	31.5300	Ensemble(C)	No object found	
J181820.10+314200.5	274.5837	31.7001	NeuN(A,B,C); Ensemble(A,B,C)	V* A0 Lyr	Variable star of Mira Cet type
J181850.55+315116.8	274.7106	31.8546	NeuN(A); Ensemble(A)	No object found	
J181811.58+325942.6	274.8476	31.7453	Ensemble(C)	No object found	
J181950.91+325747.9	274.9621	32.9633	NeuN(A); Ensemble(A,C)	No object found	
J182043.02+314006.9	275.1793	31.6686	Ensemble(A)	NVSS J182043+314006	Radio Source
J182046.97+332431.9	275.1957	33.4089	Ensemble(C)	No object found	
J182047.58+313841.0	275.1983	31.6447	Ensemble(C)	No object found	
J182138.94+323907.4	275.4122	32.6521	NeuN(A)	No object found	
J182240.79+332256.1	275.6699	33.3822	Ensemble(A)	No object found	
J182246.23+322258.9	275.6926	32.3830	NeuN(A); $Ensemble(A,C)$	NVSS J182246+322258	Radio Source
J182259.75+314404.1	275.7490	31.7345	Ensemble(A)	No object found	
J182302.37+321010.3	275.7599	32.1695	Ensemble(C)	No object found	
J182528.82 + 313045.4	276.3701	31.5126	NeuN(A); Ensemble(A)	No object found	
J182529.61 + 313305.3	276.3734	31.5515	NeuN(A,C); $Ensemble(A,C)$	V* IS Lyr	Variable star of Mira Cet type

three regions in the sky: Serpens, Hercules and Lyra. Of these, Serpens and Hercules have known brown dwarfs, previously identified by WISE.

- The NeuN classifier performs relatively better than the k-NN methods in the three regions, in the 2-class classification, identifying all the previously known dwarfs. This is followed by the ensemble classifier. The two k-NN methods do not fare as well, withk-NN-C being the better of the two.
- The 3-class classification also holds promise with its performance equalling or even exceeding that of the 2-class NeuN.
- A search for counterparts in the SIMBAD and Gaia databases was also carried out for the brown dwarf candidates from each region. This led to the identification of one of the candidates in the Serpens region as a brown dwarf which was not part of the brown dwarfs identified by WISE. A fraction of the other candidates are variable stars and other background objects.
- These methods of multi-dimensional classification based on photometric colours are expected to significantly downsize the candidate sample for follow-up studies, as compared to traditional colour and magnitude diagrams or threshold cuts.

# ACKNOWLEDGEMENTS

We thank the referee M. Marengo for useful suggestions that have improved the paper presentation. This publication makes use of data products from the Wide-Field In-

frared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. This publication also makes use of data products from 2MASS, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation. This research has made use of the Vizier and SIM-BAD databases, operated at CDS, Strasbourg, France. This work has also benefitted from the M, L, and T dwarf compendium housed at DwarfArchives.org, whose server was funded by a NASA Small Research Grant, administered by the American Astronomical Society. This work presents results from the European Space Agency (ESA) space mission Gaia (http://www.cosmos.esa.int/gaia) taken from the archive website https://archives.esac.esa.int/gaia. The data was processed by the Gaia Data Processing and Analysis Consortium (DPAC) which is funded by national institutions, in particular the institutions participating in the Gaia MultiLateral Agreement (MLA).

#### REFERENCES

Akras S., Leal-Ferreira M. L., Guzman-Ramirez L., Ramos-Larios G., 2019, MNRAS, 483, 5077

Allen L. E., et al., 2004, The Astrophysical Journal Supplement Series, 154, 363

Table 10. Brown dwarf candidates identified by NeuN and Ensemble Classifier in Serpens, the alphabets in brackets indicate the training set.

WISE ID	$lpha_{ m J2000} \  m (deg)$	$\delta_{ m J2000} \  m (deg)$	Technique	$\begin{array}{c} {\rm SIMBAD} \\ {\rm Association} \end{array}$	Category
J162918.56+033535.5	247.3273	3.5932	NeuN(A,B,C); Ensemble(A,B,C)	2MASS J16291840+0335371	Brown Dwarf
J163236.47+032927.3	248.1520	3.4909	NeuN(A,B,C); Ensemble(A,B,C)	2MASS J16323642+0329269	Brown Dwarf
J161927.53+031348.1	244.8649	3.2299	NeuN(A,B,C); Ensemble(A,B,C)	2MASS J16192751+0313507	Brown Dwarf binary
J162414.07+002915.6	246.0587	0.4877	NeuN(A,B,C); Ensemble(A,B,C)	2MASS J16241436+0029158	Brown Dwarf
J163022.90+081821.0	247.5954	8.3058	NeuN(A,B,C); Ensemble(B,C)	2MASS J16302295+0818221	Brown Dwarf
J161800.81+074046.4	244.5034	7.6795	NeuN(A)	No object found	
J161804.33+034225.6	244.5180	3.7071	Ensemble(C)	No object found	
J161817.46+073151.3	244.5728	7.5309	Ensemble(A)	No object found	
J161835.18+070629.7	244.6466	7.1083	Ensemble(A)	No object found	
J161910.21+011905.8	244.7926	1.3183	NeuN(C); Ensemble(C)	No object found	G. 1 . G.1
J161910.47+064223.3	244.7936	6.7065	NeuN(C); Ensemble(C)	SDSS J161910.48+064223.2	Starburst Galaxy
J161915.97+083555.5	244.8166	8.5987	Ensemble(A)	No object found	
J161917.17+060856.5	244.8215	6.1490	Ensemble(A)	No object found	B Df
J161928.33+005010.8	244.8681	0.8363	NeuN(A)	2MASS J16192830+0050118	Brown Dwarf
J161938.90+043203.8	244.9121	4.5344	NeuN(C); Ensemble(C)	No object found	
J161950.93+082348.4	244.9622	8.3968	Ensemble(A)	No object found	
J162004.13+002141.9	245.0172 $245.0737$	0.3616 $4.1358$	Ensemble(A) $ Ensemble(A)$	No object found No object found	
J162017.70+040808.8 J162037.66+014617.6	245.1569	1.7715	NeuN(A)	No object found	
J162055.66+070510.0	245.2319	7.0861	Ensemble(A)	No object found	
J162035.66+070310.0 J162106.03+073555.9	245.2751	7.5988	Ensemble $(C)$	No object found	
J162208.71+014252.2	245.5363	1.7145	NeuN(C); Ensemble(C)	No object found	
J162221.38+030814.8	245.5891	3.1374	Ensemble(A)	No object found	
J162308.40+002259.7	245.7850	0.3832	Ensemble(A)	No object found	
J162327.26+000643.4	245.8636	0.1120	Ensemble(C)	No object found	
J162348.61+043008.2	245.9526	4.5023	NeuN(C); Ensemble(C)	No object found	
J162350.29+053635.4	245.9595	5.6098	NeuN(A)	No object found	
J162359.75+082017.6	245.9990	8.3382	NeuN(A)	No object found	
J162407.87+072619.7	246.0328	7.4388	Ensemble(C)	No object found	
J162426.65+054218.5	246.1110	5.7051	Ensemble(A)	No object found	
J162450.40+054231.4	246.2100	5.7087	Ensemble(A)	No object found	
J162450.47+014948.1	246.2103	1.8300	Ensemble(A)	No object found	
J162454.89+025957.7	246.2287	2.9994	NeuN(C)	V* V911 Oph	Long-period variable star
J162532.03+012123.7	246.3834	1.3566	Ensemble(A)	No object found	3 1
J162532.15+055407.0	246.3839	5.9019	Ensemble(C)	No object found	
J162616.28+051700.6	246.5678	5.2835	Ensemble(A)	No object found	
J162618.85+041127.6	246.5785	4.1910	NeuN(A)	No object found	
J162623.15+020009.7	246.5965	2.0027	Ensemble(C)	No object found	
J162642.50 + 045429.5	246.6771	4.9082	Ensemble(A)	No object found	
J162655.81+083510.7	246.7325	8.5863	Ensemble(A)	No object found	
J162716.25+004031.8	246.8177	0.6755	Ensemble(A)	No object found	
J162810.89 + 003947.7	247.0454	0.6632	Ensemble(C)	No object found	
J162815.15 + 005836.8	247.0631	0.9769	NeuN(B,C); Ensemble(C)	No object found	
J162823.78 + 072620.6	247.0991	7.4390	NeuN(C); Ensemble(C)	No object found	
J162903.50 + 055218.3	247.2646	5.8717	NeuN(A,C); Ensemble(C)	No object found	
J162936.12 + 060416.8	247.4005	6.0713	NeuN(C); Ensemble(C)	No object found	
J162948.11 + 054951.4	247.4504	5.8309	NeuN(A)	No object found	
J163030.11 + 030044.7	247.6255	3.0124	Ensemble(A)	No object found	
J163045.42 + 065252.6	247.6893	6.8813	NeuN(C); Ensemble(C)	SDSS J163045.43 $+$ 065252.5	Blue object
J163053.87 + 045609.3	247.7244	4.9359	Ensemble(C)	LARCS $a2204r04-2748$	Galaxy in a cluster
J163109.27 + 033600.5	247.7887	3.6001	NeuN(C); Ensemble(B,C)	No object found	
J163114.37 + 044720.4	247.8099	4.7890	NeuN(C); Ensemble(C)	No object found	
J163157.59 + 030610.9	247.9899	3.1030	NeuN(C); Ensemble(B,C)	V* V721 Oph	Variable star (Mira Cet type)
J163236.72 + 014649.0	248.1530	1.7803	Ensemble(A)	No object found	
J163240.24 + 055427.9	248.1677	5.9077	Ensemble(A)	No object found	
J163247.31 + 021041.2	248.1971	2.1781	Ensemble(A)	No object found	
J163248.85 + 073420.3	248.2035	7.5723	Ensemble(A)	No object found	
J163255.54 + 065129.7	248.2314	6.8582	NeuN(A,B,C); $Ensemble(B,C)$	V* SS Her	Variable star (Mira Cet type)
J163304.98+051652.0	248.2707	5.2811	Ensemble(A)	No object found	
J163314.56+011850.4	248.3106	1.3140	Ensemble(A)	No object found	
J163343.00+064010.3	248.4291	6.6695	Ensemble(A)	No object found	
J163345.48+070903.2	248.4395	7.1509	Ensemble(A)	PMN J1633+0708	Radio source
J163351.17+002959.7	248.4632	0.4999	Ensemble(A)	2MASX J16335114+0029595	Galaxy
J163356.21 + 062654.2	248.4842	6.4484	Ensemble(A)	No object found	

Anders, F. et al., 2017, A&A, 597, A30

Beale R., Jackson T., 1990, Neural Computing: An Introduction. IOP Publishing Ltd., Bristol, UK, UK

Best W. M. J., et al., 2018, The Astrophysical Journal Supplement Series, 234, 1

Bogart R. S., Wagoner R. V., 1973, ApJ, 181, 609

Chen P., Liu J., Shan H., 2017a, New Astronomy, 54, 30

Chen P. S., Liu J. Y., Shan H. G., 2017b, The Astronomical Jour-

nal, 153, 218

Cho T. ., Conners R. W., Araman P. A., 1991, in [1991] Proceedings of the IEEE/ACM International Conference on Developing and Managing Expert System Programs. pp 202–209, doi:10.1109/DMESP.1991.171738

Costa E., Méndez R. A., Jao W. C., Henry T. J., Subasavage J. P., Brown M. A., Ianna P. A., Bartlett J., 2005, AJ, 130, 337

Cushing M. C., et al., 2011, The Astrophysical Journal, 743,  $50\,$ 

Table 11. The Gaia associations and their properties (parallaxes, proper motions and estimates spectral types), of the brown dwarf candidates identified by NeuN and Ensemble Classifier in all three regions.

Region	$lpha_{ m J2000} \  m (deg)$	$\delta_{ m J2000} \  m (deg)$	Gaia Source Id	Parallax (mas)	Distance (parsecs)	Estimated Spectral Type $^{l}$	Proper Motion (mas/yr)
Hercules	267.1941	16.9935	4550887629617161344	0.0972	10284.8809	>L0	0.172097
	267.0487	17.686	4550974216159379200	0.4213	2373.1620	>L0	4.523704
	267.1125	17.8537	4551012973945004032	0.6534	1530.2454	>L0	5.065651
Serpens	244.5181	3.7071	4436379751351585536	6.0044	166.5441	>L0	21.157053
	244.7926	1.3183	4408885737249657216	0.9791	1021.2935	>L0	6.5097901
	244.8681	0.8363	4408577118080317056	37.2891	26.8175	L0 - L2	96.588497
	245.0737	4.1358	4436510696314413312	1.6612	601.9658	>L0	0.726790
	245.5363	1.7145	4432757204136317440	0.1174	8515.7912	>L0	1.553490
	246.2287	2.9994	4433148282382642304	0.6089	1642.3517	>L0	6.897879
	246.3839	5.9019	4438657282313809408	0.4231	2363.6879	>L0	8.115868
	247.0991	7.4390	4439633854803047680	0.4901	2040.3025	>L0	0.436674
	247.4005	6.0713	4438724112005249152	1.2443	803.6814	>L0	0.944441
	247.6893	6.8813	4438837636580679680	0.0376	26602.8631	>L0	0.340288
	247.7886	3.6001	4433462811427426816	0.3383	2955.5194	>L0	8.287301
	248.1677	5.9077	4438472461282781696	4.1644	240.1279	>L0	31.395639
	248.1971	2.1781	4432414564529967232	2.3157	431.8293	>L0	2.881503
	248.2314	6.8582	4439156220075003520	0.5491	1821.0399	>L0	6.037068
	248.2707	5.2811	4435337375673394944	0.1840	5433.4089	>L0	0.390203
	248.4395	7.1509	4439197211243356672	0.4347	2300.4445	>L0	0.807281
Lyra	274.5483	31.5300	4591800083627498496	0.3162	3162.5501	>L0	9.983660
	275.4122	32.6521	4592724906048223616	0.3433	2912.5313	>L0	6.467676

<sup>1 &</sup>gt; means earlier spectral type

Cutri R. M., et al. 2014, VizieR Online Data Catalog, p. II/328 Faherty J. K., et al., 2016, The Astrophysical Journal Supplement SeriesRumelhart D. E., Hinton G. E., Williams R. J., 1986, 225, 10

Fischer W. J., Padgett D. L., Stapelfeldt K. L., SewiÅĆo M., 2016, The Astrophysical Journal, 827, 96

Gaia Collaboration et al., 2018, A&A, 616, A1

González R. E., Muñoz R. P., Hernández C. A., 2018, Astronomy and Computing, 25, 103

Hastie T., Tibshirani R., Friedman J., 2001, The Elements of Statistical Learning. Springer Series in Statistics, Springer New York Inc., New York, NY, USA

Heydon-Dumbleton N. H., Collins C. A., MacGillivray H. T., Monthly Notices of the Royal Astronomical Society, 238, 379

Ho Y.-C., Agrawala A. K., 1968, Proceedings of the IEEE, 56, 2101

Kirkpatrick J. D., et al., 2011, The Astrophysical Journal Supplement Series, 197, 19

Krakowski T., Małek K., Bilicki M., Pollo A., Kurcz A., Krupa M., 2016, A&A, 596, A39

Kurcz A., Bilicki M., Solarz A., Krupa M., Pollo A., Małek K., 2016, A&A, 592, A25

Luhman K. L., 2012, Annual Review of Astronomy and Astrophysics, 50,65

Marengo M., Sanchez M. C., 2009, The Astronomical Journal, 138, 63

Miettinen O., 2018, Ap&SS, 363, 197

Miller A., 1993, Vistas in Astronomy, 36, 141

Mužić K., Schödel R., Scholz A. e., Geers V. C., Jayawardhana R., Ascenso J., Cieza L. A., 2017, MNRAS, 471, 3699

Nakoneczny S., Bilicki M., Solarz A., Pollo A., Maddox N., Spiniello C., Brescia M., Napolitano N. R., 2018, arXiv eprints, p. arXiv:1812.03084

Odewahn S. C., Stockwell E. B., Pennington R. L., Humphreys R. M., Zumach W. A., 1992, AJ, 103, 318

Pecaut M. J., Mamajek 2016. Monthly Notices of the Royal Astronomical Society, 461.

Popescu M., Licandro J., Carvano J. M., Stoicescu R., de León J., Morate D., Boacă I. L., Cristescu C. P., 2018, A&A, 617, A12 Rosenblatt F., 1958, Psychological Review, pp 65-386

MIT Press, Cambridge, MA, USA, Chapt. Learning Internal Representations by Error Propagation, pp 318-362, http://dl.acm.org/citation.cfm?id=104279.104293

Skrutskie M. F., et al., 2006, AJ, 131, 1163

Spiegel D. S., Burrows A., Milsom J. A., 2011, The Astrophysical Journal, 727, 57

Stephens D. C., Leggett S. K., 2004, Publications of the Astronomical Society of the Pacific, 116, 9

Stephens D. C., et al., 2009, in Stempels E., ed., American Institute of Physics Conference Series Vol. 1094, 15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun. pp 172-177, doi:10.1063/1.3099088

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

Su Y., Fang M., Yang J., Zhou P., Chen Y., 2014, The Astrophysical Journal, 788, 122

Thompson M. A., et al., 2013, Publications of the Astronomical Society of the Pacific, 125, 809

Tinney C. G., Faherty J. K., Kirkpatrick J. D., Wright E. L., Gelino C. R., Cushing M. C., Griffith R. L., Salter G., 2012, The Astrophysical Journal, 759, 60

Wallace S. M., Burbine T. H., Sheldon D., Dyar M. D., 2019, in Lunar and Planetary Science Conference. Lunar and Planetary Science Conference. p. 1097

Warren S. J., et al., 2007, MNRAS, 381, 1400

Wright E. L., et al., 2010, The Astronomical Journal, 140, 1868 Zhang Z., et al., 2018, ApJ, 858, 41

van der Veen W. E. C. J., Habing H. J., 1988, A&A, 194, 125

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