Artificial neural networks for selection of pulsar candidates from radio continuum surveys

Naoyuki Yonemaru,1,2★ Keitaro Takahashi,1,3★ Hiroki Kumamoto,1,2 Shi Dai,2 Shintaro Yoshiura1,4 and Shinsuke Ideguchi5

1Kumamoto University, Graduate School of Science and Technology, Kumamoto, 860-8555, Japan
2CSIRO Astronomy and Space Science, PO Box 76, Epping NSW 1710, Australia
3International Research Organization for Advanced Science and Technology, Kumamoto University, Kumamoto, 860-8555, Japan
4The University of Melbourne, School of Physics, Parkville, VIC 3010, Australia
5Department of Astrophysics/IMAPP, Radboud University Nijmegen, PO Box 9010, NL-6500 GL Nijmegen, the Netherlands

Accepted 2020 March 12. Received 2020 February 21; in original form 2018 November 26

ABSTRACT
It is very computationally expensive to search for pulsars using time-domain observations, and the volume of data will be enormous with next-generation telescopes such as the Square Kilometre Array. We use artificial neural networks (ANNs), a machine learning method, for the efficient selection of pulsar candidates from radio continuum surveys; this is much cheaper than using time-domain observations. With observed quantities such as radio fluxes, sky position and compactness as inputs, our ANNs output the ‘score’ that indicates the degree of likeliness that an object is a pulsar. We demonstrate ANNs based on existing survey data by the Tata Institute for Fundamental Research (TIFR) Giant Metrewave Radio Telescope (GMRT) Sky Survey (TGSS) and the National Radio Astronomy Observatory (NRAO) Very Large Array (VLA) Sky Survey (NVSS) and we test their performance. The precision, which is the ratio of the number of pulsars classified correctly as pulsars to the number of any objects classified as pulsars, is about 96 per cent. Finally, we apply the trained ANNs to unidentified radio sources and our fiducial ANN with five inputs (the galactic longitude and latitude, the TGSS and NVSS fluxes and compactness) generates 2436 pulsar candidates from 456 866 unidentified radio sources. We need to confirm whether these candidates are truly pulsars by using time-domain observations. More information, such as polarization, will narrow the number of candidates down further.

Key words: methods: statistical – pulsars: general – radio continuum: galaxies.

1 INTRODUCTION
Pulsars are rapidly rotating neutron stars with ultra-strong magnetic fields. They emit weak radio beams from their magnetic poles, which can be seen as pulses with extremely stable periods. They are used as tools in a wide range of physical experiments: low-frequency gravitational wave detection by regular monitoring of time-of-arrival (ToA) of pulses, known as a pulsar timing array (Foster & Backer 1990; Jenet et al. 2009; Manchester et al. 2012; Kramer & Champion 2013), tests of gravitational theory (Kramer et al. 2006; Berti et al. 2015), nuclear physics inside neutron stars (Lattimer & Prakash 2004), studies of the galactic interstellar medium (ISM) and magnetic fields (Han, Ferriere & Manchester 2004; Schnizler 2012), etc. Since the discovery of the pulsar in 1968 (Hewish et al. 1968), many pulsar surveys have been performed over the last half century (Manchester et al. 2001; Cordes et al. 2006) and currently about 2500 pulsars have been found.

However, the search for pulsars in the time-domain is observationally and computationally expensive as it is necessary to resolve narrow pulses with high time resolution. For example, the Parkes Multibeam Pulsar Survey (Manchester et al. 2001) is a blind pulsar survey with an observation time per pointing of 35 min and with 2670 pointings in the region of $50^\circ \leq l \leq 260^\circ$ and $|b| \leq 5^\circ$, where $l$ and $b$ are the galactic longitude and latitude, respectively. Another example is the Arecibo Pulsar Survey using the Arecibo L-band Feed Array (Cordes et al. 2006). With an observation time per pointing of 17.1 h (32.2 h), it performed observations of 919 (865) pointings in the Galactic anti-Centre (the inner Galaxy), covering an area of 15.8 deg$^2$ (14.8 deg$^2$).

In the near future, an exceedingly large number of pulsars are expected to be discovered with the Square Kilometre Array (SKA; Keane et al. 2015), and accordingly the volume of data will be
enormous (Smits et al. 2009; Levin et al. 2018). Therefore, the
selection of pulsar candidates from radio continuum surveys, which is
much cheaper and related to other sciences, will be useful to reduce
the number of objects to perform time-domain observations.
Recently, the selections of pulsar candidates using the spectral index
and compactness (Frail et al. 2018; Maan et al. 2018) and variance
images (Dai et al. 2016) have been studied. Frail et al. (2018) have
found five new pulsars from unassociated sources of the Fermi
Large Area Telescope.
In this work, we apply artificial neural networks (ANNs) to
the selection of pulsar candidates from radio continuum survey
data. An ANN is a machine learning method that is inspired
by the structure of the human brain. Recently, machine learning
methods including ANNs have been studied and applied in the
field of astronomy. Some representative examples include the
morphological classification of galaxies (Storrie-Lombardi et al.
1992; Naim et al. 1995; Folkes, Lahav & Maddox 1996; Banerji
et al. 2010), the detection and parameter estimation of gravitational
waves with multiple interferometers (George & Huerta 2018), the
improvement of the accuracy of photometric redshift estimation
with spectroscopic and photometric data of galaxies (Collister
& Lahav 2004; Vanzella et al. 2004; Samui & Samui Pal 2017)
and the extraction of astrophysical parameters from the power spectrum
of the 21-cm line from the epoch of reionization (Shimabukuro
& Semelin 2017).
We construct ANNs that output the ‘score’ that represents the
similarity of an observed object to a pulsar from several quantities
obtained from radio continuum surveys, such as flux, spectral index,
sky position and compactness. Using the existing radio
catalogue, we select radio sources that are very likely to be
pulsars. This could make pulsar searches much more efficient
than blind surveys. The ANNs are trained with known pulsars
and non-pulsar objects and thus this approach is categorized as
supervised machine learning. Specifically, we construct our ANNs
using data from the Tata Institute for Fundamental Research (TIFR)
Giante Metrewave Radio Telescope (GMRT) Sky Survey (TGSS;
Intema et al. 2017) and the National Radio Astronomy Observatory
(NRAO) Very Large Array (VLA) Sky Survey (NVSS; Condon et al.
1998), and we demonstrate how precisely our ANNs select pulsar
candidates. Here, we note that any type of pulsar can be found by
this method, irrespective of their periods and dispersion measures
(DMs), because all available pulsars are used as training data.
There are several previous works on the selection of pulsar
candidates using ANNs (Eatough et al. 2010; Morello et al. 2014;
Bethapudi & Desai 2018). These works utilize quantities from time-
domain observations, such as the pulse profile, signal-to-noise ratio
(S/N), width and chi-square of fit to the theoretical DM–S/N curve.
Then, ANNs are used to judge if the signal is from a pulsar or
terrestrial radio frequency interference. Furthermore, several studies
(Zhu et al. 2014; Guo et al. 2017; Connor & van Leeuwen 2018;
Wang et al. 2019) adopt convolutional neural networks (CNNs)
for the classification of pulsars and fast radio bursts using time–phase
plots and frequency–phase plots as inputs. However, our ANNs
are able to pick out pulsar candidates from the continuum surveys
without time-domain observations. Therefore, our application is
in a different phase of pulsar searching from the previous works. Note
that our ANNs provide not true pulsars but their candidates, and so
time-domain observations are necessary for these candidates to be
confirmed as pulsars.
The outline of this paper is as follows. In Section 2, we introduce
the source catalogues that provide the unidentified objects and the
training data for our ANNs. In Section 3, we present the architecture
and training method of the ANNs. In Section 4, we describe the
features used as inputs for the ANNs and how to apply the ANNs
to the selection of pulsar candidates. In Section 5, we test the
performance of trained networks, we try to interpret their interiors
and we apply them to the unidentified objects. We give a summary
and discussion in Section 6.

2 RADIO SOURCE CATALOGUE

In this paper, we construct ANNs using a radio source catalogue
developed by de Gasperin, Intema & Frail (2018). The catalogue
consists of radio sources cross-matched between the TGSS
and NVSS, as described below.

(i) The TGSS Alternative Data Release 1 (Intema et al. 2017) is
a radio continuum survey at 147 MHz carried out with the GMRT.
This survey covers the north sky of $\delta = -53^\circ$ visible from the GMRT
(90 per cent of the celestial sphere). The resolution of this survey is
25 arcsec and the median rms noise is 3.5 mJy beam$^{-1}$. The overall
astrometric accuracy is better than 1 arcsec in RA and Dec., while
the flux density accuracy is estimated to be $\sim 10$ per cent for most of
the survey area. The higher resolution of the GMRT, combined with
the data reduction strategy that down-weights the short baselines,
reduced both the sensitivity of the TGSS to extended emission as
well as the presence of artefacts along the Galactic plane due to
bright, extended sources. The largest detectable angular scale in the
TGSS is of the order of a few arcmin.

(ii) The NVSS (Condon et al. 1998) is a radio continuum survey at
1.4 GHz carried out with the VLA. This survey covers the sky
north of $\delta = -40^\circ$ (82 per cent of the celestial sphere). The survey
was performed with the VLA in D and DnC configurations in full
polarization. However, for this work we used only Stokes I images.
The resolution is 45 arcsec and the background rms noise is nearly
uniform at 0.45 mJy beam$^{-1}$. The overall astrometric accuracy is
better than 1 arcsec in RA and Dec. Because of the compactness
of the VLAF configuration used, the surface brightness of extended
sources is fairly well reconstructed up to scales of $\sim 16$ arcmin.
At the same time, the extended and unmodelled surface brightness
from the Galactic plane lowers the fidelity of images at low galactic
latitude.

In de Gasperin et al. (2018), radio sources are cross-matched
and objects with a separation less than 15 arcsec are regarded as
the same object. Besides these cross-matched sources, we also use
radio sources detected only by the TGSS. This is because pulsars
with steep spectra could be dimmer than the detection limit of the
NVSS and appear only in the TGSS catalogue. For these sources,
we allocate the upper bound on the NVSS flux and spectral index.
Hereafter, we call these objects (‘S’, ‘M’ and ‘U’ in de Gasperin
et al. 2018) the ’de Gasperin catalogue’ and it has 470,052 sources.
In order to construct ANNs, we need a training data set with
radio sources that are already known to be pulsars or non-pulsars.
To extract pulsars from the de Gasperin catalogue, we cross-match
it with the Australia Telescope National Facility (ATNF) pulsar
catalogue (Manchester et al. 2005) and 127 sources are identified as
pulsars. The ATNF pulsar catalogue consists of 2253 normal pulsars
and 360 millisecond pulsars, whereas our training data include 107
normal pulsars and 20 millisecond pulsars. Although the ratio of
pulsars in the training data to all the pulsars in the ATNF pulsar
catalogue is just 4.9 per cent, there is no significant bias in the ratio
of millisecond pulsars to normal pulsars in the training data. The de
Gasperin catalogue is further cross-matched with the Million Quasar
(MILLIQUAS) catalogue (Flesch 2015), which consists of various
ANNs for pulsar candidate selection

The ANN, which is a machine learning method, is a mathematical model inspired by the human brain and it has recently been attracting much attention. The purpose of the ANN is to classify objects from input data and it is necessary to construct a suitable network by optimizing the network parameters with a training data set. In our case, the input data are observed quantities that characterize a radio source, such as flux, spectral index, sky position and compactness, while the output of the training data is unity/zero for a pulsar/non-pulsar, respectively. In this work, we employ the simplest model of the multilayer perceptron with three layers because of the relatively small number of input quantities mentioned above.

In this section, we describe the network architecture and the process of optimizing the network parameters (the training process) briefly.

### 3.1 ANN architectures

We consider ANNs that consist of three layers: the input, hidden and output layers. Each layer has neurons that are described as $x_i, y_j$.
and $z_k$, respectively. Here, a neuron is the basic element of an ANN, which generates one output from multiple inputs. An output from a neuron in the hidden layer, $y_j$, is written as

$$ y_j = f(u_j), $$ (1)

where $u_j$ is given by a linear combination of the input $x_i$, weight $w_{ij}^{(1)}$ and the bias $b_j^{(1)}$ as

$$ u_j = \sum_i x_i w_{ij}^{(1)} + b_j^{(1)}. $$ (2)

Here, $f(x)$ is the activation function and we adopt the sigmoid function, which is used commonly in the field of ANNs, given by

$$ f(u_j) = \frac{1}{1 + \exp(-u_j)}. $$ (3)

Concerning the output layer, an output $z_k$ is written as

$$ z_k = g(v_k), $$ (4)

where $v_k$ is given by a linear combination of the output from the hidden layer $y_j$, weight $w_{jk}^{(2)}$ and the bias $b_k^{(2)}$ as

$$ v_k = \sum_j y_j w_{jk}^{(2)} + b_k^{(2)}. $$ (5)

In this paper, we adopt the softmax function as the activation function in the output layer,

$$ g(v_k) = \frac{\exp(v_k)}{\sum_m \exp(v_m)}, $$ (6)

which is commonly used for classification problems. In our case, the values of $z_k$ for $k = 1$ and 2 represent the scores for the similarity of the source to a pulsar and non-pulsar, respectively.

Here we note that, although our network includes only one hidden layer, any functional form could be approximated as long as non-linear functions are used as the activation functions and as long as the hidden layer consists of a sufficient number of neurons. This fact is known as the universal approximation theorem (Cybenko 1989; Hornik 1991).

### 3.2 Training

Appropriate values of the network parameters (the weights and biases) are found by minimizing the loss function (or the cost function) and this process is called ‘training’. The loss function characterizes the difference between $z_k$ obtained from the network and the correct value $t_k$. In the classification problem, the cross-entropy error is often used and is defined as

$$ CE = -\frac{1}{N} \sum_n \sum_k t_{n,k} \log z_{n,k}. $$ (7)

where $n = 1, \ldots, N$ is the number of training data. In the process of training, we need to avoid ‘overfitting’, where a network is too closely fitted to the training data. There are several methods to suppress overfitting, and we adopt the weight decay method for our ANNs. The weight decay imposes a penalty on the weights and the loss function is given by the sum of the cross-entropy error and the squared weights,

$$ L = -\frac{1}{N} \sum_n \sum_k t_{n,k} \log z_{n,k} + \frac{1}{2} \lambda \left( \sum_{ij} |w_{ij}^{(1)}|^2 + \sum_{jk} |w_{jk}^{(2)}|^2 \right). $$ (8)

where $\lambda$ is a hyper-parameter called the weight decay term and represents the amount of the penalty. This parameter is determined by cross-validation, as explained in Section 4.2.

The network parameters are optimized by the momentum method described below. First, let $\xi(t) = [w_{ij}^{(1)}(t), b_j^{(1)}(t), w_{jk}^{(2)}(t), b_k^{(2)}(t)]$ be the network parameters at $t$th step of training. In the next step $t + 1$, they are updated as

$$ v(t + 1) = \mu v(t) - \eta \frac{\partial L}{\partial \xi} |_t, $$ (9)

$$ \xi(t + 1) = \xi(t) + v(t + 1), $$ (10)

where $\eta$ and $\mu$ are the learning rate and friction coefficient, which are fixed to 0.01 and 0.9, respectively. These are also hyper-parameters and can be determined in the same way as $\lambda$. However, they affect only the efficiency of the training and not the performance of the network. Further, the number of training steps is also a hyper-parameter and too many steps tend to induce overfitting. Thus, in addition to $\lambda$, we optimize the number of training steps by the method described in Section 4.2 fixing $\eta$ and $\mu$. Here, the initial values of $v(t)$ and $\xi(t)$ are set to $v(0) = 0$ and random values with a normal distribution of zero mean and standard deviation of 0.1, respectively. We evaluate the derivative of the loss function in equation (9) using the backpropagation algorithm (Rumelhart, Hinton & Williams 1986), which is a very computationally efficient method.

The training process is summarized as follows.

(i) Initialize the network parameters $[w_{ij}^{(1)}, b_j^{(1)}, w_{jk}^{(2)}, b_k^{(2)}]$.

(ii) Compute output $z_k$ with equations (1)-(6), and then the loss function (8).

(iii) Compute the derivative of the loss function with respect to the weights, and update the network parameters according to equations (9) and (10).

(iv) Go back to (ii) and iterate the number of times determined by the method explained in Section 4.2.

### 4 IMPLEMENTATION OF ANNS FOR PULSAR CANDIDATE SELECTION

#### 4.1 Input parameters

In this paper, we consider the following seven quantities as the inputs.

(A) Galactic longitude $l$ normalized to $[-1:1]$.

(B) Galactic latitude $b$ normalized to $[-1:1]$.

These quantities (A) and (B) represent the sky position in galactic coordinates. The majority of pulsars are expected to be located on the Galactic plane as can be seen from Fig. 1, as they are formed inside the Galaxy, while extragalactic non-pulsar objects should distribute uniformly in the sky. Thus, these quantities are potentially informative for pulsar candidate selection.

(C) Absolute value of galactic latitude $|b|$ normalized to $[0:1]$. We consider this as an alternative to (B) because, as mentioned above, pulsars are located near the Galactic plane in the sky so that $|b|$ rather than $b$ may be more useful.

(D) Logarithmic TGSS total flux (mJy) normalized so that the mean value is 0 and the standard deviation is 0.5.

(E) Logarithmic NVSS total flux (mJy) normalized in the same way as (D). Here, objects below the detection limit of the NVSS are given an upper-limit value of 2.5 mJy.

(F) Spectral index $\alpha$ normalized in the same way as (D).
As we saw in Fig. 3, pulsars tend to have steep spectra (Ivezić et al. 2002; de Gasperin et al. 2018; Jankowski et al. 2018). Although the spectral index (F) is a direct measure of the steepness, the pair of quantities (D) and (E) have more information than the index and these are adopted in our fiducial network. They are specific to surveys we use in the current paper, but the fluxes at different frequencies could also be used if other radio surveys are used. Note that we assume a single power law, but some pulsars have spectral turnover at \( O(10) \) MHz and the spectra are described by a broken power law (Bilous et al. 2016; Murphy et al. 2017).

(G) Source compactness \( C \) normalized to \([-1:1]\). This is defined in de Gasperin et al. (2018) as

\[
C = \frac{1.071 + 2 \sqrt{0.038^2 + 0.39^2 (S_{\text{peak}}/\sigma_1)^{−3.8}}}{S_{\text{peak}}},
\]

where \( S_{\text{total}}, S_{\text{peak}} \) and \( \sigma_1 \) are the total flux, peak flux and local rms noise of the TGSS, respectively. Pulsars with radii of a few tens of km are completely point sources, while non-pulsar objects can have much more extended structures than pulsars. However, there is little difference between pulsars and non-pulsars, as seen in Fig. 4. Nevertheless, we consider this feature as it could be correlated with other features.

Then, four sets of the above quantities are taken as input parameters:

(i) Case 1, (A), (B), (D), (E) and (G);
(ii) Case 2, (A), (B), (F) and (G);
(iii) Case 3, (A), (C), (D), (E) and (G);
(iv) Case 4, (A), (C), (F) and (G).

Here, Case 1 is our fiducial set and uses original quantities, rather than derived quantities such as (C) and (F). We set the number of neurons in the hidden layer as twice that of the input layer as our fiducial set-up. Thus, the input, hidden and output layers have five, ten and two neurons for Cases 1 and 3, and four, eight and two neurons for Cases 2 and 4, respectively. Later, we also investigate the networks with five and 15 neurons in the hidden layer for Case 1 (see Section 6).

4.2 Determination of hyper-parameters and performance test

In order to construct ANNs, we need to fix the values of hyper-parameters: the weight decay term \( \lambda \) and the number of training steps. In this subsection, we describe the method to determine these hyper-parameters following Eatough et al. (2010).

First, a subset is selected randomly from the whole data \((x_i, t_i)\). Here, the size of the subset is typically 10 per cent of all the data. The subset and remainder are called validation data and training data, respectively. Then, for a fixed value of \( \lambda \), the ANN is trained with the training data. At each step of training, the ANN is applied to the validation data and the cross-entropy error is calculated between the output from the ANN and the correct value of \( t_i \) and the error is minimum. We repeat this process, varying the value of \( \lambda \), until the error is minimum. We repeat this process, varying the value of \( \lambda \), and we choose both \( \lambda \) and the number of steps by comparing the minima of the cross-entropy error. We vary the value of \( \lambda \) in the range of \(-10 \leq \log_{10} \lambda \leq -2\) and we also consider the case of \( \lambda = 0 \). Finally, the ANN is trained once again with all data and hyper-parameters determined in the above way. The resultant ANN is now ready to be applied to unidentified radio sources to judge if they are likely to be a pulsar or not. It should be noted that time-domain observations are necessary to confirm whether the pulsar candidates are really pulsars or not.

In this paper, we not only apply our ANNs to unidentified sources but we also demonstrate the performance of our methodology. To do the performance test as well as cross-validation, we need to divide the data into three subsets: training data, validation data and test data. In our performance test, we first construct ANNs with training and validation data in the above way, and then the ANNs are applied to the test data. We repeat this process basically ten times, changing the choice of the data sets randomly. Consequently, we construct ten independent ANNs for each case. Note that training, validation and test data are chosen randomly every time. Finally, the performance is statistically checked and this process is commonly called the bootstrap method.

As we stated in Section 2, 127 pulsars and 13 166 non-pulsars were identified and these can be used as training, validation and test data. Although the ratio of the training data between pulsars and non-pulsars is imbalanced, we use 10 000 non-pulsars for training. We also study cases with 200 and 1000 non-pulsars in order to see the effect of the imbalance of the training data sets. The numbers of pulsars and non-pulsars in training, validation and test data are summarized in Table 1.

| Total | Training | Validation | Test |
|-------|----------|------------|------|
| Pulsar | 127      | 107        | 10   | 10 |
| Non-pulsar objects | 13 166 | 10 000 | 100 | 1000 |

5 RESULTS

5.1 Performance test

First, we show the results of the performance tests of our ANNs. Fig. 5 represents the distribution of hyper-parameters determined by the method mentioned in Section 4.2 for ten realizations of Case 1. The number of training steps is in the range of \([10^5, 10^6]\), while the weight decay term \( \lambda \) is scattered in a wide range below \(10^{-4}\).
Hyper-parameters are determined to optimize the network to given training data and the variation of hyper-parameters is attributed to the variation of training data, which are chosen randomly for each realization. Consequently, the performance of ANNs also varies among the ten realizations. We discuss the average performance below.

Next, we show the results of the performance test of trained networks. The outputs of our ANNs are the scores, $z_1$ and $z_2$, which represent the similarity to a pulsar and non-pulsar, respectively, and the sum is normalized to unity. Fig. 6 shows the histogram of $z_1$ of the test data obtained from all of the ten realizations for Case 1. As can be seen, the value of $z_1$ is almost zero or unity for most objects.

To determine the criterion of $z_1$ that classifies objects into the pulsars or non-pulsars, we use the following evaluation measures:

Recall = $\frac{TP}{TP + FN}$ \hspace{1cm} (12)

Precision = $\frac{TP}{TP + FP}$ \hspace{1cm} (13)

F1-score = $\frac{2 \times Recall \times Precision}{Recall + Precision}$ \hspace{1cm} (14)

Here, $TP$, $FN$ and $FP$ stand for true positives, false negatives and false positives, which represent the numbers of pulsars classified as pulsars, pulsars classified as non-pulsars and non-pulsars classified as pulsars, respectively. Fig. 7 shows the averaged Recall, Precision and F1 score over ten realizations as a function of the criterion of $z_1$, $z_{1c}$ for Case 1. Although the F1 score is the largest at around $z_{1c} = 0.3$, it is almost flat between 0.1 and 0.9 and drops sharply for $z_{1c} > 0.9$. Fig. 7 also shows the number of pulsar candidates that are obtained by applying the trained ANN to the unidentified objects in the de Gasperin catalogue (see Section 5.4). Because non-pulsar objects are considered to dominate unidentified objects, we choose the threshold to be $z_{1c} = 0.9$, which gives a high value of Precision and a relatively small number of pulsar candidates. The situation is similar for other cases, so we take $z_{1c} = 0.9$ as the criterion of pulsar candidates for all cases hereafter.

Table 2 shows the mean and standard deviation of the above three evaluation measures over ten realizations for the fiducial and variant ANNs.

Fixing the number of non-pulsar training data to 10,000, Case 3 has the largest average evaluation measures. Thus, Case 3 would be the best ANN of the four. Although there are uncertainties in the evaluation measures over the networks, it can also be seen from the comparison of the four cases that the absolute value of galactic latitude is a better input than the galactic latitude itself, while individual fluxes of TGSS and NVSS are better than the spectral index. Finally, comparing Case 1 with different numbers of non-pulsar training data, it is seen that its increase results in decreasing Recall, but increasing Precision and smaller variance among the networks. The network in the case of the non-pulsar training data of 200 would be the best in terms of F1 score, but high Precision (which means that the ratio of non-pulsars classified as incorrectly pulsars is small) is also very important because the number of non-pulsar objects should be much larger than that of pulsars as mentioned above.

Another common measure of the effectiveness of ANNs is the area under the receiver operating characteristic (ROC) curve (AUC). The ROC curve is a plot of the Recall (also known as the true positive rate) against the false positive rate (FPR) at various values of $z_{1c}$. Here, the FPR is given by $FPR = FP/(FP + TN)$, where $TN$ stands for true negatives. Noting that both the Recall and FPR are in a range of [0, 1], the AUC is defined as the area that is surrounded by the ROC curve and two lines of Recall = 0 and FPR = 1. The AUC can take a value from zero to unity, and a classifier with a larger value of the AUC is considered to be more powerful. The average values of the AUC are also shown in Table 2. These values are generally very close to unity and show the high performance of the trained ANNs.

### 5.2 Interpretation of weights

In this subsection, we try to interpret the behaviour of the weights and understand the interior of the trained ANNs. To do this, we neglect the activation functions for simplicity. In this approximation, the output from the hidden layer, equation (2), is given by

$$y_j = a \sum_i x_i w_{ij}^{(1)} + b_j^{(1)} \tag{15}$$

and by substituting this into equation (5), we obtain

$$v_k = a \sum_i x_i w_{ik}^{(2)} + b_k^{(2)} \tag{16}$$
Table 2. Recall, Precision and F1 score with a pulsar-candidate criterion of $z_1 \geq 0.9$, average value of the AUC and the number of pulsar candidates obtained by applying the trained ANN to the unidentified objects in the de Gasperin catalogue with $z_1 \geq 0.9$.

| Input | Case 1 | Case 2 | Case 3 | Case 4 |
|-------|--------|--------|--------|--------|
| Number of non-pulsar training data | 200 | 1000 | 10 000 | 10 000 | 10 000 |
| Recall (%) | 85.0 ± 14.9 | 79.5 ± 16.1 | 71.0 ± 17.9 | 52.0 ± 16.9 | 75.0 ± 14.3 | 65.0 ± 19.0 |
| Precision (%) | 92.6 ± 12.3 | 96.9 ± 10.9 | 96.1 ± 6.32 | 95.4 ± 7.47 | 98.8 ± 3.95 | 98.3 ± 5.27 |
| F1 score (%) | 87.9 ± 11.7 | 86.6 ± 13.0 | 80.3 ± 14.5 | 65.3 ± 17.9 | 84.6 ± 9.39 | 76.4 ± 13.4 |
| Averaged AUC | 0.974 | 0.968 | 0.976 | 0.967 | 0.973 | 0.989 |
| Number of candidates | 20 971 | 52 615 | 2 436 | 3 765 | 11 675 | 3 109 |

Here, $a \sim 0.2$ is the coefficient of the linear function, and $w_{ik}$ is the product of two weight matrices,

$$w_{ik} = \sum_j w_i^{(1)} w_j^{(2)}.$$  \hspace{1cm} (17)

Here, we ignore the bias $b_i$ since we focus on the behaviour of the weights in the networks. We can sum up the weights with respect to the hidden layer and the input layer is connected with the output layer directly by approximating the sigmoid function as the linear one. In the following, we study behaviour of this $w_{ik}$.

Because $w_{ik} = -w_{i1}$, we argue the behaviour of $w_{i1}$ only. Fig. 8 shows the mean and standard deviation of $w_{i1}$ for each case over ten realization. The weight of the longitude, $w_{i1}$, for every case is consistent with 0, which implies the longitude is not informative for the selection. However, the weights of the latitude are consistent with zero for Cases 1 and 2, which use the latitude itself, while those for Cases 3 and 4, which use the absolute value of the latitude, $|b|$, are significantly negative. This implies that $|b|$ is useful to select pulsar candidates and that pulsars tend to have small values of $|b|$; that is, pulsars are mostly located within the Galactic plane.

The negative and positive weights of the TGSS and NVSS fluxes for Cases 1 and 3 indicate that an object that is bright and dark in the TGSS and NVSS, respectively, tends to be selected as a pulsar. This is consistent with the fact that pulsars have steep spectra, as seen in Fig. 3. For Cases 2 and 4, this is seen as the negative weights of the spectral index.

The weight of the compactness is almost consistent with 0, but is slightly positive for all cases. Alghough the compactnesses of the pulsar and non-pulsar objects looks almost the same in Fig. 4, this might imply that the trained ANNs detect invisible correlation with other parameters.

Figure 8. Plot of the averaged $w_{i1}$ over realizations with the error bar for each case.
Thus, our simple interpretation of weights is consistent with our understanding of the basic properties of pulsars and non-pulsars. However, it should be noted that, in the above interpretation, possible correlations between the input quantities are marginalized and only direct correspondence between the inputs and the pulsar score is investigated. Thus, even if the average weight is consistent with zero, it does not necessarily mean the corresponding input has no effect on the pulsar selection.

5.3 Missed objects

In this subsection, we show features of the ‘missed’ objects in the test process, focusing on Case 1. Here, the missed pulsars represent true pulsars in the test data, which have $z_1 < 0.9$ and, consequently, were not selected as pulsar candidates. Conversely, missed non-pulsars represent true non-pulsars which have $z_1 \geq 0.9$ and, consequently, are selected as pulsar candidates. As explained above, we have ten realizations for each case and each ANN is tested with ten pulsars and 1000 non-pulsars (see Table 1). Thus, the total numbers of pulsars and non-pulsars in the test data are 100 and 10 000, respectively.

Among the test data, 29 pulsars and three non-pulsars were missed through ten realizations and the fraction of wrong selection is 29 per cent (false negative rate) and 0.03 per cent (false positive rate), respectively. The former fraction may look large, and this is partly because the criterion of classification is rather high ($z_{1c} = 0.9$). However, it is important to suppress the latter as low as possible, rather than the former. This is because most unknown objects are considered to be non-pulsars so that only a small value of the false positive rate can substantially increase the number of false pulsar candidates, which makes pulsar search in the time-domain very inefficient. Thus, we accept this relatively high false negative rate.

Fig. 9 shows the distribution of the missed pulsars and non-pulsars in galactic coordinates. The filled symbols represent objects that were missed multiple times. Although most pulsars are located in the Galactic plane, as seen in Fig. 1, many of the missed pulsars distribute roughly uniformly, which indicates that pulsars at high latitudes are more likely to be missed. However, because the number of missed non-pulsars is very small, it is not possible to discuss their distribution.

Fig. 10 shows the scatter plot of the missed objects in a plane of logarithmic TGSS and NVSS fluxes. We see that two missed non-pulsars, one of which is missed twice, are located out of the main region of the non-pulsar population. However, while many of missed pulsars have large values (small absolute values) of the spectral index, the steepest missed pulsars have spectral indices of about $-1.5$. Although there are more known pulsars than known non-pulsars around $\alpha \sim -1.5$, the number of non-known non-pulsars is not negligible ($\sim 10$). Therefore, $\alpha \sim -1.5$ would be the boundary of the classification and this is why some pulsars with $\alpha \sim -1.5$ are missed.

5.4 Applying ANNs to unidentified objects

We apply our trained ANNs to the unidentified objects in the de Gasperin catalogue. In this application, we use the networks of all cases individually. We choose training and validation data randomly, determine the hyper-parameters by the method mentioned in Section 4.2, train the network with those hyper-parameters, and then apply the trained network to the 456 866 unidentified objects. The number of pulsar candidates with $z_1 \geq 0.9$ for each case is shown in Table 2.

Comparing Case 1 ANNs, which were trained with different numbers of non-pulsars (200, 1000 and 10 000), the number of candidates is smallest for the network with the 10 000 non-pulsar training samples. However, comparing the four cases with 10 000 non-pulsar training samples, Case 1 has the smallest number of candidates, while Precisions are comparable within the statistical errors. Because, as stated before, the unidentified objects would be dominated by non-pulsars, we regard Case 1 with 10 000 non-pulsar training samples as the most effective ANN.

Comparing individual candidates of Cases 1–4, we find that, among the 2436 pulsar candidates of Case 1, 2047 (84 per cent), 1996 (82 per cent) and 819 (33.6 per cent) are common with Cases 2, 3 and 4, respectively. The similarity is relatively low for Case 4 compared with Cases 2 and 3. This would be because, for Case 4, more inputs are replaced from Case 1 compared with Cases 2 and 3.

Next, we describe the candidates of Case 1 in more detail. Fig. 11 shows the distribution of the known pulsars, known non-pulsars and 2436 pulsar candidates in the sky. The candidates are mainly located on the Galactic plane, but some of them are at high latitudes. This
distribution seems to be biased by the SDSS-surveyed and non-observed areas especially in the upper right area ($-180^\circ \leq l \leq -120^\circ$ and $0^\circ \leq b \leq 30^\circ$) of Fig. 11, where fewer candidates are distributed than in the upper left area ($120^\circ \leq l \leq 180^\circ$).

Fig. 12 shows a scatter plot of the known pulsars, known non-pulsars and 2436 candidates on a plane of the logarithmic TGSS and NVSS fluxes. The distribution of candidates mostly overlaps with that of known pulsars, with small NVSS fluxes and steep spectral indices.

To show the validity of our method, we checked if our candidates include newly found pulsars and candidates in Maan et al. (2018) and Frail et al. (2018). We found that our candidates cross-match 21 of 25 candidates in Maan et al. (2018) and all new pulsars and candidates in Frail et al. (2018) contained in our unidentified catalogue. Thus, our ANNs work very well for the selection of pulsar candidates.

Let us discuss the effect of the spatial bias of the training data, especially non-pulsars. As we saw in Fig. 1, the non-pulsar distribution is biased due to the limited area surveyed by the SDSS. In order to investigate the effect of the spatial bias, we performed the same analysis with 3639 non-pulsars sampled spatially uniformly in the de Gasperin catalogue area. We used 3500 for training and 100 for validation out of 3639 non-pulsars. Consequently, we obtained 3636 candidates ($z_1 \geq 0.9$) by applying the ANN to the unidentified objects. The spatial distribution of the training data and pulsar candidates is shown in Fig. 13. Compared with Fig. 11, the distribution of the candidates extends to the outside of the Galactic plane, although the concentration on the plane can still be seen. This implies that the spatial distribution of candidates is affected by the spatial bias of the training data. Further study on this effect is beyond the scope of the current work and will be presented elsewhere.

6 SUMMARY AND DISCUSSION

We applied ANNs for efficient selection of pulsar candidates from continuum surveys. From the input quantities such as radio fluxes, sky position and compactness, ANNs were constructed to output a score that represents a degree of likelihood for an object to be a pulsar. We demonstrated ANNs based on existing survey data by the TGSS and the NVSS and tested their performance by using the input parameters and the number of training data. Finally, we obtained pulsar candidates by applying the trained ANNs to unidentified radio sources. For the validation, our candidates should be confirmed if they are truly pulsars by using time-domain observations. This is ongoing and will be presented elsewhere.

We have evaluated our trained networks with test data that consist of known pulsars and known non-pulsars. As a result, it is indicated that the trained networks have high classification performance and Precision (which is the ratio of the number of pulsars classified correctly as pulsars to the number of objects classified as pulsars) is basically higher than 95 per cent. This implies that the fraction of non-pulsars among pulsar candidates is less than 5 per cent, although non-pulsar objects are considered to dominate radio point sources. Our ANNs are also tested with pulsar candidates and newly found pulsars in Maan et al. (2018) and Frail et al. (2018). We found that our candidates generated by the Case 1 ANN include 21 of 25 candidates in Maan et al. (2018) and all new pulsars and candidates in Frail et al. (2018) contained in our unidentified catalogue. Thus, our ANNs work very well for the selection of pulsar candidates.

To show the validity of our method, we checked if our candidates include newly found pulsars and candidates in Maan et al. (2018) and Frail et al. (2018). We found that our candidates cross-match 21 of 25 candidates in Maan et al. (2018), while three of five new pulsars and three of five candidates in Frail et al. (2018) are cross-matched with our candidates. In fact, these four non-cross-matched objects in Frail et al. (2018) are not included in our catalogue of unidentified objects. Thus, our ANN selects all of the new pulsars and candidates in Frail et al. (2018) included in our catalogue, which shows the effectiveness of our method for pulsar candidate selection.
As mentioned in Section 3.1, ANNs with a larger number of neurons in the hidden layer are expected to work better, while the computational cost becomes larger. Here, we briefly compare the performance of three Case 1 ANNs with 5, 10 and 15 neurons in the hidden layer. Table 3 shows Recall, Precision, F1 score, the average AUC and the number of candidates \(z_{1c} \geq 0.9\) for the three ANNs. These characteristic numbers are almost within the statistical fluctuations, while the case with 5 neurons has slightly low performance. Hence, we conclude that it is reasonable to set the number of neurons in the hidden layer to 10 as our fiducial network.

Besides ANNs, there are several other machine learning methods such as the support vector machine and random forest, which exhibit excellent performance in pattern recognition. It is worth applying other methods to the selection of pulsar candidates and comparing the results. This is beyond the scope of the current paper and will be pursued elsewhere in future.

In this work, we used objects in the de Gasperin catalogue cross-matched with the ATNF pulsar catalogue and the MILLIQUAS catalogue as our training data. Although these are currently the largest available catalogues, the number of cross-matched pulsars is relatively small and we need a further analysis with future larger catalogues to consider if ANNs are effective for the selection of pulsar candidates.

Other observable quantities, such as the rotation measure and polarization fraction, could be useful as inputs. We did not adopt these because the number of radio sources with them is currently very limited. If we can have a sufficient number of samples with polarization data, as is expected in future large surveys, they will make ANNs more effective and narrow the pulsar candidates down further.

### ACKNOWLEDGEMENTS

We thank Shiro Ikeda, Shuhei Mano, Shinto Eguchi and Hayato Shimabukuro for useful discussions. NY and SY are financially supported by the Grant-in-Aid from the Overseas Challenge Program for Young Researchers of JSPS. KT is partially supported by a Grant-in-Aid from the Ministry of Education, Culture, Sports, and Science and Technology (MEXT) of Japan (Nos 15H05896, 16H05999 and 17H01110), and Bilateral Joint Research Projects of JSPS. SY is supported by JSPS KAKENHI Grant Numbers JP16J01585.

### REFERENCES

Abolfathi B. et al., 2018, *ApJS*, 235, 42
Banerji M. et al., 2010, *MNRAS*, 406, 342
Berti E. et al., 2015, *Class. Quantum Gravity*, 32, 3001
Bethapudi S., Desai S., 2018, *A&A*, 23, 15
Bilous A. et al., 2016, *A&A*, 591, A134
Collister A. A., Lahav O., 2004, *PASP*, 116, 345
Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, *AJ*, 1065, 1693
Connor L., van Leeuwen J., 2018, *AJ*, 156, 256
Cordes J. M. et al., 2006, *ApJ*, 637, 446
Cybenko G., 1989, *Mathematics of Control, Signals and Systems*, 2, 303
Dai S., Johnston S., Bell M. E., Coles W. A., Hobbs G., Ekers R. D., Lent E., 2016, *MNRAS*, 462, 3115
de Gasperin F., Intema H. T., Frail D. A., 2018, *MNRAS*, 474, 5008
Eatough R. P., Molkenthin N., Kramer M., Noutsos A., Keith M. J., Stappers B. W., Lyne A. G., 2010, *MNRAS*, 407, 2443
Flesch E., 2015, *PASA*, 32, 10
Folkes S. R., Lahav O., Maddox S. J., 1996, *MNRAS*, 283, 651
Foster R. S., Backer D. C., 1990, *ApJ*, 361, 300
Frail D. A. et al., 2018, *MNRAS*, 475, 942
George D., Huerta E. A., 2018, *Phys. Lett. B*, 778, 64
Guo P., Fuqing D., Wang P., Yao Y., Xin X., 2017, preprint (arXiv:1711.0339)
Han J. L., Ferriere K., Manchester R. N., 2004, *ApJ*, 610, 820
Hewish A., Bell S. J., Pilkington J. D. H., Scott P. F., Collins R. A., 1968, *Nature*, 217, 709
Hornik K., 1991, *Neural Networks*, 4, 251
Intema H. T., Jagannathan P., Mooley K. P., Frail D. A., 2017, *A&A*, 598, A78
Ivezić Z. et al., 2002, *AJ*, 124, 2364
Jankowski F., van Straten W., Keane E. F., Bailes M., Barr E. D., Johnston S., Kerr M., 2018, *MNRAS*, 473, 4436
Jenet F. et al., 2009, preprint (arXiv:0909.1058)
Keane E. F. et al., 2015, preprint (arXiv:1501.00056)
Kramer M., Champion D. J., 2013, *Class. Quantum Gravity*, 30, 4009
Kramer M. et al., 2006, *Science*, 314, 97
Lattimer J. M., Prakash M., 2004, *Science*, 304, 536
Levin L. et al., 2018, in Proc. IAU Symp. 337, Pulsar Astrophysics the Next Fifty Years. Kluwer, Dordrecht, p. 171
Maan Y., Bassa C., van Leeuwen J., Krishnakumar M. A., Joshi B. C., 2018, *ApJ*, 864, 16
Manchester R. N. et al., 2001, *MNRAS*, 328, 17
Manchester R. N., Hobbs G. B., Teoh A., Hobbs M., 2005, *AJ*, 129, 1993
Manchester R. N. et al., 2012, *PASA*, 30, e017
Morello V., Barr E. D., Bailes M., Flynn C. M., Keane E. F., van Straten W., 2014, *MNRAS*, 443, 1651
Murphy T. et al., 2017, *PASA*, 34, e020
Naim A., Lahav O., Sorde L., Jr, Storrie-Lombardi M. C., 1995, *MNRAS*, 275, 567
Rumelhart D. E., Hinton G. E., Williams R. J., 1986, *Nature*, 323, 533
Samui S., Samui Pal S., 2017, *New Astron.*, 51, 169
Schnitzeler D. H. F. M., 2012, *MNRAS*, 427, 664
Shimabukuro H., Semelin B., 2017, *MNRAS*, 468, 3869
Smits R., Kramer M., Stappers B., Lorimer D. R., Cordes J., Faulkner A., 2009, *A&A*, 493, 1161
Storrie-Lombardi M. C., Lahav O., Sorde L., Jr, Storrie-Lombardi L. J., 1992, *MNRAS*, 259, 8
Vanzella E. et al., 2004, *A&A*, 423, 761
Wang H. et al., 2019, *Science China Physics, Mechanics, and Astronomy*, 62, 95907
Zhu W. W. et al., 2014, *ApJ*, 781, 117

This paper has been typeset from a \TeX/\LaTeX file prepared by the author.