New Astronomy 85 (2021) 101561

Contents lists available at ScienceDirect

New Astronomy

journal homepage: www.elsevier.com/locate/newast

# Learning from imbalanced pulsar data by combine DCGAN and PILAE algorithm

Mohammed A.B. Mahmoud [*,a], Ping Guo [a,b]

[a] School of Computer Science and Technology, Beijing Institute of Technology, China
[b] School of Systems Science, Beijing Normal University, China

ARTICLE INFO

*Keywords:*
Pulsar
DCGAN
SMOTE
Pseudoinverse learning (PIL)

ABSTRACT

A pulsar is a rapidly rotating neutron star and transmits periodic oscillations of power to the earth. We introduce a novel method for pulsar candidate classification. The method contains two major steps: (1) make strong representations for pulsar candidate in the image domain by extracting deep features with the deep convolutional generative adversarial Networks (DCGAN) and (2) develop a classifier defined by multilayer perceptron (MLP) neural networks trained with pseudoinverse learning autoencoder (PILAE) algorithm. We utilized the synthetic minority over-sampling technique (SMOTE) to handle the imbalance in the dataset. We report a variety of measure scores from the output of the PILAE method on datasets utilized in the experiments. The PILAE training process does not have to determine the learning control parameters or indicate the number of hidden layers. Therefore, the PILAE classifier can fulfil superior execution in terms of training effectiveness and accuracy. Empirical results from the high time resolution universe (HTRU) mid-latitude dataset, MNIST dataset and CIFAR-10 have demonstrated that the presented framework achieves excellent results with other models and reasonably low complexly.

## 1. Introduction

Learning from imbalanced data is a typical and challenging scenario in today's machine learning applications (He and Garcia, 2008). The imbalance refers to the case that the distribution of different data classes is significantly unequal, where traditional learning models will fail to provide accurate predictions. This situation occurs commonly in areas such as security (e.g. spammer detection, where spammers are much fewer in number than normal users) and medicine (e.g. cancer diagnosis, where cancerous patients are much fewer in number than healthy ones). In this paper, we investigate the problem of learning from imbalanced data, in a novel application of pulsar candidate selection. The task to select a small minority of positive data samples as pulsar signals, from a large set of radiation signal data, where other signals are noise, becomes the first challenge. In astronomy, searching for real pulsars is important because they are useful in many astronomy tasks, such as the gravitational wave detection and spacecraft navigation. Therefore, the aim of the paper is to propose a novel framework for classification problems with imbalanced data in general cases, as well as the verification in the specific application of pulsar candidate selection.

Pulsar explores from present-day radio overviews include filtering through candidates distinguished by pulsar search channels, comprising of either single-beat searches or periodicity. These pulsar-explore algorithms are regularly computationally over the top expensive (despite enhancements to their speed and affectability e.g. Cameron et al., 2017; Smith, 2016). The yield of these channels has a large number of candidates, out of which only a small portion comprises pulsars with remaining candidates emerging from radio frequency interference (RFI) or different origins of noise (Keith et al., 2010). Significant numbers of these candidates are evidently investigated and physically screened by space specialists. For current creation pulsar studies, it takes approximately 1–300 s to vet every candidate (Eatough et al., 2010). More than 70,000 h would be needed to examine the million or so candidates. Such a manual visual classification of the pulsar candidates becomes difficult during the Square Kilometre Array (SKA) period, where we hope to find approximately 20,000 new pulsars (Kramer and Stappers, 2015). Despite the fact that the filtering of genuine pulsar signals from noise can be encouraged with graphical utilities, for example in Keith et al. (2009), these signals have impediments, and one is inclined to commit errors (Eatough et al., 2010; Bates et al., 2012). Accordingly, to augment the location of pulsars in the SKA period, the computational expenses amid every one of the means of the pulsar-explore channel ought to be

* Corresponding author.
*E-mail address:* mahmoud.mohammed@bit.edu.cn (M.A.B. Mahmoud).

https://doi.org/10.1016/j.newast.2020.101561
Received 16 August 2020; Received in revised form 1 October 2020; Accepted 9 December 2020
Available online 13 December 2020

decreased and human intercession should be limited at each progression, comprising the second challenge. An imperative advance in this procedure is robotized the sifting of pulsar candidate obtained from pulsar-explore channels however much as could reasonably be expected.

Among the main approaches in solving the problem of imbalanced data, the sampling-based method is a competitive stream that is the domain of this paper. Studies have presented many base classifiers, in which the total classification execution is improved by utilizing a dataset that is balanced compared with an imbalanced one (Weiss and Provost, 2001; Laurikkala, 2001; Estabrooks et al., 2004). These outcomes legitimize the utilization of sampling techniques for imbalanced learning. Regarding synthetic sampling, the synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002) is a robust method that has produced an incredible number of achievement in different applications. The SMOTE method makes synthetic data dependent on the feature space likenesses between existing minority samples. SMOTE has been recently utilized for the classification of obscure point sources from the Fermi-LAT index (Abdo et al., 2013), as well as for classification of variable stars from Kepler Bass and Borne (2016). It has additionally been utilized in a pulsar search to handle the class imbalance issues (Devine et al., 2016). Two features selection methods which are Recursive Feature Elimination (RFE), and Grid Search (GS) were used to select as few as possible features of pulsar candidates with SMOTE algorithm for handling the imbalance issue in the dataset (Lin et al., 2020). Hence, the first challenge can be taken care of utilizing SMOTE calculation.

In this paper, we consider the case of pulsar candidate in the image domain. Fig. 1 illustrates the pulsar to non-pulsar candidate's image in terms of candidate sub-bands, which is two-dimensional array containing the candidate sub-bands or phase-frequency diagram. Recently, learning strategies utilizing deep-learning features have been efficiently connected to resolve multiple computer vision issues (Lee et al., 2009; 2011; Bengio et al., 2013) that prove troublesome. The core idea of deep learning is using forthright representations to learn hierarchical feature representation and then proceeding to develop more complex ones from the preceding level. Compared with the handcrafted models, deep learning design is capable of encoding data that is multilevel from their initial simple nature to a more intricate one. Therefore, for feature learning, this method is highly encouraging because (1) it does not necessitate ground truth; (2) It induces intricate non-linear connections using deep architecture that is hierarchical in nature; (3) It does not depend on chosen handcrafted features but rather is completely driven by data and (4) owing to its trained hierarchical deep network, it is capable of effectively and rapidly determining the feature representation of low-level images. Guo et al. (2019a) utilized the DCGAN model for feature extraction and synthesis of the minor class to overcome the issue of imbalance in the HTRU dataset followed by L2-SVM classifier.

A pseudoinverse learning algorithm (PIL) Guo and Lyu (2004); Wang et al. (2018); Deng et al. (2019), it is a multilayer perceptron (MLP) learning algorithm composed of stacked generalization connected such that it dominates the neural networks' (NNs) degradation predictive accuracy. Its structure possesses the identical number of hidden neurons as the number of samples that are to be learned. PIL overcomes learning errors by performing the addition of hidden layers. It had been fully automated, feed-forward and does not contain critical user subordinate parameters, for example, learning rate, the maximum epoch and momentum constants. PIL has been demonstrated to be an efficient algorithm and by far much better than the standard back propagation (BP) and other algorithms of gradient descent. Wang et al. (2017) asserted that PILAE was a fully and fast automated framework that uses deep neural networks to train stacked autoencoders (Hinton and Salakhutdinov, 2006). PILAE trains the stacked autoencoder by embracing the PIL algorithm with a low-rank approximation; PILAE is not a gradient descent method and does not have the shortcoming of gradient vanishing. It also does not have the problem of saturation activation as a matrix that is multiplied by its pseudoinverse. Thus PILAE can be used for handling the second challenge.

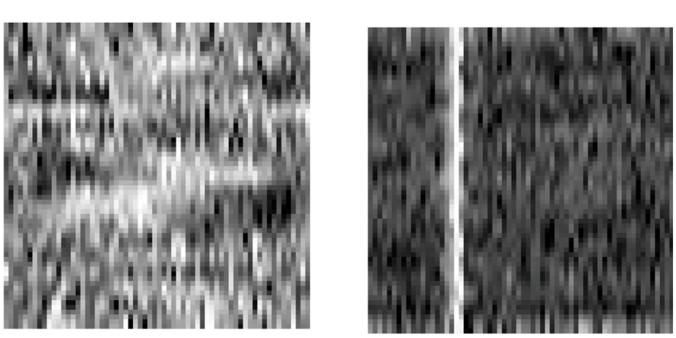

Fig. 1. An illustration of non-pulsar and pulsar candidate representations. The left and right images show a 2D graphical representation of a phase frequency diagram for the non-pulsar and pulsar candidates, respectively.

Different from our previous work (Guo et al., 2019a) we propose a DCGAN-PILAE model that uses SMOTE algorithm to overcome the imbalance problem and the DCGAN model for extraction deep features in an unsupervised manner. In the PILAE classifier, weight parameters are calculated with the pseudoinverse solution and do not have to be modified further, thus having an important effect on the performance of the model. We investigate the performance of the introduced model on the high time resolution universe (HTRU) dataset.

The paper proceeds to evaluate Pulsar candidate classification related issued in Section 2. Section 3 provides a detailed description of the proposed model. Section 4 presents Implementation details, experimental results and the discussion. Section 5 concludes the paper.

## 2. Related work

Many algorithms have been employed for pulsar candidate classification. For example, an artificial neural network (ANN) was utilized in the pulsar selection by Eatough et al. (2010) to handle 16 million pulsar candidates acquired by reprocessing the Parkes multi-pillar dataset. Bates et al. (2012) additionally utilized an ANN in the HTRU dataset. They had the capacity to dismiss 99% of the noise candidates and identify 85% of the pulsars through a visually impaired investigation.

In Zhu et al. (2014), the authors utilized a blend of three diverse methods—support vector machine (SVM), convolution neural network (CNN) and ANN—in their recognition of image based on PICS algorithm. The PALFA dataset is utilized to train PICS (Pulsar Image-based Classification System) algorithm where it is tested on the dataset from the green bank north celestial cap survey (GBNCC). From the test set, PALFA had the capacity to rank 100% of the pulsars in the top 1% from all candidates, while 80% were positioned higher than any impedance occasions or noise. Lyon et al. (2013) contemplated the execution of different stream classifiers like VFDT (Hulten et al., 2001) on HTRU dataset. The authors demonstrated the vulnerability of the pulsar data to the imbalanced learning issue and how the imbalance seriously diminishes recall ratio of pulsar.

In Lyon et al. (2014) presented another classification method for the data that involves imbalance utilizing Hellinger distance measure, which they tested the HTRU dataset. The authors had the capacity to show that the employed method can successfully improve minority class rates of recall on a dataset that is imbalanced. Morello et al. (2014) utilized neural network in a pulsar positioning method named SPINN. This method had the ability to distinguish every pulsar in the HTRU dataset with 0.64% of false positive rate (FPR) and furthermore decreased the number of possibilities to check by up to four requests of extent.

Lyon et al. (2016) introduced another strategy for online sifting of pulsars utilizing a classifier called Gaussian Hellinger Very Fast Decision Tree. This strategy was able to process up to one million pulsars in few seconds and had rates near 98% of recall when the HTRU-1 and LOTAAS datasets were used; it concluded with >90% recall rates and <0.5% of FPR. Twenty new pulsars from the LOTAAS (Lyon et al., 2016) dataset were discovered using this strategy. Wagstaff et al. (2016) fused random forests algorithm in their radio transient discovery method, entitled V-FASTR. This approach had the capacity to consequently sift through realized occasion types (pulsars and noise) with 98.6% training accuracy and accomplished a 99–100% accuracy on test data.

Devine et al. (2016) utilized six diverse frameworks (e.g. random forests and SVM) to characterize scattered pulsar bunches in the second phase of their single-pulse seek framework. They used a dataset comprised of more than 300 pulsars and approximately 9600 noise pulsar utilizing perceptions from the Green Bank Telescope. They found that the most efficient learning method was the multiclass ensemble tree. Mohamed (2018) proposed strategy relied upon fuzzy KNN classifier that was trained on HTRU-1 dataset (Lyon et al., 2016) and accomplished accuracy of 97.8%. In Bethapudi and Desai (2018), the authors presented different machine learning methods such as Adaboost, GBC, and XGBoost for Pulsar candidate classification in which the SMOTE method was applied to overcome the class imbalance problem. But the key problem of this method is that the accuracy of radio frequency interference classification is very sensitive to feature selection.

Wang et al. (2019) traded the CNN in PICS algorithm by the ResNet model, which achieved a recall rate on the GBNCC dataset of 96%. The authors in Li et al. (2018) employed a various-leveled display for pulsar classification that involves collecting a significant number of prepared base classifiers. They used the PIL rather than gradient descent (GD) method for the introduced method training process. This technique had the ability for classification of the pulsar in both HTRU and PMPS-26k (Manchester et al., 2001) datasets with a recall rate of 95.74% and 87.50%, respectively. Wang et al., 2020 presented a swift model for eliminating the RFI in pulsar data. They used the PILAE to learn the RFI signatures and eliminate them from fast-sampled spectra, leaving real pulsar signals

In Yao et al. (2016), Yao et al. proposed Hierarchical Candidate-Sifting Model (HCSM) to deal with imbalance issue of three pulsar datasets (HTRU, HTRU-1 and LOTAAS 1) by asserting the cost of misclassified of positive samples and collecting numerous classifiers prepared with various weighting parameters. The authors accomplished recall rates of 97.49%, 84.52% and 100%, respectively for the utilized datasets.

Comparing to our previous work (Guo et al., 2019a; Li et al., 2018; Wang et al., 2020; Yao et al., 2016), we use the SMOTE method to overcome the imbalance problem in the dataset, DCGAN for feature extraction, and the PILAE for classification. In Guo et al. (2019a), we employed the two functions of the DCGAN that are feature extraction and synthesis of the minor class samples followed by the L2-SVM classifier.

## 3. Approach

In this section, we illustrate the pipeline of the suggested model. As shown in Fig. 2, the dataset is given to the DCGAN model for extracting features. Then, SOMTE is applied to synthesis the data, and finally, the PILAE algorithm is trained over these data also as features extraction and classifier. In the following subsections, we will demonstrate the concept of the DCGAN, SMOTE and PILAE. Our algorithm uses the same framework of DCGAN (Radford et al., 2015) without any amendments.

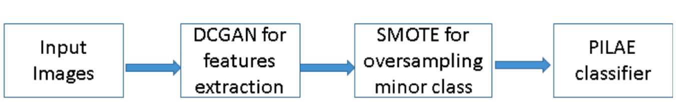


**Fig. 2.** The proposed model pipeline. First, there is features extraction using *D* network of DCGAN model. Second, the extracted features are given to SMOTE to synthesis the minor class (pulsar samples). Third, the PILAE algorithm is given the balanced data and trained also as features extraction and classifier to give the final test accuracy.

### 3.1. Feature learning

Inside our framework, DCGAN is utilized for the extracting features of a total of visible images. Produced from the GAN, DCGAN is suggested to entrench the GAN, which is designed from the convolutional composition to the training model. From that, DCGAN can make use the benefit of convolutional networks and achieve in the region of image processing and computer vision.

The DCGAN includes the generator predicated on a transposed convolutional composition and the discriminator predicated on a convolutional composition. The generator's role is to create a random sample, that is steadily changed by the correct execution of the image what we wish, predicated on the probability distribution and adversarial learning. The discriminator also distinguishes whether the insight image is genuine or fraudulent. In this framework, this technique is trained to raise the discriminate rate, using the fraudulent image that is nearly the same as the genuine image and produced by the generator model. It is worth mentioning that the generator must create an image with no memorization of prior images. When the generator makes an image predicated on the memorisation, it says that the generator just learns the one-on-one mapping method due to the overfitting. There is no image generation method predicated on features. Moreover, whenever we transfer to the latent space (an insight space of the generator), an even transition should be performed, rather than a sharpened transition.

Inside the proposed framework, we utilize the DCGAN model for learning the features of images. In the learning process, discriminator and generator are trained concurrently to produce a more exact discriminator.

### 3.2. Synthetic sampling with SMOTE

The SMOTE algorithm provides out an oversampling method of rebalancing the main training set. Rather than implementing a straightforward reproduction of the minority class samples, the main notion of SMOTE is to expose synthetic samples. This new data are established by interpolation among several minority class samples that are within a precise neighborhood. Hence, the task is reported to be centered on the 'feature space' alternatively than on the 'data space'; quite simply, the algorithm is depends on the ideals of the features and their relationship, instead of considering the data points altogether.

### 3.3. Pseudoinverse learning autoencoder (PILAE)

PIL Guo and Lyu (2004); Guo et al. (2019b); Feng et al. (2019) is a swift supervised learning algorithm for feedforward neural network; it depends on generalized linear algebra. PIL differs from techniques that are gradient based containing BP, in that PIL is not necessarily required to adapt affined variables such as momentum, learning epoch and step length; instead, these variables are normally hard selected by the users.

PIL is utilized to train SAE with layer-wise learning framework that is greedy based; PILAE is utilized as an SAE building block. The number of hidden neurons is specified by the equation in Wang et al. (2017). To enable the learning of data feature, the hidden neuron's number is a slightly larger than the input network rank's and in general smaller than the input vector's dimension. The rank of the input matrix is computed by the singular value decomposition method (SVD). Truncated SVD is specifically utilized to compute the input matrix pseudoinverse, which is then utilized as the encoder weight network. On the other hand, PIL is used to compute the decoder weight matrix. Furthermore, to reduce the degree of independence parameter, the decoder and encoder weights are tied up, allowing the weight of encoder network to be equal to the decoder weight that is transposed. The autoencoder is able to map the data to high- or low-rank approximation and is used to map the data to low-rank dimensions.

$\mathbf{X} \in R^{MxN}$ is a given training set, $\mathbf{X} = [x_1, x_2, \ldots, x_N]$ where the $i-th$ training sample represents as vector $x^i = [x^{(1)}, x^{(2)}, \ldots, x^{(m)}]^T$, we utilize the SVD method to compute the pseudoinverse of the input matrix,

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^{\mathbf{T}} \tag{1}$$

The pseudoinverse of the input matrix can be calculated depend on the output of the SVD:

$$\mathbf{X}^{+} = \mathbf{V}\Sigma^{'}\mathbf{U}^{\mathbf{T}}, \tag{2}$$

$\Sigma^{'}$ performs the diagonal matrix which is transposed that consisted of the mutual of nonzero elements in matrix $\Sigma$. The first hidden layer neuron numbers are adjusted to equal to the input matrix rank.

The number of hidden neurons are assigned as

$$p = \beta Dim(x), \beta \in (0, 1], \tag{3}$$

Dim is the function which gives the input matrix dimension, on the other hand $\beta$ can be an empirical parameter which is based upon the degree of the dimension to be reduced. This will assure that dimension reduction is boosted.

$\widehat{\mathbf{X}}^{+}$ is the low-rank approximation of the matrix $\mathbf{X}^{+}$ which calculating as follows:

$$\widehat{\mathbf{X}}^{+} = \widehat{\mathbf{V}}\Sigma^{'}\mathbf{U}^{\mathbf{T}}, \tag{4}$$

where $\widehat{\mathbf{V}}$ is shaped of the initial $p$ rows of a singular matrix $\mathbf{V}$.

Since the autoencoder's input data is approximately equal to the output, we append the restriction $\mathbf{X} = \mathbf{O}$, where $\mathbf{O}$ represents the output matrix. $\mathbf{Q_d}$ is the decoder weight matrix and $\mathbf{Q_e}$ is the encoder weight matrix, we assign $\mathbf{Q_e} = \widehat{\mathbf{X}}^{+}$. Though, the optimization objective function can rewrite as follows:

$$\min \| \mathbf{ZQ} - \mathbf{X} \|^{2}. \tag{5}$$

where the $\mathbf{ZQ}$ is the output of the last hidden layer and the weight matrix product. The following pseudoinverse approximate solution is used to solve the optimization problem:

$$\mathbf{Q} = \mathbf{Z}^{+}\mathbf{X}. \tag{6}$$

According to the work of Hoerl and Kennard (1970) that in the case of multiple linear regression where $b = Ax + \varepsilon$, if $A^TA$ is not a unit matrix, the least squares estimation is sensitive to the error in $A$, and $x$ isn't meaningful. consequently, the following equation is used to represent.

$$x = \left(A^TA + k\right)^{-1}A^Tb, k > 0. \tag{7}$$

where $k$ is regularization variable. When matrix column rank is full, the orthogonal projection method is utilized to solve the pseudoinverse:

$$\mathbf{Q}^{\mathbf{L}} = \left(\mathbf{Z}^{\mathbf{T}}\mathbf{Z}\right)^{-\mathbf{1}}\mathbf{Z}^{\mathbf{T}}\mathbf{X}. \tag{8}$$

The $\mathbf{Q}^{\mathbf{L}}$ is the weight matrix of layer $L$.

The weight decay regularization parameter is used to avert ill-posed problem:

$$\mathbf{Q} = \left(\mathbf{Z}^{\mathbf{T}}\mathbf{Z} + \mathbf{kI}\right)^{-\mathbf{1}}\mathbf{Z}^{\mathbf{T}}\mathbf{X}. \tag{9}$$

where $\mathbf{k}$ is regularization restriction which the user specifies, It could be determined with the form generated in Guo et al. (2003). In this manner, the value of $k$ is tune to get it the trade-off among the accuracy and generalization.

## 4. Experiments

### 4.1. Dataset

Three datasets are utilized to evaluate the proposed approach: (1) the HTRU medlat[1] (Morello et al., 2014); (2) the MNIST dataset[2] (LeCun et al., 1998) and (3) the CIFAR-10[3] (Krizhevsky and Hinton, 2009).

The HTRU medlat dataset is utilized to evaluate the proposed approach. The HTRU medlat is the first labeled candidate dataset produced that is overtly obtainable. It contains 1196 pulsar and 89,996 non-pulsar candidates.

The MNIST dataset contains 70,000 handwritten digital images of 0–9, where 60,000 images are utilized as training samples and the remaining 10,000 images are the test samples.

CIFAR-10 is an affirmed computer-vision dataset utilized for image recognition. It contains 60,000 color images consisting one of ten object classes, with 6000 images per class. *Data pre-processing* In the HTRU medlat dataset, the size of images is altered to 64x64 pixels through bicubic interpolation before DCGAN training. We performed our framework in MATLAB, utilizing the MatConvNet library (Vedaldi and Lenc, 2015) for our execution of DCGAN. Our experiments used a PC with a hardware configuration as follows: (1) CPU: Intel core i7-6800k; (2) Memory: 128 GB and (3) GPU: 1 x TITAN XP.

### 4.2. Evaluation on HTRU dataset

Every candidate in the dataset has 22 attributes come in one file of XML or PHCX format. We are using the sub bands attribute, which is a 2D array containing the candidate sub-bands, also termed a phase frequency diagram; after extraction of the phase frequency diagram for every candidate within the dataset, we get 2D grayscale image for every pulsar and non-pulsar (Fig. 1). The dataset has a ratio of pulsar to non-pulsar equal to 1:75, which making it high-class imbalance problem. Therefore, we over-sampled the positive samples (pulsar candidate) initially to avoid an imbalance of the dataset. We obtain a different ratio of pulsar to non-pulsar (within the range 1:2 to 1:4) to choose the best ratio. We found that the ratio of 1:2 is the best according to a result of the PILAE classifier.

Our approach has three steps. First, there is features extraction using DCGAN model: we feed the dataset images to DCGAN for extracting features. The fourth layer of the discriminator network is selected out as extractor features because the DCGAN is not used to perform classification but is instead employed to extract deep features. An 8192-dimensional space for each image is then formed by concatenating and flattening these features. The hyperparameters of DCGAN are set to the same work of Radford et al. (2015), after preliminary experiments, we found it is the best to set the learning rate to 1e3 with a batch size of 64 and we train for 15 epochs where each epoch take 42 s.

Second, the extracted features from DCGAN model are given to SMOTE to synthesis the minor class (pulsar samples). We use SMOTE to artificially balance the dataset (i.e. to have the ratio of 1:1 of minor to major). Because the ratio of 1:2 gave the best result, it became (after applying SMOTE) 1:1 where each class (pulsar and non-pulsar) has 2392 candidates. We perform SMOTE with a parameter of $U = 100$ and $F = 5$, where $U$ and $F$ are the amount of oversampling and nearest neighbors, respectively.

Third, the PILAE algorithm is given the balanced data and trained also as features extraction and classifier to give the final test accuracy. The dataset is split into training and testing with different ratios (80%:20%, 70%:30% and 60%:40% where the first number refers to training and the second to testing ratio).

[1] http://astronomy.swin.edu.au/~vmorello/
[2] http://yann.lecun.com/exdb/mnist/
[3] https://www.cs.toronto.edu/~kriz/cifar.html

*Evaluation measures*

To similarly rank the execution of the proposed strategy, we ascertain different execution scores; we have utilized seven distinct scores to measure the model execution as in Lyon et al. (2016). They are Accuracy, G-Mean, F-Score, Recall, Precision, Specificity and FPR. On binary classification problem such as our case, every component of the confusion matrix relates to TP, TN, FP and FN, which are characterized in Table 1.

Presently, we characterize every one of the seven scores used to assess the proposed method utilized in this paper.

**Accuracy** is the ultimate prominent score in any classification Problem, where the classes have approximately the same ratio of samples (balanced) and is defined by (TN + TP)/(FP + FN + TP + TN); the superior accuracy value is 1, denoting there are no misclassification of samples in the dataset, and 0 is the worst case.

**Recall** is given as TP/(TP+FN), depicts the positive samples ratio recovered in the test set, A perfect classifier will have a recall estimation of 1 and the lower rate for an outrageous classifier would be 0.

**Specificity** is defined as TN/(TN+FP), which is a ratio of how well the classifier can mark a negative class.

**G-Mean** id identified as $\sqrt{Specificity * Recall}$, and it is the geometric mean of Specificity and Recall. The worst and best values of G-Mean are 0 and 1, respectively.

**Precision** is identified as TP/(TP+FP). Precision is the ratio of TRUE class perceptions classified effectively to the absolute number of perceptions that were classified as positive. Similar to recall, even here the worst rate is 0 and the best rate is 1.

**F-score** is given as 2*(Precision*Recall)/(Precision + Recall), reflecting recall and precision together. F-score gives a mutual degree of accuracy and has the worst value at 0 and the best value at 1.

**FPR** is given as FP/(FP+TN). It is the negative samples that are misclassified as Positive. The worst and best values of FPR are 1 and 0, respectively.

In Table 2, we can investigate the efficiency of the proposed model where different split ratios of a train to test are used to reduce variability. As the ratio of training size increases, the more time is consumed (time contains extract features by DCGAN, syntheses minor class samples using the SMOTE algorithm and the PILAE classifier) for classification. The best results are for a 60:40 ratio, and we therefore utilize this ratio for comparing with other state-of-the-art methods that used the same dataset.

Table 3 illustrates the robustness and capacity of DCGAN-PILAE model compared with other best results. DCGAN-PILAE can be ranked to be the optimum solution in terms of test accuracy and time complexity. In the previous work which utilized the handcrafted features followed by PILAE classifier, it has training time lower than the present work due to DCGAN-PILAE is handled the imbalance problem, thus increasing the training time. Spinn has the same ratio of recall with the proposed method, but the rate of FPR score is higher than for our method, indicating there are a significant number of non-pulsar candidates being misclassified as a pulsar.

In Fig. 4a the impact of $\beta$ values regarding the accuracy of prediction on the dataset is given by $\beta = 0.2$. By examining the accuracy curve, It is clear there are moderate changes as the parameters of $\beta$ vary. This result implies that the empirical parameter $\beta$ necessitates a small influence on

**Table 1**
Confusion matrix for binary classification.

| Outcomes | Positive prediction | Negative prediction |
|---|---|---|
| Pulsar | True Positive (TP) | False Negative (FN) |
| Non-Pulsar | False Positive (FP) | True Negative (TN) |

**Table 2**
Results for proposed method over HTRU dataset with difference split ratio.

| Train-test | Training accuracy | Testing accuracy | Time min. |
|---|---|---|---|
| 80-20 | 99.97 | 100 | 7.49 |
| 70-30 | 100 | 100 | 7.39 |
| 60-40 | 100 | 100 | 7.29 |

prediction performance. We compute the ratio between rank and the input data dimension in each layer in Fig. 4b; the test error curve on the dataset is illustrated in Fig. 4c. The test error is a classification error over the test set. From Fig. 4c, it can be identified that the test error reduces while the network goes more deeply. Fig. 3 shows a batch of generated images from DCGAN's generator network of pulsar images.

### 4.3. Evaluation on MNIST dataset

To demonstrate the viability of our proposed technique, we trained our method on MNIST dataset and contrasted the execution result with other presented techniques. The MNIST dataset is balanced with 6000 images per class for marks 0–9; we make a few changes on it to ensure it is imbalanced: we pick digits from 5 to 9 as the minority classes and randomly pick 300 images from each class (1,500 images). The other classes from 0 to 4 are picked as majority class where 3000 images are chosen randomly from each class (15,000 images). Finally, the dataset has ratio of a minor: major equal to 1:10. For DCGAN hyperparameters, we run the same number of epochs as a first dataset (15 epochs, 80 s per epoch) with a batch size of 256 and learning rate set to 2E-3. The SMOTE algorithm then applied to oversample the minor class to have a ratio to the major class equal to 9:10; the $k$ value is set to 5 as HTRU dataset, although the final dataset will have 28,500 samples. We run the experiment on different data split ratios to reduce variability, as in Table 4.

In Table 5, we can observe the results of DCGAN-PILAE algorithm over the MNIST dataset compared with other reported methods. There are three methods—DOS (Ando and Huang, 2017), CoSen CNN (Khan et al., 2018) and DQNimb (Lin et al., 2019)—where in Ando and Huang (2017) presents oversampling to the space of deep features created by CNN in their Deep Over-sampling (DOS) structure. This strategy is broadly assessed by creating imbalanced datasets from five prominent benchmark datasets, such as MNIST and CIFAR-10. The DOS structure comprises two synchronous learning strategies, improving both the upper- and lower- layer parameters independently. The upper layers determine how to differentiate between classes utilizing the created embeddings, whereas the lower layers are accountable for gaining the embedding function; in contrast, Khan et al. (2018) Presented a cost-sensitive (CoSen) deep-learning technique that mutually learns the weight parameters of network and the misclassification costs of every class through learning. The suggested technique, CoSen CNN, is assessed against six image datasets; for example, MNIST and CIFAR-100 (Krizhevsky and Hinton, 2009). The pre-trained VGG16 over ImageNet dataset is utilized as a feature extractor and the standard CNN all through the examinations. The cost matrix that is learned by the CoSen CNN is utilized to change the yield of the VGG16 CNN's last layer, giving higher significance to images with a greater cost. In the last method (Ando and Huang, 2017), the classification of imbalanced datasets issue is viewed as a speculating competition that can be disintegrated into a consecutive decision-making operation. At each time step, the operator obtains a domain state that is performed by a training sample and after that proceeds to a classification efficiency under the direction of the strategy. On the off chance that the operator plays out a correct classification efficiency, it will be given a positive reward, else, it will be given a negative reward. The reward from the majority class is lower than that of the minority class. The objective of the operator is to get as many increasingly combined rewards as conceivable amid the procedure of consecutive decision-making, that is, to accurately perceive the samples

**Table 3**
Result comparisons for HTRU dataset with best method. Time in minutes.

| Method | G-Mean | F-Score | Recall | Precision | Specificity | FPR | Accuracy | Training time |
|---|---|---|---|---|---|---|---|---|
| GH-VFDT Lyon et al. (2016) | 96.10 | 94.10 | 92.80 | 95.50 | 99.50 | 5E-3 | 98.80 | – |
| PIL& HOG + PILAE Li et al. (2018) | – | 94.65 | 95.74 | 93.95 | – | – | – | 5.3 |
| SPINN Morello et al. (2014) | – | – | 100 | – | – | 0.64 | – | – |
| GS + AdaBoost Lin et al. (2020) | – | 80 | 99 | 67 | – | 0.651 | – | – |
| REF + GBoost Lin et al. (2020) | – | 95 | 99 | 93 | – | 0.102 | – | – |
| ANN (MLP) Bethapudi and Desai (2018) | 99.80 | 97.90 | 99.80 | 96.10 | – | 0.055 | 99.90 | – |
| AdaBoost Bethapudi and Desai (2018) | 99.80 | 98.60 | 99.70 | 97.60 | – | 0.032 | 99.90 | – |
| GBC Bethapudi and Desai (2018) | 99.70 | 99.90 | 99.50 | 98.60 | – | 0.020 | 99.90 | – |
| XGBoost Bethapudi and Desai (2018) | 99.80 | 98.50 | 99.70 | 97.40 | – | 0.036 | 99.90 | – |
| HCSM Yao et al. (2016) | 98.47 | 98.31 | 97.49 | 99.15 | 100 | 0 | 99.99 | – |
| DCGAN + L2-SVM-2 Guo et al. (2019a) | – | 96.40 | 96.30 | 96.50 | – | – | – | – |
| DeepF + SVM Guo et al. (2019a) | – | – | 100 | – | – | 0 | – | – |
| **DCGAN-PILAE** | 100 | 100 | 100 | 100 | 100 | 0 | 100 | 7.29 |

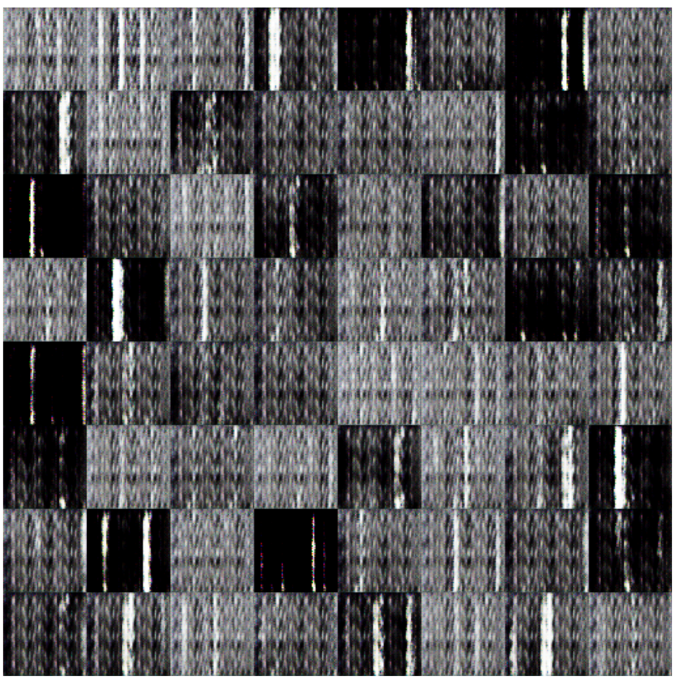

**Fig. 3.** Generated pulsars images after fifteen epochs of training.

as much as conceivable. From Table 4, the best results are for a 70:30 ratio, and we subsequently utilize this ratio for comparing with other best methods that used the same dataset. In Table 5, the PILAE has achieved the best accuracy compared to other methods, proving evidence for the efficiency of our proposed method. The suggested model becomes balanced between evaluation scores and time complexity.

### 4.4. Evaluation on CIFAR-10 dataset

Cifar-10 is a more complicated image dataset than HTRU and MNIST datasets. It includes 32x32 color images with ten classes of physical objects. The standard train: test ratio for every class is 5000:1,000. We use the same procedure of the MNIST dataset to make the CIFAE-10 dataset imbalance. For the DCGAN hyperparameters, we run for twenty epochs where each epoch takes one minute to train and learning rate equals to 2E-3. Table 6 shows the training and test accuracy with total consuming time for a different train: test ratios.

The measure scores of the experimental results are given in Table 7. Despite the different comparing strategies, the DCGAN-PILAE method obtains the optimal classification model. The F-measure scores and G-mean scores of the DCGAN-PILAE model are better than those of the comparing model.

### 4.5. Discussion

From the above results over the three datasets HTRU, MNIST and CIFAR-10, we can state that the proposed method achieves a good balance between computation complexity and test accuracy rate. The DCGAN model is used to extract features from the discriminator ($D$) network. We use the fourth layer from the $D$ network, which is the top layer before output. The intuition is that these features are linearly separable because the top layer is just a logistic regression. In general, the highest-level features are extracted in the last convolution layer (Ren et al., 2012; Athiwaratkun and Kang, 2015). The SMOTE algorithm is used to oversample the minor class in the three datasets used in the proposed method. SMOTE can be applied to high-dimensional data such as for our situation (8192-dimensional space for each sample) and gives good results (Blagus and Lusa, 2013). The above experiments have demonstrated that the suggested combination features leaned by DCGAN and PILAE supplemented by the softmax classifier method attains an excellent performance of prediction accuracy as compared to other methods.

The computation complexity of our model has three terms. First, for DCGAN, it is $O(B^2)$ where $B$ is the batch size of discriminator and generator networks. Second, the SMOTE computational time can be given as $O(|X_R^+| * V * C)$ (Ertekin, 2013) where $|X_R^+|$ is the minor's training samples number, $V$ is the number of artificially generated samples for minor class at each iteration and $C$ is the method that used to find the $K$ nearest neighbors that involves use of the K-nearest neighbor (KNN) algorithm. Third, the PILAE algorithm can be approximately clarified as $\mathcal{O}(dN^2)$ (Guo et al., 2018) where $N$ indicates the number of samples in dataset and $d$ is the dimension of the sample. Although our method has many parts in terms of computational complexity, it achieves good results comparing to other models. For HTRU dataset (60%:40%) one epoch of DCGAN model takes 42 s, the SMOTE method consuming one minute to generate artificial samples of minor class (pulsar) and finally, the PILAE algorithm only takes 26.50 s to classify pulsar to non-pulsar.

PILAE possesses a fast training time because (1) it does not require fine-tuning; (2) PILAE weights can be analytically identified, in contract to traditional autoencoders where iterative algorithms are essentially required; and (3) it learns to signify features through singular values, unlike autoencoders where the representation of data is learned. In addition, the high classification ratios in suitable processing times, are enforced by the fact that our model has been executed effectively using GPUs; the utilization of GPUs allow us to realize the full probability of DCGAN approach for feature extraction followed by PILAE structure. The proposed method thus achieves a good balance between computational performance and prediction ability as compared to other classical methods.

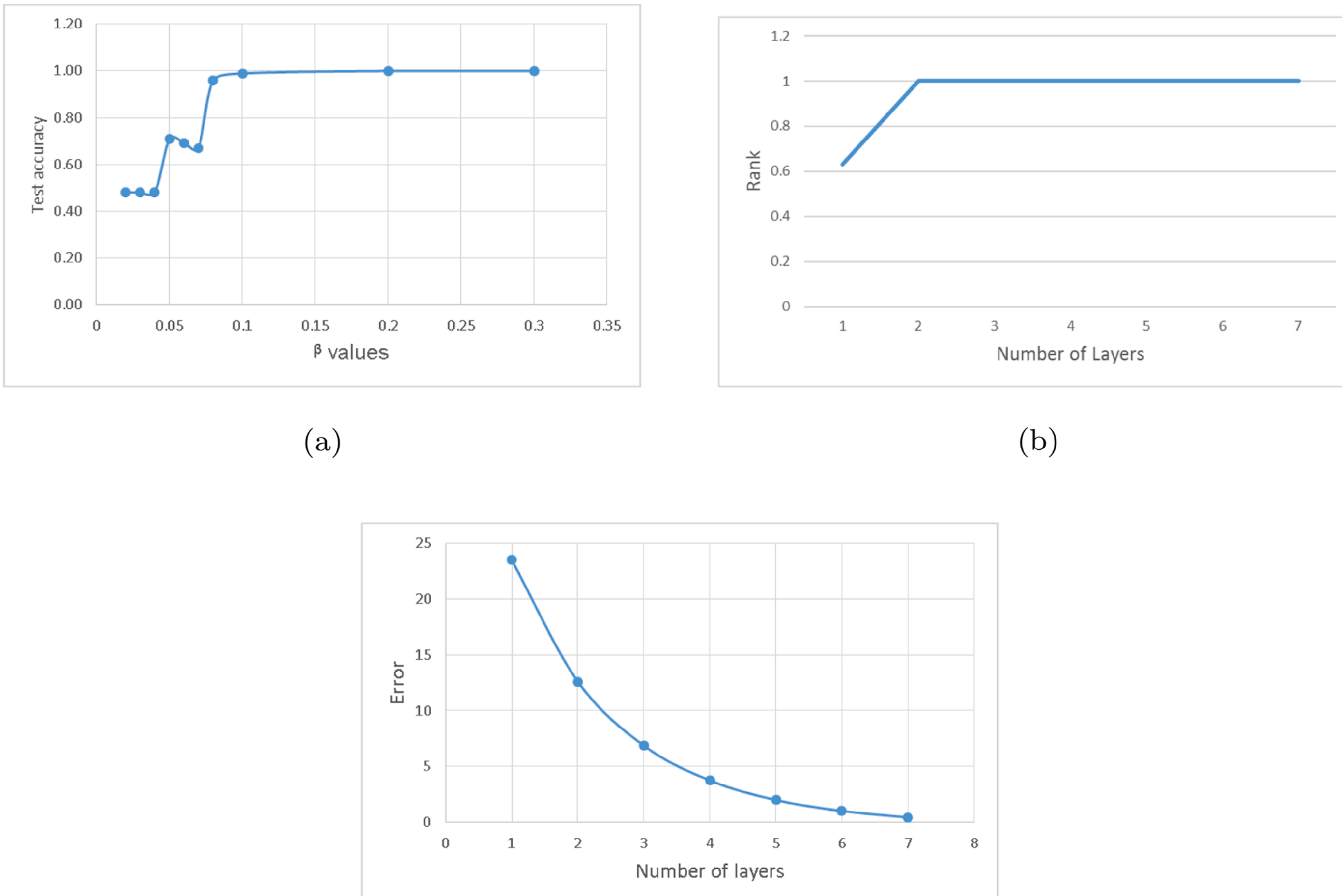


**Fig. 4.** The PILAE curves on HTRU dataset: (a) Test accuracy curve with regard to $\beta$, (b) The rank ratio as the model layers increased, and (c) The error as the model layers increased

**Table 4**
Results for proposed method over MNIST dataset with difference split ratio.

| Train-test | Training accuracy | Testing accuracy | Time min. |
|---|---|---|---|
| 80-20 | 98.34 | 95.95 | 21 |
| 70-30 | 99.89 | 98.97 | 21.3 |
| 60-40 | 99.64 | 98.29 | 22 |

**Table 5**
Result comparisons for MNIST dataset with best method.

| Method | G-Mean | F-Score | Recall | Precision |
|---|---|---|---|---|
| DOS Ando and Huang (2017) | N/A | 98 | 97 | 99 |
| CoSen CNN Khan et al. (2018) | 99.2 | 49.30 | N/A | N/A |
| DQNimb Lin et al. (2019) | 99.70 | 99.20 | N/A | N/A |
| **DCGAN-PILAE** | 98.16 | 98.09 | 97.50 | 98.96 |
| **Method** | **Specificity** | **FPR** | **Accuracy** | **Training time** |
| DOS Ando and Huang (2017) | N/A | N/A | N/A | N/A |
| CoSen CNN Khan et al. (2018) | N/A | N/A | 98.60 | 2.5 min/epoch |
| DQNimb Lin et al. (2019) | N/A | N/A | N/A | N/A |
| **DCGAN-PILAE** | 98.83 | 1e-2 | 98.97 | 21.3 min. |

**Table 6**
Results for proposed method over CIFAR-10 dataset with difference split ratio.

| Train-test | Training accuracy | Testing accuracy | Time min. |
|---|---|---|---|
| 80-20 | 99.89 | 100 | 21.2 |
| 70-30 | 99.66 | 100 | 21.4 |
| 60-40 | 99.95 | 99.92 | 23 |

**Table 7**
Result comparisons for CIFAR-10 dataset with best method.

| Method | G-Mean | F-Score | Recall | Precision |
|---|---|---|---|---|
| DOS Ando and Huang (2017) | N/A | 64 | N/A | N/A |
| DQNimb Lin et al. (2019) | 96.70 | 95 | N/A | N/A |
| **DCGAN-PILAE** | 100 | 100 | 100 | 100 |
| **Method** | **Specificity** | **FPR** | **Accuracy** | **Training time** |
| DOS Ando and Huang (2017) | N/A | N/A | N/A | N/A |
| DQNimb Lin et al. (2019) | N/A | N/A | N/A | N/A |
| **DCGAN-PILAE** | 100 | 2.74E19 | 100 | 21.2 min. |

## 5. Conclusion

To make up for the shortcomings of the learning from imbalanced data, this paper establishes a new model that can optimize the speed and prediction of the learning process, utilizing the SMOTE algorithm for tackling the imbalance problem. The DCGAN-PILAE framework is employed, in which DCGAN functions as an extraction feature. A PILAE classifier is then trained on the extracted features of DCGAN. The deep features can, therefore, be fully tuned with the PILAE classifier generalization performance, occasioning in a recognition accuracy that is satisfying, without the requirement for additional and complicated DCGAN frameworks. More so with other techniques, the DCGAN-PILAE method can achieve perfect results with a forthright structure which alleviates training process, which is frequently time-consuming.

## CRediT authorship contribution statement

**Mohammed A.B. Mahmoud:** Conceptualization, Project administration, Writing - original draft, Writing - review & editing, Validation. **Ping Guo:** Conceptualization, Methodology, Funding acquisition, Validation.

## Declaration of Competing Interest

Authors declare that they have no conflict of interest.

## Acknowledgments

This work is fully supported by the grants from the National Natural Science Foundation of China (61375045), and the Joint Research Fund in Astronomy (U1531242) under cooperative agreement between the National Natural Science Foundation of China (NSFC) and Chinese Academy of Sciences (CAS ).

## References

Abdo, A., Ajello, M., Allafort, A., Baldini, L., Ballet, J., Barbiellini, G., Baring, M., Bastieri, D., Belfiore, A., Bellazzini, R., et al., 2013. The second fermi large area telescope catalog of gamma-ray pulsars. Astrophys. J. Suppl. Ser. 208 (2), 17.
Ando, S., Huang, C.Y., 2017. Deep over-sampling framework for classifying imbalanced data. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 770–785.
Athiwaratkun, B., Kang, K., Feature representation in convolutional neural networks. arXiv preprint arXiv:1507.02313.
Bass, G., Borne, K., 2016. Supervised ensemble classification of kepler variable stars. Mon. Not. R. Astron. Soc. 459 (4), 3721–3737.
Bates, S., Bailes, M., Barsdell, B., Bhat, N., Burgay, M., Burke-Spolaor, S., Champion, D., Coster, P., D'Amico, N., Jameson, A., et al., 2012. The high time resolution universe pulsar survey-vi. an artificial neural network and timing of 75 pulsars. Mon. Not. R. Astron. Soc. 427 (2), 1052–1065.
Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. 35 (8), 1798–1828.
Bethapudi, S., Desai, S., 2018. Separation of pulsar signals from noise using supervised machine learning algorithms. Astron. Comput. 23, 15–26.
Blagus, R., Lusa, L., 2013. Smote for high-dimensional class-imbalanced data. BMC Bioinform. 14 (1), 1–16.
Cameron, A., Barr, E., Champion, D., Kramer, M., Zhu, W., 2017. An investigation of pulsar searching techniques with the fast folding algorithm. Mon. Not. R. Astron. Soc. 468 (2), 1994–2010.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.
Deng, X., Sun, X., Guo, P., Yin, Q., 2019. Image recognition based on combined filters with pseudoinverse learning algorithm. IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, pp. 202–209.
Devine, T.R., Goseva-Popstojanova, K., McLaughlin, M., 2016. Detection of dispersed radio pulses: a machine learning approach to candidate identification and classification. Mon. Not. R. Astron. Soc. 459 (2), 1519–1532.
Eatough, R.P., Molkenthin, N., Kramer, M., Noutsos, A., Keith, M., Stappers, B., Lyne, A., 2010. Selection of radio pulsar candidates using artificial neural networks. Mon. Not. R. Astron. Soc. 407 (4), 2443–2450.
Ertekin, Ş., 2013. Adaptive oversampling for imbalanced data classification. Information Sciences and Systems. Springer, pp. 261–269.
Estabrooks, A., Jo, T., Japkowicz, N., 2004. A multiple resampling method for learning from imbalanced data sets. Comput. Intell. 20 (1), 18–36.
Feng, S., Deng, X., Guo, P., Zhao, B., Yin, Q., Wang, H., 2019. An ensemble model for error modeling with pseudoinverse learning algorithm. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, pp. 1893–1898.
Guo, P., Duan, F., Wang, P., Yao, Y., Yin, Q., Xin, X., Li, D., Qian, L., Wang, S., Pan, Z., et al., 2019a. Pulsar candidate classification using generative adversary networks. Mon. Not. R. Astron. Soc. 490 (4), 5424–5439.
Guo, P., Lyu, M.R., 2004. A pseudoinverse learning algorithm for feedforward neural networks with stacked generalization applications to software reliability growth data. Neurocomputing 56, 101–121.
Guo, P., Lyu, M.R., Chen, C.L.P., 2003. Regularization parameter estimation for feedforward neural networks. IEEE Trans. Syst. Man. Cybern. Part B 33 (1), 35–44.
Guo, P., Zhao, D., Han, M., Feng, S., 2019b. Pseudoinverse learners: new trend and applications to big data. INNS Big Data and Deep Learning Conference. Springer, pp. 158–168.
Guo, P., Zhou, X., Wang, K., Pilae: a non-gradient descent learning scheme for deep feedforward neural networks. arXiv preprint arXiv:1811.01545.
He, H., Garcia, E.A., 2008. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. (9), 1263–1284.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.
Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.
Hulten, G., Spencer, L., Domingos, P., 2001. Mining time-changing data streams. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 97–106.
Keith, M., Eatough, R., Lyne, A., Kramer, M., Possenti, A., Camilo, F., Manchester, R., 2009. Discovery of 28 pulsars using new techniques for sorting pulsar candidates. Mon. Not. R. Astron. Soc. 395 (2), 837–846.
Keith, M., Jameson, A., Van Straten, W., Bailes, M., Johnston, S., Kramer, M., Possenti, A., Bates, S., Bhat, N., Burgay, M., et al., 2010. The high time resolution universe pulsar survey–I. System configuration and initial discoveries. Mon. Not. R. Astron. Soc. 409 (2), 619–627.
Khan, S.H., Hayat, M., Bennamoun, M., Sohel, F.A., Togneri, R., 2018. Cost-sensitive learning of deep feature representations from imbalanced data. IEEE Trans. Neural Netw. Learn. Syst. 29 (8), 3573–3587.
Kramer, M., Stappers, B., Pulsar science with the ska. arXiv preprint arXiv:1507.04423.
Krizhevsky, A., Hinton, G., 2009. Learning multiple layers of features from tiny images. Technical Report. Citeseer.
Laurikkala, J., 2001. Improving identification of difficult small classes by balancing class distribution. Conference on Artificial Intelligence in Medicine in Europe, pp. 63–66.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.
Lee, H., Grosse, R., Ranganath, R., Ng, A.Y., 2009. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th Annual International Conference on Machine Learning. ACM, pp. 609–616.
Lee, H., Grosse, R., Ranganath, R., Ng, A.Y., 2011. Unsupervised learning of hierarchical representations with convolutional deep belief networks. Commun. ACM 54 (10), 95–103.
Li, S., Feng, S., Guo, P., Yin, Q., 2018. A hierarchical model with pseudoinverse learning algorithm optimazation for pulsar candidate selection. 2018 IEEE Congress on Evolutionary Computation (CEC). IEEE, pp. 1–6.
Lin, E., Chen, Q., Qi, X., Deep reinforcement learning for imbalanced classification. arXiv preprint arXiv:1901.01379.
Lin, H., Li, X., Luo, Z., 2020. Pulsars detection by machine learning with very few features. Mon. Not. R. Astron. Soc. 493 (2), 1842–1854.
Lyon, R., Brooke, J., Knowles, J., Stappers, B., 2013. A study on classification in imbalanced and partially-labelled data streams. 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, pp. 1506–1511.
Lyon, R., Brooke, J., Knowles, J., Stappers, B., 2014. Hellinger distance trees for imbalanced streams. 2014 22nd International Conference on Pattern Recognition. IEEE, pp. 1969–1974.
Lyon, R.J., Stappers, B., Cooper, S., Brooke, J., Knowles, J., 2016. Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. Mon. Not. R. Astron. Soc. 459 (1), 1104–1123.
Manchester, R.N., Lyne, A.G., Camilo, F., Bell, J., Kaspi, V., D'Amico, N., McKay, N., Crawford, F., Stairs, I., Possenti, A., et al., 2001. The parkes multi-beam pulsar survey–I. Observing and data analysis systems, discovery and timing of 100 pulsars. Mon. Not. R. Astron. Soc. 328 (1), 17–35.
Mohamed, T.M., 2018. Pulsar selection using fuzzy KNN classifier. Future Comput. Inform. J. 3 (1), 1–6.
Morello, V., Barr, E., Bailes, M., Flynn, C., Keane, E., van Straten, W., 2014. Spinn: a straightforward machine learning solution to the pulsar candidate selection problem. Mon. Not. R. Astron. Soc. 443 (2), 1651–1662.
Radford, A., Metz, L., Chintala, S., Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
Ren, J.S., Wang, W., Wang, J., Liao, S., 2012. An unsupervised feature learning approach to improve automatic incident detection. 2012 15th International IEEE Conference on Intelligent Transportation Systems, pp. 172–177.
Smith, K. M., New algorithms for radio pulsar search. arXiv preprint arXiv:1610.06831.
Vedaldi, A., Lenc, K., 2015. Matconvnet: Convolutional neural networks for matlab. Proceedings of the 23rd ACM International Conference on Multimedia. ACM, pp. 689–692.
Wagstaff, K.L., Tang, B., Thompson, D.R., Khudikyan, S., Wyngaard, J., Deller, A.T., Palaniswamy, D., Tingay, S.J., Wayth, R.B., 2016. A machine learning classifier for fast radio burst detection at the vlba. Publ. Astron. Soc. Pac. 128 (966), 084503.

Wang, H., Yuan, M., Yin, Q., Guo, P., Zhu, W., Li, D., Feng, S., 2020. Radio frequency interference mitigation using pseudoinverse learning autoencoders. Res. Astron. Astrophys. 20 (8), 114.
Wang, H., Zhu, W., Guo, P., Li, D., Feng, S., Yin, Q., Miao, C., Tao, Z., Pan, Z., Wang, P., et al., 2019. Pulsar candidate selection using ensemble networks for fast drift-scan survey. Sci. China Phys. Mech. Astron. 62 (5), 959507.
Wang, J., Guo, P., Xin, X., 2018. Review of pseudoinverse learning algorithm for multilayer neural networks and applications. International Symposium on Neural Networks. Springer, pp. 99–106.
Wang, K., Guo, P., Xin, X., Ye, Z., 2017. Autoencoder, low rank approximation and pseudoinverse learning algorithm. Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on. IEEE, pp. 948–953.
Weiss, G. M., Provost, F., 2001. The effect of class distribution on classifier learning: an empirical study.
Yao, Y., Xin, X., Guo, P., 2016. Pulsar candidate selection by assembling positive sample emphasized classifiers. 2016 12th International Conference on Computational Intelligence and Security (CIS). IEEE, pp. 120–124.
Zhu, W., Berndsen, A., Madsen, E., Tan, M., Stairs, I., Brazier, A., Lazarus, P., Lynch, R., Scholz, P., Stovall, K., et al., 2014. Searching for pulsars using image pattern recognition. Astrophys. J. 781 (2), 117.