Efficient Fermi Source Identification with Machine Learning Methods

H. B. Xiao\textsuperscript{a,b,d,e}, H. T. Cao\textsuperscript{c,d,e,*}, J. H. Fan\textsuperscript{b,e,++}, D. Costantin\textsuperscript{b,f,g}, G. Y. Luo\textsuperscript{e}, Z. Y. Pei\textsuperscript{a,b,d,e}

\textsuperscript{a}Department of Physics and Astronomy "G. Galilei", University of Padova, Via Marzolo, 8, Padova PD, 35131, Italy.
\textsuperscript{b}Center for Astrophysics, Guangzhou University, No. 230, Wai Huan Xi Road, Guangzhou Higher Education Mega Center, Guangzhou, 510006, China.
\textsuperscript{c}Department of Information Engineering, University of Padova, Via Gradenigo, 6/B, Padova PD, 35131, Italy.
\textsuperscript{d}Istituto Nazionale di Fisica Nucleare, Via Marzolo, 8, Padova PD, 35131, Italy.
\textsuperscript{e}School of Physics and Electronic Engineering, Guangzhou University, No. 230, Wai Huan Xi Road, Guangzhou Higher Education Mega Center, Guangzhou, 510006, China.
\textsuperscript{f}Astronomy Science and Technology Research Laboratory of Department of Education of Guangdong Province, Guangzhou, 510006, China.
\textsuperscript{g}Department of Statistical Science, University of Padova, Via Battisti, 241, Padova PD, 35121, Italy.

Abstract

In this work, Machine Learning (ML) methods are used to efficiently identify the unassociated sources and the Blazar Candidate of Uncertain types (BCUs) in the Fermi-LAT Third Source Catalog (3FGL). The aims are twofold: 1) to distinguish the Active Galactic Nuclei (AGNs) from others (non-AGNs) in the unassociated sources; 2) to identify BCUs into BL Lacertae objects (BL Lacs) or Flat Spectrum Radio Quasars (FSRQs). Two dimensional reduction methods are presented to decrease computational complexity, where Random Forest (RF), Multilayer Perceptron (MLP) and Generative Adversarial Nets (GAN) are trained as individual models. In order to achieve better performance, the ensemble technique is further explored. It is also demonstrated that grid search method is of help to choose the hyper-parameters of models and decide the final predictor, by which we have identified 748 AGNs out of 1010 unassociated sources, with an accuracy of 97.04%. Within the 573 BCUs, 326 have been identified as BL Lacs and 247 as FSRQs, with an accuracy of 92.13%.

Keywords: Fermi source, AGNs, Blazar, dimensionality reduction, ensemble method, grid search

1. Introduction

Active Galactic Nuclei (AGNs) have been a hot topic in astrophysics since their discovery in 1960s, while their nature still remains uncertain. The study of AGNs is always one of the paramount issues. For instance, it provides important clues for the evolution and structure of galaxies (Hopkins et al., 2006) and the whole Universe in the reionization era. AGNs have also been used to constrain the cosmological constant and the baryon density $\Omega_B$ (Varshalovich et al., 2012).

In some of AGNs, they contain an energetic jet, which emits from the center of a Supermassive Black Hole (SMBH) and is surrounded by an accretion disk (Urry & Padovani, 1995). The jets emissions are dominated by nonthermal radiation, with spectrum spanning from radio to $\gamma$-ray band. The Spectral Energy Distribution (SED) of jet emissions shows a two-hump structure. The lower-energy hump, peaked between millimeter band and soft X-ray band, is attributed to the synchrotron emission produced by electrons in the jet. And, the higher-energy hump, peaked in the MeV-GeV range, is mainly due to inverse Compton scattering according to leptonic model. While in a hadronic scenario, protons are assumed to be accelerated along with electrons, and contribute much of the high-energy component via photopion interactions (Dimitrakoudis et al., 2012; Böttcher et al., 2009).

AGNs standard model also shows that different types of AGNs result from different viewing angles between the line of sight and the jet direction. In particular, the sources can be classified as blazars if their jets are pointing towards the observer.
Blazar is an extreme subclass of AGNs, since it shows high and fast flux variability, coupled with rapidly varying polarization. It can be divided into two subclasses: BL Lacertae objects (BL Lacs) and Flat Spectrum Radio Quasars (FSRQs). The BL Lac shows weak or no emission lines, while the FSRQ shows strong emission lines.

*Fermi*-LAT (Large Area Telescope on-board Gamma-Ray Space Telescope, see Atwood et al. 2009) is a powerful imaging high-energy γ-ray telescope launched in 2008 and spans over the energy ranges from about 20 MeV to more than 300 GeV. Its field of view covers about 20% of the sky at any time and the whole sky in every three hours through a continuous scan. With respect to the previous telescope (e.g., EGRET, the Energetic Gamma-Ray Experiment Telescope, see Fichtel et al. 1993), *Fermi*-LAT shows better energy resolution, better angular resolution and wider effective area in both low-energy and high-energy bands\(^1\). And *Fermi*-LAT collaboration provides point source catalogues, which have been used to study the properties of AGNs and their subclasses.

Our work focused on utilizing the third *Fermi*-LAT source catalog (3FGL), which covers the four-year observation data of the *Fermi*-LAT. 3FGL contains 3034 sources, among which 1744 belong to AGN class. Within the AGN sources, there are 573 Blazars that haven’t been tagged as BL Lacs or FSRQ. These uncertain Blazars are also named as Blazar Candidate of Uncertain types (BCUs). Besides, the 3FGL remains 1010 unassociated sources (Acero et al., 2015). These BCUs and unassociated sources result in an incomplete catalog, which decreases the significance of observations. However, a certain source sample can be of great help to study its high energy properties, e.g., variability and radiative process. Therefore, efficient algorithms are desirable to handle the incompleteness problem of 3FGL.

Many efforts have been dedicated to this issue (Hassan et al., 2013; Doert et al., 2014; Cavuoti et al., 2014; Kang, 2019). For instance, Saz Parkinson et al. (2016) used two Machine Learning (ML) models, i.e., Random Forest (RF) and Logistic Regression (LR) to identify 1008 unassociated 3FGL sources, of which 893 sources were in agreement with both methods. But the authors probably ignored the precision of predicted AGNs, which will be one of the contributions in our work.

Since variability is one of the characterizing properties of Blazar (Paggi et al., 2011), Chiaro et al. (2016) utilized Blazar Flaring Patterns (B-FlaP) as the inputs of an Artificial Neural Networks (ANN) to identify the 573 BCUs in 3FGL, while they have 77 still remained uncertain. Similarly, Lefaucheur et al. (2017) also left 45 BCUs unprocessed. We will further deal with these remained BCUs and discussed via comparisons in this paper.

However, the increasing number of sources, coming both from current and next-generation telescopes, requires more powerful classifiers. In this work, different ML methods are explored to design an efficient algorithm for further identifying 3FGL unassociated sources and BCUs. The improvements in identification efficiency mainly focus on two aspects:

1) We introduce two dimensionality reduction methods acting as the data preprocessing step to remove the redundancy and noise of data and decrease computational complexity.

2) The independently trained models are aggregated to be the ensemble model for strong generalization capability and better performance on unseen samples.

Furthermore, grid search method is used to record the hyperparameters achieving the best performance and to choose the final predictors acting on the identification tasks. Finally, the evaluation of models on multi-splitting datasets demonstrates the effectiveness of proposed method on the identification of the unassociated sources and BCUs in 3FGL.

The paper is organized as follows. Section 2 presents a brief introduction of ML methods we used. Section 3 describes our proposed method in detail. The experiments and results are reported in Section 4. Further discussions and conclusions are presented in Section 5 and 6, respectively.

---

\(^1\)Check [http://www.slac.stanford.edu/exp/glast/groups/canda/lat_Performance.htm](http://www.slac.stanford.edu/exp/glast/groups/canda/lat_Performance.htm) for more performance of *Fermi*-LAT.
2. Machine Learning Methods

As ML increases success in many computer science and engineering applications, it is gaining popularity in the astronomical subjects as well, for instance to complete the 3FGL data catalog (Saz Parkinson et al., 2016; Chiaro et al., 2016). Similar to the image classification in computer science, the source identification task could be conducted by building a classifier.

2.1. Dimensionality Reduction

High dimensionality is one of the best-known problems when processing and analyzing big data, due to the high costs of the hardware resources and computing time. Researches have shown that the experimental data are highly redundant, leaving room for improvement through the design of efficient algorithms and the reduction of the dimensionality (Bakshi et al., 1993; Bengio et al., 2009; Glorot et al., 2011). Lower dimensionality will also lead to a decreased computational complexity, which has become a preferred avenue nowadays (He et al., 2018; Sellami, 2018; Kuang et al., 2018).

One of the widely used techniques for reducing dimensionality is Feature Selection (FS). With this method, we select a subset of features that are most relevant to the task. It improves computational efficiency and reduces the generalization error of model by removing irrelevant features or noise (Sebastian, 2015). Feature Importance (FI) is commonly used to measure the feature significance. It is usually calculated from the average impurity decrease of the criterion brought by the particular feature in RF models. We pick out the features with high FI values, while abandon others whose FI values are low. RF was firstly proposed by Kam (1995) and since then it has become a powerful ML method for classification and prediction tasks despite its simple mechanism (Breiman, 2001; Saz Parkinson et al., 2016). A RF consists of many Decision Trees (DTs) and manages to reduce both the bias and variance of the tree models, which is the idea of Ensemble Learning (EL) (Dietterich, 2002).

Another strategy for dimensionality reduction is Feature Extraction (FE). This method transforms or projects the original data into a new feature space, in which the data will be represented by other types of variables. Principal Component Analysis (PCA) is an unsupervised linear transformation technique for dimensionality reduction. With PCA, the data is transformed from its original coordinate system to a new coordinate system. The first new axis is chosen in the direction of the largest variance in the data. Then the second axis is orthogonal to the first axis and in the direction of the second largest variance. This procedure is repeated until the number of features is met. The majority of the variance is contained along the first few axes. Therefore, the rest of the axes can be neglected and as a result the dimensionality of data can be reduced (Peter, 2012).

2.2. Neural Network Classifiers

After data preprocessing, classifiers are built for data analysis. Neural Network (NN) is an important aspect of ML methods. It has been successfully applied in different classification tasks within the astrophysical content (Chiaro et al., 2016; Salvetti et al., 2017). It usually consists of one input layer, several hidden layers, and one output layer. One-time NN training process includes Forward Propagation (FP) and Back Propagation (BP), but a good NN model always requires training many times. In FP the data flows from input to output layer, while in BP the flow has a converse direction.

FP aims to calculate the model prediction errors based on a predefined loss function, such as the classical cross-entropy function. A lower error always means a higher accuracy score for the classification task, and indicates a better classifier. In addition to the cross-entropy function, in our analysis we also considered some other non-classical loss functions listed in Janocha et al. (2016).

BP algorithm was proposed by Rumelhart et al. (1988) to update the parameters of NN model. Starting from the predicted error obtained in output layer, the errors of each hidden layer are successively calculated. With these errors, the model parameters are then updated by different optimization algorithms (Duchi et al., 2011; Sutskever et al., 2012).

The layers of a Multiplayer Perceptron (MLP) model are
often fully connected to each other, which are also known as Fully-connected Layers (FcLs) (Cavuoti et al., 2014). Another type of NN model, Generative Adversarial Nets (GAN) that firstly proposed by Goodfellow et al. (2014), can also be constructed by FcLs. It includes a discriminator and a generator, between which an adversarial game takes place: the discriminator is trained to tell the samples coming from the generator, while the generator is trained to produce samples that cannot be recognized by the discriminator.

Odena et al. (2017) and Salimans et al. (2016) successfully built a semi-supervised classifier with GAN model. They labeled the samples produced from the generator as one class, namely the fake class, while the samples that came from the dataset were labeled as true. As the generator was trained to approximate true samples, the discriminator was trained to identify fake samples. Moreover, since the true samples had distinct known classes as well, the discriminator was additionally trained to predict these known classes. Their work obtained good results on some public image datasets.

2.3. Ensemble Methods

Traditionally, dataset is divided into two sub-datasets: training dataset for model training, and test dataset for model evaluation. The so-called over-fitting, referred to the phenomenon that a model performs well on training dataset while badly on test dataset, is a common issue in ML field. It actually indicates the weak generalization capability of a model. That is, a model with strong generation capability will achieve satisfactory performance on unseen samples, as expected from classification tasks. The Ensemble Method (EM), proposed by Freund et al. (1995), aggregates independently trained ML models into an ensemble predictor and reduces both the bias and the variance of individual models, which is of help to alleviate the over-fitting problem.

Two strategies are usually used to aggregate models: hard and soft. The former counts the prediction of each individual model and gives the class that gets the most votes (Zhang et al., 2014). The drawback is that it cannot estimate the likelihood of predicted class. In other words, it cannot tell how certain the class of the predicted result is. Therefore, this strategy will not be considered in our work.

While the latter identifies samples based on the weighted likelihood of predictions coming from the individual models (Jiménez, 1998). These weights indicate the contributions of models and they are decided by the performance of individual models, i.e., the model with higher accuracy score will hold heavier weight. Obviously, the soft ensemble strategy requires every individual model to have the ability of estimating the likelihood of prediction. This aggregation strategy will be used in our analysis.

3. Proposed Method

There are two efforts in our work: Mission A and Mission B, shown in the top panel of Figure 1(a). We aim to build a Best Model A able to identify AGNs versus non-AGNs for Mission A, and a Best Model B able to distinguish FSRQs and BL Lacs for Mission B. The Best Model in each mission is chosen from five candidate models: ‘RF\_reduced’, ‘RF\_pipe’, ‘MLP’, ‘GAN’ and ‘Ensemble Model’, shown in the bottom panel of Figure 1(a). The candidate model achieving best performance on test dataset will be taken as the Best Model used to accomplish the identification task.

Besides, only one-time dataset split may result in a biased model, thus we will randomly split the dataset into training dataset and test dataset several times and repeat the procedure, which is the so-called hand-out evaluation method. In this way, the model will be evaluated more than once and it will achieve strong generalization capability if its performance on each splitting test dataset is comparable.

3.1. Data Preparation

If we assume that ML models, in particular, the NN models, have the capability of finding underlying function relationship among different features, then it is not needed to use some predefined parameters such as hardness ratio in Saz Parkinson et al.
and flux ratio in Mirabal et al. (2016). We will not con-
sider these parameters in our work. Instead, we directly use FS
method to select the subset of features that are most relevant to
our tasks. Experimental Section will give more details on the
feature characteristics.

3.2. Method Scheme

Training the ML models is a challenge. One of the main
difficulties comes from the choice of so many hype-parameters,
such as the depth, criterion and the number of trees in RF model
and the number of hidden layers in MLP and GAN models.
There is not a general rule to choose the suitable set of hyper-
parameters. In our work, grid search will be used in every step.
The procedures for two missions are the same. We will consider
different candidate models coming from RF, MLP, GAN and
ensemble models.

Following the scheme of Figure 1(a), the first step of the
procedure is FS. Based on different levels of FI values, several
subsets of features are generated. If the FI values are higher
than a set level then these features are selected into a subset.
The data within the features of a subset will be fed to next step.

The scheme is then divided into two possibilities: normal-
ization or not. Normalization ensures that values of different
features range in the same scale. We will use four normal-
ization methods, namely ‘Standardize’, ‘Normalize’, ‘Minmax’
and ‘MaxAbs’ (Pedregosa et al., 2011). Therefore, RF has two
different processes for input features: the one not normalized,
i.e., ‘RF_reduced’, or the normalized one followed by FE step,
i.e., ‘RF_pipe’.

Whereas MLP and GAN models are fed just with the nor-
malized features. The reasons why we don’t consider FE step
here come from two sides. On one hand, PCA is a linear dimen-
sionality reduction technique. It is not appropriate for complex
problems with nonlinear correlation structures (Zhang, 2000).
While NN models have the nonlinear fitting capability for dea-
ling with these problems. If the FE step is added ahead of NN
models, the nonlinear correlation structures of data may be de-
stroyed and cannot be seen by NN models. As a result, the iden-
tification performance will degrade. On the other hand, when
performing dimensionality reduction with PCA, the discriminative
information that distinguishes one class from another one
may exist in the low variance components and may be removed
improperly, which makes performance worse.

The last candidate model comes from the soft ensemble
model, which is aggregated by the four weighted individual
models. The sum of weights is 1. All of the possible combina-
tions of weights in two decimal places will be tested and the
one which gives the best performance will be recorded.

3.3. Evaluation of Performance

For identification tasks, there exist three metrics for perf or-
ance evaluation: Sensitivity, Precision, and Accuracy. They
can be defined by the form of confusion matrix, which consists
of four prediction cases:

- True Positive (TP): the number of positive samples cor-
correctly predicted to be positive;
- False Positive (FP): the number of negative samples wrongly
predicted to be positive;
- False Negative (FN): the number of positive samples wrongly
predicted to be negative;
- True Negative (TN): the number of negative samples cor-
correctly predicted to be negative.

Then the aforementioned three metrics can be written as:

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]
\[
\text{Precision} = \frac{TP}{TP + FP}
\]
\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}
\]

In our experiments, the performance of models in Mission B
will be evaluated by accuracy. While in Mission A, both the
accuracy and precision will be considered.
4. Experiments and Results

The experiments were set up on the Ubuntu 18.04 system. To speed up the computing, RF programs were distributed on multi-CPUs with scikit-learn (version 0.20) (Pedregosa et al., 2011). While NN models were accelerated by GPU with TensorFlow (version 1.10) (Abadi et al., 2016). In the experiments, the CPU (Central Processing Unit) was Intel Core i5-8400 and the GPU (Graphic Processing Unit) was NVIDIA GTX1050Ti with 768 CUDA cores.

4.1. Source Samples

In this work, we used the 3FGL data downloaded from Space Science Data Center (SSDC)\(^2\). To proceed with ML methods, we randomly split all the known entries into two datasets: training dataset (70% of the entries) and test dataset (the remaining 30%). As already mentioned in Section 3, to test the generalization ability of models, we repeatedly split the dataset five times. As the top panel of Figure 1(b) shows, Dataset A had 2024 associated sources and it was divided into the two sub-classes: 1744 AGNs and 280 non-AGNs. 70% of them, i.e., 1220 AGNs and 196 non-AGNs, formed the training dataset. While the remaining 30%, with 608 sources (524 AGNs and 84 non-AGNs) was used to evaluate the performance of trained models. The same methodology was applied to Dataset B, as shown in the bottom panel of Figure 1(b). In the experiments, 21 features were picked out from the 32 attributes of the SSDC 3FGL data for the ML methods, as shown in Table 1.

4.2. Data Analysis

In FS step, the FI values of 21 features were calculated with a RF model whose hyper-parameters can be found in Table 2. With two optional number of DTs and two criterions, we got four FI results.

Figure 2(a) displays the FI values in the third split case of Dataset A when we chose 30000 DTs and set ‘entropy’ as the criterion. Empirically, the features were divided into four groups roughly based on three importance levels. The first few groups that contain most unimportant features could be removed and then feature-removed datasets were prepared for next step. In Figure 2(a), there were three groups of features could be removed, i.e., (part 1), (part 1, part 2) and (part 1, part 2, part 3). Moreover, the case of the whole 21 features without any removing was also tested in our work for comparison. All the removable features in each split case are listed in the second and third columns of Table 3.

Due to empiricism, the feature division can be different and some features may be ambiguously divided. For instance, feature ‘F1’ would move from part 1 to part 2 when we changed the criterion from ‘gini’ to ‘entropy’ in the case of 10000 DTs. But we found it didn’t affect the predicted results a lot, thus we ignored the differences of different feature divisions.

Similarly, in FE step, the components projected from the feature-removed dataset were also divided into several parts according to different variance levels. Only the parts characterized with high variance values were kept. In the experiments we found that, for a particular feature-removed dataset, the number of principal components was the same in spite of different normalization methods ahead of FE step. In other words, only the FS method influenced the number of principal components. However, we didn’t remove the normalization step since it differed the results of ‘RF\_pipe’ models.

Figure 3(a) shows the variance values of components after projecting the Dataset A without removing any features in the first split case. It gave three groups of principal components: (part 1), (part 1, part 2) and (part 1, part 2, part 3). Thus we had three choices on the number of retainable principal components: [14, 7, 4]. All of the retainable principal components for each feature-removed dataset are listed in the fourth column of the upper panel of Table 3.

With the same philosophy, we display the similar items in Figure 2(b) and Figure 3(b) for Dataset B, and the retainable principal components for each feature-removed dataset are listed in the fourth column of the lower panel of Table 3.

As for a particular ML model: ‘RF\_reduced’, ‘RF\_pipe’,
‘MLP’ or ‘GAN’, our idea was to span over a wide range of hyper-parameters values and took the model that had the best performance on the test dataset to be the candidate model. Even though it was impossible to test all the hyper-parameters, we supposed that the results in our work were good enough with the optional hyper-parameters listed in Table 4.

4.3. Identification Results

The identification results and model hyper-parameters for each split case in our two missions are listed in Table 5 and 6 respectively, where the Best Models are highlighted by gray color. The bold results are the highest scores under the respective metric.

Since in Mission A we required precise AGN candidates identification, we focused both on precision and accuracy metrics. The candidate models (‘RF_reduced’, ‘RF_pipe’, ‘MLP’, ‘GAN’ and soft ensemble model) were evaluated on five split test datasets. The results showed that the MLP and soft ensemble model had the highest accuracy both in the first and third split cases, and the MLP model in the first split case had the highest precision. Under the consideration that the soft ensemble model has stronger generalization ability and the individual models in the third split case had less parameters, we chose the soft ensemble model of the third split case as Best Model A, where the precision was 97.03% and accuracy was 97.04%. While in Mission B, the accuracy was taken as the metric, then the soft ensemble model in the second split case was seen as Best Model B, where the accuracy is 92.13%.

We also noticed that the results in different split cases were comparable. This means that the Best Models were unbiased and could be employed to identify the unassociated and uncertain sources effectively, as reported in Table 7 and 8, respectively. In order to further understand the model performance, we also recorded the confusion matrices of the Best Model A and B in Table 9 and 10, respectively. The few misclassified sources in the tables implied the strong capability of Best Models on 3FGL source identification.

4.4. Physical Analysis

The variability index is a statistical parameter that indicates the confidence level on the variability of a γ-ray source (Nolan et al., 2012; Acero et al., 2015). Considering that different types of sources should show different variability patterns, in order to cross-check our identification results, we used the variability index as a physical metric to compare the results in Mission A with the associated 3FGL results. The average AGN logarithm variability index \( \log V^{TW} \) of this work is \( 1.66 \pm 0.12 \) base on our 748 AGN candidates, while the average log\( \gamma \)--ray variability index \( \log V^{3FGL} \) of 1744 3FGL known AGNs is \( 1.91 \pm 0.42 \), as shown in Figure 4(a). The average non-AGNs logarithm variability index \( \log V^{TW} \) of this work is \( 1.65 \pm 0.09 \) base on our 262 non-AGN candidates, while the average log\( \gamma \)-ray variability index \( \log V^{3FGL} \) of 280 3FGL known non-AGNs is \( 1.68 \pm 0.13 \), see Figure 4(b).

Since variability index is not a proper choice to distinguish an FSRQ from BL Lacs, we used the spectral index to cross-check our results. The averaged FSRQ spectral index \( \alpha^{TW} \) of this study is \( 2.47 \pm 0.17 \) base on our 247 FSRQ candidates, while the averaged spectral index \( \alpha^{3FGL} \) of 484 3FGL known FSRQs is \( 2.42 \pm 0.22 \), see Figure 5(a). The average BL Lac spectral index \( \alpha^{TW} \) of this work was \( 1.98 \pm 0.22 \) base on our 326 BL Lac candidates, while the average spectral index \( \alpha^{3FGL} \) of 660 3FGL known BL Lacs is \( 2.02 \pm 0.25 \), see Figure 5(b).

5. Discussion

5.1. Discussion of our results

The main goal of this study is to propose a powerful and efficient method to identify the unassociated and uncertain sources in 3FGL. Two dimensionality reduction methods were used to decrease the computational complexity, and the ensemble method was applied to the independently training models for better performance. As for a number of hyper-parameters in each step, grid search method was utilized to decide the final best predictor.
The feature importance distributions of two missions were shown in Figure 2, which indicated different identification criteria. In Mission A, the most important feature is F20 (Variability Index). While in Mission B, it is F8 (Power-law Index). This is obviously in line with the AGN’s properties and exactly what we expected, for the reason that one of the distinctive properties of AGN is broadband variability, and the different radiation mechanisms of FSRQs and BL Lacs could explain their spectrum differences which behave as power-law index differences.

The likelihood distributions of identified AGN candidates and non-AGN candidates in Mission A are shown in Figure 6(a). In this figure, there are 22 AGN candidates and 21 non-AGN candidates with likelihood value lower than 0.6. And, there are 12 AGN candidates and 16 non-AGN candidates with latitude |b| < 10°. It can be attributed that the γ-ray emissions from the Galactic plane, especially from PSRs and supernova remnants (SNRs), are significant.

For Mission B, the likelihood distribution of identified FSRQ candidates and BL Lac candidates are shown in Figure 6(b). From the figure, there are 37 FSRQ candidates and 20 BL Lac candidates whose likelihood are lower than 0.6. In addition, there are 7 FSRQ candidates and 3 BL Lac candidates with latitude |b| < 10°. The best predictor for FSRQs and BL Lacs is the optical spectroscopy observation of emission lines. But for BCUs the optical spectrum is not available or not sensitive enough, then a ML method with outstanding performance has proven to be promising. Indeed, we will plan an optical observation for those FSRQ candidates and BL Lac candidates with likelihood less than 0.7 in our Mission B.

A comparison of variability index and the spectral index and 3FGL is shown in Table 11. We found that our results were consistent with 3FGL except for the average variability index for AGNs. Referring to Figure 4(a), the difference of average variability index is caused by the ‘long tail’ in the distribution of 3FGL AGNs. And this ‘long tail’ can be attributed by some variant AGNs, especially Blazars. These qualitative comparison showing that our results was not wrong, at least without significant mistakes, and implying that our method is functional and efficient to this identification.

5.2. Comparison with other studies

There are 708 3FGL unassociated sources can be found in 4FGL. Out of them, 537 sources were identified as AGNs and 171 sources were identified as non-AGNs in our Mission A. In the 537 predicted AGNs, 256 sources were confirmed as AGNs (1 AGN, 191 BCUs, 42 BL Lacs, 19 FSRQs, 3 radio galaxies (RDGs)) by 4FGL. While 18 sources were other certain types (14 pulsars (PSRs), 2 SNRs, 1 Supernova remnant / Pulsar wind nebula (SPP), 1 High-mass binary (HMB)) and 263 sources were still unknown in 4FGL. In the 171 predicted non-AGNs, there were 9 AGNs (8 BCUs and 1 BL Lac), 26 other certain types (1 Magellanic cloud (MC), 18 PSRs, 1 pulsar wind nebula (PWN), 1 SNR, 5 SPP) and 136 unknown in 4FGL.

Moreover, there are 544 3FGL BCUs can be found in 4FGL. Out of them, 221 sources and 323 sources were identified as FSRQs and BL Lacs in our Mission B, respectively. In the 221 predicted FSRQs, 19 sources were confirmed as FSRQs by 4FGL. While 25 sources were other certain types (2 AGN, 21 BL Lacs, 1 HMB, 1 PSR), 11 sources were unknown and 166 sources stayed BCUs in 4FGL. In the 323 predicted BL Lacs, 149 sources were confirmed as BL Lacs by 4FGL. While 9 sources were other certain types (5 FSRQs, 4 RDGs), 13 sources were unknown and 152 sources stayed BCUs in 4FGL.

The detailed comparison results were enumerated in Table 12. Through the comparison between our results and 4FGL catalogue, we have an accuracy of 91.3% for Mission A and an accuracy of 83.2% for Mission B. Besides, we noticed that our methods were outstanding to identify AGNs in Mission A and identify BL Lacs in Mission B, with precision 93.4% and 94.3%, respectively. While the lower precision for non-AGNs and FSRQs identification may result from the quantity of 4FGL confirmed sources with certain types.

In Saz Parkinson et al. (2016), they collected 3021 3FGL sources, including 1738 AGNs and 166 PSRs. Among the 3021 sources, there were 1008 3FGL sources remain unassociated.
These 1008 unassociated sources were identified as 631 AGNs and 377 PSRs with their algorithms. In our Mission A, we identified 748 AGNs and 262 non-AGNs from 1010 3FGL unassociated sources. Through a comparison listed in Table 13, we found that there were 608 AGN candidates from their work consistent with ours. But in our work, we focused more on the precision metric. Hence, the identified AGNs with our methods can be considered to be more reliable.

Chiari et al. (2016) identified 573 BCU sources into 154 FSRQ candidates, 342 BL Lac candidates and 77 stay uncertain. In comparison, we predicted 139 FSRQ candidates and 285 BL Lac candidates in common. Lefaucheur et al. (2017) also worked on this identification through ML methods. In their work, 486 BCUs were identified into 146 FSRQ candidates, 295 BL Lac candidates and 45 uncertain sources. Through a comparison, we found 144 FSRQ candidates and 282 BL Lac candidates were in common with theirs, as shown in Table 14.

In addition, both Chiari et al. (2016) and Lefaucheur et al. (2017) remained several BCUs uncertain which were thoroughly divided into FSRQ and BL Lac in our Mission B. There were 77 sources staying uncertain in Chiari et al. (2016) and 45 in Lefaucheur et al. (2017), with 9 common sources included. We summarized these uncertain sources and predicted them to be 74 FSRQs and 39 BL Lacs, as listed in Table 15. With our method, the uncertain BCUs were further identified.

These comparisons indicated that our method was functionally working well as those applied in other work, but we achieved better performance for both missions.

6. Conclusion

ML methods has proven to be a promising approach to process astronomical data and they provide classification based on high-dimensionality patterns that human investigation may miss in the first place. In this paper, we used ML methods to complete the source identification in 3FGL for further astrophysical studies. The FS and PCA techniques indeed helped to develop a more efficient algorithm, and ensemble models performed better on unseen samples. Grid search method was demonstrated to be of help to choose the hyper-parameters. With these ML methods, we have successfully identified 748 AGNs out of 1010 unassociated sources, with an accuracy of 97.04%. Within 573 BCUs, 326 have been identified as BL Lacs and 247 as FSRQs, with an accuracy of 92.13%.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China (NSFC 11733001, NSFC U1531245), Natural Science Foundation of Guangdong Province (2017 A030313011), supports for Astrophysics Key Subjects of Guangdong Province and Guangzhou City. Authors would like to acknowledge the scholarship sponsored by Guangzhou University, China and the excellent research facilities provided by Istituto Nazionale di Fisica Nucleare (INFN) and department of Physics and Astronomy of the University of Padova, Padova, Italy. We also thank Riccardo Rando, Alessandro Paccagnella and Denis Bastieri from Padova university for their constructive and valuable suggestions and comments.

References

Abadi, M., Barham, P., Chen, J., et al. 2016, USENIX Symposium on Operating Systems Design and Implementation, 16, 265
Abdo, A. A., Ackermann, M., Ajello, M., et al., 2010, ApJS, 188, 405
Abdo, A. A., Ackermann, M., Ajello, M., et al., 2010b, ApJ, 710, 810
Acero, F., Ackermann, M., Ajello, M., Albert, A., et al., 2015, ApJS, 218, 23
Atwood, W. B., Abdo, A. A., Ackermann, M., et al., 2009, ApJ, 697, 2, 1071
Bakshi, B. R., Stephanopoulos, G., 1993, AIChE, 39, 1, 57
Bengio, Y., et al., 2009, Foundations and trends in Machine Learning, 2, 1, 1
Böttcher, M., Reimer, A., & Marscher, A. P. 2009, ApJ, 703, 1168
Breiman, L. 2001, Machine Learning, 45, 5
Cavuoti, S., Brescia, M., D’Abrusco, R., et al., 2014, MNRAS, 968, 975
Chiari, G., Salvetti, D., La Mura, G., et al. 2016, MNRAS, 462, 1830
Dietterich, T. G., 2002, The handbook of brain theory and neural networks, MIT press, 405
Dimitrakoudis, S., Mastichiadis, A., Protheroe, R. J., Reimer, A. 2012, A&A, 546, 120
Doert, M., Errando, M., 2014, APJ, 782, 1, 41
Duchi, J., Hazan, E., Singer, Y., 2011, JMLR, 12, 2121
Fan, J. H., Yang, J. H., Liu, Y., et al. 2016, ApJS, 226, 20
Table 1: 21 features in 3FGL.

| Index | Feature | Description |
|-------|---------|-------------|
| F0    | LII     | Galactic longitude |
| F1    | BII     | Galactic latitude |
| F2    | Significance100MeV–300GeV | source significance in $\sigma$ units over the 100 MeV to 300 GeV band |
| F3    | $F_{1-100\text{GeV}}$ | integral photon flux from 1 to 100 GeV in units of $\text{ph} \cdot \text{cm}^{-2} \cdot \text{s}$ |
| F4    | Pivot Energy | energy at which error on differential flux is minimal |
| F5    | Spectral index | best fit photon number power-law index |
| F6    | Beta     | curvature parameter, for LogParabola |
| F7    | Cutoff Energy | cutoff energy for PL(Super)ExpCutoff |
| F8    | Power law index | best fit power-law index |
| F9    | $F_{100-300\text{MeV}}$ | integral photon flux from 100 to 300 MeV |
| F10   | Significance100–300MeV | significance on 100-300 MeV |
| F11   | $F_{0.3-1\text{GeV}}$ | integral photon flux from 0.3 GeV to 1 GeV |
| F12   | Significance0.3–1GeV | significance on 0.3-1 GeV |
| F13   | $F_{1-3\text{GeV}}$ | integral photon flux from 1 GeV to 3 GeV |
| F14   | Significance1–3GeV | significance on 1-3 GeV |
| F15   | $F_{3-10\text{GeV}}$ | integral photon flux from 3 GeV to 10 GeV |
| F16   | Significance3–10GeV | significance on 3-10 GeV |
| F17   | $F_{10-100\text{GeV}}$ | integral photon flux from 10 GeV to 100 GeV |
| F18   | Significance10–100GeV | significance on 10-100 GeV |
| F19   | Sign.Curve | fit improvement between power law and LogParabola or PLExpCutoff |
| F20   | Variability Index | sum of $2A\cdot\log(\text{Likelihood})$ difference between the flux fitted in each time interval and the average flux over the full catalog interval |

Table 2: RF model used to calculate the FI. Other parameters were set as default in scikit-learn package.

| Number of DTs | Criterions |
|---------------|------------|
| 10000         | entropy   |
| 30000         | gini      |
Figure 1: (a) Illustration of missions (top) and models (bottom). (b) Datasets of 3FGL sources split for Mission A (top) and B (bottom).
Figure 2: The descending FI values in the (a) third split case of Dataset A, and (b) second split case of Dataset B.
Figure 3: The variance values of components projected from (a) Dataset A in the third split case, and (b) Dataset B in the second split case.
Table 3: The removable features and retainable principal components in FS and FE step for two datasets, respectively.

| Dataset A | Group | Indices of removable features | Number of retainable principal components |
|-----------|-------|-------------------------------|------------------------------------------|
| Split 1   | 0     | None (keep 21 features)       | 14 / 7 / 4                               |
|           | 1     | (6, 0, 17, 8)                 | 10 / 4                                   |
|           | 2     | (6, 0, 17, 8, 16, 4, 12, 10, 14, 2, 1, 18) | 5 / 3                                   |
|           | 3     | (6, 0, 17, 8, 16, 4, 12, 10, 14, 2, 1, 18, 5, 9, 15, 3, 11) | 4                                         |
| Split 2   | 0     | None (keep 21 features)       | 14 / 7 / 4                               |
|           | 1     | (16, 17, 6, 8, 0)             | 10 / 5                                   |
|           | 2     | (16, 17, 6, 8, 0, 4, 12, 14, 10, 2, 1, 18) | 5                                         |
|           | 3     | (16, 17, 6, 8, 0, 4, 12, 14, 10, 2, 1, 18, 5, 9, 15, 11, 3) | 4                                         |
| Split 3   | 0     | None (keep 21 features)       | 14 / 9 / 4                               |
|           | 1     | (6, 0, 8, 17, 16)             | 10 / 5                                   |
|           | 2     | (6, 0, 8, 17, 16, 10, 12, 4, 1, 14, 2, 15, 18) | 5                                         |
|           | 3     | (6, 0, 8, 17, 16, 10, 12, 4, 1, 14, 2, 15, 18, 9, 5, 11, 3) | 4                                         |
| Split 5   | 0     | None (keep 21 features)       | 14 / 7 / 4                               |
|           | 1     | (14, 1, 3, 6, 7)              | 10 / 5                                   |
|           | 2     | (7, 6, 3, 0, 13, 14, 1, 15, 2, 12, 16) | 8 / 6                                   |
|           | 3     | (7, 6, 3, 0, 13, 14, 1, 15, 2, 12, 16, 11, 10, 9, 17, 19) | 5                                         |

| Dataset B | Group | Indices of removable features | Number of retainable principal components |
|-----------|-------|-------------------------------|------------------------------------------|
| Split 1   | 0     | None (keep 21 features)       | 12 / 6 / 4                               |
|           | 1     | (14, 1, 3, 6, 7)              | 10 / 5                                   |
|           | 2     | (14, 1, 3, 6, 7, 10, 16, 12, 2, 13, 15, 0) | 7 / 4                                   |
|           | 3     | (14, 1, 3, 6, 7, 10, 16, 12, 2, 13, 15, 0, 17, 19, 9, 11) | 5                                         |
| Split 2   | 0     | None (keep 21 features)       | 12 / 9 / 6                               |
|           | 1     | (7, 6)                        | 11 / 6 / 4                               |
|           | 2     | (7, 6, 3, 0, 13, 14, 1, 15, 2, 12, 16) | 8 / 6                                   |
|           | 3     | (7, 6, 3, 0, 13, 14, 1, 15, 2, 12, 16, 11, 10, 9, 17, 19) | 5                                         |
| Split 3   | 0     | None (keep 21 features)       | 13 / 9 / 6                               |
|           | 1     | (15, 2, 1, 13, 0, 14, 3, 6, 7) | 9 / 6 / 4                                |
|           | 2     | (15, 2, 1, 13, 0, 14, 3, 6, 7, 10, 16, 12) | 7 / 5                                   |
|           | 3     | (15, 2, 1, 13, 0, 14, 3, 6, 7, 10, 16, 12, 11, 9, 17, 19) | 5                                         |
| Split 4   | 0     | None (keep 21 features)       | 12 / 9 / 6                               |
|           | 1     | (1, 3, 6, 7)                  | 10 / 7 / 5                               |
|           | 2     | (1, 3, 6, 7, 12, 13, 2, 14, 15, 0) | 9 / 6                                   |
|           | 3     | (1, 3, 6, 7, 12, 13, 2, 14, 15, 0, 17, 9, 11, 19, 10, 16) | 5                                         |
| Split 5   | 0     | None (keep 21 features)       | 12 / 10 / 6                              |
|           | 1     | (14, 13, 3, 6, 7)             | 11 / 9 / 6                               |
|           | 3     | (1, 3, 6, 7, 12, 13, 2, 14, 15, 0, 17, 9, 11, 19, 10, 16) | 5                                         |

Table 4: The optional hyper-parameters of RF, MLP and GAN models.

| Hyper-parameter         | RF           | MLP             | GAN             |
|------------------------|--------------|-----------------|-----------------|
| Criterion              | gini / entropy |                 |                 |
| Number of DTs          | 50 / 100 / 300 / 500 |        |                 |
| Number of units        |               |                 |                 |
| hidden layer 1:        | 50 / 30       |                 |                 |
| hidden layer 2:        | 20 / 10       |                 |                 |
| Activation function    | reLU / elu / sigmoid / tanh / ReLU / softplus |      |                 |
| Loss function          | l2 / expectation / reg_expectation / chebyshev / hinge / hinge2 / hinge3 / log / log2 / tan / dcs | |
Table 5: Identification results and hyper-parameters of different candidate models in five split cases for Mission A. We combined the precision and accuracy as the metric. The hyper-parameters of soft ensemble models were consistent with the respective individual models. $w_1$, $w_2$, $w_3$, $w_4$ represent the weights of ‘RF_reduced’, ‘RF_pipe’, ‘MLP’ and ‘GAN’, respectively.

| Dataset A | Models    | Hyper-parameters                                                                 | Sensitivity | Precision | Accuracy |
|-----------|-----------|----------------------------------------------------------------------------------|-------------|-----------|----------|
|           | RF_reduced| removed features: Group 1 criterion: entropy number of trees: 50                   | 99.05%      | 96.83%    | 96.38%   |
| Split 1   | RF_pipe   | removed features: Group 1 number of retained principal components: 10 normalization: L2 Normalizer criterion: gini number of trees: 300 | 99.43%      | 96.48%    | 96.38%   |
|           | MLP       | removed features: Group 1 normalization: Standardizer structure: [17 30 20 10 2] activation function: tanh loss function: hinge3 | 99.24%      | 97.38%    | 97.04%   |
|           | GAN       | removed features: Group 1 normalization: Standardizer structure of generator: [10 10 20 30 17] structure of discriminator: [17 50 20 10 3] activation function: leaky relu loss function: expectation | 99.43%      | 97.02%    | 96.88%   |
|           | Soft ensemble | $w_1 = 0.05, w_2 = 0.07, w_3 = 0.31, w_4 = 0.57$ | 99.62%      | 97.03%    | 97.04%   |
| Split 2   | RF_reduced| removed features: Group 1 criterion: entropy number of trees: 50                   | 99.43%      | 96.48%    | 96.38%   |
|           | RF_pipe   | removed features: Group 1 number of retained principal components: 10 normalization: L1 Normalizer criterion: entropy number of trees: 300 | 98.47%      | 96.63%    | 95.72%   |
|           | MLP       | removed features: Group 0 normalization: Standardizer structure: [21 30 20 10 2] activation function: leaky relu loss function: expectation | 98.09%      | 96.98%    | 95.72%   |
|           | GAN       | removed features: Group 1 normalization: Standardizer structure of generator: [10 10 20 50 16] structure of discriminator: [16 50 20 10 3] activation function: leaky relu loss function: chebyshev | 98.09%      | 96.80%    | 95.56%   |
|           | Soft ensemble | $w_1 = 0.34, w_2 = 0.25, w_3 = 0.15, w_4 = 0.26$ | 99.43%      | 96.84%    | 96.71%   |
| Split 3   | RF_reduced| removed features: Group 1 criterion: gini number of trees: 100                    | 99.81%      | 96.49%    | 96.71%   |
|           | RF_pipe   | removed features: Group 2 number of retained principal components: 5 normalization: Max Normalizer criterion: entropy number of trees: 300 | 98.86%      | 96.28%    | 95.72%   |
| Model   | Removed features          | Normalization       | Structure                  | Activation function | Loss function  | Accuracy   | Macro F1  | Micro F1  |
|---------|---------------------------|---------------------|----------------------------|---------------------|---------------|-----------|-----------|-----------|
| MLP     | Group 1                   | Standardizer        | [16 50 10 10 2]            | relu                | reg, expectation | 98.47%    | 97.36%    | 96.38%    |
| GAN     | Group 1                   | Standardizer        | [10 10 10 30 16]           | leaky relu          | expectation   | 99.62%    | 97.03%    | 97.04%    |
| Soft ensemble | $w_1 = 0.01, w_2 = 0.01, w_3 = 0.01, w_4 = 0.97$ |            |                           |                     |               | 99.62%    | 97.03%    | 97.04%    |
| Reduced | Group 1                   |                     |                            |                     |               | 99.43%    | 95.95%    | 95.89%    |
| Split 4 | Group 2                   |                     |                            |                     |               | 99.24%    | 95.94%    | 95.72%    |
| RFpipe  |                            | Max Normalizer      | [10 20 50 10 3]            | gini                |               | 99.43%    | 95.63%    | 95.89%    |
| Soft ensemble | $w_1 = 0.58, w_2 = 0.26, w_3 = 0.09, w_4 = 0.07$ |            |                           |                     |               | 99.81%    | 96.49%    | 96.71%    |
| Reduced | Group 1                   |                     |                            |                     |               | 99.62%    | 96.13%    | 96.22%    |
| Split 5 | Group 1                   |                     |                            |                     |               | 99.43%    | 95.42%    | 95.40%    |
| RFpipe  |                            | MaxAbsScaler        | [21 30 10 10 2]            |                     |               | 99.43%    | 95.95%    | 95.89%    |
| Soft ensemble | $w_1 = 0.01, w_2 = 0.24, w_3 = 0.16, w_4 = 0.59$ |            |                           |                     |               | 99.43%    | 96.48%    | 96.38%    |
| MLP     | Group 0                   | Standardizer        | [21 30 10 10 2]            | elu                 | hinge3        | 98.86%    | 96.64%    | 96.05%    |
| GAN     | Group 0                   |                     | [10 10 10 50 21]           |                     | log2          | 99.43%    | 95.95%    | 95.89%    |
| Soft ensemble | $w_1 = 0.01, w_2 = 0.01, w_3 = 0.01, w_4 = 0.97$ |            |                           |                     |               | 99.43%    | 96.48%    | 96.38%    |
Table 6: Identification results and hyper-parameters of different candidate models in five split cases for Mission B. We just took the accuracy score as the metric. The hyper-parameters of soft ensemble models were consistent with the respective individual models. \( w_1, w_2, w_3, w_4 \) represent the weights of 'RF\_reduced', 'RF\_pipe', 'MLP' and 'GAN', respectively.

| Dataset B | Models       | Hyper-parameters                                                                 | Sensitivity | Precision | Accuracy |
|-----------|--------------|----------------------------------------------------------------------------------|-------------|-----------|----------|
| Split 1   | RF\_reduced  | removed features: Group 3 criterion: entropy number of trees: 50                | 89.90%      | 89.45%    | 88.05%   |
|           | RF\_pipe     | removed features: Group 0 criterion: StandardScaler number of retained principal components: 6 normalization: StandardScaler | 91.41%      | 89.16%    | 88.63%   |
|           | MLP          | removed features: Group 2 normalization: Standardizer structure: [9 50 10 10 2] activation function: leaky relu loss function: chebyshev | 91.92%      | 90.55%    | 89.80%   |
|           | GAN          | removed features: Group 1 normalization: Standardizer structure of generator: [10 10 20 30 16] structure of discriminator: [16 30 10 10 3] activation function: leaky relu loss function: expectation | 87.88%      | 94.05%    | 89.80%   |
| Soft ensemble | \( w_1 = 0.01, w_2 = 0.29, w_3 = 0.52, w_4 = 0.18 \) |                                                                      | 91.41%      | 92.35%    | 90.67%   |
| Split 2   | RF\_reduced  | removed features: Group 0 criterion: entropy number of trees: 500                | 91.41%      | 91.88%    | 90.38%   |
|           | RF\_pipe     | removed features: Group 1 criterion: Standardizer number of retained principal components: 11 normalization: Standardizer | 90.40%      | 92.27%    | 90.09%   |
|           | MLP          | removed features: Group 3 normalization: Standardizer structure: [5 50 10 10 2] activation function: tanh loss function: chebyshev | 92.42%      | 93.37%    | 91.84%   |
|           | GAN          | removed features: Group 3 normalization: Standardizer structure of generator: [10 10 20 50 5] structure of discriminator: [5 30 10 10 3] activation function: elu loss function: l2 | 92.42%      | 92.89%    | 91.55%   |
| Soft ensemble | \( w_1 = 0.01, w_2 = 0.19, w_3 = 0.73, w_4 = 0.07 \) |                                                                      | 92.42%      | 93.85%    | 92.13%   |
| Split 3   | RF\_reduced  | removed features: Group 0 criterion: entropy number of trees: 100                | 91.41%      | 90.05%    | 89.21%   |
|           | RF\_pipe     | removed features: Group 1 criterion: Max Normalizer number of retained principal components: 6 normalization: Max Normalizer | 92.93%      | 88.89%    | 89.21%   |
| Model   | removed features | normalization | structure | activation function | loss function | Accuracy 1 | Accuracy 2 | Accuracy 3 |
|---------|------------------|---------------|-----------|---------------------|--------------|------------|------------|------------|
| MLP     | Group 2          | Standardizer  | [9 30 10 10 2] | sigmoid            | hinge2       | 91.41%    | 91.41%    | 90.09%    |
| GAN     | Group 2          | Standardizer  | [10 10 10 50 9] | elu                | log2         | 91.41%    | 91.88%    | 90.38%    |
| Soft ensemble | $w_1 = 0.01, w_2 = 0.43, w_3 = 0.02, w_4 = 0.54$ | | | | | **93.43%** | 91.13% | 90.96% |
| RF reduced | Group 2         | gini          | | | | 91.41% | 90.96% | 89.80% |
| RF pipe | Group 1         | entropy       | [10 50 10 3] | | | 87.88% | 92.55% | 88.92% |
| MLP     | Group 3          | Standardizer  | [5 30 20 10 2] | tanh               | hinge2       | 90.91%    | 91.84%    | 90.09%    |
| GAN     | Group 1          | Standardizer  | [10 10 10 50 17] | relu               | tan          | 91.41%    | 92.35%    | 90.67%    |
| Soft ensemble | $w_1 = 0.02, w_2 = 0.01, w_3 = 0.17, w_4 = 0.80$ | | | | | 91.92% | 92.86% | 91.25% |
| RF reduced | Group 1         | gini          | | | | 92.42% | 88.41% | 88.63% |
| RF pipe | Group 2         | Standardizer  | [21 50 10 10] | | | 92.42% | 87.56% | 88.05% |
| MLP     | Group 0          | Standardizer  | [21 50 10 10] | relu               | reg         | 89.39%    | 91.24%    | 88.92%    |
| GAN     | Group 0          | Standardizer  | [10 10 10 30 21] | relu               | hinge3      | 87.88%    | 91.58%    | 88.34%    |
| Soft ensemble | $w_1 = 0.56, w_2 = 0.38, w_3 = 0.05, w_4 = 0.01$ | | | | | 91.41% | 90.50% | 89.50% |
Figure 4: The (a) AGN and (b) non-AGN logarithm variability index distributions of this work (red) and 3FGL (blue).
Figure 5: The (a) FSRQ and (b) BL Lac spectral index distributions of this work (red) and 3FGL (blue).
Figure 6: The probability distributions in (a) Mission A and (b) Mission B.
Table 7: Identification results of the 1010 unassociated sources with Best Model A.

| 3FGL name     | $l$($^\circ$) | $b$($^\circ$) | Spectral index | Variability index | Prediction | Likelihood |
|---------------|---------------|---------------|----------------|-------------------|------------|------------|
| J0000.1+6545  | 117.69        | 3.4           | 2.411          | 40.754            | non-AGN    | 0.56142    |
| J0000.2-3738  | 345.41        | -74.95        | 1.867          | 44.845            | AGN        | 0.98165    |
| J0001.6+3535  | 111.66        | -26.19        | 2.351          | 32.394            | AGN        | 0.95838    |
| J0002.0-6722  | 310.13        | -49.07        | 1.946          | 46.125            | AGN        | 0.97277    |
| J0002.6+6218  | 117.3         | -0.04         | 2.086          | 48.024            | AGN        | 0.81137    |

* The full version of this table is available online.

Table 8: Identification results of the 573 BCUs with Best Model B.

| 3FGL name     | $l$($^\circ$) | $b$($^\circ$) | Spectral index | Variability index | Prediction | Likelihood |
|---------------|---------------|---------------|----------------|-------------------|------------|------------|
| J0002.2-4152  | 334.07        | -72.15        | 2.089          | 56.351            | BL Lac     | 0.80919    |
| J0003.2-5246  | 318.97        | -62.83        | 1.895          | 45.28             | BL Lac     | 0.90905    |
| J0003.8-1151  | 84.43         | -71.09        | 2.024          | 48.492            | BL Lac     | 0.84678    |
| J0009.6-3211  | 1.26          | -79.61        | 2.31           | 42.534            | BL Lac     | 0.52062    |
| J0124.4+7040  | 119.68        | 8.04          | 2.484          | 51.576            | FSRQ       | 0.73085    |

* The full version of this table is available online.

Table 9: Confusion matrix of Best Model A on the third split test dataset.

| Number of Predicted Results | Predicted |
|-----------------------------|-----------|
|                             | AGN       | non-AGN   |
| Actual                      | 522 (TP)  | 2 (FN)    |
| non-AGN                     | 16 (FP)   | 68 (TN)   |

Table 10: Confusion matrix of Best Model B on the second split test dataset.

| Number of Predicted Results | Predicted |
|-----------------------------|-----------|
|                             | BL Lac    | FSRQ     |
| Actual                      | 183 (TP)  | 15 (FN)  |
| FSRQ                        | 12 (FP)   | 133 (TN) |

Table 11: Comparison of parameters from this work and 3FGL.

| Mission | Parameter | Class       | This Work          | 3FGL          |
|---------|-----------|-------------|--------------------|---------------|
| A       | Variability index | AGN        | 1.66 ± 0.12        | 1.91 ± 0.42   |
|         |           | non-AGN     | 1.65 ± 0.09        | 1.68 ± 0.13   |
| B       | Spectral index | FSRQ       | 2.47 ± 0.17        | 2.42 ± 0.22   |
|         |           | BL Lac      | 1.98 ± 0.22        | 2.02 ± 0.25   |

23
Table 12: Comparison of our results with 4FGL.

| Prediction of Mission A | 4FGL Confirmation | Precision | Accuracy |
|-------------------------|--------------------|-----------|----------|
| 537 AGNs                | 256 AGNs           | 263 Un    | 256/256 = 93.4% |
| 171 non-AGNs            | 9 AGNs             | 26 Others | 26/256 = 10.5%  |

| Prediction of Mission B | 4FGL Confirmation | Precision | Accuracy |
|-------------------------|--------------------|-----------|----------|
| 221 FSRQs               | 19 FSRQs           | 11 Un + 166 BCUs | 19/19 = 100% |
| 323 BL Lacs             | 149 BL Lacs        | 9 Others   | 149/149 = 100% |

1 ‘Un’ means unknown type of sources.

Table 13: Comparisons of our results with Saz Parkinson et al. (2016) in Mission A.

| Class       | Our result | P2016 | Common |
|-------------|------------|-------|--------|
| AGN         | 748        | 631   | 608    |
| non-AGN     | 262        |       |        |
| PSR         | 377        |       |        |

In total 1010 1008

Table 14: Comparisons of our results with Chiaro et al. (2016) and Lefaucheur et al. (2017) in Mission B.

| Class | Our result | C2016 | Common | L2017 | Common |
|-------|------------|-------|--------|-------|--------|
| FSRQ  | 247        | 154   | 139    | 146   | 144    |
| BL Lac| 326        | 342   | 285    | 295   | 282    |
| Uncertain | 0       | 77    | 45     |       |        |

In total 573 573 486

Table 15: 113 uncertain BCUs from Chiaro et al. (2016) and Lefaucheur et al. (2017).

| 3FGL name | C2016 | L2017 | TW | Likelihood |
|-----------|-------|-------|----|------------|
| J0002.2-4152 | Unc  | BL Lac |    | 0.80919    |
| J0021.6-6835 | Unc  | FSRQ  |    | 0.8107     |
| J0050.0-4458 | Unc  | FSRQ  |    | 0.78408    |
| J0055.2-1213 | Unc  | FSRQ  |    | 0.69895    |
| J0059.1-5701 | Unc  | FSRQ  |    | 0.82588    |

* The full version of this table is available online;
1 ‘Unc’ means ‘Uncertain type of BCUs’;
2 The last two columns indicate the source prediction and likelihood in our work.