The Astrophysical Journal Supplement Series, 275:11 (16pp), 2024 November https://doi.org/10.3847/1538-4365/ad7730

Chasing the Neutrino Blazar Candidates

Jingtian Zhu[1,2], Haitao Cao[3], Hubing Xiao[4], Zhiyuan Pei[2], JunHui Fan[2,5,6], and Denis Bastieri[1,2,7]
[1] Department of Physics and Astronomy "G. Galilei", University of Padova, Padova I-35131, Italy; denis.bastieri@unipd.it
[2] Center for Astrophysics, Guangzhou University, Guangzhou 510006, People's Republic of China; fjh@gzhu.edu.cn
[3] School of Information Engineering, Guangzhou Panyu Polytechnic, Guangzhou 511483, People's Republic of China; caoht@gzpyp.edu.cn
[4] Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai 200234, People's Republic of China
[5] Astronomy Science and Technology Research Laboratory of Department of Education of Guangdong Province, Guangzhou 510006, People's Republic of China
[6] Greater Bay Brand Center of the National Astronomical Data Center, Guangzhou 510006, People's Republic of China
[7] INFN—Istituto Nazionale di Fisica Nucleare, Sez. Padova, Padova I-35131, Italy
Received 2023 September 27; revised 2024 August 30; accepted 2024 August 30; published 2024 November 5

Abstract

In our study of the correlations between IceCube-detected neutrino events and γ-ray properties of blazars, we recognize the inherent challenges posed by the limited detection of neutrinos. In this paper, we explore few-shot learning to deal with the class imbalance and few-shot issues presented in the incremental version of the 12 yr Fermi-LAT γ-ray source catalog (4FGL_ DR3). Specifically, we train a triplet network to transform the blazars with neutrino emission (NBs) and nonblazar samples into an embedding space where their similarities can be measured. With two-way three-shot learning, 199 out of 3708 blazars without neutrino emission (non-NBs) are considered as the potential blazars emitting neutrinos (NB candidates, or NBCs for short), with a similarity score against NBs exceeding 98%. Moreover, the Kolmogorov–Smirnov test supports our identification of NBCs.

Unified Astronomy Thesaurus concepts: Blazars (164)

Materials only available in the online version of record: machine-readable tables

1. Introduction

1.1. Neutrino Detection and Blazars

Observation of neutrinos from extraterrestrial sources used to be limited to the Kamiokande II detector (K. Hirata et al. 1987) observing the neutrinos produced by the Sun and by the famous supernova 1987A, which emitted neutrinos with energy in the tens of MeV. In 2013, the IceCube Neutrino Observatory (hereafter IceCube; IceCube Collaboration 2005) detected 28 high-energy neutrinos between 30 TeV and 1.2 PeV. These extraterrestrial neutrinos were observed between 2010 and 2012, with no clear patterns in time or space. Moreover, this limited data makes conclusions unfirm (IceCube Collaboration 2013).

Further investigation was performed on six years worth of IceCube data, and the results were reported in M. G. Aartsen et al. (2016), which excluded that the detected neutrinos were produced in the atmosphere. The authors found no significant correlation between the directions of reconstructed neutrino events and the Galactic plane, concluding that the dominant fraction of the high-energy neutrino flux is isotropic. Additionally, an analysis of the arrival directions of neutrinos with reconstructed muon energies above 200 TeV found no correlation with known γ-ray sources. However, this could also be attributed to the limited data.

The discovery of high-energy neutrinos from extragalactic sources has heralded a new epoch in the field of neutrino astronomy (IceCube Collaboration 2013). It has been widely accepted that high-energy neutrinos are generated by interacting energetic cosmic rays with the surrounding matter or photon fields within the source. These interactions give rise to the creation of charged muons, which subsequently decay, producing neutrinos. Within the realm of potential sources for these high-energy neutrinos and cosmic rays, a range of extragalactic entities, including starburst galaxies (R.-Y. Liu et al. 2014; X.-C. Chang et al. 2015), tidal disruption events (C. Lunardini & W. Winter 2017; N. Senno et al. 2017), and active galactic nuclei (AGN; F. W. Stecker 2013; K. Murase et al. 2014; P. Padovani et al. 2015), are commonly postulated.

1.2. Blazars in Neutrino Astronomy

Blazars are distinguished among AGN subclasses by their unique characteristics (B. J. Wills et al. 1992; C. M. Urry & P. Padovani 1995; J.-H. Fan 2002; M. Villata et al. 2006; J.-H. Fan et al. 2014; H. B. Xiao et al. 2015; A. C. Gupta et al. 2016; H. Xiao et al. 2019; S. Abdollahi et al. 2020; J. H. Fan et al. 2021; J. Fan et al. 2023), including dominant radio and γ-ray emissions, rapid variability across wavelengths, and high linear polarization in radio and optical bands. Their nonthermal spectrum extends from radio to X-ray bands, marked by strong radiative luminosity. A classic categorization within blazars is based on the equivalent width (EW) of their emission lines: flat spectrum radio quasars (FSRQs) feature emission lines with EW $\geqslant$ 5 Å, whereas BL Lacertae objects (BL Lacs) typically lack emission lines or present an EW $\leqslant$ 5 Å (M. Stickel et al. 1991).

Blazars constitute the primary source of the extragalactic diffuse γ-ray background (M. Ajello et al. 2015). If the emitted γ-ray photons originate solely from hadron interactions, blazars could also make a substantial contribution to the diffuse neutrino background (A. Atoyan & C. D. Dermer 2001; C. R. Zhu et al. 2013; P. Padovani et al. 2016; A. Palladino et al. 2019). By reconstructing their arrival direction, muon tracks help locate γ-ray point sources related to neutrinos. The arrival directions of ν_μ (muon neutrinos) observed by IceCube for eight years with energies higher than 200 TeV (C. Haack et al. 2017) are almost isotropic, indicating that most high-energy neutrinos may come from extragalactic sources. Notably, while hadron interactions that produce neutrinos also emit γ photons, the IceCube data suggest a lack of significant correlation between the arrival directions of neutrinos and the known γ-ray point sources (C. Haack et al. 2017; F. Halzen 2017). By comparing the arrival positions of ν_μ with those associated with various classes of high-energy objects, the contribution of each object type to the diffuse neutrino background can be effectively constrained. Stacking analyses reveal that γ-ray bursts, at most, contribute only 1% to the neutrino background (M. G. Aartsen et al. 2017a), while blazars contribute less than 30% (M. G. Aartsen et al. 2017b).

Later, IceCube announced a high-energy neutrino event (IceCube-170922A) coincident in direction and time with a γ-ray flare. The event was also observed by the Fermi Large Area Telescope (hereafter Fermi-LAT; W. B. Atwood et al. 2009) with blazar TXS 0506+056 (IceCube Collaboration et al. 2018a), albeit with a significance of $<3\sigma$. Between 2014 September and 2015 March, there was an excess of events toward TXS 0506+056 above the atmospheric neutrino background detected with a significance of 3.5σ. This suggests that blazars with neutrino emission (NBs) are identifiable (IceCube Collaboration et al. 2018b).

However, understanding the mechanisms behind neutrino production in blazars is crucial. In the single-zone model, electromagnetic cascade and neutrino production occur within the same photon field. Studies of extremely high-energy track alerts from IceCube suggest that this model predicts a maximum annual detection rate of neutrinos of only 0.03 (IceCube Collaboration et al. 2018b; A. Keivani et al. 2018; S. Gao et al. 2019). Such a low detection rate implies a probability of less than 1% of detecting neutrinos from months-long blazar flares with IceCube, posing a significant challenge (K. Murase et al. 2018; H. Zhang et al. 2019; R. Xue et al. 2019a).

To address this challenge, a two-zone model has been proposed. In this model, neutrino production requires an external photon field provided by the broad-line region (BLR). The radiation region is divided into two distinct zones: the inner zone, located inside or close to the BLR, and the outer zone, situated at a considerable distance from the BLR (N. Sahakyan 2018; R.-Y. Liu et al. 2019). In the inner zone, relativistic protons interact with BLR clouds, utilizing the photons emitted by the BLR for both the $p\gamma$ process and the inverse Compton (IC) scattering process. The BLR clouds may also induce the *pp* process within the jet. In contrast, the outer zone faces challenges in obtaining an adequate external photon field for the $p\gamma$ process and a sufficient supply of target protons for the *pp* process.

Compared to the single-zone model, the two-zone model separates the radiation regions for the low-energy peak of blazars and neutrino emission in the spectral energy distribution (SED). This allows the inner zone to yield an order of magnitude higher neutrino flux, while the outer zone can meet observational constraints on the X-ray flux (K. Murase et al. 2018; R. Xue et al. 2019a; H. Zhang et al. 2019).

During the regular and increased activity, blazar emissions form a characteristic two-hump SED: the lower-energy bump is attributed to the synchrotron emission, while the process producing the higher-energy bump might be either one or a mixture of two different classes of processes: the leptonic model where the higher-energy bump is due to an IC process (R. D. Blandford & A. Koenigl 1979; M. Sikora et al. 1994; A. Sokolov & A. P. Marscher 2005; R. Xue et al. 2019b; G. Wang et al. 2022a), and the hadronic model where the bump can be interpreted as a secondary particle cascade initiated, typically, by high-energy protons (A. Mücke & R. J. Protheroe 2001; S. Dimitrakoudis et al. 2012; M. Cerruti et al. 2015; S. Gao et al. 2019; R. Xue et al. 2021; Z.-R. Wang et al. 2022b). In addition, hadronic models favor the production in the jet of neutrinos alongside high-energy protons and γ-rays (e.g., K. Mannheim 1995; F. Halzen & E. Zas 1997; A. Mücke et al. 2003; C. Guépin & K. Kotera 2017), so that a significant correlation of an event like IceCube-170922A with the direction of a blazar could tip the balance toward them.

1.3. Searching for Potential Neutrino-emitting Blazars with AI

In the meantime, some blazars have currently been identified as potential blazars emitting neutrinos (NB candidates, or NBCs for short), e.g., PKS 1424-41 (M. Kadler et al. 2016) and GB6 J1040+0617 (S. Garrappa et al. 2019), which display an increased γ-ray activity in temporal and spatial coincidence with high-energy neutrino events. Several other blazars are also listed as NBCs, like PKS 1502+106 in spatial coincidence with the event IceCube-190730A (S. Garrappa et al. 2022) and NVSS J065844+063711 with the alert IceCube-201114A (S. Garrappa et al. 2021; R. de Menezes et al. 2022). Moreover, A. Galván et al. (2022) reported 23 NBCs by using a spatial correlation of angular distance between the high-energy neutrino events detected by IceCube and the γ-ray blazars reported by Fermi-LAT.

However, current studies of NBs still need to answer if blazars account for most of the high-energy neutrino diffuse flux and what the lepton-to-hadron ratio in blazar jets is, primarily because of the limited number of events. Under the hypothesis that high-energy neutrinos are produced in the jets, we expect to find more significant correlations between neutrinos detected by IceCube and, for example, the Fermi-LAT catalogs (3FGL, F. Acero et al. 2015; 4FGL, S. Abdollahi et al. 2020), allowing us to increase the number of NBs and corroborate the hadronic scenario.

In recent years, artificial intelligence methods have been employed in many fields of astronomy (G. Chiaro et al. 2016; S.-J. Kang et al. 2019; H. B. Xiao et al. 2020). G. Chiaro et al. (2016) utilized blazar flaring patterns and designed an artificial neural network (ANN) to identify blazar candidates of uncertain types (BCUs) in 3FGL. H. B. Xiao et al. (2020) employed ensemble machine learning (ML) methods to identify AGNs among 3FGL unassociated sources and BCUs. More recently, J. T. Zhu et al. (2023) proposed a supervised ML method to sift TeV γ-ray candidates and provide a sample for ground-based Cherenkov telescopes. H. Cao et al. (2024) performed an in-depth analysis of the physical attributes of the sources and completed the categorization of the incremental version of the 12 yr Fermi-LAT γ-ray source catalog (4FGL_DR3) based on fractal dimension theory and wavelet transform.

In this work, we explore the few-shot learning (FSL) based on the ANN to search for the NBCs from the blazars without neutrino emission (non-NBs) of 4FGL_DR3. The paper is arranged as follows: In Section 2, we introduce the ANN and present the class imbalance and few-shot problems of 4FGL_DR3. The introduction to FSL and our method are detailed in Section 3. The analysis results are reported in

Section 4. Furthermore, the statistical characteristics of the found NBCs are analyzed in Section 5. Finally, the paper concludes in Section 6.

2. Background

2.1. Artificial Neural Networks

In ML, the term a sample (or a data point) is commonly utilized to denote an instance that contains multiple features. Let $\boldsymbol{x} \in \mathbb{R}^D$ denote a sample with D features, which typically include the source's physical attributes in astronomical data processing, such as redshift and flux density. The ground-truth label $y \in \mathcal{C}$ denotes the category to which the source belongs, where $|\mathcal{C}| = N$ is the cardinality of $\mathcal{C}$. For instance, in a binary classification task, we have $N = 2$ where 1 indicates the positive category, and 0 indicates the negative category. The whole data set typically comprises a collection of features and labels with M samples, i.e., $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{M}$.

ANNs, inspired by the biological neural networks in animal brains (W. S. McCulloch & W. Pitts 1943), have achieved breathtaking success in deep learning (DL). Comprising multiple layers populated by artificial neurons or nodes, ANNs adeptly facilitate complex data processing from the input layer through many hidden layers to the output layer, thus enabling intricate pattern recognition and predictive capabilities (Y. Bengio 2009; Y. LeCun et al. 2015; J. Ahmad et al. 2019).

The preference for employing DL methodologies based on ANNs over traditional ML techniques is substantiated by the universal approximation theorem. This theorem ensures that ANNs can approximate any continuous function, which gives the network a strong learning capability to mine the crucial information from the raw input data in complex tasks without requiring manual feature extraction (K. Hornik et al. 1989). Many achievements in real-life applications have demonstrated that DL can extract more discriminate features from the original data, thereby enhancing the effectiveness of ML (Y. Bengio et al. 2013).

The operational essence of ANNs is encapsulated in two fundamental processes: forward propagation and backward propagation (I. Goodfellow et al. 2016). These processes enable ANNs to learn from data dynamically and progressively refine their predictive accuracy. The forward process commences with the features of input samples. Each feature is linked to a node in the input layer; subsequent processing involves linear combinations of these features, adjusted by weights and bias. If we denote $\boldsymbol{a}^l \in \mathbb{R}^{H_l}$ as the input data of layer l, then the output data of layer $l + 1$ can be expressed as

$$\boldsymbol{a}^{l+1} = \delta(\boldsymbol{a}^l \boldsymbol{W}^l + b^l), \tag{1}$$

where $\boldsymbol{W}^l \in \mathbb{R}^{H_l \times H_{l+1}}$ and $b^l \in \mathbb{R}$ are the weights and bias of the hidden layer $l + 1$, respectively, and H_l is the number of nodes of layer l. After the linear projection, an activation function $\delta(\cdot)$ acts as a nonlinear transform for the final output. This transformation facilitates the advancement of data through the network until the output layer ultimately yields a prediction. In particular, $\boldsymbol{a}^l = \boldsymbol{x}$ for the input layer $l = 0$.

In summary, we can obtain the output of an ANN model for an input sample: $\hat{y} = F_\theta(\boldsymbol{x}) = \delta(\cdots\delta(\delta(\boldsymbol{x}\boldsymbol{W}^0 + \boldsymbol{b}^0)\boldsymbol{W}^1 + \boldsymbol{b}^1)\cdots)$, after data forwarding through multiple layers. The ANN can be seen as a nested nonlinear function that maps the input data to an embedding space, with learnable parameters θ (i.e., weights and bias): F_θ: $\mathbb{R}^D \to \mathbb{R}^E$, $\quad x \mapsto \hat{y}$. E is the number of nodes of the output layer. Thus, here we define a E-dimensional embedding space.

Following this, the backward propagation phase targets network optimization by evaluating the loss value $\mathcal{L} = G(\boldsymbol{x}, \hat{y})$, a measure of the discrepancy between the predicted outputs and the ground-truth labels. $G(\cdot)$ is a predefined loss function that guides the adjustment of weights and bias to minimize $\mathcal{L}$ through a process known as gradient descent. This iterative adjustment, governed by the learning rate, substantially enhances the network's ability to extract features from the training data, thereby improving its predictive performance. The backward propagation, seen as an optimization process, can be expressed as

$$\Theta = \arg\min_\theta \sum_{i=1}^{M} G(\boldsymbol{x}_i, F_\theta(\boldsymbol{x}_i)), \tag{2}$$

where $\boldsymbol{x}_i$ is the ith sample of the data set $\mathcal{D}$ containing M samples, and F_θ is the mapping function of ANN model. After the training process finished, we obtained a converged ANN model with optimal parameters Θ.

Typically, the data set $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{M}$ is divided into a training set $\mathcal{D}_t = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{T}$ and a validation set $\mathcal{D}_v = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{V}$. The former leads the model to learn features from data, while the latter is often used to monitor the troublesome overfitting phenomenon during training. Overfitting means the ANN model excessively learns the details instead of general patterns from the training data while generalizing poorly to the validation data. In this paper, the $\mathcal{D}$ was randomly partitioned into $\mathcal{D}_t$ and $\mathcal{D}_v$ 20 times. The model demonstrating the lowest loss on the validation set was selected as the optimal model. Dropout, which randomly omits a subset of nodes during training, is one of the most widely used techniques to alleviate overfitting and enhance the generalization ability of models (N. Srivastava et al. 2014).

Most ANN-based classifiers presuppose a balanced distribution of classes within the data set (J. Ortigosa-Hernández et al. 2016). Nevertheless, this assumption does not always hold in practical scenarios. The volume of one category within a data set may be substantially smaller than others, which is referred to as the class imbalance problem (M. Kubat et al. 1998; H. He & X. Shen 2007; H. He & E. A. Garcia 2009; J. Ortigosa-Hernández et al. 2017). Many pieces of research have established that the class imbalance problem skews the model toward the majority class, degrading the performance of traditional DL methods (N. Japkowicz & S. Stephen 2002; C. Drummond & R. C. Holte 2005; M. A. Mazurowski et al. 2008; M. Buda et al. 2018). Besides, another problem exists in our data set, i.e., there are few samples in each class, which will be detailed in the next section.

2.2. Samples of 4FGL_DR3

The Fermi-LAT Collaboration has recently released 4FGL_DR3 (S. Abdollahi et al. 2022), which comprises 6659 γ-ray sources and represents a comprehensive data set of high-energy astrophysical objects. This data set can be divided into subsets based on whether the sources are categorized into blazar samples (or blazars) and nonblazar samples (or nonblazars). As discussed in Section 1, in this work, we divide blazars into two subsets: NBs and non-NBs.

It is noteworthy that there are some unassociated and unknown sources among the nonblazars. The latter consists of

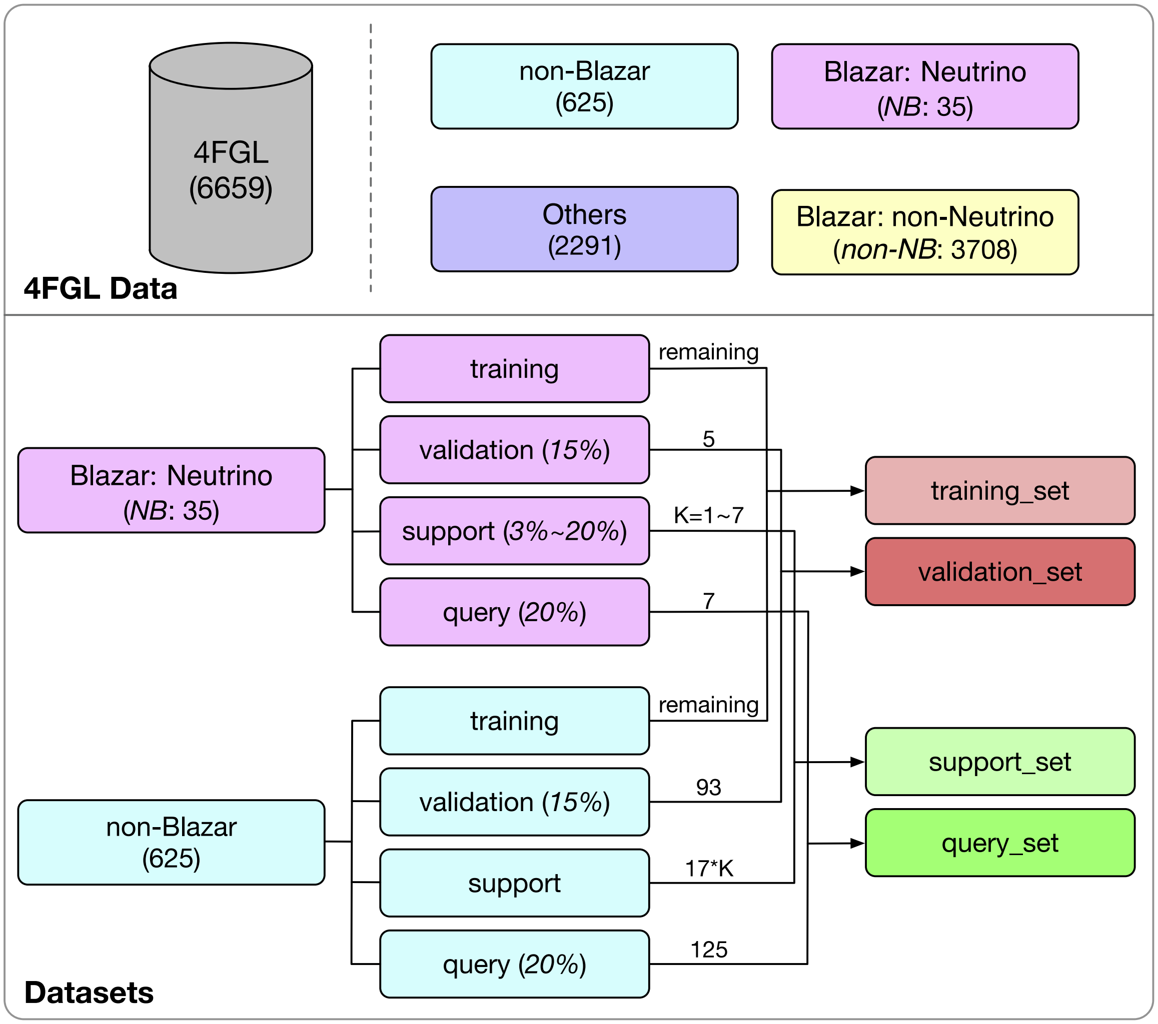

Figure 1. The top panel shows the samples of 4FGL_DR3: 35 NBs, 3708 non-NBs, 625 nonblazars, and 2291 other sources. The bottom panel shows the 35 NBs and 625 nonblazars divided into four data sets: the training set consisting of 304–412 sources for pretraining, the validation set consisting of 98 sources for monitoring performance during training, the support set consisting of 18–126 sources for providing embedding templates, and the query set consisting of 132 sources for determining the K value in our two-way K-shot task.

γ-ray sources located at Galactic latitudes $|b| \leqslant 10^\circ$ and are solely associated through the likelihood-ratio method with large radio and X-ray surveys. Given their ambiguous classifications, these sources do not fulfill the standard requirements for supervised learning data sets containing clearly labeled categories. Thus, we omitted these two subsets from further consideration for this discussion when referring to nonblazars (henceforth known as "others"). The division for NBs, non-NBs, nonblazars, and others are shown in the top panel of Figure 1. More details are given as follows.

Blazars. The 4FGL_DR3 presents a comprehensive compilation of 3743 blazars (hereafter 4FGL_DR3 blazars, S. Abdollahi et al. 2022), including 794 FSRQs, 1456 BL Lacs, and 1493 BCUs. The catalog provides essential astrophysical information for each source, enabling further investigations into the properties and energetics of these high-energy γ-ray emitters. This release represents a rich inventory of our expectations of the NBCs in the blazar population.

Nonblazars. After excluding 3743 blazars, 2916 sources remain in 4FGL_DR3. Subsequently, 2157 unassociated sources and 134 sources with unknown classifications (i.e., other sources) are removed, resulting in 625 nonblazars with 20 definitive classes: nine nonblazar active galaxies (nonblazar AGNs), seven binaries, five compact steep-spectrum sources, six normal galaxies, one galactic center, 35 globular clusters, 11 high-mass binaries, eight low-mass binaries, 155 millisecond pulsars identified by pulsations, eight narrow-line Seyfert 1s, four novae, 137 young pulsars identified by pulsations, 20 pulsar wind nebulae, 45 radio galaxies, eight starburst galaxies, two Seyfert galaxies, five star-forming regions, 43 supernova remnants, 114 supernova remnant/pulsar wind nebulae, and two steep-spectrum radio quasars.

NBs and non-NBs. Numerous investigations have been dedicated to identifying blazar counterparts of neutrinos, extensively leveraging the blazar catalogs obtained by Fermi-LAT (D. Gasparrini et al. 2012; P. L. Nolan et al. 2012; M. Ajello et al. 2017; B. Lott et al. 2020; M. Ajello et al. 2020; S. Abdollahi et al. 2020, 2022). We curate a compilation of eight works (e.g., M. Kadler et al. 2016; S. Garrappa et al. 2019, 2021; A. Galván et al. 2022; N.-H. Liao et al. 2022; R. de Menezes et al. 2022; R.-L. Li et al. 2022; S. Garrappa et al. 2022) derived from the research mentioned above regarding the blazar counterparts of neutrinos. These works present a comprehensive analysis of 35 blazars reported in 4FGL_DR3,

Table 1
The 35 NBs in 4FGL_DR3

4FGL Name (1)	Class in 4FGL_DR3 (2)	Class in J.-H. Fan (2022) (3)	Reference (4)
4FGL J0006.4+0135	BLL		R.-L. Li et al. (2022)
4FGL J0118.7-0848	BCU	FSRQ	R.-L. Li et al. (2022)
4FGL J0148.6+0127	BLL		A. Galván et al. (2022)
4FGL J0206.4-1151	FSRQ		S. Garrappa et al. (2022)
4FGL J0244.7+1316	BCU	FSRQ	A. Galván et al. (2022)
4FGL J0258.1+2030	BLL		A. Galván et al. (2022)
4FGL J0420.3-3745	BCU	BLL	A. Galván et al. (2022)
4FGL J0428.6-3756	BLL		A. Galván et al. (2022)
4FGL J0509.4+0542	BLL		A. Galván et al. (2022)
4FGL J0525.6-2008	BLL		A. Galván et al. (2022)
4FGL J0609.5+1402	BCU	FSRQ	A. Galván et al. (2022)
4FGL J0649.5-3139	BLL		A. Galván et al. (2022)
4FGL J0658.6+0636	BCU	BLL	S. Garrappa et al. (2021), A. Galván et al. (2022), R. de Menezes et al. (2022)
4FGL J0725.8-0054	BCU	BLL	A. Galván et al. (2022)
4FGL J0738.1+1742	BLL		S. Garrappa et al. (2021)
4FGL J0946.2+0104	BLL		S. Garrappa et al. (2022)
4FGL J1003.4+0205	BCU	BLL	S. Garrappa et al. (2022)
4FGL J1039.6+0535	BCU	BLL	A. Galván et al. (2022)
4FGL J1040.5+0617	BLL		S. Garrappa et al. (2019), A. Galván et al. (2022)
4FGL J1043.6+0654	BLL		A. Galván et al. (2022)
4FGL J1210.3+3928	BLL		R.-L. Li et al. (2022)
4FGL J1220.1+3432	BLL		A. Galván et al. (2022)
4FGL J1231.5+1421	BLL		A. Galván et al. (2022)
4FGL J1342.7+0505	BLL		S. Garrappa et al. (2022)
4FGL J1359.1-1152	BCU	BLL	A. Galván et al. (2022)
4FGL J1427.0+2348	BLL		R.-L. Li et al. (2022)
4FGL J1427.9-4206	FSRQ		M. Kadler et al. (2016)
4FGL J1504.4+1029	FSRQ		S. Garrappa et al. (2022)
4FGL J1505.0-3433	BLL		A. Galván et al. (2022)
4FGL J1543.0+6130	BLL		R.-L. Li et al. (2022)
4FGL J1744.9-1727	BCU	BLL	A. Galván et al. (2022)
4FGL J1751.6-1750	BCU	BLL	A. Galván et al. (2022)
4FGL J1808.8+3522	BLL		A. Galván et al. (2022)
4FGL J2113.9+1120	BCU	FSRQ	N.-H. Liao et al. (2022)
4FGL J2227.9+0036	BLL		A. Galván et al. (2022)

(This table is available in machine-readable form in the online article.)

showcasing a notable temporal and spatial correlation with neutrino emissions. In this paper, these 35 blazars are taken as NBs, while the remaining 3708 blazars that are not certified as NBs are recorded as non-NBs. The basic information of the 35 NBs is listed in Table 1, where column (1) gives the NB names in 4FGL_DR3, column (2) is the SED class in 4FGL_DR3, column (3) shows the classification results of BCUs according to J.-H. Fan et al. (2022), and column (4) lists the references.

As described, we deal with 35 NBs, 625 nonblazars, and 3708 non-NBs in this work. This data set suffers from a class imbalance problem since there is a significant disparity in the proportion, namely, NBs:nonblazars:non-NBs ≈1:18:106. In addition, there are few samples in each class, especially the NBs, which we call the "few-shot" problem. In this case, learning the crucial features from this limited data is challenging with traditional DL methods.

2.3. Class Imbalance Solutions

The solutions to reducing class imbalance problems have been developed through much work (G. M. Weiss 2004, 2005), and can be roughly categorized into two levels. Data-level methods, also known as resampling methods, aim to refine the training data so that standard learning algorithms can be effectively applied. These methods include undersampling the majority class and oversampling the minority class. Techniques such as NearMiss, edited nearest neighbors, and Tomeklink are examples of undersampling, whereas SMOTE, ADASYN, and Borderline-SMOTE represent oversampling methods (D. L. Wilson 1972; I. Tomek 1976; N. V. Chawla et al. 2002; J. Zhang & I. Mani 2003; H. Han et al. 2005; H. He et al. 2008). After resampling, the volumes of different classes become comparable for DL methods.

Moreover, data augmentation and generative techniques also play an important role. These include adding random noise to original data and generating synthetic samples through models like variational autoencoding networks and generative adversarial networks (I. Goodfellow et al. 2014; L. Pinhero Cinelli et al. 2021). In addition, G. Batista et al. (2004) utilized a hybrid method to combine oversampling and undersampling for better performance. However, these generative techniques require the model to learn the crucial features of the original data and generate samples with a distribution that is as close to the original data as possible. Therefore, generating a highquality sample from a data set containing few samples per class remains challenging.

As the second type of method to alleviate class imbalance, the algorithm-level methods involve adapting existing DL algorithms to be sensitive to class proportions. Cost-sensitive learning, which assigns higher misclassification costs to the minority class, exemplifies this approach (Y. Freund & R. E. Schapire 1997; C. Elkan 2001). Prioritizing the minority class during training helps alleviate the inherent bias toward the majority class. Additionally, there has been investigative work into how ANNs address the class imbalance issue, including the development of novel loss functions specifically tailored for the training of ANNs (S. Wang et al. 2016), as well as a two-phase training method for convolutional neural networks that initially trains on a class-balanced data set followed by fine-tuning of the output layer (M. Havaei et al. 2017).

3. Methodology

3.1. Few-shot Learning

This paper investigates the FSL method for class imbalance and few-shot problems. FSL represents a significant paradigm shift within the field of DL, specifically addressing the challenge of deriving meaningful inferences from a limited data set size. Contrary to the traditional DL methods that leverage abundant data to minimize generalization errors and achieve superior performance, FSL incorporates meta-learning principles to optimize its learning strategy across diverse tasks, thus acquiring a broadly applicable and generalized effective strategy even in data-scarce environments.

In the related work, training, and validation sets defined in DL are referred to as support and query sets respectively in FSL, collectively called a few-shot episode (G. S. Dhillon et al. 2019). Let $\mathcal{D}_s = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{S}$ and $\mathcal{D}_q = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{Q}$ denote the support and query sets, respectively, where $y_i \in \mathcal{C}_f$ for some set of classes $\mathcal{C}_f$. The number of ways, or classes, is $|\mathcal{C}_f| = N$. The set $\{\boldsymbol{x}_i | y_i = k, (\boldsymbol{x}_i, y_i) \in \mathcal{D}_s\}$ is the support of class k and its cardinality is s support shots. Similarly, the set $\{\boldsymbol{x}_i | y_i = k, (\boldsymbol{x}_i, y_i) \in \mathcal{D}_q\}$ is the query of class k and its cardinality is q query shots. The numbers s and q are small positive values, leading to the term "few-shot."

The support set lays the foundation for the model's learning, enabling it to discern among the N classes from minimal data exposure. This set is meticulously designed to impart the necessary knowledge to the model with sparse data. Then, the model's generalization capability learned from this limited data is evaluated using the query set. In this phase, the model is believed to identify the unseen samples of the query set based on the insights garnered from the support set. Each class in both the support set and query set is represented by K samples, resulting in the so-called "N-way K-shot" task. In this paper, we ranged K from 1 to 7 to explore the effectiveness of FSL.

We follow a DL approach to do the FSL, where a specific ANN architecture (called triplet network, see Section 3.2) is constructed to transform input data into an embedding space through multilayer nonlinear mappings (Y. LeCun et al. 2015; J. Ahmad et al. 2019). Based on the analysis of the 4FGL_DR3 samples in Section 2.2, we prepared the data sets for three steps in our FSL method. Specifically, in the first pretraining step, the training and validation sets featured two classes (NBs and nonblazar) used to pretrain and choose a best-performing ANN model, respectively. Second, the support and query data sets were transformed into the embedding space for determining a specific K value in our N-way K-shot task under the F-score metric. Finally, the degrees of similarity between the embeddings of non-NBs and NBs were computed to consider the NBCs in the identification process. In these steps, the model deals with two types of samples, i.e., NBs and nonblazar, or non-NBs and NBs. Thus, our method is considering a two-way K-shot task. These three steps are illustrated in the bottom panel of Figure 2. In subsequent sections, we will detail each processing stage, explaining the methods and techniques to achieve these objectives.

3.2. Triplet Networks

A type of novel architecture named a Siamese network was introduced by J. Bromley et al. (1993) to solve signature verification problems. In recent years, it has evolved into a robust framework for applications requiring an assessment of similarity or relationship between input samples. These networks employ a unique architecture of twin networks joined at their outputs. The twin networks are identical, sharing the same parameters, ensuring that two comparable input vectors yield similar output vectors. Siamese networks are particularly effective when the samples are limited or fine-grained differentiation is required.

The triplet network originates from Siamese networks. However, it employs a triplet-based structure to learn embeddings such that similar samples are placed closer in the embedding space while dissimilar samples are placed further apart. Instead of learning the discriminate features of data and categorizing the input sample into a distinct class in traditional ML or DL methods, triplet networks focus on learning similarities or dissimilarities between pairs of inputs. This characteristic is particularly crucial in scenarios where the available training data is insufficient to train an ANN without overfitting (D. Chicco 2021).[8]

The key of the triplet network is the triplet-loss function, which was introduced to the FaceNet system for face recognition, extending the concept of learning from comparisons by using triplets of samples (F. Schroff et al. 2015). A triplet consists of an anchor sample, a positive sample (similar to the anchor), and a negative sample (dissimilar from the anchor), as shown in the top panel of Figure 2. Minimizing the triplet loss means simultaneously bringing the anchor and positive samples closer in the embedding space while pushing the anchor and negative samples further apart.

The triplet loss is computed with Equation (3) (F. Schroff et al. 2015). By feeding the samples into the triplet network, the objective of the loss function is to train the model such that the anchor $\boldsymbol{x}_a$ is closer to the positive sample $\boldsymbol{x}_p$ than the negative sample $\boldsymbol{x}_n$ in the embedding space. $F_\theta(\cdot)$ is the triplet network and m is the margin quantifying the distance between two resulting clusters:

$$\mathcal{L}_0 = \max(0, m + \|F_\theta(\boldsymbol{x}_a) - F_\theta(\boldsymbol{x}_p)\|_2^2 - \|F_\theta(\boldsymbol{x}_a) - F_\theta(\boldsymbol{x}_n)\|_2^2). \tag{3}$$

Moreover, in the data analysis phase, we used a weighting trick during training to further alleviate the class imbalance problem. We added a weight $w_a > 1$ to the anchor sample if it is a positive sample (NB in our task), as formulated in Equation (4). By adjusting this weighting factor, the model can

[8] https://www.cs.utoronto.ca/~rsalakhu/papers/oneshot1.pdf

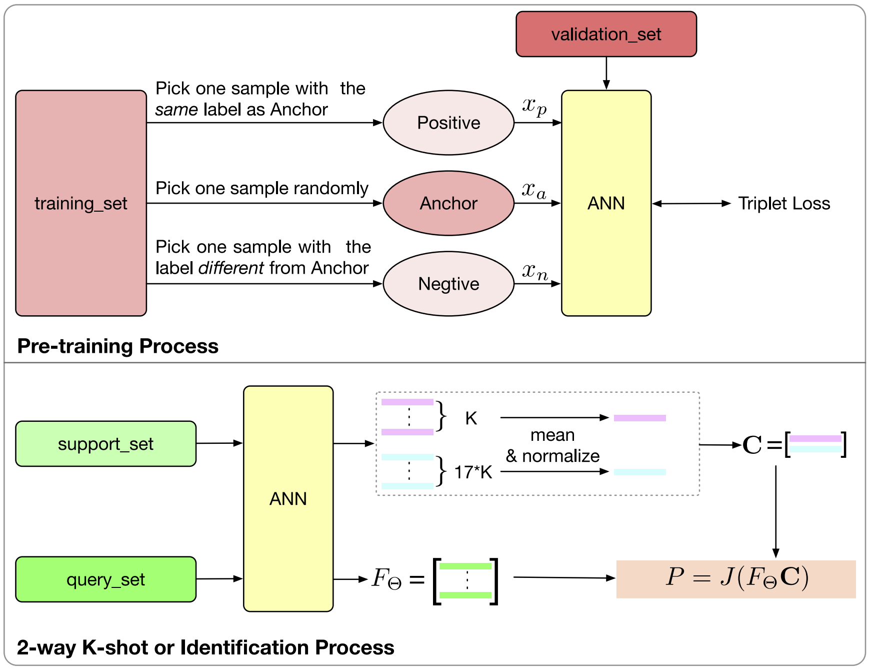

Figure 2. The top panel shows the ANN (triplet network) pretraining process with triplet loss. Random triplets are selected from the training set, and the ANN with the smallest triplet loss on the validation set is chosen. The bottom panel demonstrates the use of the trained ANN to transform the support set and query set into the embedding space, where their similarity score is calculated. In the two-way K-shot process, the softmax function is taken as $J(\cdot)$ to calculate the similarities, as shown in Equation (6). While in the identification process, $J(\cdot)$ becomes cosine similarity for further analysis, as shown in Equation (7).

be guided to focus more on reducing the distance to positive samples.

$$\tilde{\mathcal{L}}_0 = \mathcal{L}_0 \cdot w_a. \tag{4}$$

3.3. Feature Engineering

Before dividing the samples into different data sets for our method, the original samples of 4FGL_DR3 need to be preprocessed, which is often called feature engineering (G. Dong & H. Liu 2018) and involves four steps in our work:

1. *Feature discretization*. In 4FGL_DR3, the feature *nuFnu_Band* for a given source is a tuple of eight values: *nuFnu_Band*1 to *nuFnu_Band*8. To analyze it, we split the data into eight separate features, corresponding to eight SED bands: *nuFnu_band*1 (50–100 MeV), *nuFnu_band*2 (100–300 MeV), *nuFnu_band*3 (300–1000 MeV), *nuFnu_band*4 (1–3 GeV), *nuFnu_band*5 (3–10 GeV), *nuFnu_band*6 (10–30 GeV), *nuFnu_band*7 (30–100 GeV), and *nuFnu_band*8 (100–1000 GeV).
2. *Feature selection*. First, the string features were not considered. Simultaneously, we select only one from the physic view when multiple features show strong correlations. For example, we chose the latter from the integrated photon flux (*Flux_Band*) and SED (*nuFnu_Band*) per energy interval in 4FGL_DR3. Moreover, although 4FGL_DR3 provides three distinct spectral types, i.e., *PowerLaw*, *LogParabola*, and *PLSuperExpCutoff* (see details in Section 3.4 of S. Abdollahi et al. 2022), it is more appropriate to use *LogParabola* to describe the large sample of blazars in Fermi-LAT, since it is a good description for many γ-ray blazar spectra. So, we standardized the spectral type as *LogParabola* for all samples. In the end, we selected 16 features for our data analysis phase, as listed in Table 2. Columns (1) and (2) denote the names and units of the features in 4FGL_DR3, respectively; column (3) provides the descriptions.
3. *Feature cleaning*. Some samples may have missing values where we filled in the mean value. For instance, if a BL Lac source misses redshift *nuFnu_band*8, we will fill it in with the mean of other BL Lacs. Among the 16 selected features, three have missing values: each one of the variability index (V), fractional variability (V_{F}), and nuFnu Band8 ($\log F_8$) has one missing value. Therefore, replacing so few missing values with the mean of other available data is unlikely to significantly impact the data distribution or introduce bias. This approach allows us to utilize other sources' features while maintaining the samples' integrity.
4. *Feature normalization*. We standardized all features by removing the mean and scaling to unit variance, ensuring the individual values would not dominate the data set.

3.4. Flowchart and Data Sets

As mentioned, our FSL method involves three steps. The strategies of data set division, shown in the bottom panel of Figure 1, are to be detailed according to these steps.

1. *Pretraining*. Pretraining is an essential step in FSL, where a triplet network is initially trained on a more reasonable and diverse data set. This phase helps the model learn general features and representations, providing a solid base for further utilization of more specific data. During pretraining, the model undergoes multiple epochs with techniques like dropout and learning-rate adjustments to enhance robustness. The pretraining phase involves the training set and validation set.

   The training set is dedicated to leading the model to learn general features and representations, as formulated

Table 2
The 16 Selected Features in 4FGL_DR3

Feature (1)	Unit (2)	Description (3)
Pivot Energy (E_P)	GeV	Energy at which the error on differential flux is minimal
Energy Flux100 ($\log F$)	erg cm^{-2} s^{-1}	Energy flux from 100 MeV to 100 GeV obtained by spectral fitting
LP Flux Density ($\log f$)	cm^{-2} MeV^{-1} s^{-1}	Differential flux at Pivot Energy in LogParabola fit
LP_Index (α_2)		Photon index at Pivot Energy when fitting with LogParabola
LP_beta (β)		Curvature parameter when fitting with LogParabola
LP_Epeak (E_{peak})		Peak energy in νF_ν estimated from the LogParabola model
Variability Index (V)		The variability index is the sum of 2 × log(Likelihood) differences between the flux fitted in each 1 yr interval and the average flux over the entire catalog interval. A value greater than 24.72 across 12 intervals suggests there is less than a 1% chance of the source being steady.
Frac. Variability (V_F)		It measures each source's excess variance beyond statistical and systematic fluctuations, calculated as $\frac{\delta F}{F_{av}}$. It typically ranges between 50% and 90%. Most blazars exhibit fractional variability exceeding 10%
nuFnu Band1 ($\log F_1$)	erg cm^{-2} s^{-1}	Spectral energy distribution in 50–100 MeV
nuFnu Band2 ($\log F_2$)	erg cm^{-2} s^{-1}	Spectral energy distribution in 100–300 MeV
nuFnu Band3 ($\log F_3$)	erg cm^{-2} s^{-1}	Spectral energy distribution in 300–1000 MeV
nuFnu Band4 ($\log F_4$)	erg cm^{-2} s^{-1}	Spectral energy distribution in 1–3 GeV
nuFnu Band5 ($\log F_5$)	erg cm^{-2} s^{-1}	Spectral energy distribution in 3–10 GeV
nuFnu Band6 ($\log F_6$)	erg cm^{-2} s^{-1}	Spectral energy distribution in 10–30 GeV
nuFnu Band7 ($\log F_7$)	erg cm^{-2} s^{-1}	Spectral energy distribution in 30–100 GeV
nuFnu Band8 ($\log F_8$)	erg cm^{-2} s^{-1}	Spectral energy distribution in 100–1000 GeV

in Equation (2). It usually contains most of the data to ensure comprehensive learning. Due to the ranging K, the training set comprised 55%–62% of the original NBs and nonblazars, i.e., 16–22 NBs and 288–390 nonblazars, which resulted in a training set with a total of 304–412 sources.

During pretraining, the validation set serves as a checkpoint to evaluate the model's performance, helping to avoid critical overfitting. It remains separate from the training set to ensure a reliable assessment of the model's generalization capability. The best-performing model F_Θ, obtaining a low loss value with Equation (4) on the validation set, is chosen for the next steps. In our data analysis phase, the validation set was constructed by taking 15% of the original NBs and nonblazars, yielding five NBs and 93 nonblazars.

2. *Two-way K-shot*. In this step, the best-forming triplet network F_Θ transforms the support set and query set into the embedding space for similarity computing. In the context of FSL, the support set plays a pivotal role. It contains a limited number of samples for each class and is used to derive the embedding templates. The query set evaluates the network's ability to identify unseen samples based on the embedding templates.

As expressed in Section 3.1, for each class k of the support set $\mathcal{D}_s$, we calculate the mean vector of the support points $\{\boldsymbol{x}_i | y_i = k, (\boldsymbol{x}_i, y_i) \in \mathcal{D}_s\}$ in the embedding space:

$$\boldsymbol{c}_k = \frac{1}{s_k} \sum_{(\boldsymbol{x}_i, y_i) \in \mathcal{D}_s} F_\Theta(\boldsymbol{x}_i), \tag{5}$$

where $\boldsymbol{c}_k \in \mathbb{R}^E$, E is the dimension of embedding space, as defined in Section 2.1. s_k is the support shots of class k. Therefore, the embedding templates $\boldsymbol{C} \in \mathbb{R}^{E\times N}$ can be written as the concatenation of $\boldsymbol{c}_k$: $\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, ... \boldsymbol{c}_N\}$, where N is the number of classes. In our two-way K-shot task, $N = 2$.

Then, an inner product between the embedding templates $\boldsymbol{C}$ and the query set is performed in the embedding space, producing a similarity distribution over classes for a query sample $\boldsymbol{x}_i$, which can be expressed as

$$\boldsymbol{p}_i = \mathrm{Softmax}(\mathrm{Norm}(F_\Theta(\boldsymbol{x}_i)) \cdot \mathrm{Norm}(\boldsymbol{C})), \tag{6}$$

where Norm(·) is an L2-normalization along the embedding axis. Softmax(·) outputs a vector of probabilities that sum to 1, representing the likelihood of each class being the true class. For an output node, the higher the probability is, the more similarly the query sample belongs to the corresponding support class. Therefore, the index with the highest probability in the vector $\boldsymbol{p}_i$ is taken as the predicted class of the query sample $\boldsymbol{x}_i$. This operation is widely implemented as an argmax function. This highest probability is used to determine the optimal model during training.

To build the support set, 3%–20% of the original NBs and nonblazars were selected, translating to one to seven NBs and $625/35 \times K$ nonblazars, which led to a set containing 18–126 sources. The query set was formed by taking 20% of the original NBs and nonblazars, resulting in seven NBs and 125 nonblazars. In this step, the K value of the two-way K-shot task was determined according to a high F-score on the query set.

3. *Identification*. Based on the best-performing model F_Θ and the determined K value, we calculate the similarity scores with Equation (7), which can be expressed as

$$\boldsymbol{s}_i = \frac{F_\Theta(\boldsymbol{x}_i) \cdot \boldsymbol{C}}{\mathrm{Norm}(F_\Theta(\boldsymbol{x}_i)) \cdot \mathrm{Norm}(\boldsymbol{C})}, \tag{7}$$

while the support set and query set become 35 NBs and 3708 non-NBs respectively. The similarity scores lie within the interval [−1, 1], where scores closer to −1 indicate higher similarity to non-NBs, and scores closer to 1 indicate higher similarity to NBs. Finally, the non-NBs with a similarity score higher than a threshold are considered as NBCs.

Table 3
Hyperparameters for Triplet Network Training

Hyperparameter (1)	Value (2)	Description (3)
Layers and Nodes	16, 1024, 128, 128, 128, 128, 1024	Number of nodes in the input layer (16 nodes), five hidden layers, and output layer (1024-dimensional embedding space).
Dropout Probability	0.7	Randomly omit nodes during training to reduce overfitting.
Activation Function δ	Tanh	Nonlinear function mapping inputs between -1 and 1, reducing the risk of vanishing gradients.
Margin m	15	Minimum distance between positive and negative samples in the triplet loss.
Loss Weighting w_a	7	Weighting factor for positive samples, allowing cost-sensitive adjustments in the triplet loss.
Learning Rate	0.0005	Controls step size for updating neural network weights during backpropagation.
Optimizer	Adam	Adjust network weights to minimize the loss function during training. It controls the learning rate and ensures efficient convergence.
Learning-rate Decay Step	30	Number of epochs after which the learning rate is decreased.
Learning-rate Decay Factor	0.95	The rate at which the learning rate is reduced over time.
Epochs	5000	Number of complete passes through the entire training set during the training process.
Batch Size	256	Number of samples used in each training iteration.

3.5. Evaluation Metrics

Besides the metric of accuracy that is widely used in the classification tasks, the precision, defined as $\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$, and recall, defined as $\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$, provide a clearer insight into a model's performance. They are particularly crucial as they help understand a model's ability to predict positive cases and identify all relevant instances correctly. The terms true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) come from the definition of a more fundamental tool, i.e., confusion matrix (J. Davis & M. Goadrich 2006; D. M. W. Powers 2011). To comprehensively evaluate our model facing the class imbalance and few-shot problems, we extend the F-score to harmonize the balance between recall and precision metrics, resulting in the F_β score, where β indicates the importance of recall relative to precision. The F_β score is defined as

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}} \tag{8}$$

This score is instrumental in scenarios with class imbalance, as traditional accuracy metrics may not adequately reflect the performance due to their tendency to be inflated by the majority class (H. He & E. A. Garcia 2009). In this paper, we investigated β ranging from 1 to 5, where we assumed that recall is more important than precision.

4. Data Analysis

4.1. Hyperparameters

Hyperparameters are external configurations used to manage the training of triplet networks. They are manually configured before training, which differs from the model parameters (i.e., weights and bias) that automatically update during training. The details of the hyperparameters in our analysis are listed in Table 3.

4.2. Results

The flowchart of our FSL method has three processes (see details in Figure 2 and Section 3.4), and the corresponding results are presented as follows.

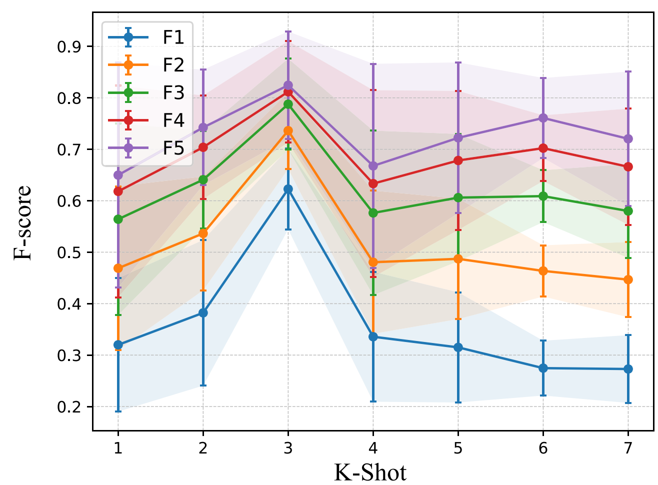

Figure 3. The F_β scores for different K-shot configurations, with β ranging from 1 to 5, represented by blue, orange, green, red, and purple lines, respectively. The shaded areas around the lines indicate the standard deviation.

1. *Pretraining*. As described in Section 3.4, we randomly divided the 4FGL_DR3 data into training, validation, support, and query sets 20 times for investigating the uncertainty of the results. Therefore, 20 best-performing models were selected across these 20 validation sets after the pretraining process.
2. *Two-way K-shot*: Considering the β ranges from 1 to 5 and the K ranges from 1 to 7, the F_β scores of the 20 best-performing models were further evaluated independently across 20 query sets and averaged across different β values and K values. The results are compared in Figure 3. We found that, when $K=3$, the highest F_β score were achieved with $F_1 = 0.62 \pm 0.08$, $F_2 = 0.74 \pm 0.07$, $F_3 = 0.79 \pm 0.09$, $F_4 = 0.81 \pm 0.1$, and $F_5 = 0.82 \pm 0.1$. Therefore, we fixed $K=3$ for the next step.
3. *Identification*. Finally, we took the 35 NBs as the support set and the 3708 non-NBs as the query set. Under the determined $K=3$, their similarity scores were calculated with Equation (7). Ultimately, 199 non-NBs were identified as NBCs, with their similarity scores higher than a threshold of 98%. The mean and standard deviation of the similarity scores obtained by the 20

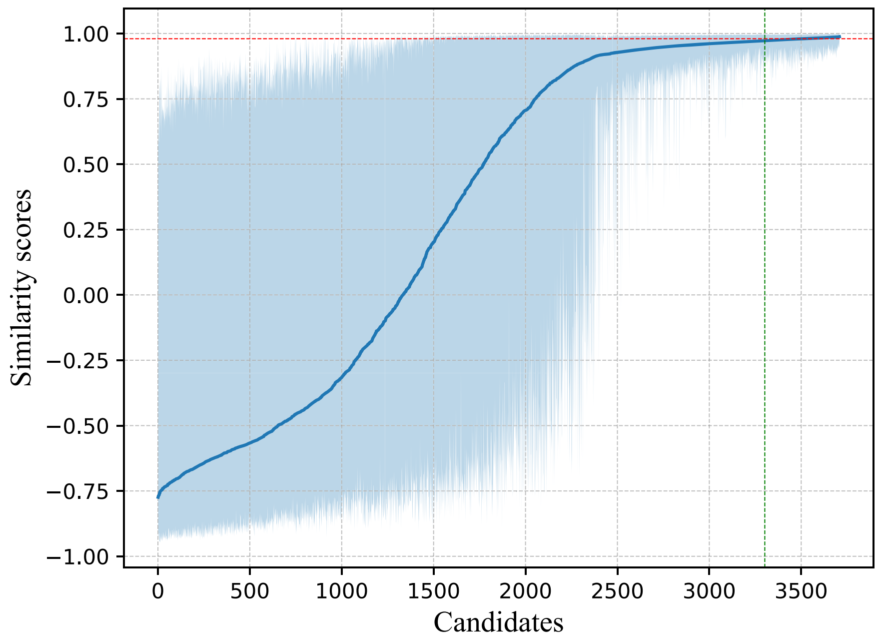

Figure 4. Probability of each blazar being an NBC. The *x*-axis represents all blazars, while the *y*-axis shows the probability of each blazar being an NB. Error bars marked in blue indicate the uncertainty in the probability estimates. The red dashed line represents a similarity score of 0.98, marking the threshold for selecting NBCs. The green dashed line indicates the starting point, with 199 sources to the right identified as NBCs.

Table 4
Mean and Standard Deviation of Similarity Scores for 199 NBCs

Source (1)	Mean of the Similarity Scores (2)	Standard Deviation (3)
J0006.3-0620	0.984	0.016
J0013.4+0950	0.982	0.017
J0029.0-7044	0.982	0.017
J0031.3+0726	0.983	0.02
J0040.3+4050	0.982	0.016

(This table is available in its entirety in machine-readable form in the online article.)

best-performing models are shown in Table 4 and Figure 4. Figure 4 demonstrates that as we move to the right, the uncertainty decreases, indicating more confident predictions. The rightmost 199 blazars, with the highest similarity scores and smallest uncertainties, are identified as NBCs.

5. Discussion

Our comprehensive study also sought to discern the distributional differences in features across three distinct nonoverlapping categories: 35 NBs, 3509 non-NBs (without NBCs), and 199 NBCs. Building upon the foundational work of J. H. Yang et al. (2022), which conducted SED fitting for 2709 blazars from the 4FGL_DR3 catalog spanning radio to X-ray wavelengths, we incorporated crucial features such as the peak frequency ($\log \nu_{\rm p}^{\rm s}$) and peak luminosity ($\log L_{\rm p}^{\rm s}$) of the synchrotron spectral component. Furthermore, in a subsequent study, J. Yang et al. (2023) determined the peak frequency ($\log \nu_{\rm p}^{\rm IC}$) and peak luminosity ($\log L_{\rm p}^{\rm IC}$) of the IC component for 3743 blazars from the 4FGL_DR3. By meticulously cross-matching the samples from J. H. Yang et al. (2022), S. Abdollahi et al. (2022), and J. Yang et al. (2023), we curated an enhanced sample of 2366 4FGL_DR3 blazars consisting of 35 NBs, 2244 non-NBs, and 87 NBCs, integrating four novel features: $\log \nu_{\rm p}^{\rm s}$, $\log L_{\rm p}^{\rm s}$, $\log \nu_{\rm p}^{\rm IC}$, and $\log L_{\rm p}^{\rm IC}$, so we now have 20 features. To this end, for the 20 features, we employ a two-sample Kolmogorov–Smirnov test (KS test; F. J. Massey 1951), setting the significance level at $\alpha = 0.05$. We also calculate the mean and standard deviation for each sample. The *p*-values obtained from the KS test and the mean and standard deviations of each feature are listed in Table 5, where column (1) represents the names of the features, columns (2)–(4) provide the *p*-values from both the KS tests, conducted pairwise across the three samples, and column (5) delineates the mean and stand deviation for each sample. The distribution histograms of the 20 features for NBs, non-NBs, and NBCs are also compared in Figure 5.

Among the features, several showed significant separation between the pairs of categories (NBs, non-NBs, and NBCs), as indicated by $p \leqslant 0.05$. Specifically, for the pair of NBs and non-NBs, features such as $E_{\rm peak}$, β, $\log F_7$, and $\log \nu_{\rm IC}$ exhibited good separation. For the pair of NBs and NBCs, features including $E_{\rm P}$, $\log F$, β, $\log f$, V, $V_{\rm F}$, $\log F_2$, $\log F_1$, $\log F_4$, $\log F_5$, $\log F_6$, $\log \nu_{\rm syn}$, and $\log L_{\rm IC}^{\rm p}$ showed significant separation. All features of the pair of non-NBs and NBCs showed significant differences with $p \leqslant 0.001$. Detailed results and *p* for each feature are provided in Table 5.

The 199 identified NBCs exhibit particular properties in the γ-ray band, which are potential indicators of neutrino emission. For instance, NBCs have higher $E_{\rm peak}$, indicating efficient particle acceleration; they show greater β, suggesting complex acceleration and cooling processes; they exhibit higher $\log F_7$, which is linked to hadronic processes; and they have higher $\log \nu_{\rm IC}$, indicating more energetic electron populations. Additionally, a significant proportion of NBCs are found to be FSRQs, suggesting a potential link between this blazar subtype and neutrino emission. This finding aligns with previous studies that suggest FSRQs are more likely to be associated with high-energy neutrino events (P. Padovani et al. 2016;

Table 5
Statistical Results of the Kolmogorov–Smirnov Test

Features (1)	NBs versus Non-NBs (2)	NBs versus NBCs (3)	Non-NBs versus NBCs (4)	Mean ± std (NBs, Non-NBs, NBCs) (5)
$E_{\rm P}$	$p = 0.094$	$p = 0.022$	$p \leqslant 0.001$	2.8 ± 3.19, 2.08 ± 2.06, 2.77 ± 1.06
$\log F$	$p = 0.242$	$p = 0.003$	$p \leqslant 0.001$	−11.16 ± 0.72, −11.31 ± 0.47, −11.48 ± 0.22
$\log f$	$p = 0.295$	$p \leqslant 0.001$	$p \leqslant 0.001$	−12.68 ± 1.22, −12.56 ± 0.99, −13.29 ± 0.44
α	$p = 0.049$	$p = 0.138$	$p \leqslant 0.001$	2.04 ± 0.36, 2.15 ± 0.35, 1.98 ± 0.15
β	$p = 0.044$	$p = 0.007$	$p \leqslant 0.001$	0.13 ± 0.16, 0.16 ± 0.17, 0.06 ± 0.05
$E_{\rm peak}$	$p = 0.012$	$p = 0.685$	$p \leqslant 0.001$	41.2 ± 87.4, 121.1 ± 2020.99, 21.75 ± 57.42
V	$p = 0.336$	$p = 0.007$	$p \leqslant 0.001$	1079.7 ± 3477.07, 364.21 ± 2858.19, 24.48 ± 11.62
$V_{\rm F}$	$p = 0.451$	$p = 0.026$	$p \leqslant 0.001$	0.37 ± 0.32, 0.48 ± 0.4, 0.4 ± 0.14
$\log F_1$	$p = 0.19$	$p = 0.01$	$p = 0.027$	−12.97 ± 2.13, −13.04 ± 1.82, −13.25 ± 1.48
$\log F_2$	$p = 0.195$	$p = 0.001$	$p \leqslant 0.001$	−12.08 ± 1.21, −12.55 ± 1.63, −12.42 ± 0.57
$\log F_3$	$p = 0.381$	$p \leqslant 0.001$	$p \leqslant 0.001$	−12.03 ± 0.96, −12.18 ± 0.94, −12.35 ± 0.26
$\log F_4$	$p = 0.232$	$p = 0.001$	$p \leqslant 0.001$	−11.97 ± 0.76, −12.15 ± 0.58, −12.33 ± 0.25
$\log F_5$	$p = 0.183$	$p = 0.004$	$p = 0.002$	−12.06 ± 0.79, −12.32 ± 0.66, −12.32 ± 0.26
$\log F_6$	$p = 0.043$	$p = 0.05$	$p \leqslant 0.001$	−12.45 ± 1.55, −12.96 ± 1.63, −12.34 ± 0.31
$\log F_7$	$p = 0.001$	$p = 0.091$	$p \leqslant 0.001$	−13.23 ± 2.2, −14.14 ± 2.39, −12.75 ± 1.13
$\log F_8$	$p = 0.035$	$p = 0.209$	$p \leqslant 0.001$	−15.09 ± 3.0, −15.89 ± 2.33, −14.66 ± 2.07
$\log \nu_{\rm syn}$	$p = 0.111$	$p = 0.001$	$p \leqslant 0.001$	14.42 ± 0.99, 14.27 ± 1.26, 15.21 ± 1.04
$\log L_{\rm syn}^{\rm p}$	$p = 0.22$	$p = 0.004$	$p \leqslant 0.001$	45.48 ± 0.77, 45.39 ± 0.82, 44.86 ± 1.01
$\log \nu_{\rm IC}$	$p = 0.004$	$p = 0.004$	$p \leqslant 0.001$	22.98 ± 0.95, 22.71 ± 1.31, 23.57 ± 0.65
$\log L_{\rm IC}^{\rm p}$	$p = 0.842$	$p \leqslant 0.001$	$p \leqslant 0.001$	45.58 ± 1.11, 45.46 ± 1.11, 44.45 ± 0.98

IceCube Collaboration et al. 2018a). Several features exhibit significant differences between NBCs and other blazars, highlighting their unique properties:

1. $E_{\rm peak}$. The $E_{\rm peak}$ feature shows significant separation between NBs and non-NBs ($p = 0.012$). High peak energies suggest efficient particle acceleration mechanisms, potentially driven by shock acceleration or magnetic reconnection processes within the jet. These high-energy electrons are also likely to produce high-energy neutrinos through interactions with ambient photons or matter (K. Murase et al. 2012; S. Gao et al. 2019). The significant difference in $E_{\rm peak}$ between NBs and non-NBs highlights its potential as a diagnostic tool for identifying neutrino-emitting blazars.
2. β. The spectral curvature parameter β also shows significant differences between NBs and non-NBs ($p = 0.044$) and between NBs and NBCs ($p = 0.007$). A higher curvature can indicate more complex acceleration processes or varying cooling mechanisms. Complex acceleration processes, such as multiple shock fronts and magnetic reconnection, lead to diverse particle energy distributions, affecting the curvature of the spectrum. Different cooling mechanisms, like synchrotron cooling and IC scattering, can create variations in the energy distribution of electrons and protons, which in turn influence neutrino production. The significant difference in β suggests that blazars with higher spectral curvature might be more efficient at producing neutrinos (G. Ghisellini & F. Tavecchio 2010, M. Cerruti et al. 2015).
3. $\log F_7$. The $\log F_7$ feature ($p = 0.001$) measures the energy flux in the 30–100 GeV band. High flux values in this band indicate significant high-energy γ-ray emission, which can result from hadronic processes in the jet. In such scenarios, proton–proton or proton–photon interactions produce pions, which decay into γ-rays and neutrinos. The significant difference in $\log F_7$ between NBs and non-NBs suggests that higher γ-ray fluxes in this band are associated with neutrino emission (P. Padovani et al. 2016; A. Palladino et al. 2019).
4. $\log \nu_{\rm IC}$. The IC peak frequency $\log \nu_{\rm IC}$ also exhibits good separation, with $p = 0.004$ for both NBs versus non-NBs and NBs versus NBCs. This parameter indicates the peak frequency of the IC emission, which is a result of high-energy electrons scattering off low-energy photons. The location of this peak can provide insights into the energy distribution of electrons and the seed photon field. Higher peak frequencies are indicative of more energetic electron populations, which are also capable of producing high-energy neutrinos through hadronic interactions. The significant separation in $\log \nu_{\rm IC}$ underscores the role of energetic electrons in both γ-ray and neutrino production (M. Böttcher et al. 2013, Gao et al. 2019).
5. $E_{\rm P}$. The $E_{\rm P}$ feature shows significant separation between NBs and NBCs ($p = 0.022$). This indicates that the energy at which the flux measurement is most precise differs between neutrino-emitting and nonemitting blazars, possibly reflecting differences in their underlying particle acceleration mechanisms or environmental conditions.
6. $\log F$. The energy flux ($\log F$) from 100 MeV to 100 GeV shows significant separation between NBs and NBCs ($p = 0.003$). Higher-energy fluxes can be indicative of more intense high-energy processes within the blazar jets, which could be associated with neutrino production. These processes might involve more efficient particle acceleration or a denser target photon field for hadronic interactions.
7. $\log f$. The $\log f$ feature shows significant separation between NBs and NBCs ($p \leqslant 0.001$). This suggests that the intensity of γ-ray emission at the pivot energy is different for neutrino-emitting blazars, possibly reflecting variations in the energy distribution of accelerated particles.
8. V and $V_{\rm F}$. The variability index (V) and fractional variability ($V_{\rm F}$) both show significant separation between NBs and NBCs, with p-values of 0.007 and 0.026,

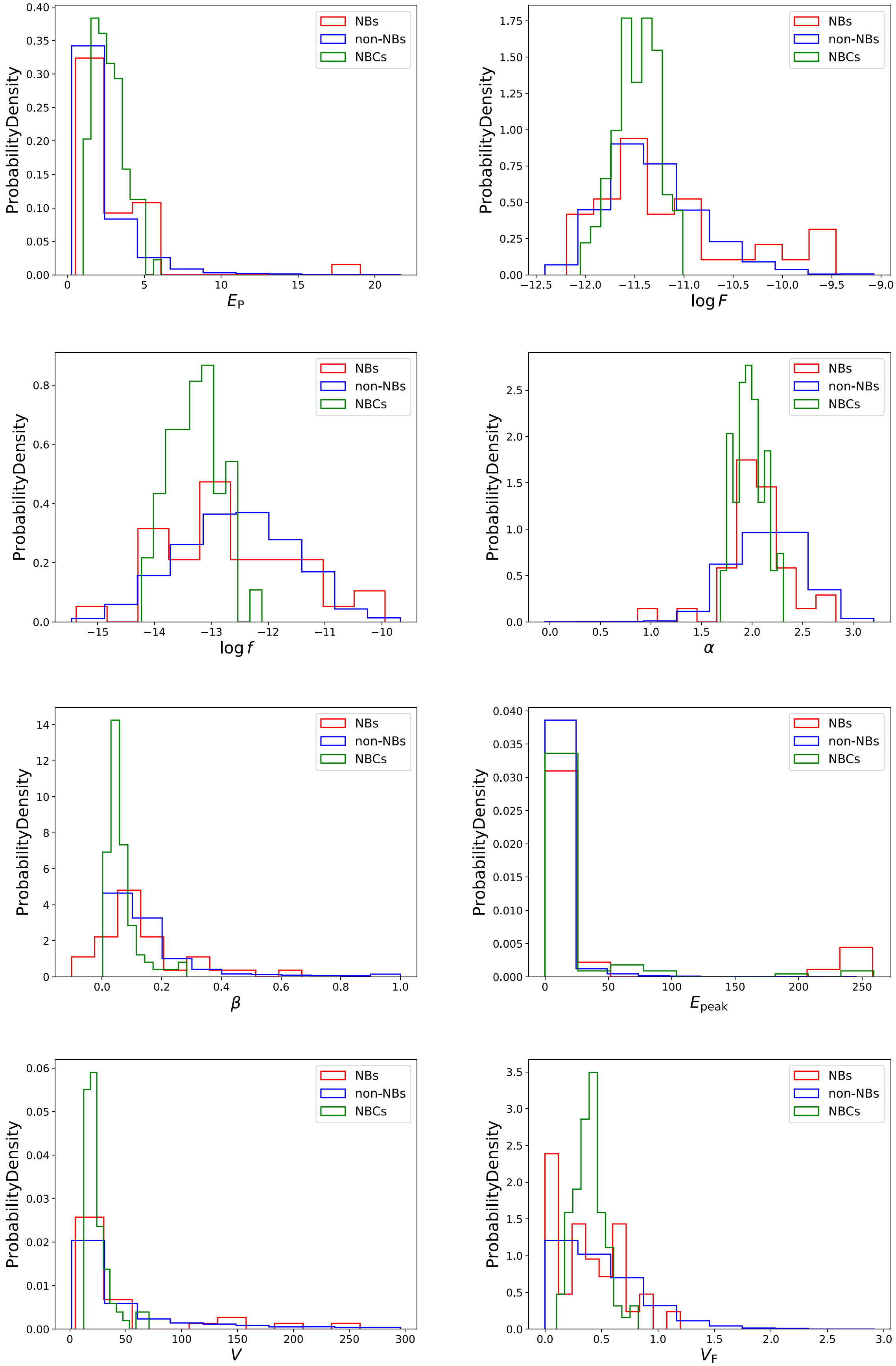

Figure 5. The distribution histograms for the 35 NBs are red, blue for the 2244 non-NBs, and green for the 87 NBCs.

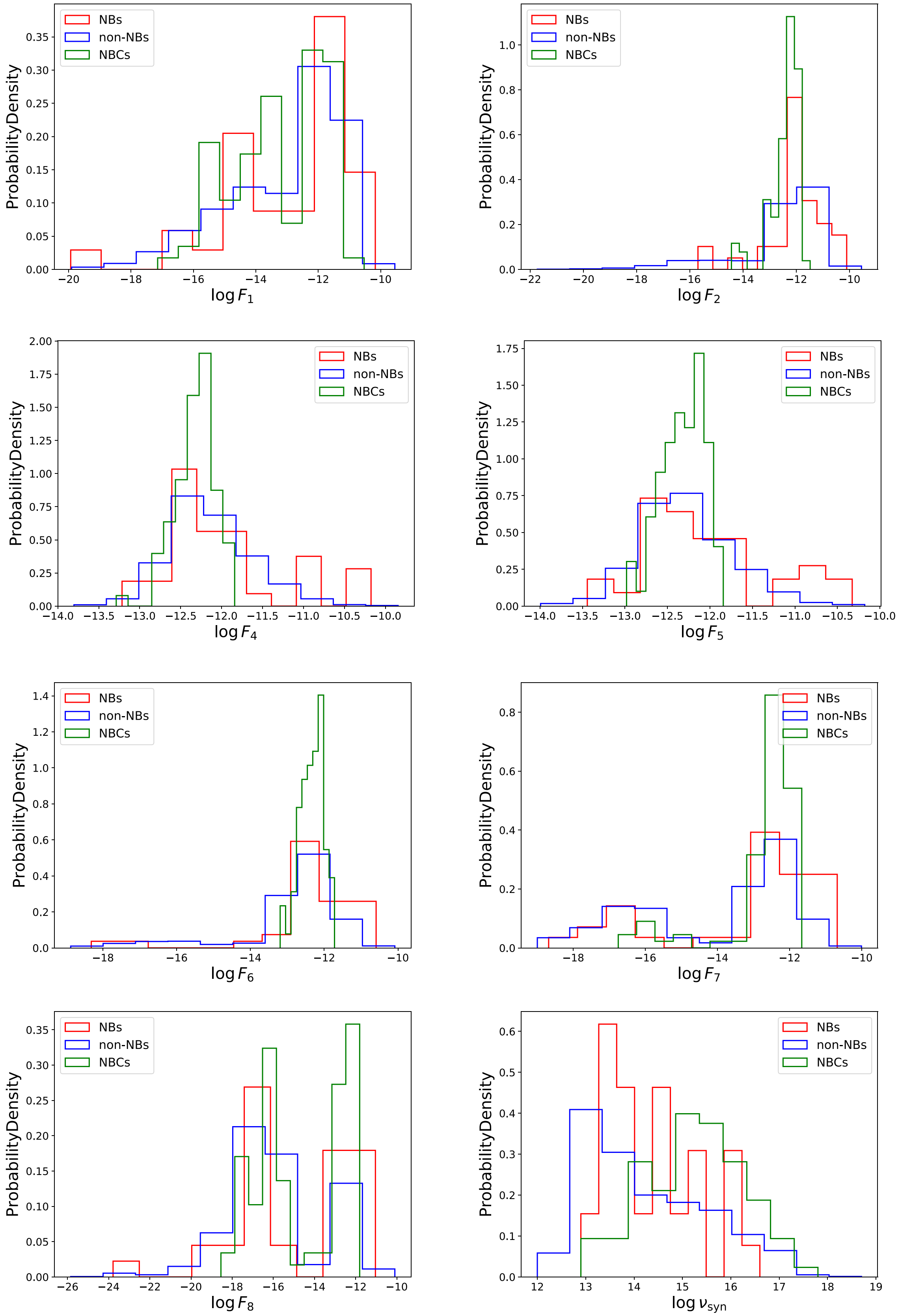

Figure 5. (Continued.)

respectively. Variability in γ-ray emission is often linked to changes in the jet environment or acceleration processes, which could also influence neutrino production. Higher variability might indicate more dynamic conditions within the jet, leading to more frequent or intense hadronic interactions.

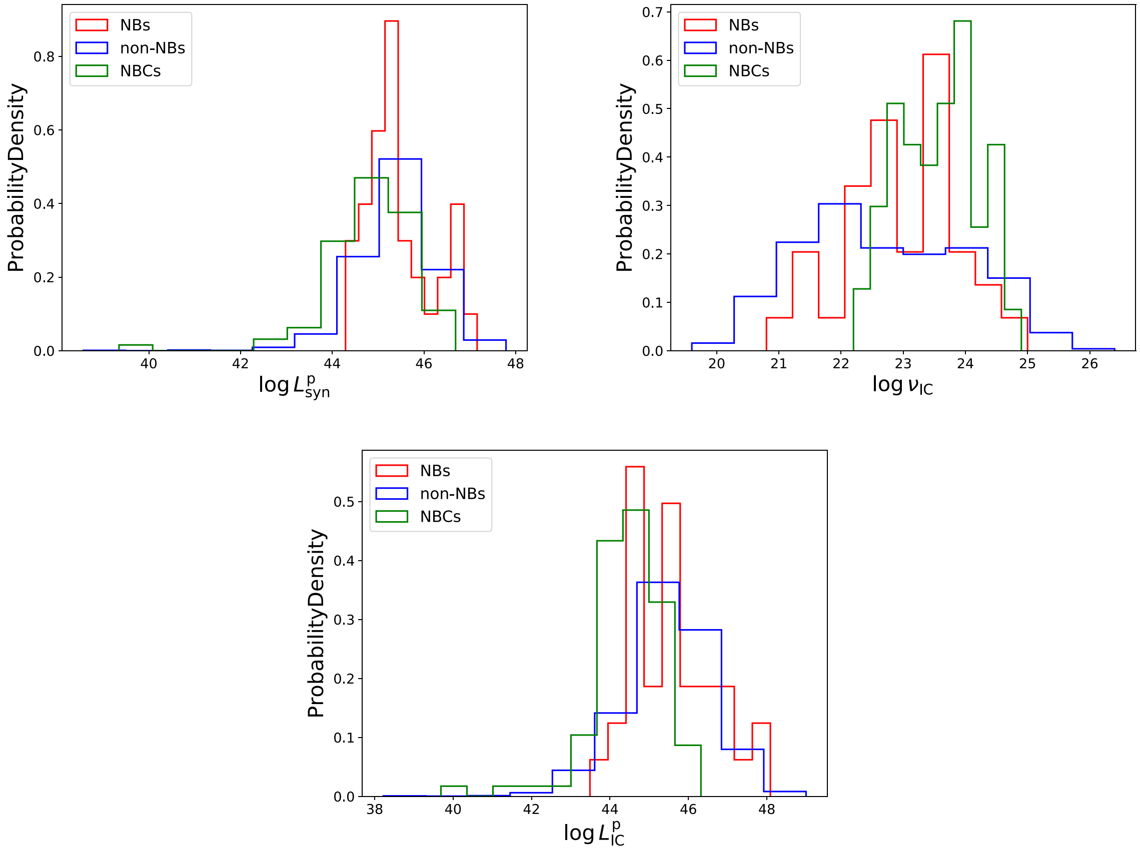

Figure 5. (Continued.)

9. $\log F_2$, $\log F_1$, $\log F_4$, $\log F_5$, $\log F_6$. Several energy fluxes in different γ-ray bands show significant separation between NBs and NBCs. Specifically, $\log F_2$ (100–300 MeV, $p = 0.001$), $\log F_1$ (50–100 MeV, $p = 0.010$), $\log F_4$ (1–3 GeV, $p = 0.001$), $\log F_5$ (3–10 GeV, $p = 0.004$), and $\log F_6$ (10–30 GeV, $p = 0.050$). These differences in flux across various energy bands suggest that neutrino-emitting blazars have distinct γ-ray spectral properties, possibly due to differences in their particle acceleration and emission processes.
10. $\log \nu_{\mathrm{syn}}$. The synchrotron peak frequency ($\log \nu_{\mathrm{syn}}$) shows significant separation between NBs and NBCs ($p = 0.001$). This parameter reflects the peak of the synchrotron emission, which is related to the maximum energy of electrons in the jet. Higher synchrotron peak frequencies can indicate more efficient acceleration processes, which might also enhance neutrino production.
11. $\log L_{\mathrm{IC}}^{\mathrm{p}}$. The peak luminosity of the IC component ($\log L_{\mathrm{IC}}^{\mathrm{p}}$) shows significant separation between NBs and NBCs ($p \leqslant 0.001$). This suggests that neutrino-emitting blazars tend to have different luminosities in the IC component, which could be linked to differences in their high-energy electron populations and the target photon fields for IC scattering.

6. Conclusion

FSL is preferable when traditional DL methods struggle with limited data from underrepresented and imbalanced classes. With FSL, we constructed a triplet network to transform the input data into an embedding space where the similarity can be measured. Using a two-way three-shot approach, we achieved a strong F_β score on the query set. This performance could be transferred to identify 199 NBCs out of 3708 non-NBs, each with a similarity score exceeding 98%. Additionally, we found substantial differences in the distributions of features among NBs, non-NBs without NBCs, and NBCs. Specifically, the KS test indicates significant differences across all features between non-NBs and NBCs, further supporting the successful separation of NBCs from non-NBs.

Acknowledgments

J.Z acknowledges the support from the University of Padova, Italy, and Guangzhou University, China. H.X acknowledges the support from the National Natural Science Foundation of China (NSFC) under grant No. 12203034 and from the Shanghai Science and Technology Fund under grant No. 22YF1431500. Z.P. acknowledges support from the National Science Foundation for Young Scientists of China (grant 12103012). This work is partially supported by the National Natural Science Foundation of China (NSFC 12433004, NSFC U2031201, NSFC 11733001, NSFC 12203034), Guangdong Major Project of Basic and Applied Basic Research (2019B030302001), Scientific and Technological Cooperation Projects (2020–2023) between the People's Republic of China and the Republic of Bulgaria, and Astrophysics Key Subjects of Guangdong Province and Guangzhou City.

ORCID iDs

Jingtian Zhu https://orcid.org/0000-0002-8206-5080
Haitao Cao https://orcid.org/0000-0002-8626-8686
Hubing Xiao https://orcid.org/0000-0001-8244-1229
Zhiyuan Pei https://orcid.org/0000-0002-4970-3108
JunHui Fan https://orcid.org/0000-0002-5929-0968
Denis Bastieri https://orcid.org/0000-0002-6954-8862

References

Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2016, ApJ, 833, 3
Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2017b, ApJ, 835, 45
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2017a, ApJ, 843, 112
Abdollahi, S., Acero, F., Ackermann, M., et al. 2020, ApJS, 247, 33
Abdollahi, S., Acero, F., Baldini, L., et al. 2022, ApJS, 260, 53
Acero, F., Ackermann, M., Ajello, M., et al. 2015, ApJS, 218, 23
Ahmad, J., Farman, H., & Jan, Z. 2019, in Deep Learning: Convergence to Big Data Analytics, ed. M. Khan et al. (Singapore: Springer), 31
Ajello, M., Angioni, R., Axelsson, M., et al. 2020, ApJ, 892, 105
Ajello, M., Atwood, W. B., Baldini, L., et al. 2017, ApJS, 232, 18
Ajello, M., Gasparrini, D., Sánchez-Conde, M., et al. 2015, ApJL, 800, L27
Atoyan, A., & Dermer, C. D. 2001, PhRvL, 87, 221102
Atwood, W. B., Abdo, A. A., Ackermann, M., et al. 2009, ApJ, 697, 1071
Batista, G., Prati, R., & Monard, M.-C. 2004, ACM SIGKDD Explor. Newsl., 6, 20
Bengio, Y. 2009, Found. Trends Mach. Learn., 2, 1
Bengio, Y., Courville, A., & Vincent, P. 2013, ITPAM, 35, 1798
Blandford, R. D., & Koenigl, A. 1979, ApL, 20, 15
Böttcher, M., Reimer, A., Sweeney, K., & Prakash, A. 2013, ApJ, 768, 54
Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. 1993, Advances in Neural Information Processing Systems 6 (NIPS 1993), ed. J. Cowan et al. (NeurIPS), https://proceedings.neurips.cc/paper/1993/hash/288cc0ff022877bd3df94bc9360b9c5d-Abstract.html
Buda, M., Maki, A., & Mazurowski, M. A. 2018, NN, 106, 249
Cao, H., Xiao, H., Luo, Z., Zeng, X., & Fan, J. 2024, ApJ, 961, 91
Cerruti, M., Zech, A., Boisson, C., & Inoue, S. 2015, MNRAS, 448, 910
Chang, X.-C., Liu, R.-Y., & Wang, X.-Y. 2015, ApJ, 805, 95
Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002, J. Artif. Intell. Res., 16, 321
Chiaro, G., Salvetti, D., La Mura, G., et al. 2016, MNRAS, 462, 3180
Chicco, D. 2021, Artificial Neural Networks (Humana, NY: Springer US), 73
Davis, J., & Goadrich, M. 2006, in ICML '06: Proc. of the 23rd Int. Conf. on Machine Learning, ed. W. Cohen & A. Moore (New York: ACM), 233
de Menezes, R., Buson, S., Garrappa, S., et al. 2022, ICRC (Berlin), 37, 955
Dhillon, G. S., Chaudhari, P., Ravichandran, A., & Soatto, S. 2019, arXiv:1909.02729
Dimitrakoudis, S., Mastichiadis, A., Protheroe, R. J., & Reimer, A. 2012, A&A, 546, A120
Dong, G., & Liu, H. 2018, Feature Engineering for Machine Learning and Data Analytics (1st ed.; Boca Raton, FL: CRC Press)
Drummond, C., & Holte, R. C. 2005, in Machine Learning: ECML 2005, ed. J. Gama et al. (Berlin: Springer), 539
Elkan, C. 2001, Proc. of the Seventeenth Int. Conf. on Artificial Intelligence (San Francisco, CA: Morgan Kaufmann), 973, https://eva.fing.edu.uy/pluginfile.php/63457/mod_resource/content/1/Elkan_2001_The_foundations_of_cost-sensitive_learning.pdf
Fan, J., Xiao, H., Yang, W., et al. 2023, ApJS, 268, 23
Fan, J.-H. 2002, PASJ, 54, L55
Fan, J.-H., Bastieri, D., Yang, J.-H., et al. 2014, RAA, 14, 1135
Fan, J.-H., Chen, K.-Y., Xiao, H.-B., et al. 2022, Univ, 8, 436
Fan, J. H., Kurtanidze, S. O., Liu, Y., et al. 2021, ApJS, 253, 10
Freund, Y., & Schapire, R. E. 1997, J. Comput. Syst. Sci., 55, 119
Galván, A., Fraija, N., Aguilar-Ruiz, E., et al. 2022, ICRC (Berlin), 37, 1009
Gao, S., Fedynitch, A., Winter, W., & Pohl, M. 2019, NatAs, 3, 88
Garrappa, S., Buson, S., Franckowiak, A., et al. 2019, ApJ, 880, 103
Garrappa, S., Buson, S., Franckowiak, A., et al. 2022, ICRC (Berlin), 37, 956
Garrappa, S., Buson, S., Sinapius, J., Kadler, M. & Fermi-LAT Collaboration 2021, GCN, 31194, 1
Gasparrini, D., Cavazzuti, E., Cutini, S., et al. 2012, in AIP Conf. Proc. 1505, High Energy Gamma-Ray Astronomy: 5th Int. Meeting on High Energy Gamma-Ray Astronomy, ed. F. A. Aharonian, W. Hofmann, & F. M. Rieger (Melville, NY: AIP), 478
Ghisellini, G., & Tavecchio, F. 2010, MNRAS, 409, L79
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning, Adaptive Computation and Machine Learning (Cambridge, MA: MIT Press)
Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, Advances in Neural Information Processing Systems 27, ed. Z. Ghahramani et al. (NeurIPS), https://proceedings.neurips.cc/paper_files/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html
Guépin, C., & Kotera, K. 2017, A&A, 603, A76
Gupta, A. C., Agarwal, A., Bhagwan, J., et al. 2016, MNRAS, 458, 1127
Haack, C., Wiebusch, C. & IceCube Collaboration 2017, ICRC (Busan), 35, 1005
Halzen, F. 2017, NatPh, 13, 232
Halzen, F., & Zas, E. 1997, ApJ, 488, 669
Han, H., Wang, W., & Mao, B. 2005, in Advances in Intelligent Computing, ed. D. S. Huang et al. (Berlin: Springer), 878
Havaei, M., Davy, A., Warde-Farley, D., et al. 2017, Med. Image Anal., 35, 18
He, H., Bai, Y., Garcia, E. A., & Li, S. 2008, in 2008 IEEE Int. Joint Conf. on Neural Networks (IEEE World Congress on Computational Intelligence) (Piscataway, NJ: IEEE), 1322
He, H., & Garcia, E. A. 2009, IEEE Trans. Knowl. Data Eng., 21, 1263
He, H., & Shen, X. 2007, Int. Conf. on Artificial Intelligence, ICAI 2007, ed. H. R. Arabnia, M. Q. Yang, & J. Y. Yang, (CSREA Press), 358
Hirata, K., Kajita, T., Koshiba, M., et al. 1987, PhRvL, 58, 1490
Hornik, K., Stinchcombe, M., & White, H. 1989, NN, 2, 359
IceCube Collaboration 2005, NuPhS, 138, 179
IceCube Collaboration 2013, Sci, 342, 1242856
IceCube Collaboration, Aartsen, M. G., Ackermann, M., et al. 2018a, Sci, 361, 147
IceCube Collaboration, Aartsen, M. G., Ackermann, M., et al. 2018b, Sci, 361, eaat1378
Japkowicz, N., & Stephen, S. 2002, Intell. Data Anal., 6, 429
Kadler, M., Krauß, F., Mannheim, K., et al. 2016, NatPh, 12, 807
Kang, S.-J., Li, E., Ou, W., et al. 2019, ApJ, 887, 134
Keivani, A., Murase, K., Petropoulou, M., et al. 2018, ApJ, 864, 84
Kubat, M., Holte, R. C., & Matwin, S. 1998, Mach. Learn., 30, 195
LeCun, Y., Bengio, Y., & Hinton, G. 2015, Natur, 521, 436
Li, R.-L., Zhu, B.-Y., & Liang, Y.-F. 2022, PhRvD, 106, 083024
Liao, N.-H., Sheng, Z.-F., Jiang, N., et al. 2022, ApJL, 932, L25
Liu, R.-Y., Wang, K., Xue, R., et al. 2019, PhRvD, 99, 063008
Liu, R.-Y., Wang, X.-Y., Inoue, S., Crocker, R., & Aharonian, F. 2014, PhRvD, 89, 083004
Lott, B., Gasparrini, D., & Ciprini, S. 2020, arXiv:2010.08406
Lunardini, C., & Winter, W. 2017, PhRvD, 95, 123001
Mannheim, K. 1995, APh, 3, 295
Massey, F. J. 1951, J. Am. Stat. Assoc., 46, 68
Mazurowski, M. A., Habas, P. A., Zurada, J. M., et al. 2008, NN, 21, 427
McCulloch, W. S., & Pitts, W. 1943, Bull. Math. Biophys., 5, 127
Mücke, A., & Protheroe, R. J. 2001, APh, 15, 121
Mücke, A., Protheroe, R. J., Engel, R., Rachen, J. P., & Stanev, T. 2003, APh, 18, 593
Murase, K., Dermer, C. D., Takami, H., & Migliori, G. 2012, ApJ, 749, 63
Murase, K., Inoue, Y., & Dermer, C. D. 2014, PhRvD, 90, 023007
Murase, K., Oikonomou, F., & Petropoulou, M. 2018, ApJ, 865, 124
Nolan, P. L., Abdo, A. A., Ackermann, M., et al. 2012, ApJS, 199, 31
Ortigosa-Hernández, J., Inza, I., & Lozano, J. A. 2016, arXiv:1608.08984
Ortigosa-Hernández, J., Inza, I., & Lozano, J. A. 2017, PaReL, 98, 32
Padovani, P., Petropoulou, M., Giommi, P., & Resconi, E. 2015, MNRAS, 452, 1877
Padovani, P., Resconi, E., Giommi, P., Arsioli, B., & Chang, Y. L. 2016, MNRAS, 457, 3582
Palladino, A., Rodrigues, X., Gao, S., & Winter, W. 2019, ApJ, 871, 41
Pinhero Cinelli, L., Araújo Marins, M., Barros da Silva, E., & Lima Netto, S. 2021, in Variational Methods for Machine Learning with Applications to Deep Networks, ed. L. Pinheiro Cinelli (Cham: Springer), 111
Powers, D. M. W. 2011, J. Mach. Learn. Technol., 2, 37
Sahakyan, N. 2018, ApJ, 866, 109
Schroff, F., Kalenichenko, D., & Philbin, J. 2015, in 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Piscataway, NJ: IEEE), 815
Senno, N., Murase, K., & Mészáros, P. 2017, ApJ, 838, 3
Sikora, M., Begelman, M. C., & Rees, M. J. 1994, ApJ, 421, 153
Sokolov, A., & Marscher, A. P. 2005, ApJ, 629, 52
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, JMLR, 15, 1929
Stecker, F. W. 2013, PhRvD, 88, 047301
Stickel, M., Padovani, P., Urry, C. M., Fried, J. W., & Kuehr, H. 1991, ApJ, 374, 431

Tomek, I. 1976, ITSMC, SMC-6, 769
Urry, C. M., & Padovani, P. 1995, PASP, 107, 803
Villata, M., Raiteri, C. M., Balonek, T. J., et al. 2006, A&A, 453, 817
Wang, G., Fan, J., Xiao, H., & Cai, J. 2022a, PASP, 134, 104101
Wang, S., Liu, W., Wu, J., et al. 2016, in 2016 Int. Joint Conf. on Neural Networks (IJCNN) (New York, NY: IEEE), 4368
Wang, Z.-R., Liu, R.-Y., Petropoulou, M., et al. 2022b, PhRvD, 105, 023005
Weiss, G. M. 2004, SIGKDD Explor. Newsl., 6, 7
Weiss, G. M. 2005, in Mining with Rare Cases, ed. O. Maimon & L. Rokach (Boston, MA: Springer US), 765
Wills, B. J., Wills, D., Breger, M., Antonucci, R. R. J., & Barvainis, R. 1992, ApJ, 398, 454
Wilson, D. L. 1972, ITSMC, SMC-2, 408
Xiao, H., Fan, J., Yang, J., et al. 2019, SCPMA, 62, 129811
Xiao, H. B., Cao, H. T., Fan, J. H., et al. 2020, A&C, 32, 100387
Xiao, H. B., Pei, Z. Y., Xie, H. J., et al. 2015, Ap&SS, 359, 39
Xue, R., Liu, R.-Y., Petropoulou, M., et al. 2019a, ApJ, 886, 23
Xue, R., Liu, R.-Y., Wang, X.-Y., Yan, H., & Böttcher, M. 2019b, ApJ, 871, 81
Xue, R., Liu, R.-Y., Wang, Z.-R., Ding, N., & Wang, X.-Y. 2021, ApJ, 906, 51
Yang, J., Fan, J., Liu, Y., et al. 2023, SCPMA, 66, 249511
Yang, J. H., Fan, J. H., Liu, Y., et al. 2022, ApJS, 262, 18
Zhang, H., Fang, K., Li, H., et al. 2019, ApJ, 876, 109
Zhang, J., & Mani, I. 2003, Proc. of the ICML'2003 Workshop on Learning from Imbalanced Datasets, https://www.site.uottawa.ca/~nat/Workshop2003/jzhang.pdf
Zhu, C. R., Wang, G., Liu, B. L., et al. 2013, PhRvB, 88, 121301
Zhu, J. T., Lin, C., Xiao, H. B., et al. 2023, ApJ, 950, 123