The Astrophysical Journal, 950:123 (19pp), 2023 June 20 https://doi.org/10.3847/1538-4357/acca85
© 2023. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

Exploring TeV Candidates of Fermi Blazars through Machine Learning

J. T. Zhu[1,2,3,4], C. Lin[1,3,4], H. B. Xiao[5], J. H. Fan[1,3,4], D. Bastieri[1,2], and G. G. Wang[1,3,4]
[1] Center for Astrophysics, Guangzhou University, Guangzhou 510006, People's Republic of China; fjh@gzhu.edu.cn
[2] Department of Physics and Astronomy "G. Galilei," University of Padova, Padova I-35131, Italy
[3] Astronomy Science and Technology Research Laboratory of Department of Education of Guangdong Province, Guangzhou 510006, People's Republic of China
[4] Key Laboratory for Astronomical Observation and Technology of Guangzhou, Guangzhou 510006, People's Republic of China
[5] Shanghai Key Lab for Astrophysics, Shanghai Normal University Shanghai, 200234, People's Republic of China
Received 2020 August 21; revised 2023 March 14; accepted 2023 March 16; published 2023 June 15

Abstract

In this work, we make use of a supervised machine-learning algorithm based on Logistic Regression (LR) to select TeV blazar candidates from the 4FGL-DR2/4LAC-DR2, 3FHL, 3HSP, and 2BIGB catalogs. LR constructs a hyperplane based on a selection of optimal parameters, named features, and hyperparameters whose values control the learning process and determine the values of features that a learning algorithm ends up learning, to discriminate TeV blazars from non-TeV blazars. In addition, it gives the probability (or logistic) that a source may be considered a TeV blazar candidate. Non-TeV blazars with logistics greater than 80% are considered high-confidence TeV candidates. Using this technique, we identify 40 high-confidence TeV candidates from the 4FGL-DR2/4LAC-DR2 blazars and we build the feature hyperplane to distinguish TeV and non-TeV blazars. We also calculate the hyperplanes for the 3FHL, 3HSP, and 2BIGB. Finally, we construct the broadband spectral energy distributions for the 40 candidates, testing for their detectability with various instruments. We find that seven of them are likely to be detected by existing or upcoming IACT observatories, while one could be observed with extensive air shower particle detector arrays.

Unified Astronomy Thesaurus concepts: Blazars (164)

Supporting material: figure sets, machine-readable tables

1. Introduction

Blazars, an extreme subclass of active galactic nuclei (AGNs), are known for prominent observation properties, such as high-energy gamma-ray emissions, rapid and significant amplitude variability, high luminosity, high and variable polarization, and superluminal motions, etc. (Wills et al. 1992; Urry & Padovani 1995; Fan 2002; Villata et al. 2006; Fan et al. 2014, 2021; Xiao et al. 2015, 2019, 2020; Gupta et al. 2016; Abdollahi et al. 2020). Blazars are historically subdivided into two main categories based on the equivalent width (EW) of the optical emission lines: flat spectrum radio quasars (FSRQs) and BL Lacertae objects (BL Lacs). In general, FSRQs show an EW greater than 5 Å, while BL Lacs illustrate no or weak emission lines, with EW less than 5 Å. Meanwhile, the spectral energy distributions (SEDs) of blazars are generally characterized by two well-separated bumps, a low-energy one is in the infrared to soft X-ray energy range due to synchrotron emission, and a high-energy one is in the region between hard X-ray to gamma ray that is associated with an inverse Compton (IC) radiation according to leptonic model (IC; e.g., Sikora et al. 1994). The seed photons undergoing IC scattering could be from the same electron population producing the synchrotron bump in the so-called synchrotron self-Compton (SSC) model (Ghisellini et al. 1985; Maraschi et al. 1992; Bloom & Marscher 1996), e.g., 1ES 0347-121, 1ES 0229+200 (Aharonian et al. 2007a, 2007b; Costamante et al. 2018), or from external regions (External Compton model, EC), e.g., from the accretion disk (Dermer & Schlickeiser 1993), broad line region (Sikora et al. 1994), and dust torus (Błażejowski et al. 2000). While the hadronic process could also contribute to the high-energy bump by high-energy cosmic rays via the photohadronic reactions (Dimitrakoudis et al. 2012; Tavecchio 2014). SEDs are also used to make classifications of blazars. In the original BL Lac classification of Padovani & Giommi (1995), BL Lacs are classified as two subclasses: "low synchrotron-peaked (LSP) BL Lac" or LBL and "high synchrotron-peaked sources (HSP) BL Lac" or HBL, depending on their broadband radio to X-ray spectral index is larger or smaller than 0.75. Abdo et al. (2010) later extended the classification to all blazars. They suggested to group blazars into three subclasses based on synchrotron-peak frequency, LSP, $\log \nu_{\rm p}^{\rm s} \leqslant 14.0$; ISP (intermediate synchrotron-peaked sources), $14 < \log \nu_{\rm p}^{\rm s} \leqslant 15$; and HSP, $\log \nu_{\rm p}^{\rm s} > 15$. While, Fan et al. (2016) proposed slightly different criteria: $14 < \log \nu_{\rm p}^{\rm s} \leqslant 15.3$ for ISP, and $\log \nu_{\rm p}^{\rm s} > 15.3$ for HSP. Very recently, J. H. Yang et al. (2022) also gave a similar classification.

Additionally, Costamante et al. (2001) proposed a fourth subclass namely extreme HBLs (EHBLs) for BL Lacs having $\log \nu_{\rm p}^{\rm s} > 17$. The EHBLs can be divided into three subclasses according to their IC bump peak (Foffano et al. 2019) with the subclass with IC bump peak between 0.1 TeV and 1 TeV being a continuation of HSP, while the subclass with IC bump peak >10 TeV named "hard-TeV blazars" is an independent category featuring high power, very stable flux, and hard-TeV spectral behavior at TeV energies, posing a challenge to the SSC. With an IC bump peaking at a few TeV, the remaining subclass behaves as a transition class with a flat TeV spectral slope. Thus, the HSPs and HBLs are often considered as the potential emitters of very high-energy (VHE, above 300 GeV) radiations, among which, the sources with small redshift have a significant fraction of TeV photons to be

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

observed. The emissions of VHE photons with energy >300 GeV from blazars reveals new phenomena, especially photons in the TeV band raise the challenge of particle acceleration in jets, and they are also essential clues for indirectly measuring the extragalactic background light, estimating the intergalactic magnetic field, and probing the possible origin of high-energy extragalactic neutrinos (Foffano et al. 2019). For instance, the Very Long Baseline Array provided 23 images of six TeV blazars and proved that apparent jet bending is a common property of TeV blazars (Piner et al. 2010).

The emitted flux from most astronomical sources is very low in the TeV band. In addition, the extragalactic background light (EBL) absorbs the most energetic gamma-ray emissions (Primack et al. 1999; Kneiske et al. 2004; Fermi-LAT Collaboration et al. 2018) via the interaction $\gamma + \gamma \rightarrow e^+ + e^-$, which is relative to many fundamental astrophysics problems (Domínguez et al. 2019). Therefore, observations at TeV energies require large collection areas, which are only affordable for ground-based detectors like atmospheric Cherenkov telescopes (IACTs) and extended arrays of particle detectors (EAS arrays). The gamma-ray detection techniques of IACTs and EAS arrays are different. IACTs detect Cherenkov photons in the atmosphere generated by the atmospheric extended shower (EAS) of the secondary particles initiated by the primary gamma rays, and popular examples are the High Energy Stereoscopic System[6] (H.E.S.S.; Aharonian et al. 2004), the Major Atmospheric Gamma Imaging Cherenkov Telescopes[7] (MAGIC; MAGIC Collaboration 2000), the Very Energetic Radiation Imaging Telescope Array System[8] (VERITAS; VERITAS Collaboration et al. 2005) and the next-generation IACT array: the Cherenkov Telescope Array Observatory[9] (CTAO; The CTA Consortium 2011). While EAS arrays detect secondary particles from the cosmic ray surviving down to the ground, such as the High Altitude Water Cherenkov Observatory[10] (HAWC; Abeysekara et al. 2017a), and the Large High Altitude Air Shower Observatory[11] (LHAASO; Cao et al. 2019).

Up to 2022 August, only about 90 extragalactic TeV sources (Wakely & Horan 2008),[12] out of which are 81 TeV blazars, have been verified by IACTs and EAS arrays. Finding TeV blazar candidates is an exciting and challenging work, since the EBL and the sensitivity of the detector limit the number of TeV sources with high redshift. Therefore, under the constraints of EBL, an improved sensitivity will expand the sample of TeV blazars in terms of the sheer number and will find sources with higher redshift. Briefly, the current and next generation of IACTs and EAS arrays that are sensitive to photons at TeV energies should increase the TeV blazar population, helping us to explore this classification further and understand the mechanism of VHE emission. (Costamante & Ghisellini 2002; Massaro et al. 2013; Chang et al. 2017; Foffano et al. 2019). The current TeV gamma-ray detection is offered by a large area of ground-based detectors.

Most of the TeV blazars are HSPs (almost all BL Lacs), thus the TeV blazar candidate searches are often limited to HSPs/BL Lacs (e.g., Costamante & Ghisellini 2002; Massaro et al. 2013; Chang et al. 2017; Chiaro et al. 2019; Foffano et al. 2019). Typically, BL Lacs are HSPs, whereas FSRQs are mostly LSPs and ISPs. LSP/ISP often have EC components where the electrons see a strong photon field that is Doppler shifted. In principle they are capable of producing strong gamma-ray emissions, sometimes entering the VHE regime. But in the case of LSP/ISPs, the interaction often happens in the Klein–Nishina regime and the same strong photon field often induces a very strong absorption of the produced gamma-ray photons. Compared to LSP/ISPs, the VHE emission of HSPs is thought to originate from the low-energy photons produced by synchrotron radiation by ultrarelativistic electrons in the jet according to the SSC process. Additionally, HSPs have a higher $\log \nu_{\rm p}^{\rm s}$, making their synchrotron bump more easily scattered into the TeV band by the relativistic electrons. However, it is not obvious that extreme $\log \nu_{\rm p}^{\rm s}$ values lead to the emission of TeV gamma rays by itself, as it also depends on many other properties of the emission region, such as electron distribution, magnetic field strength, internal absorption, and the redshift of the source (Nievas Rosillo et al. 2022).

Lin & Fan (2016) gathered 662 Fermi BL Lacs, including 47 TeV sources, and compared the multiwavelength observation properties of the TeVs with those of the non-TeVs. They discovered that TeVs have a smaller average redshift, a higher flux density, and a harder gamma-ray spectrum. Flux variability in all wave bands is very common among blazars (Fan et al. 2017; Majumder et al. 2019; J. H. Yang et al. 2022; W.-X. Yang et al. 2022), with some blazars, such as 1ES 0229+200, exhibiting moderate variability and others displaying violent and high amplitude variability, with timescales ranging from hours to years. In addition, the $\log \nu_{\rm p}^{\rm s}$ is proportional to the variability of blazars. Gupta et al. (2016) analyzed 50 observations of 12 LSP blazars from XMM-Newton and discovered that LSP blazars vary more slowly in the X-ray bands than in the IR/optical bands because the IC mechanism dominates the X-ray radiation in this case. In contrast, HSP blazars are predicted to exhibit more extreme X-ray band variability than LSP blazars. According to the SSC model, relativistic electrons upscatter X-rays into the TeV region, and observations have confirmed correlations between X-ray and TeV emissions (Costamante & Ghisellini 2002; Osorio et al. 2019; Singh et al. 2019). One can therefore anticipate a correlation between X-ray and TeV emissions. Notably, Markarian 501 and 1ES 1959+650's synchrotron-peak frequencies in the X-ray band shifted to the higher energy region when a flare was observed in the TeV band (Sambruna et al. 2000; Kapanadze et al. 2018; Singh et al. 2019). In contrast, there exists an intriguing counterexample. Foffano et al. (2019) observed two groups of EHBLs with the same range of $\log \nu_{\rm p}^{\rm s}$ but opposite spectral slope in the TeV band, implying that $\log \nu_{\rm p}^{\rm s}$ and the TeV emission are independent of one another.

Beyond all doubt, the Fermi Large Area Telescope (Fermi-LAT) brought prosperity to the study of blazars after its launch in 2008, and the progress of high-energy gamma-ray astronomy of space-borne instruments is gradually dominated by it. Fermi-LAT detected thousands of blazars and published them in their works (e.g., Abdollahi et al. 2020), which provided a large gamma-ray blazar sample. The Fermi-LAT 10 yr source catalog (4FGL-DR2, Abdollahi et al. 2020; Ballet et al. 2020), and the Fermi-LAT 10 yr AGN catalog (4LAC-DR2, Ajello et al. 2020; Lott et al. 2020) listed 80 out of the 81 TeV blazars (hereafter

[6] https://www.mpi-hd.mpg.de/hfm/HESS/
[7] https://magic.mpp.mpg.de/
[8] https://veritas.sao.arizona.edu/
[9] https://www.cta-observatory.org/
[10] https://www.hawc-observatory.org/
[11] http://english.ihep.cas.cn/lhaaso/
[12] http://TeVCat.uchicago.edu

Fermi blazars). Thus, we have a convincing reason to believe that Fermi catalogs should contain many TeV blazar candidates.

Machine-learning (ML) techniques have become popular among astronomers (Ball & Brunner 2010; Chiaro et al. 2019; Kang et al. 2019; Xiao et al. 2022). Based on the Fermi-LAT third source catalog (3FGL), Chiaro et al. (2019) applied the Artificial Neural Network (ANN) to 573 uncertain types and 559 unassociated catalog sources of the 3FGL blazars and found 80 HSPs candidates, 16 of them are proposed as TeV candidates with the highest confidence. Kang et al. (2019) collected 1312 blazar candidates of uncertain type (BCUs) out of 3137 blazars recorded in Fermi-LAT 4th source catalog, then employed random forest (RF), support vector machine (SVM), and ANN to predict the category of the BCUs. Finally, 724 BL Lac candidates and 332 FSRQ candidates are predicted by combined classification results of the ML methods. Xiao et al. (2022) compiled radio-loudness $\log R$ and radio 6 cm luminosity $\log L_{6\,\mathrm{cm}}$ of 2943 AGNs, and got $\log R = \langle 1.37 \pm 0.02 \rangle$ as the separation of radio-loud and radio-quiet AGNs by the Gaussian Mixture Model. Furthermore, a more advanced double-criterion dividing boundary $\log L_{6\,\mathrm{cm}} = -2.7 \log R + 44.3$ was obtained by the SVM.

From a more comprehensive perspective, this work aims to increase the number of TeV blazar candidates, and break through the limitation of searching for candidates only in HSPs. Using the machine-learning method, we look for the physical properties that truly distinguish TeV from non-TeV blazars, and quantify the performance of distinguishing boundaries. We also calculate the probability that a source can be called a TeV source based on these physical properties and find sources with a higher probability of being called a TeV source from non-TeV sources. The data are compiled from four catalogs: 4FGL-DR2/4LAC-DR2, the 3rd Catalog of Hard Fermi-LAT Sources (3FHL; Ajello et al. 2017), the 3rd catalog of HSP blazars (3HSP; Chang et al. 2019), and The Second Brazil-ICRANet Gamma-ray Blazars (2BIGB; Arsioli et al. 2020). Then make sure how many candidates could be detected by the ground-based Cherenkov telescopes. The paper is organized as follows: Section 2 introduces the ML method; Section 3 gives the experiments and results; and we make discussion and draw the conclusion in Section 4.

2. Supervised Machine-learning Method

We intend to use the so-called supervised machine learning (SML) to select TeV blazar candidates among four blazar catalogs. In ML, samples are often referred to as data sets. SML considers a labeled data set as a set of features and a target variable, based on example input–output pairs. The SML method constructs an inferred function (often called a model) to map the features to dispersed target variables from the known data set in a classification task. Moreover, the model can predict labels for the unknown data set. The data set will be divided into three sets: a training set acts as known data to train the model, which is further decomposed into a smaller training set and a validation set which is used to improve the model, and, finally, the test set, acts as unclassified data to evaluate the generalization of the model. The dispersed target variables are often called labels, which mark the class of data (Ball & Brunner 2010; Baron 2019). Each source catalog can be regarded as a data set in a matrix. The parameter columns are the features that characterize the source's physical properties. Additionally, binary labels distinguish TeV blazars from non-TeV blazars. Our work gets the model from training sets and then employs them in the whole data set. Consequently, the non-TeV sources with the same predicted labels as the TeV sources are considered TeV blazar candidates. It is taken as a high-confidence candidate if the possibility of being a TeV source is higher than 80%.

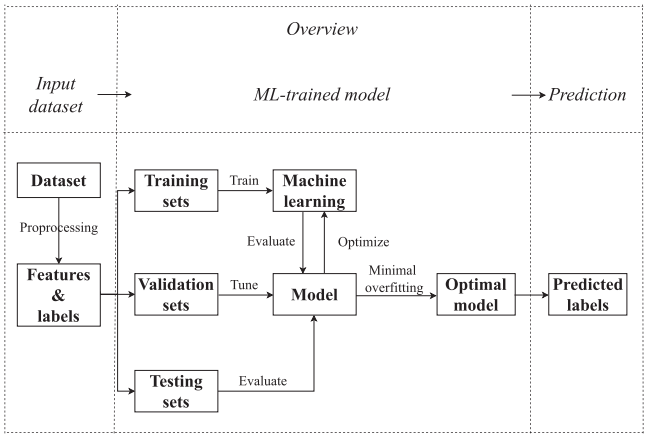

Figure 1. Flowchart of SML.

There are several steps to accomplish this goal, which are demonstrated in Figure 1.

The process of SML classification can be roughly divided into three parts: input data set, ML-trained model, and prediction. Next, we dive into details.

2.1. Data Preprocessing

Data preprocessing is an essential step in the data set before starting the training model. By manipulating or deleting the data, preprocessing of data can change natural features into a more available representation for the downstream model. It contains the following steps:

Data Discretization: In some of the data sets, one feature is represented by multiple columns. For example, in 4FGL-DR2, such as Flux_band, seven columns represent integral photon flux in each spectral band. We divide Flux_band as *Flux_band*1, *Flux band*2, ... *Flux_band*7, respectively.

Data cleaning: There may be some useless and missing data, which will reduce model performance. We discard error features, string features, and most missing value features.

Data standardization: To make all features dimensionless, we use *sklearn.preprocessing.StandardScaler* to convert all features into the standard normal distribution.

Data partition: We randomly divide the data set into a 4:1 training set and testing set using the Python package *StratifiedKFold.split* of *sklearn.model_selection*, and each training/testing set preserves the percentage of the data set for each class. We repeat the division five times with five different random seeds: 0, 1, 2, 3, 4 to get five training sets and five testing sets. We mark the training/testing sets as training 1/testing 1, training 2/testing 2, ..., training 5/testing 5.

2.2. Selecting SML Model

We chose the SML method based on the following considerations:

1. We request a highly interpretable model that is not overly dependent on computer performance, which outputs an explicitly linear boundary to separate TeV blazars and non-TeV blazars in the feature space.
2. Once we have the physical properties of a blazar, that are, the features, we want to get the conditional probability that the blazar is a TeV one under the features.

There are many highly developed SML classification methods, such as Logistic Regression (Korsós et al. 2021, LR), Support Vector Machines (Hearst 1998, SVM), artificial neural network (McCulloch & Pitts 1943, ANN), random forest (Kam Ho 1995), etc. In this work, we choose the LR model over other methods like SVM, ANN, and the like for two main reasons. Mathematically, we assumed that the binary labels of blazars follow a Bernoulli distribution, and, as the logistic function is the expectation of a Bernoulli distribution, we believe it is the most natural way to map the linear combination of physical properties of the source into a Bernoulli distribution. Algorithmically, SVM, RF, and ANN are more complicated than LR, resulting in higher calculation costs and an overall loss in the ability to interpret the model, moreover the output of SVM needs to go through the LR model to obtain a conditional probability. The same holds true for ANN, although other models may be used to obtain conditional probability. In addition, ANN and RF cannot guarantee a linear boundary.

To illustrate LR, we introduce the odds ratio: $\frac{p}{1-p}$, where p represents the probability of the positive event labeled 1. Then we further define the logit function (log-odds), which can be written as $\mathrm{logit}\,(p) = \log\frac{p}{1-p}$. We can express a linear relationship between features and the log-odds:

$$\mathrm{logit}\,(p(y=1|\boldsymbol{x})) = w_0 + w_1x_1 + \cdots + w_mx_m = \boldsymbol{w}^T\boldsymbol{x}, \quad (1)$$

here y is the label; the positive event is marked as "1," with "0" for the negative events; $p(y=1|\boldsymbol{x})$ is the conditional probability that an individual is labeled as "1," and $\boldsymbol{w}$ is the weight; $p(y=1|\boldsymbol{x})$ (or logistic) is the inverse function of the logit function called the logistic function (or sigmoid function), which can be expressed as:

$$\mathrm{logistic}\,(\mathrm{logit}(p)) = \frac{1}{1+e^{-\mathrm{logit}(p)}}, \quad (2)$$

where $\mathrm{logit}(p)$ means $\mathrm{logit}\,(p(y=1|\boldsymbol{x}))$.

For the implementation, we used *scikit-learn* (sklearn; Pedregosa et al. 2011), an ML library in Python, which provides many useful tools. In particular, we used a sklearn API (Buitinck et al. 2013): *sklearn.linear_model.LogisticRegression* which implements the LR model, to train the training set. Eventually, we used the LR to classify the data set by evaluating the logit value of each source in the data set and calculating the conditional probability logistic that a source emits TeV radiation. Note that the number of TeV sources in our sample is small. To increase the weight of TeV sources, we set the weight parameter *class_weight* to *balance* in *sklearn.linear_model.LogisticRegression*.

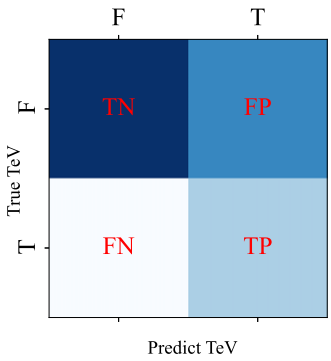

Figure 2. A plot of the confusion matrix, where the coordinate value "T" stands for "True," and "F" stands for "False."

2.3. Evaluating Performance

A confusion matrix is applied to visualize the model performance. True positive (TP) represents the number of positive events that are correctly classified, true negative (TN) represents the number of negative events that are correctly classified, false positive (FP) represents the number of misclassified positive events, false negative (FN) represents the number of misclassified negative events, as shown in Figure 2.

The accuracy ($\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{FP}+\mathrm{FN}+\mathrm{TP}+\mathrm{TN}}$) is one of the most common indexes for evaluating model performance. However, we have an extreme imbalance of the two classes, which means non-TeV sources dominate the results. We also want the model to identify as many TeV sources as possible while reducing the number of non-TeVs misjudged as TeVs. To meet these two requirements, we introduce a larger area under the curve (AUC; Fawcett 2006) as the metric to evaluate LR performance, which is based on the receiver operating characteristic curve (ROC; Fawcett 2006) with the true positive rate (TPR) on the y-axis and the 1- false positive rate (FPR) on the x-axis, where $\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$ and $\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}$. Furthermore, we introduce Youden's J statistic (or Youden's index; Youden 1950), which is TPR − FPR. The optimal p_{thre} corresponds to the maximum Youden index, which means that when we fix TPR, FPR reaches a minimum, which indicates that our screening of TeV candidates is very stringent. So we consider a non-TeV blazar as a TeV one if its logistics are beyond the optimal p_{thre}.

2.4. Training and Tuning Model

Now we can train the models on the five training sets and further optimize the model performance in two ways:

Feature selection: It may degrade model performance if all features are used, which is called the *curse of dimensionality* (Taylor 2019). Data becomes sparse in high-dimensional space, resulting in no more extended similarities between data, which hinders efficient organization and data processing.

The *sklearn.feature_selection.SequentialFeatureSelector* (or SFS for short) of Python could filter useless features, where the basic idea is to sequentially remove features from the full feature space until the new feature subspace contains the desired number of features.

Hyperparameters optimization: Some parameters in the model cannot be obtained from the training model, which is often called hyperparameters; they are not the $\boldsymbol{w}$ in Formula (1).

Instead, we need to specify them manually before the training model.

The *sklearn.grid_search.GridSearchCV* (or GS for short) is helpful in the selection of optimal hyperparameters, which performs a brute-force search on a list of values that we specify for hyperparameters, and picks out the optimal hyperparameters. For *sklearn.linear_model.LogisticRegression*, we optimize six hyperparameters[13]: *penalty, C, solver, tol, max_iter, l1_ratio*.

The optimization strategy is inspired by the k-fold cross validation (Ball & Brunner 2010), which randomly divides the training set into k-folds without replacement. Then $k-1$ folds are used for the training model, and the remaining one fold, namely the validation set, evaluates the model performance. This process was repeated k times.

Based on k-fold cross validation, we do both feature selection and hyperparameter optimization at the same time using nested cross validation (Varma & Simon 2006; Raschka 2015), which is characterized by a nested loop. The data set in the outer loop is divided into a training set and a testing set. The training set in the inner loop undergoes k-fold cross validation selecting optimal hyperparameters, which is performed by GS. Next, the SFS in the outer loop picks out the optimal features, and the testing set evaluates the model generalization. We implement nested cross validation on five partitions of the data set and get five models with optimal features and hyperparameters.

However, a bias–variance trade-off is one of the fundamental rules of machine learning. The bias measures the deviation between the predicted value and the actual value on the known data set, and the variance measures the performance of the model on the unknown data set. Overfitting is a common problem in machine learning, and it is mainly due to high variance, making the performance on the testing set worse than that on the training set. Therefore, for each of the five partitions, we calculate the gap that equals the AUC of the training set minus the AUC of the testing set and selects one optimal model corresponding to the smallest positive gap.

2.5. Constructing Linear Boundary

Now we could build linear boundaries to distinguish TeV-emitting sources from non-TeV-emitting sources. When the logistic of a source is greater than a threshold value ($p_{\rm thre}$), the LR determines the source as a TeV source, otherwise as a non-TeV one. A $p_{\rm thre}$ corresponds to a logit threshold value: $\rm logit_{thre}$ (see Formula (2)). So we could express the linear boundary as $\rm logit' = 0$, where $\rm logit' = logit - logit_{thre}$.

2.6. Predicting Labels

To better understand the process of predicting labels for the unknown data sets, we define some transition concepts. According to Section 2.1, we call the data set that has only undergone data discretization as the initial set. In contrast, the data set that has undergone complete preprocessing is called the learning set. We get the optimal model from the learning set, and the optimal features are also determined. In the initial set, a subset containing the optimal features is called a prediction set. The learning set requires that all features have no missing values, but the prediction set only needs the full optimal parameters, containing more samples that will help to find more TeV candidates. Consequently, we could calculate the logistic for each blazar in the prediction set, and consider the non-TeV blazars with logistic $\geqslant 80\%$ as high-confidence TeV candidates.

[13] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

3. LR Model to Distinguish TeVs from Non-TeVs

3.1. Samples

We used four catalogs: 4FGL-DR2/4LAC-DR2, 3FHL, 3HSP, and 2BIGB to complete our tasks. 4FGL-DR2, based on 10 yr Fermi-LAT data, contains over 5700 sources, 3511 of which are AGNs (3436 are blazars). In the observer frame, 4LAC-DR2 includes synchrotron-peak frequency ($\log \nu_{\rm p}^{\rm s}$), the corresponding intensity ($\log \nu_{\rm p}^{\rm s} f_{\nu_{\rm p}}^{\rm s}$), and redshift ($z$), which are critical supplements to 4FGL-DR2. 3FHL (Ajello et al. 2017) is a 7 yr based catalog of 1556 VHE sources. 3HSP (Chang et al. 2019) includes 2013 HSP blazars containing a critical feature: the peak flux in units of the faintest HSP blazar peak flux ($2.5 \times 10^{-12}\,\rm erg\,cm^{-2}\,s^{-1}$) that has been detected in the TeVCat (Wakely & Horan 2008) and Figure of Merit (FOM). 2BIGB (Arsioli et al. 2020) is the result of gamma-ray likelihood analysis for 1160 3HSP sources yielded photon flux from 500 MeV to 500 GeV: $F_{0.5-500\,\rm GeV}$ ($F^{\rm ph}$).

TeVCat is an online interactive catalog containing VHE sources. As of now, it contains 251 sources detected mainly by IACTs and EAS arrays in the TeV energy region, and at least 91 are extragalactic sources. Among those sources, 80 blazars are detected by Fermi-LAT, thus we assume that blazars emitting TeV radiation are usually detected by Fermi-LAT. We crossmatched the four catalogs with TeVCat, respectively. We marked labels of their blazars in common with TeVCat as "1," while the others were "0." According to the SML workflow introduced in Section 2, we got four groups of initial, learning, and prediction sets:

4FGL-DR2/4LAC-DR2: We supplement 4FGL-DR2 with $\log \nu_{\rm p}^{\rm s}$, $\log \nu_{\rm p} f_{\nu_{\rm p}}$ and z from 4LAC-DR2, obtaining a new catalog: 4FGL-DR2/4LAC-DR2. Initial set: 3436 blazars including 80 TeV ones, with 88 features. Learning set: 861 blazars including 71 TeV ones, with 32 features. 688/173 blazars for each training/testing set. Prediction set: 1459 blazars including 73 TeV ones, with five features.

3FHL: Initial set: 1207 blazars including 74 TeV ones, with 49 features. Learning set: 506 blazars including 63 TeV blazars, with 18 features. 404/102 blazars for each training/testing set. Prediction set: 540 blazars including 67 TeV ones, with two features.

3HSP: Initial set: 2013 blazars including 57 TeV ones, with 13 features. Learning set: 1411 blazars including 53 TeV ones, with five features. 1128/283 blazars for each training/testing set. Prediction set: 1771 blazars including 54 TeV ones, with two features.

2BIGB: Initial set: 1160 blazars including 57 TeV ones, with 31 features. Learning set: 1040 blazars including 54 TeV ones, with seven features. 832/208 blazars for each training/testing set. Prediction set: 1040 blazars including 54 TeV ones, with three features.

3.2. Logit, Logistic, and Performance of LR

Processing the SML method we ascertain the result meeting condition of logistic $\geqslant 80\%$ which gives 40 high-confidence candidates out of 150 4FGL-DR2/4LAC-DR2 candidates (see Section 4.3). Among the 40 high-confidence candidates, 24 sources are in common with 3FHL candidates, 11 with 3HSP candidates, and 14 with 2BIGB candidates. We report the main results of SML as follows:

4FGL-DR2/4LAC-DR2:

$$\begin{aligned}\mathrm{logit}' = {} & 4.808\Gamma + 2.809V_F + 3.889\log f_7^{\mathrm{ph}} \\ & - 3.34z + 0.857\log \nu_{\mathrm{p}}^{\mathrm{s}} + 20.244, \end{aligned} \quad (3)$$

where $\mathrm{logit}' = 0$ is ideal for $p_{\mathrm{thre}} = 40\%$. The optimal LR model is built on training 3, with AUC being 97% on training 3 and 96% on testing 3. 150 TeV blazar candidates are obtained from 1459 blazars in the prediction set, with AUC: 98%, FPR: 11%, and TPR: 99%. Only one TeV blazar is mistaken as a non-TeV one.

3FHL:

$$\mathrm{logit}' = 0.116\log f_2^{\mathrm{ph}} - 0.628z + 1.028, \quad (4)$$

where $\mathrm{logit}' = 0$ is ideal for $p_{\mathrm{thre}} = 52\%$, the optimal LR model is built on training 5, with AUC being 89% on training 5 and 87% on testing 5. 126 TeV candidates are obtained from 540 blazars, with AUC: 88%, FPR: 27%, and TPR: 88%. Only eight TeV blazars are mistaken as non-TeV ones.

3HSP:

$$\mathrm{logit}' = 0.306\mathrm{FOM} - 3.861z - 0.173, \quad (5)$$

where $\mathrm{logit}' = 0$ is ideal for $p_{\mathrm{thre}} = 61\%$, the optimal LR model is built on training 5, with AUC being 99% on training 5 and 94% on testing 5, 40 TeV candidates are obtained from 1771 blazars, with AUC: 98%, FPR: 2%, and TPR: 91%, only five TeV blazars are mistaken as non-TeV ones.

2BIGB:

$$\mathrm{logit}' = -0.016E_{\mathrm{P}} + 0.248\mathrm{FOM} - 4.395z - 0.122 \quad (6)$$

where $\mathrm{logit}' = 0$ is ideal for $p_{\mathrm{thre}} = 48\%$, the optimal LR model is built on training 4, with AUC being 97% on training 4 and 97% on testing 4. 83 TeV candidates are obtained from 1040 blazars, with AUC: 97%, FPR: 8%, and TPR: 96%, only two TeV blazars are mistaken as non-TeV ones.

The features of the four learning sets are reported in Table 1, where Col. (1) indicates the catalog information; Col. (2) gives how many columns the features contain, and each column represents a feature; Col. (3) and Col. (4) show the names and units of the features from the FITS version of the catalogs; Col. (5) is the description of the feature. The performance of the LR model on four learning sets is also shown in Tables 2–5, where Col. (1) shows the data sets name; Col. (2) the AUC, where the top is for training sets while the bottom is for testing sets; Col. (3) AUC difference between training sets and testing sets; Col. (4) is the p_{thre}; Col. (5) is optimal features; Col. (6) is optimal hyperparameters, note that the bold corresponds to minimum overfitting data sets, which represents the data and parameters of the optimal LR model.

We also visualize part of the results in figures. Figure 3 shows feature selection. Figure 4 is AUC of the training/testing sets. Figure 5 illustrates ROC curve, AUC, and p_{thre} of the prediction sets. Figure 6 visualizes the linear boundary of 3FHL, 3HSP, and 2BIGB. Figure 7 is the confusion matrix.

4. Potential Targets for Ground-based Cherenkov Detectors

Among the 40 high-confidence TeV candidates, we consider the potential detectable targets of IACTs and EAS arrays for the whole year of 2023. We employ two criteria to determine whether a source is detectable: (i) Its flux density is higher than the detector sensitivity curve at the energy band $\geqslant 1$ TeV; (ii) It could pass by the detectable sky region of the detectors.

4.1. SEDs Considering EBL Correction

The tactic is to correct the data by considering EBL first and then fitting SEDs on the intrinsic data. The SED contains two bumps, a low-energy one in the infrared to soft X-ray energy range due to synchrotron emission and a high-energy one in the region between hard X-ray to gamma ray that is associated with IC emission. To do so, we use an online tool[14] provided by the space science data center of the Italian Space Agency to collect data from radio to TeV from many missions and experiments together with catalogs and archival data (Aharonian et al. 2001, 2003, 2009; Giommi et al. 2002; Amenomori et al. 2003; Daniel et al. 2005; Schroedter et al. 2005; Acciari et al. 2008, 2011; Godambe et al. 2008; Chandra et al. 2010, 2012; HESS Collaboration et al. 2010, 2013; Aliu et al. 2011, 2015; Bartoli et al. 2011, 2012; Abramowski et al. 2013, 2015; Archambault et al. 2013, 2014; Arlen et al. 2013; Biteau & Williams 2015; Sharma et al. 2015). When we fit the SEDs, we discard the data bins whose flux error is bigger than the flux upper limits. Note that there are TeV photons with zero error of the TeV candidates. Those TeV band photons are not certified by IACTs or EAS array (see Figure 6). Likewise, we also discard data in the low radio energy range ($\log \nu(\mathrm{Hz}) < 9$), since there is an asymmetry in the synchrotron bump of some sources. In this paper, we do not take into account the simultaneous observation of the data. EBL photons interact with VHE photons via pair production, a process that induces an attenuation of the observed gamma-ray spectra from blazars at cosmological distances, which may have a large impact on SED in IC bump. Assuming the EBL optical depth ($\tau(E, z)$) to depend on the photon energy (E) and redshift (z), Saldana-Lopez et al. (2021) provided a 2D grid of $\tau(E, z)$ [15] in the range of 0.001 TeV $\leqslant E \leqslant$ 100 TeV, and $0 \leqslant z \leqslant 6$. Subsequently, the two bumps of the observed SED were fitted separately with a log parabola as follows (Fan et al. 2016):

$$\log(\nu f_\nu) = c(\log \nu - \log \nu_{\mathrm{p}})^2 + \log \nu_{\mathrm{p}} f_{\nu_{\mathrm{p}}}, \quad (7)$$

where $|2c|$ is the spectral curvature, $\log \nu_{\mathrm{p}}$ is the logarithm of the peak frequency, $\log \nu_{\mathrm{p}} f_{\nu_{\mathrm{p}}}$ is the logarithm of peak flux, and $\log(\nu f_\nu)$ is the flux density.

we successfully fit the synchrotron bump for 40 TeV candidates and the IC bump for 20 TeV candidates and returned the errors of the fitting $|2c|$, $\log \nu_{\mathrm{p}}$ (in units of Hz) and $\log \nu_{\mathrm{p}} f_{\nu_{\mathrm{p}}}$ (in units of erg cm^{-2} s^{-1}). The fitting process

[14] http://tools.asdc.asi.it/SED/
[15] https://www.ucm.es/blazars/ebl

Table 1
Features from Four Catalogs for Learning Sets

Catalog (1)	Column (2)	Feature (3)	Unit (4)	Description (5)
4FGL-DR2/ 4LAC-DR2	1	Pivot_Energy ($E_{\rm P}$)	GeV	Energy at which error on differential flux is minimal
	1	Flux ($\log F^{\rm ph}$)	$\rm cm^{-2}\ s^{-1}$	Integral photon flux from 1 to 100 GeV
	1	Energy_Flux ($\log F$)	$\rm erg\ cm^{-2}\ s^{-1}$	Energy flux from 100 MeV to 100 GeV
	1	PL_Flux_Density ($\log f_1$)	$\rm cm^{-2}\ MeV^{-1}\ s^{-1}$	Differential flux at Pivot_Energy in PowerLaw fit
	1	PL_Index (α_1)		Photon index when fitting with PowerLaw
	1	LP_Flux_Density ($\log f_2$)	$\rm cm^{-2}\ MeV^{-1}\ s^{-1}$	Differential flux at Pivot_Energy in LogParabola fit
	1	LP_Index (α_2)		Photon index at Pivot_Energy when fitting with LogParabola
	1	LP_beta (β)		Curvature parameter when fitting with LogParabola
	1	PLEC_Flux_Density ($\log f_3$)	$\rm cm^{-2}\ MeV^{-1}\ s^{-1}$	Differential flux at Pivot_Energy in PLSuperExpCutoff fit
	1	PLEC_Index (Γ)		Low-energy photon index when fitting with PLSuperExpCutoff
	1	PLEC_Expfactor (a)		Exponential factor when fitting with PLSuperExpCutoff
	1	PLEC_Exp_Index (b)		Exponential index when fitting with PLSuperExpCutoff
	7	Flux_Band ($\log f_1 \sim \log f_7$)	$\rm cm^{-2}\ s^{-1}$	Integral photon flux in each spectral band
	7	nuFnu_Band ($\log f_1^{\rm ph} \sim \log f_7^{\rm ph}$)	$\rm erg\ cm^{-2}\ s^{-1}$	Spectral energy distribution over each spectral band
	1	Variability_Index (V)		Likelihood difference between the flux fitted in each time interval and the average flux
	1	Frac_Variability ($V_{\rm F}$)		Fractional variability computed from the fluxes in each year
	1	Redshift (z)		Redshift
	1	nu_syn ($\log \nu_{\rm p}^{\rm s}$)	Hz	Synchrotron-peak frequency in observer frame
	1	nuFnu_syn ($\log \nu_{\rm p}^{\rm s} f_{\nu_{\rm p}}^{\rm s}$)	$\rm erg\ cm^{-2}\ s^{-1}$	Spectral energy distribution at synchrotron-peak frequency
	1	Highest_energy ($E_{\rm H}$)	GeV	Highest energy among events probably coming from the source
3FHL	1	Pivot Energy ($E_{\rm P}$)	GeV	Energy at which error on differential flux is minimal
	1	Flux Density ($\log f$)	$\rm cm^{-2}\ GeV^{-1}\ s^{-1}$	Differential flux at Pivot Energy
	1	Flux ($\log F^{\rm ph}$)	$\rm cm^{-2}\ s^{-1}$	Integral photon flux from 10 GeV to 1 TeV erg $\rm cm^{-2}\ s^{-1}$
	1	Energy Flux ($\log F$)	$\rm erg\ cm^{-2}\ s^{-1}$	Energy flux from 10 GeV to 1 TeV
	1	PowerLaw Index (α_1)		Photon index when fitting with power law
	1	Spectral Index (α_2)		Photon index at Pivot Energy when fitting with LogParabola
	1	beta (β)		Curvature parameter when fitting with LogParabola
	4	Flux Band ($\log f_1 \sim \log f_4$)	$\rm cm^{-2}\ s^{-1}$	Integral photon flux in each spectral band
	4	nuFnu_Band ($\log f_1^{\rm ph} \sim \log f_4^{\rm ph}$)	$\rm erg\ cm^{-2}\ s^{-1}$	Spectral energy distribution over each spectral band
	1	HEP energy ($E_{\rm H}$)	GeV	Highest energy among events probably coming from the source
	1	Redshift (z)		Redshift
	1	NuPeak obs ($\log \nu_{\rm p}^{\rm s}$)	Hz	Synchrotron-peak frequency in observer frame
3HSP	1	radio flux density ($\log f_{\rm R}$)	mJy	Radio flux density from the NVSS or FIRST catalog
	1	X-ray flux density ($\log f_{\rm X}$)	μJy	X-ray flux density at 1 keV
	1	nu_syn ($\log \nu_{\rm p}^{\rm s}$)	Hz	Synchrotron-peak frequency in observer frame
	1	Redshift (z)		Redshift
	1	FOM (FOM)		The figure of merit parameter, which is related to the likelihood of GeV/TeV detectability
2BIGB	1	N0 ($\log f$)	$\rm cm^{-2}\ MeV^{-1}\ s^{-1}$	Differential flux at Pivot Energy in PowerLaw fit
	1	Gamma (α)		Photon index when fitting with PowerLaw
	1	$\rm F_{0.5-500GeV}$ ($F^{\rm ph}$)		Integrated photon flux from 500 MeV to 500 GeV
	1	E0 ($E_{\rm P}$)	GeV	Energy at which error on differential flux is minimal
	1	nu_syn ($\log \nu_{\rm p}^{\rm s}$)	Hz	Synchrotron-peak frequency in observer frame from 3HSP
	1	Redshift (z)		Redshift from 3HSP
	1	FOM (FOM)		The same as in 3HSP

executes on *scipy.optimize.curve_fit* of Python. The results are also listed in Table 6, in which Col. (1) gives the 4FGL name, Col. (2) to Col. (4) are parameters and errors for fitting synchrotron bump, Col. (5) to Col. (7) are parameters and errors for fitting IC bump, while the right side of the slash is the data without EBL absorption. For each high-confidence TeV candidate, the SEDs fitting results are shown in Figure 11.

4.2. Visibility of TeV Blazar Candidates

Being different from the all-sky scanning characteristics of space detectors, ground detectors can only scan part of the sky due to location constraints. IACTs need a dark sky for observations, and the field of view (FoV) of the current IACTs is small (3°–5°). Additionally, the IACT's duty cycle is restricted by the need to observe only during clear-sky, moonless nights, and, in addition, further constraints stem

Table 2
LR Model and Performance for the Learning Data Set of 4FGL-DR2/4LAC-DR2

Partition (1)	AUC (2)	Overfitting (3)	$p_{\rm thre}$ (4)	Features (5)	Hyperparameters (6)
1	96.9% 95.6%	0.9%	53.6%	$E_{\rm p}$, $\log F_1^{\rm ph}$, $\log F_{\rm l}$, Γ, V, $V_{\rm f}$, $\log f_1^{\rm ph}$, $\log f_2^{\rm ph}$, $\log f_4^{\rm ph}$, $\log f_6^{\rm ph}$, $\log f_7^{\rm ph}$, z, $\log \nu_{\rm p}^{\rm s}$	C: *1*, max_iter: *500*, penalty: *l2*, solver: *sag*, tol: 10^{-4}
2	97.2% 96.0%	1.2%	53.5%	$E_{\rm p}$, F_1, α_1, Γ, V, $V_{\rm f}$, $f_7^{\rm ph}$, z, $\log \nu_{\rm p}^{\rm s}$	C: *1*, max_iter: *100*, penalty: *l1*, solver: *liblinear*, tol: 10^{-6}
3	96.8% 96.0%	0.8%	53.4%	Γ, $V_{\rm f}$, $\log f_7^{\rm ph}$, z, $\log \nu_{\rm p}^{\rm s}$	C: *1*, max_iter: *500*, penalty: *l1*, solver: *saga*, tol: 10^{-6}
4	96.0% 98.0%	−1.8%	53.0%	$\log F_{\rm l}$, α_1, Γ, a, $V_{\rm f}$, $\log f_7$, $\log F_7^{\rm ph}$, z, $\log \nu_{\rm p}^{\rm s}$	C: *0.1*, max_iter: *100*, penalty: *l2*, solver: *newton-cg*, tol: 10^{-6}
5	97.5% 92.5%	5.0%	52.2%	$E_{\rm p}$, α_1, α_3, $V_{\rm f}$, $\log f_3$, $\log f_5$, $\log f_7$, $\log f_7^{\rm ph}$, z, $\log \nu_{\rm p}^{\rm s}$	C: *1*, max_iter: *100*, penalty: *l2*, solver: *newton-cg*, tol: 10^{-6}

Table 3
LR Model and Performance for the Learning Set of 3FHL

Partition (1)	AUC (2)	Overfitting (3)	$p_{\rm thre}$ (4)	Features (5)	Hyperparameters (6)
1	87.8% 89.7%	−1.9%	53.6%	$\log f_2^{\rm ph}$, z	C: *0.01*, l1_ratio: *0.4*, max_iter: *100*, penalty: *elasticnet*, solver: *saga*, tol: 10^{-6}
2	88.0% 95.5%	−7.4%	53.5%	$\log f_2^{\rm ph}$, z	C: *0.01*, l1_ratio: *0.3*, max_iter: *100*, penalty: *elasticnet*, solver: *saga*, tol: 10^{-6}
3	88.7% 83.1%	5.5%	53.4%	$\log f_2^{\rm ph}$, z	C: *0.01*, l1_ratio: *0.4*, max_iter: *100*, penalty: *elasticnet*, solver: *saga*, tol: 10^{-6}
4	89.6% 86.0%	3.6%	52.9%	$\log f_2^{\rm ph}$, z	C: *0.01*, l1_ratio: *0.2*, max_iter: *100*, penalty: *elasticnet*, solver: *saga*, tol: 10^{-6}
5	89.2% 87.0%	2.3%	52.2%	$\log f_2^{\rm ph}$, z	C: *0.01*, l1_ratio: *0.4*, max_iter: *100*, penalty: *elasticnet*, solver: *saga*, tol: 10^{-6}

from the zenith angle: while the zenith angle increases, the sensitivity worsens and the energy threshold increases. This results in limiting the portion of the sky available for observation. At the same time, CTAO, as the next generation of IACT, consists of two arrays located in the Northern (28° 45′ N) and the Southern (24° 41′S) hemispheres so that the FoV covers most of the sky, but not all, since CTAO suffers the same limitations affecting the current generation IACTs observations at high Zenith Angles (Z). Conversely, the EAS arrays can work under all weather conditions and have large FoV (∼2 sr), which can continuously monitor a significant fraction of the sky every day.

We compile the sensitivity curves of IACTs and EAS arrays from (van Eldik et al. 2015; Aleksić et al. 2016; Abeysekara et al. 2017a, 2017b; Cao et al. 2019) and online database: CTAO,[16] H.E.S.S.,[17] VERITAS.[18] The FoV of these sensitivity curves are as follows: CTAO north (north site and $0° \leqslant Z \leqslant 20°$), CTAO south (south site and $0° \leqslant Z \leqslant 20°$), H.E.S.S. zenith ($Z \sim 0°$), H.E.S.S. medium ($12° \leqslant Z \leqslant 22°$), MAGIC low ($0° \leqslant Z \leqslant 30°$), MAGIC medium ($30° \leqslant Z \leqslant 45°$), VERITAS ($0° \leqslant Z \leqslant 20°$), LHAASO ($-11° \leqslant$ decl. $\leqslant 69°$), and HAWC ($-20° \leqslant$ decl. $\leqslant 60°$). The exposure times are 507 days for HAWC, 1 yr for LHAASO, 25 hr for H.E.S.S. near the zenith, and 50 hr for the rests.

Since EAS arrays can observe all sources in the FoV simultaneously according to the source decl., whereas IACTs can only track one source at a time and require dark sky, we employ two different strategies to infer the visibility of a given source. A source could pass the observation window for EAS arrays if its decl. is within the most observable sky. On the other hand, for IACTs, we evaluate observation windows by two tools: the H.E.S.S. online visibility tool[19] and a Python package: astroplan. Input the celestial coordinates of the source and the observation information of IACTs to these two

[16] https://www.cta-observatory.org/science/ctao-performance/
[17] https://www.mpi-hd.mpg.de/hfm/HESS/pages/home/proposals/sc_sens.pdf
[18] https://veritas.sao.arizona.edu/about-veritas/veritas-specifications
[19] https://www.mpi-hd.mpg.de/hfm/HESS/pages/home/visibility/

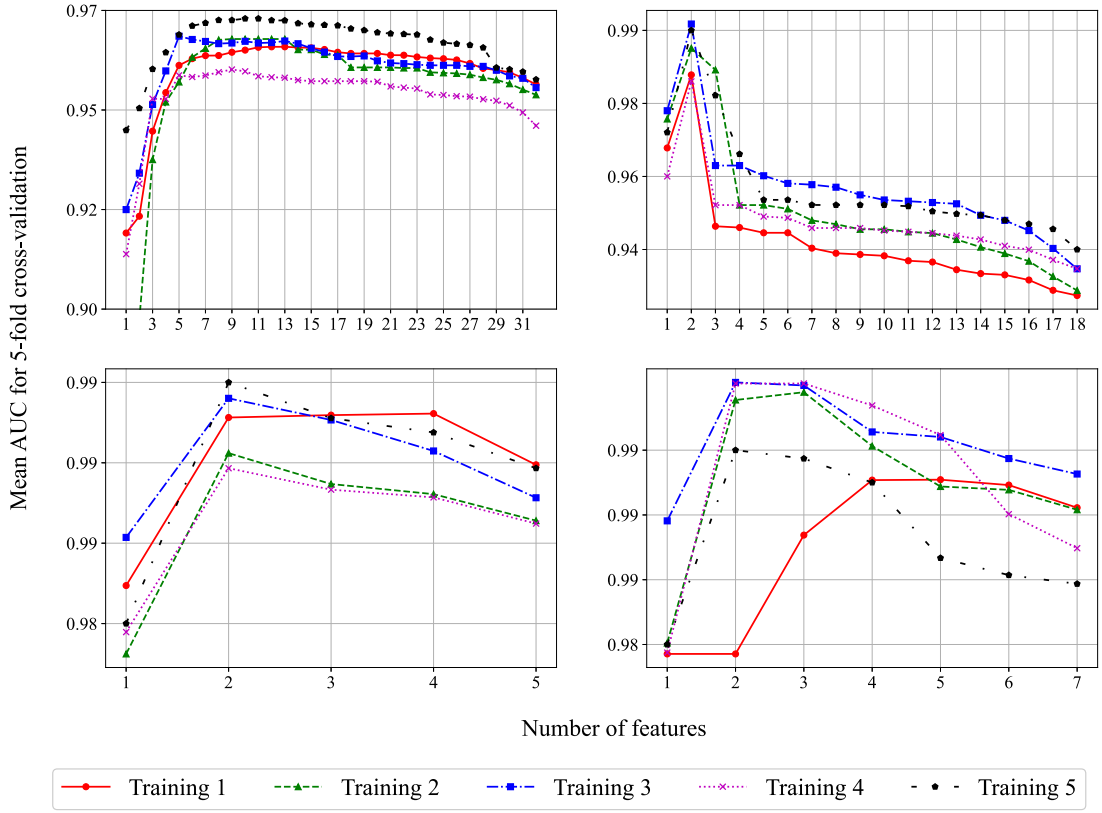

Figure 3. SFS for four learning sets, where the x-axis is the number of parameters, and the y-axis is the concentration of AUC on the five validation sets. five dotted lines in different colors indicate training $1 \sim 5$. Top left panel: 4FGL-DR2/4LAC-DR2. Top right panel: 3FHL. Bottom left panel: 3HSP. Bottom right panel: 2BIGB.

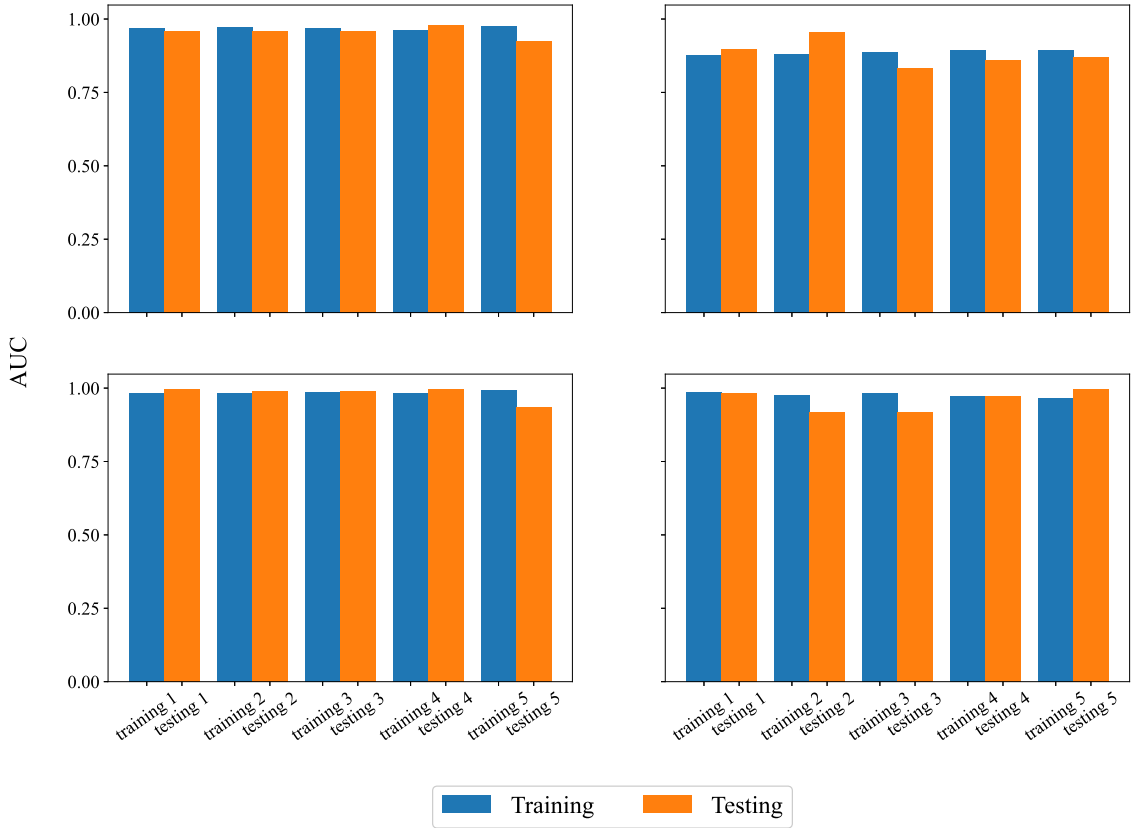

Figure 4. AUC for four learning sets, where blue histograms denote the training sets, and orange histograms represent the testing sets. Top left panel: 4FGL-DR2/4LAC-DR2. Top right panel: 3FHL. Bottom left panel: 3HSP. Bottom right panel: 2BIGB.

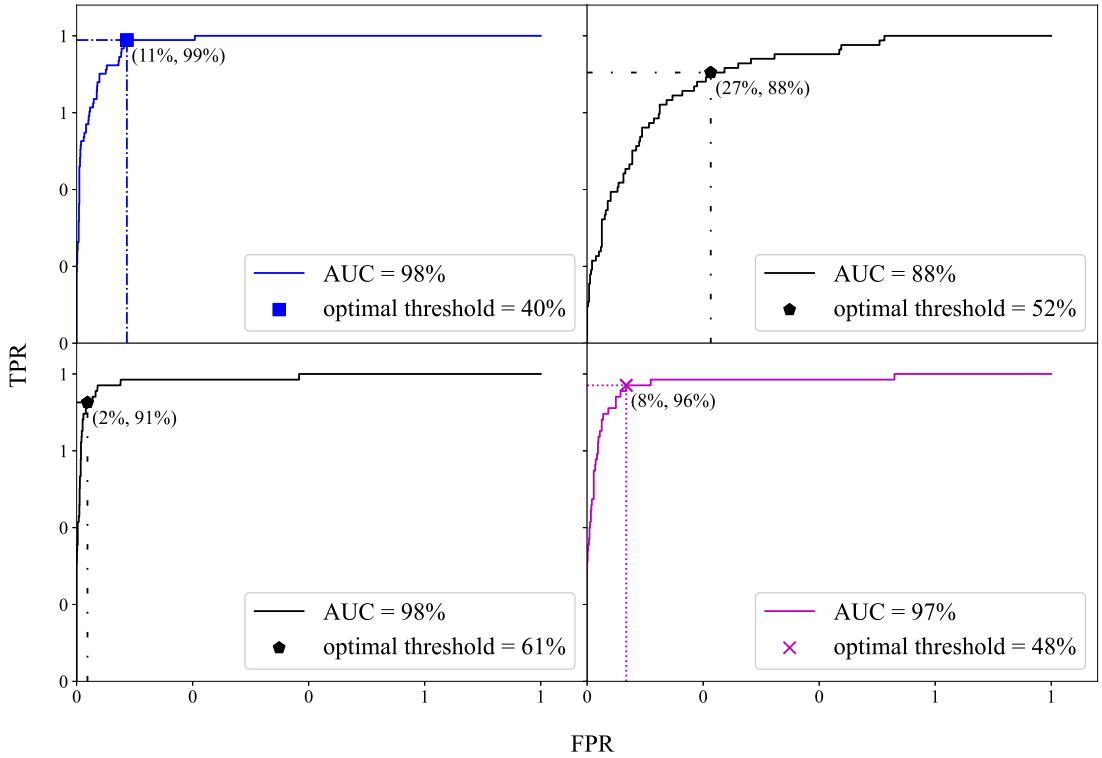

Figure 5. ROC curve for four predicting data sets. Top left panel: 4FGL-DR2/4LAC-DR2. Top right panel: 3FHL. Bottom left panel: 3HSP. Bottom right panel: 2BIGB.

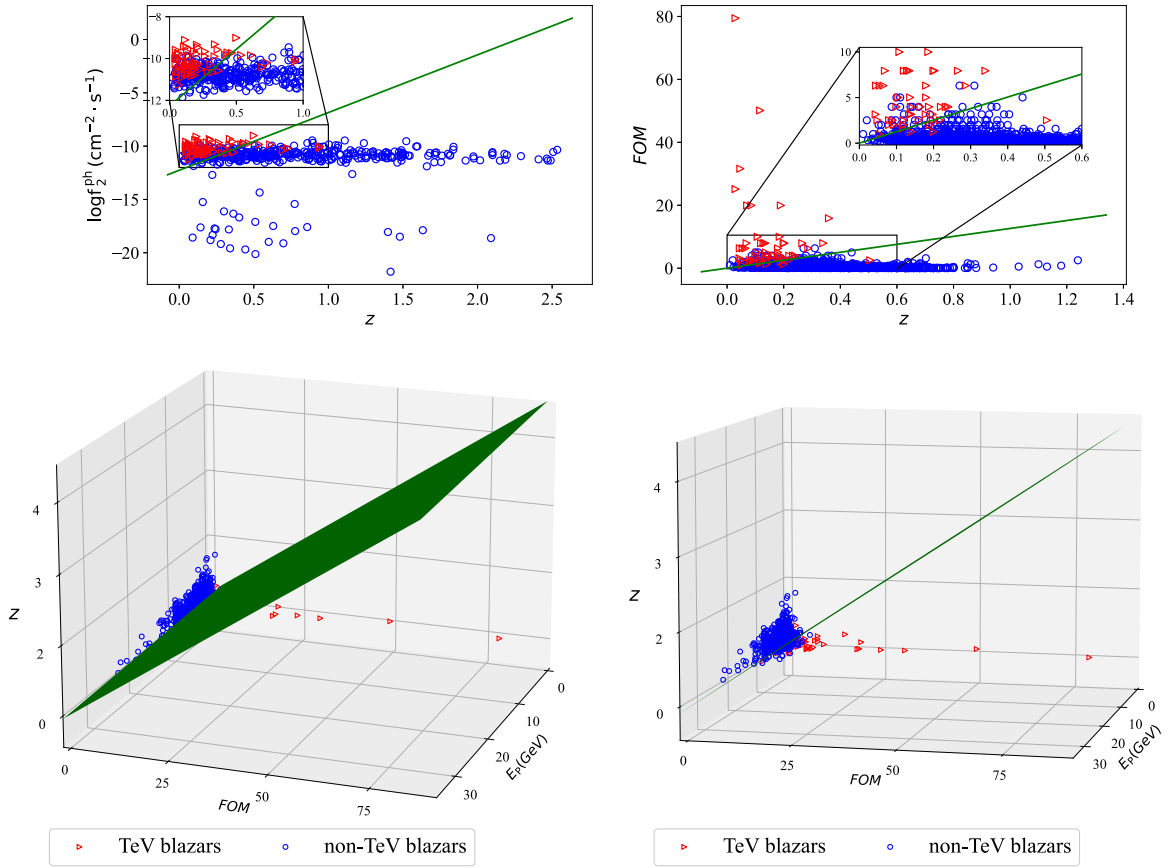

Figure 6. Classify boundary and part of predicting data sets of 3FHL, 3HSP, and 2BIGB, where the red triangle stands for TeV blazars, the blue circle stands for non-TeV ones, and the green line/plane stands for classification boundary. Top left panel: 3FHL. Top right panel: 3HSP. Bottom left panel: 2BIGB on the view of non-TeVs side. Bottom right panel: 2BIGB on the view along the edge of the green plane.

Table 4
LR Model and Performance for the Learning Set of 3HSP

Partition (1)	AUC (2)	Overfitting (3)	$p_{\rm thre}$ (4)	Features (5)	Hyperparameters (6)
1	98.3% 99.5%	−1.2%	46.9%	$\log f_{\rm X}$, $\log \nu_{\rm p}^{s}$, FOM, z	C: *0.01*, l1_ratio: *0.7*, max_iter: *100*, penalty: *elasticnet*, solver: *saga*, tol: 10^{-6}
2	98.4% 98.9%	−0.5%	60.2%	FOM, z	C: *0.01*, max_iter: *100*, penalty: *l2*, solver: *liblinear*, tol: 10^{-6}
3	98.6% 99.1%	−0.5%	47.6%	FOM, z	C: *0.001*, max_iter: *100*, penalty: *l2*, solver: *newton-cg*, tol: 10^{-6}
4	98.4% 99.8%	−1.4%	58.2%	FOM, z	C: *0.1*, max_iter: *100*, penalty: *l2*, solver: *newton-cg*, tol: 10^{-6}
5	99.2% 93.5%	5.7%	60.8%	FOM, z	C: *0.01*, max_iter: *100*, penalty: *l2*, solver: *liblinear*, tol: 10^{-6}

Table 5
LR Model and Performance for the Learning Set of 2BIGB

Partition (1)	AUC (2)	Overfitting (3)	$p_{\rm thre}$ (4)	Features (5)	Hyperparameters (6)
1	98.6% 98.4%	0.2%	44.7%	α, F, $\log \nu_{\rm p}^{s}$, FOM, z	C: *0.01*, max_iter: *100*, penalty: *l2*, solver: *newton-cg*, tol: 10^{-6}
2	97.7% 91.7%	6.0%	52.4%	$E_{\rm p}$, FOM, z	C: *0.01*, max_iter: *100*, penalty: *l2*, solver: *newton-cg*, tol: 10^{-6}
3	98.2% 91.8%	6.3%	49.1%	FOM, z	C: *0.01*, max_iter: *100*, penalty: *l2*, solver: *newton-cg*, tol: 10^{-6}
4	97.2% 97.1%	0.1%	48.1%	$E_{\rm p}$, FOM, z	C: *0.01*, l1_ratio: *0.1*, max_iter: *100*, penalty: *elasticnet*, solver: *saga*, tol: 10^{-6}
5	96.7% 99.7%	−3.0%	50.3%	FOM, z	C: *0.01*, max_iter: *100*, penalty: *l2*, solver: *newton-cg*, tol: 10^{-6}

tools. We can get visible months during darkness at a specific elevation angle ($1 - Z$), as shown in Figure 8 and Table 7. The two tools have slight differences in the specific observable months. The observable months of 40 high-confidence TeV blazar candidates for the two tools are also shown in Table 7, where Col. (1) gives the source name, and Col. (2) and Col. (7) give the observable months for IACTs and EAS arrays. The numbers represent the observable months, i.e., "1" for January, "2" for February, and so on.

4.3. High-confidence TeV Blazar Candidates

We calculate the $\tau(E, z)$ for 40 TeV candidates with E in the range of 0.001–100 TeV, and get the EBL-absorbed IC bumps. By comparing the 20 IC bumps and the sensitivity curves of IACTs and EAS arrays in the energy range of 1 TeV $\leqslant E \leqslant$ 100 TeV, we obtain seven out of 40 high-confidence TeV blazar candidates that satisfy the two detectable criteria. The source number can be detected by the TeV facilities are (see Table 8 for detail): six CTAO (north or south) targets, two H.E.S.S targets, five MAGIC targets, and one VERITAS target, and EAS arrays: one LHAASO targets, zero HAWC targets. The SML and detectability results of the 40 high-confidence TeV blazar candidates are shown in Table 8, where Col. (1) gives the 4FGL name; Col. (2) and Col. (3) give the galactic coordinates (in degrees); Col. (4) synchrotron-peak frequency ($\log \nu_{\rm p}^{s}$); Col. (5) redshift (z); Col. (6) flux density EBL absorption of sources at 1 TeV ($f_{1\ \rm TeV}$) which can be computed using Formula (7) the parameters of the IC bump; Col. (7) to Col. (10) logistic in four catalogs, which is the likelihood of a source belonging to TeV sources given its optimal features Col. (11) the SED class in 4FGL-DR2/4LAC-DR2, where "bll" stands for BL Lac, "fsrq" for FSRQ, and "bcu" for blazar candidate of uncertain type; Col. (12) to Col. (14) candidates compared with those of 3FHL, 3HSP, and 2BIGB, a candidate in common is marked as "Y;" Col. (15) a candidate comparing with those of other literature, "C & G" represents the candidate is the same with Costamante & Ghisellini (2002), while "M" for Massaro et al. (2013), "F" for Foffano et al. (2019) "Chi" for Chiaro et al. 2019; Col. (16) IC bump of candidates compare with the sensitivity of IACTs and particle detectors arrays in the range of 1 TeV $\leqslant E \leqslant$ 100 TeV. "CN," "CS," "HZ," "HL," "ML," "MM," "V," "L," and "H," stand for CTAO north, CTAO south, H.E.S.S zenith, H.E.S.S low, MAGIC low, MAGIC medium, VERITAS, LHAASO, and HAWC, respectively.

We also want to know how far away the source becomes undetectable in the TeV energy band, i.e., the redshift upper

Table 6
SEDs Fitting Results in the Observed Frame for 40 TeV Candidates[a]

4FGL Name (1)	$\lvert c^s\rvert$ (2)	$\log \nu_p^s$ (3)	$\log \nu_p^s f_p^s$ (4)	$\lvert c^{IC}\rvert$ (5)	$\log \nu_p^{IC}$ (6)	$\log \nu_p^{IC} f_p^{IC}$ (7)
J0037.8+1239	0.11 ± 0.02	15.35 ± 0.46	−11.17 ± 0.07	0.02 ± 0.02	23.15 ± 2.42	−11.74 ± 0.1
J0051.2-6242	0.16 ± 0.01	15.72 ± 0.06	−11.23 ± 0.03	0.08 ± 0.07	25.57 ± 1.2	−11.37 ± 0.13
J0110.1+6805	0.1 ± 0.01	14.99 ± 0.29	−10.89 ± 0.09			
J0110.7-1254	0.05 ± 0.01	17.76 ± 1.09	−11.81 ± 0.22			
J0115.8+2519	0.09 ± 0.01	15.67 ± 0.26	−11.54 ± 0.04			
J0123.7-2311	0.05 ± 0.01	17.47 ± 0.88	−11.72 ± 0.19			
J0159.5+1046	0.09 ± 0.0	15.63 ± 0.18	−11.66 ± 0.03	0.03 ± 0.02	22.26 ± 0.67	−11.6 ± 0.11
J0209.3-5228	0.11 ± 0.02	15.8 ± 0.09	−11.08 ± 0.05	0.07 ± 0.04	24.99 ± 0.57	−11.43 ± 0.04
J0211.2+1051	0.12 ± 0.0	14.8 ± 0.1	−10.57 ± 0.02	0.06 ± 0.05	22.53 ± 1.04	−11.04 ± 0.07
J0244.6-5819	0.12 ± 0.03	16.17 ± 0.49	−11.15 ± 0.12			
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...

Note.
[a] Only 10 items are displayed. A complete listing of this table is available in the online version.
(This table is available in its entirety in machine-readable form.)

limit of the source beyond which it will become undetectable. For this purpose, in the range of $0 \leqslant z \leqslant 6$, we compare the EBL-absorbed IC bumps for 20 blazars and the sensitivity curves of detectors, at $E = 1$ TeV, 10 TeV, and 100 TeV, respectively. Redshifts upper limits for the 20 blazars for each IACT and EAS array are listed in Table 9, where Col. (1) gives the 4FGL name; Col. (2) to Col. (10) are the redshift upper limits of the 20 blazars for each, where the values in each column are redshift upper limits at $E = 1$ TeV, $E = 10$ TeV, and $E = 100$ TeV, respectively. The value "0" means the IC bump cannot be detected at the specific E; Col. (11) redshifts; Col. (12) is the detectability for Cherenkov detectors. One can see the VHE photons could survive from EBL easier than those with higher redshift. This can also be seen from Formula (3) ∼ (6), the negative coefficient of z leads logit to be inversely proportional to z. A larger redshift will result in a smaller logistic, indicating a source is less likely to be a TeV candidate.

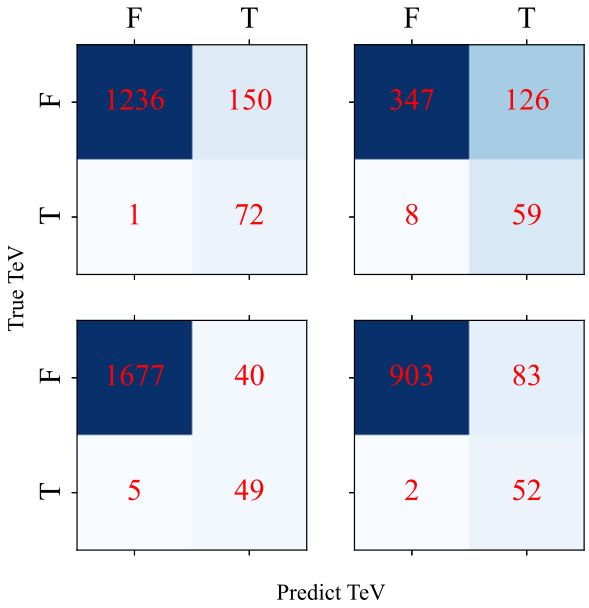

Figure 7. Confusion matrix for four learning sets, where the coordinate value "T" stands for "True," and "F" stands for "False." Top left panel: 4FGL-DR2/4LAC-DR2. Top right panel: 3FHL. Bottom left panel: 3HSP. Bottom right panel: 2BIGB. For each panel: True positives (top left), false negatives (top right), false positives (bottom left), and true negatives (bottom right).

5. Discussions and Conclusions

5.1. Discussions

The SFS identified five features in 4FGL-DR2/4LAC-DR2 that are most critical to distinguish TeVs from non-TeVs: flux density, FOM, synchrotron-peak frequency, redshift, spectral index, and variability, which are compatible with Lin & Fan (2016). Further considering X-ray emission and the $\log \nu_p^s$ shifts in behavior during a flare should be further analyzed in future work since blazars that are not generally detected at TeV energies could be detected during flaring states. Therefore, it is possible that all blazars can emit TeV photons. The blazars with TeV emission detected are only because their flux density in the TeV energy band is relatively high in specific periods, such as the flaring period. Note that 4FGL-DR2 includes features such as *Variability_Index*, which mirrored the difference between the flux fitted in each time interval and the average flux over the entire catalog interval. Furthermore, *Frac_Variability* is the fractional variability computed from the fluxes in each year. The logit′ of 4FGL-DR2/4LAC-DR2 contains the *Frac_Variability* making sure the LR is sensitive to the sources with synchrotron-peak shifting during flaring states.

According to the classification proposed by Abdo et al. (2010) or Fan et al. (2016), sources with $\log \nu_p^s > 15$ (or $\log \nu_p^s > 15.3$) are HSPs. In TeVCat, if we considered only HSP/HSP BL Lacs, we would have ignored more than 11% of TeV sources. We presented here an alternative way, where we have not applied many initial cuts by filtering on SED class or other properties. To compare the performance of two criteria: HSPs ($\log \nu_p^s \geqslant 15.3$ Hz Fan et al. 2016) and the LR model, we calculate the metric and confusion matrix on the testing sets of 4FGL-DR2/4LAC-DR2 and 3FHL. For 4FGL-DR2/4LAC-DR2, the AUC: 73% and TPR: 62% are both lower than those predicted by LR, up to 27 TeV blazars are misclassified as non-

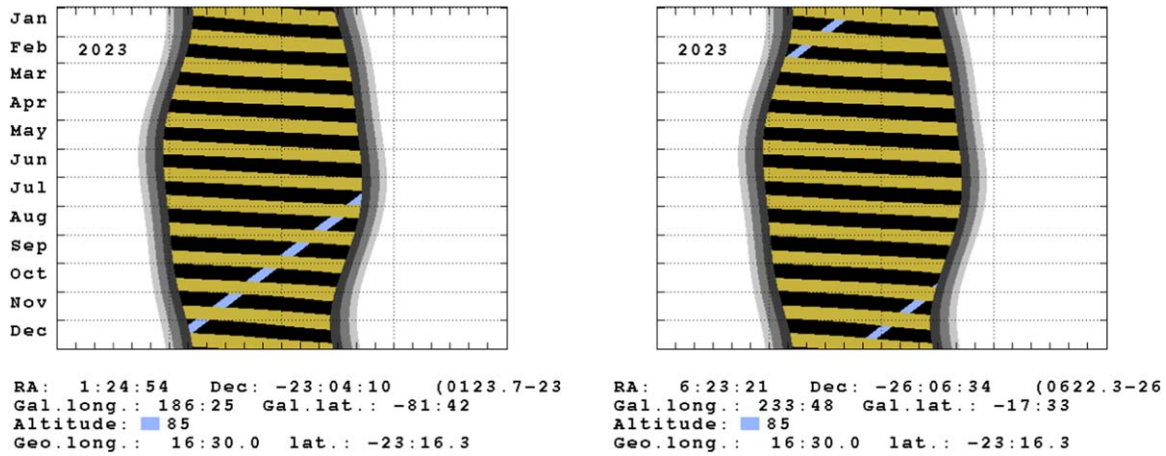

Figure 8. Four sources that could pass by the observation window of H.E.S.S near the zenith in the whole year of 2023. They are 4FGL J0123.7-2311, 4FGL J0622.3-2605. A complete display of the visibility of the source for other IACTs and EAS arrays (97 images) is available in a figure set in the online Journal. The figures are generated by H.E.S.S. online visibility tool. The Sun is up during the times indicated by the white areas. Gray levels correspond to civil, naval, and astronomical twilight, respectively. The moon is up or making twilight in the yellow areas. The blue colors indicate the times when the object is above given altitudes. R.A./decl. positions listed below the plot are for the current epoch, not J2000.

(The complete figure set (97 images) is available.)

Table 7
Observable Months of IACTs and EAS Arrays[a]

4FGL Name (1)	CTAO North (2)	CTAO South (3)	H.E.S.S. Zenith (4)	H.E.S.S. Low (5)	MAGIC Low (6)	MAGIC Medium (7)	VERITAS (8)
J0037.8+1239	7, 8, ..., 12				1, 7, ..., 12	1, 6, ..., 12	7, 8, ..., 12
J0051.2-6242							
J0110.1+6805						1, 7, ..., 12	
J0110.7-1254		7, 8, …, 12		6[h], 7, ..., 12		1, 7, ..., 12	
J0115.8+2519	1, 7, ..., 12				1, 7, ..., 12	1, 2, 6[h], ..., 12	1, 7[a], ..., 12
J0123.7-2311		7, 8, ..., 12	7, 8, ..., 12	6[h], 7, ..., 12			
J0159.5+1046	1, 8, ..., 12				1, 7, ..., 12	1, 2, ..., 12	
J0209.3-5228							
J0211.2+1051	1, 8, ..., 12				1, 8, ..., 12	1, 2, 7, ..., 12	
J0244.6-5819							
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

Notes. "h" indicates the month is only visible for the H.E.S.S. tool. "a" indicates the month is only visible for astroplan.
[a] Only 10 items are displayed. A complete listing of this table is available in the online version.

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

TeVs. For 3FHL, taking off four sources without $\log \nu_p^s$ data, the AUC: 68% and TPR: 63% are also lower than those predicted by LR, up to 25 TeV blazars are misclassified as non-TeVs. The relevant confusion matrix is shown in Figures 9 and 10. The results illustrated that our selection criterion is more effective.

So far, TeV candidate selection techniques privileged targets with high predicted TeV fluxes. In our analysis, instead, we identified targets that, according to their broadband features, have high chances of becoming promising TeV targets in specific activity states and that may have been missed so far due to their current status or to the lack of the proper instruments and observing strategies. As a data-driven algorithm, ML is inevitably sensitive to the sample composition, the ratio of the training set and test set, the value of k in cross validation, etc. Our result is to ensure its generalization based on existing data. There is a potential bias in the fact that a significant fraction of TeVCat sources are Fermi-LAT detected blazars, as Fermi-LAT catalogs are browsed to list potential candidates for IACTs, and because Fermi-LAT monitoring is used to trigger Target of Opportunity programs. Also, our observation proposal does not consider the weather or available time of detectors. A feasible observation proposal should be more reliable and agreed upon by the observatory that considers more influencing factors.

Detecting extragalactic photons in the TeV band needs to face two major limitations, which are EBL and the performance of detectors (sensitivity, effective area, FoV, etc.) The relationship between the improvement of detector sensitivity and the detection of more TeV sources is far from simple, because if it is true that an improvement in sensitivity indeed helps detect more sources but, since the attenuation rapidly increases with energy and redshift, the overall effect is rather mild (see for example Gilmore et al. 2012). The redshifts of the

Table 8
Predicting Results for 40 TeV Blazar Candidates[a]

4FGL Name	GLON	GLAT	$\log \nu_{\rm p}^{\rm s}$	z	$\log f_{1\ \rm TeV}$	Logistic				Class	Other Catalogs			Common	Detectability
						4FGL	3FHL	3HSP	2BIGB		3FHL	3HSP	2BIGB		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
J0037.8+1239	117.77	−50.08	15.05	0.09	−12.32	82.50%	54.60%			bll	Y				CN, CS, ML, MM
J0051.2-6242	302.96	−54.43	15.96	0.3	−12.91	84.10%				bll					
J0110.1+6805	124.69	5.29	14.85	0.29		81.80%				bll					
J0110.7-1254	141.53	−75.08	17	0.23		82.40%				bll					
J0115.8+2519	129.85	−37.21	15.75	0.36		97.60%				bll					
J0123.7-2311	186.41	−81.7	17.96	0.4		91.50%				bll					
J0159.5+1046	148.75	−48.66	15.8	0.2	−13.04	87.70%	53.60%			bll	Y				
J0209.3-5228	278.35	−60.77	16.14	0.16	−12.28	95.60%		67.70%	60.20%	bll		Y	Y	M	
J0211.2+1051	152.59	−47.37	14.22	0.2	−12.86	88.90%	55.50%			bll	Y				
J0244.6-5819	278.45	−53.09	17.03	0.27		88.00%	53.30%		48.50%	bll	Y		Y	F	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Notes.
[a] Only 10 sources are displayed. A complete listing of this table is available in the online version.
[b] The source can be detected in 1 TeV $\leqslant E \leqslant$ 10 TeV and 10 TeV $\leqslant E \leqslant$ 100 TeV.

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

Table 9
Redshift Uplimit of the 20 Blazars for Each IACT and EAS Array[a]

4FGL Name (1)	$z_{\rm uplimit}$ of CN (2)	$z_{\rm uplimit}$ of CS (3)	$z_{\rm uplimit}$ of HZ (4)	$z_{\rm uplimit}$ of HL (5)	$z_{\rm uplimit}$ of ML (6)	$z_{\rm uplimit}$ of MM (7)	$z_{\rm uplimit}$ of V (8)	$z_{\rm uplimit}$ of L (9)	$z_{\rm uplimit}$ of H (10)	z (11)	Detectability (12)
J0037.8+1239		0.2, 0.05, 0.01		0.02, 0.01, 0.01		0.09, 0.01, 0.01		0, 0.01, 0.01	0, 0, 0.01	0.09	CN, CS, ML, MM
J0051.2-6242		0.29, 0.07, 0.01		0.14, 0.03, 0.01	0.20, 0.02, 0.01	0.19, 0.04, 0.01	0.17, 0.01, 0.01	0.1, 0.04, 0.01	0, 0.02, 0.01	0.3	
J0110.1+6805										0.29	
J0110.7-1254										0.23	
J0115.8+2519										0.36	
J0123.7-2311										0.4	
J0159.5+1046								0, 0, 0.01		0.2	
J0209.3-5228		0.27, 0.07, 0.01		0.11, 0.02, 0.01	0.17, 0.01, 0.01	0.17, 0.02, 0.01		0.07, 0.03, 0.01	0, 0.01, 0.01	0.16	
J0211.2+1051								0, 0, 0.01		0.2	
J0244.6-5819										0.27	
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...

Note.
[a] Only 10 items are displayed. A complete listing of this table is available in the online version.

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

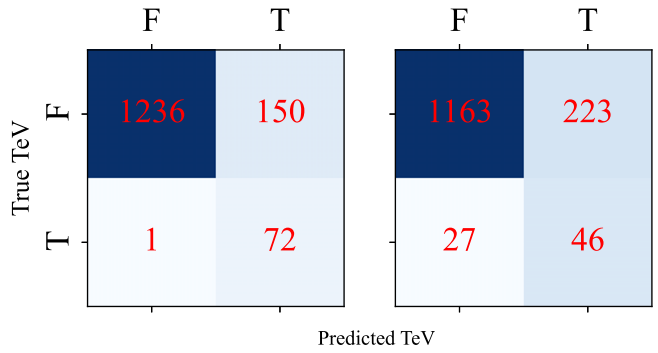

Figure 9. Confusion matrix for SML and HSPs, where the coordinate value "T" stands for "True" and "F" stands for "False." Left panel: confusion matrix for 4FGL-DR2/4LAD-DR2, with five FOFs: Γ V_F $\log f_7^{\rm ph}$ z. Right panel: confusion matrix for $\log \nu_{\rm p}^{\rm s} \geqslant 15.3$ Hz. True positives (top left), false negatives (top right), false positives (bottom left), and true negatives (bottom right) for both panels.

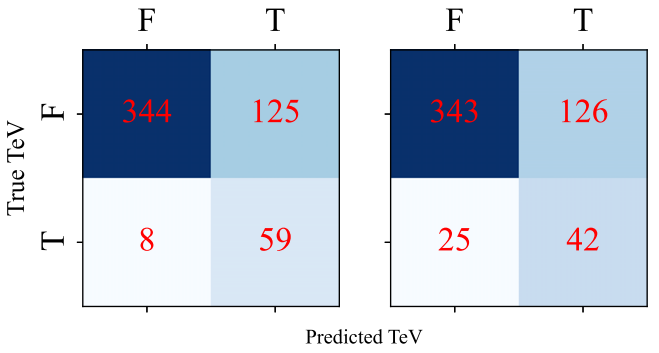

Figure 10. Confusion matrix for SML and HSPs, where the coordinate value "T" stands for "True," and "F" stands for "False." Left panel: confusion matrix for 3FHL, with two FOFs: $\log f_2^{\rm ph}$, z. Right panel: confusion matrix for $\log \nu_{\rm p}^{\rm s} \geqslant$ 15.3 Hz. True positives (top left), false negatives (top right), false positives (bottom left), and true negatives (bottom right) for both panels.

81 blazars listed in TeVCat are taken from the four catalogs and literature, noting that for 4FGL J2243.9+2021 (or RGB J2243 +203) we took the upper limit of the redshift range of $0.75 \leqslant z \leqslant 1.1$ (Sahu et al. 2019). Then we calculate the $\tau(E, z)$ fixing $E = 1$ TeV and we evaluate $\tau(E, z) \leqslant 13.1$ for all 81 blazars. The results are also listed in Table 10, where Col. (1) and Col. (2) give the TeVCat name and 4FGL-DR2 name; Col. (3) and Col. (4) are the celestial equator coordinates (in degrees); Col. (5) the redshift; Col. (6) the $\tau(E, z)$ provided by Saldana-Lopez et al. (2021) at $E = 1$ TeV.

All in all, the validation of TeV candidates is counting on IACTs and EAS arrays, while each of the IACTs and EAS arrays has advantages and disadvantages. IACTs have better energy and angular resolution. Additionally, the energy threshold of IACTs could reach the GeV band while EAS arrays are generally sensitive above ~10 TeV, but at ~30 TeV, LHAASO sensitivity is comparable to that of CTAO, and LHAASO at ~100 TeV is the most sensitive instrument (Cao et al. 2019). On the other hand, EAS array observation time is not limited to moonless nights and has a much larger FoV, and the exposure time of IACTs is typically quoted for 50 hr whereas for 1 or 5 yr of EAS arrays. Although IACTs are better at capturing transient events, larger FoVs and longer exposure time make up the gap for the EAS array. Overall, for IACTs and EAS arrays, the differences in their operation modes and performance parameters allow them to complement each other. Furthermore, The sensitivity of LHAASO and the next-generation IACT array: CTAO, has been improved by an order of magnitude compared with the current IACTs and EAS arrays. They have pushed the detection limit of the TeV band to higher than 100 TeV (The CTA Consortium 2011; Cao et al. 2019), enabling us to detect the whole energy range of the TeV sky in the future.

5.2. Conclusions

In this paper, we implement a machine-learning algorithm called *Logistic Regression* to search the TeV blazar candidates from 4FGL-DR2/4LAC-DR2, 3FHL, 3HSP, and 2BIGB. Furthermore, we filter out potential observation targets for IACTs and EAS arrays. The main conclusions are enumerated below:

Table 10
$\tau(E, z)$ at $E = 1$ TeV for the 81 TeV Blazars[a]

TeVCat Name (1)	4FGL Name (2)	R.A. (3)	Decl. (4)	z (5)	τ (6)
J0013-188	J0013.9-1854	3.47	−18.89	0.095	0.92
J0033-193	J0033.5-1921	8.4	−19.35	0.61	7.61
J0035+598	J0035.9+5950	8.82	59.79	0.467	5.7
J0112+227	J0112.1+2245	18.02	22.74	0.265	2.96
J0136+391	J0136.5+3906	24.14	39.1	0.75	9.35
J0152+017	J0152.6+0147	28.14	1.78	0.08	0.76
J0214+517	J0214.3+5145	33.57	51.75	0.049	0.45
J0218+359	J0221.1+3556	35.27	35.94	0.944	11.54
J0222+430	J0222.6+4302	35.67	43.04	0.444	5.39
J0232+202	J0232.8+2018	38.22	20.27	0.139	1.4
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...

Note.
[a] Only 10 items are displayed. A complete listing of this table is available in the online version.

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

1. For the four catalogs, *Logistic Regression* provides an empirical formula to find blazars that could be detected at TeV energies with $\text{logit}' \geqslant 0$.

   For 4FGL-DR2/4LAC-DR2:

$$\text{logit}' = 4.808\Gamma + 2.809V_F + 3.889\log f_7^{\rm ph} - 3.34z + 0.857\log \nu_{\rm p}^{\rm s} + 20.244,$$

   with AUC = 98%, and the FPR = 11%, TPR = 99%, and the $p_{\rm thre}$ = 40%. 40 out of 150 non-TeV blazars have logistic ⩾ 80% and are thus expected to be high-confidence candidates.

   For 3FHL:

$$\text{logit}' = 0.116\log f_2^{\rm ph} - 0.628z + 1.028,$$

   with AUC = 88%, FPR = 27%, TPR = 88%, and the $p_{\rm thre}$ = 52%. 24 out of 126 TeV candidates are common with the high-confidence ones of 4FGL-DR2/4LAC-DR2.

   For 3HSP:

$$\text{logit}' = 0.306\text{FOM} - 3.861z - 0.173,$$

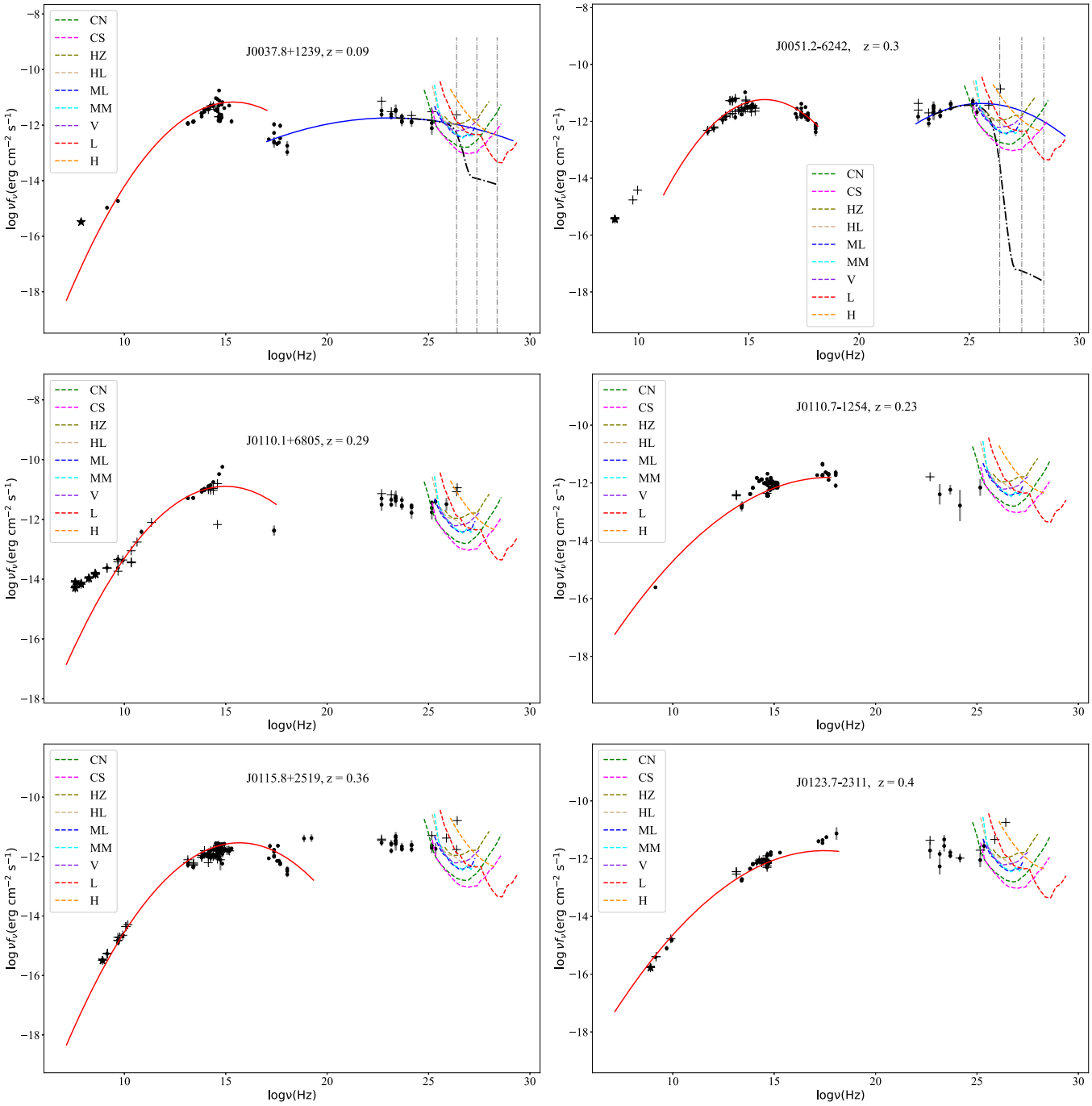

Figure 11. SEDs for 40 TeV candidates. Only six items are displayed. A complete listing of the SEDs (40 images) is available as a figure set in the online Journal. Gray error bars stand for data points. Note that the TeV band photons are not certificated by IACTs and EAS arrays. Red and solid blue curves stand for optimal-fitting synchrotron bump and IC bump, while black dashed–dotted curves are EBL-absorbed IC bumps. Three vertical dashed–dotted curves are $\log \nu$ (Hz) corresponding to 1, 10, and 100 TeV. Green, magenta, olive, tan, blue, cyan, blue-violet, red, and dark-orange dotted curves represent the sensitivity of CTAO north (north site, $0^\circ \leqslant Z \leqslant 20^\circ$), CTAO south (south site, $0^\circ \leqslant Z \leqslant 20^\circ$), H.E.S.S zenith ($0^\circ \leqslant Z \leqslant 5^\circ$), H.E.S.S low ($12^\circ \leqslant Z \leqslant 22^\circ$), MAGIC low ($0^\circ \leqslant Z \leqslant 30^\circ$), MAGIC medium ($30^\circ \leqslant Z \leqslant 45^\circ$), VERITAS ($0^\circ \leqslant Z \leqslant 20^\circ$), LHAASO, and HAWC. For each source redshifts (z) are also shown. Additionally, the pentagon stands for data whose flux error is bigger than flux upper limits, and the cross is data in the low radio energy range ($\log \nu(\mathrm{Hz}) < 9$).

(The complete figure set of 40 images is available.)

with AUC = 98%, FPR = 2%, TPR = 91%, and $p_{\mathrm{thre}} = 61\%$. 11 out of 40 TeV candidates are common with the high-confidence ones of 4FGL-DR2/4LAC-DR2.

For 2BIGB,

$$\mathrm{logit}' = -0.016E_{\mathrm{P}} + 0.248\mathrm{FOM} - 4.395z - 0.122,$$

with AUC = 97%, FPR = 8%, TPR = 96%, and $p_{\mathrm{thre}} = 48\%$. 14 out of 83 TeV candidates are common with the high-confidence ones of 4FGL-DR2/4LAC-DR2.

2. For the 40 high-confidence candidates, we independently fit the two bumps of the SEDs (see Figure 11) and infer the visibility in the sky. For EAS arrays and IACTs in

2023, seven candidates are potential targets: six CTAO targets, two H.E.S.S targets, five MAGIC targets, one VERITAS, one LHAASO target, and zero HAWC targets, respectively.

3. We get two common sources with Costamante & Ghisellini (2002), nine with Massaro et al. (2013), four with Foffano et al. (2019), and none with Chiaro et al. (2019), respectively.

Acknowledgments

Thanks are given to the excellent reviewer for the constructive comments and suggestions! The work is partially supported by the National Natural Science Foundation of China (NSFC U1831119, NSFC U2031201, NSFC 11733001, NSFC 12203034), Guangdong Major Project of Basic and Applied Basic Research (2019B030302001), Shanghai Science and Technology Fund (22YF1431500), Scientific and Technological Cooperation Projects (2020–2023) between the People's Republic of China and the Republic of Bulgaria, Astrophysics Key Subjects of Guangdong Province and Guangzhou City, the Large High Altitude Air Shower Observatory collaboration, the Cherenkov Telescope Array, the High Altitude Water Cherenkov Observatory, the Major Atmospheric Gamma Imaging Cherenkov Telescopes, the Very Energetic Radiation Imaging Telescope Array System, and the UK Swift Science Data Centre at the University of Leicester.

ORCID iDs

J. T. Zhu https://orcid.org/0000-0002-8206-5080
J. H. Fan https://orcid.org/0000-0002-5929-0968
D. Bastieri https://orcid.org/0000-0002-6954-8862
G. G. Wang https://orcid.org/0000-0002-8032-4640

References

Abdo, A. A., Ackermann, M., Agudo, I., et al. 2010, ApJ, 716, 30
Abdollahi, S., Acero, F., Ackermann, M., et al. 2020, ApJS, 247, 33
Abeysekara, A. U., Albert, A., Alfaro, R., et al. 2017a, ApJ, 843, 39
Abeysekara, A. U., Albert, A., Alfaro, R., et al. 2017b, ApJ, 843, 40
Abramowski, A., Acero, F., Aharonian, F., et al. 2013, PhRvD, 88, 102003
Abramowski, A., Aharonian, F., Ait Benkhali, F., et al. 2015, ApJ, 802, 65
Acciari, V. A., Aliu, E., Arlen, T., et al. 2011, ApJ, 738, 169
Acciari, V. A., Aliu, E., Beilicke, M., et al. 2008, ApJL, 684, L73
Aharonian, F., Akhperjanian, A., Beilicke, M., et al. 2003, A&A, 406, L9
Aharonian, F., Akhperjanian, A. G., Anton, G., et al. 2009, ApJL, 696, L150
Aharonian, F., Akhperjanian, A. G., Aye, K. M., et al. 2004, APh, 22, 109
Aharonian, F., Akhperjanian, A. G., Barres de Almeida, U., et al. 2007a, A&A, 473, L25
Aharonian, F., Akhperjanian, A. G., Barres de Almeida, U., et al. 2007b, A&A, 475, L9
Aharonian, F. A., Akhperjanian, A. G., Barrio, J. A., et al. 2001, A&A, 366, 62
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
Aleksić, J., Ansoldi, S., Antonelli, L. A., et al. 2016, APh, 72, 76
Aliu, E., Archer, A., Aune, T., et al. 2015, ApJ, 799, 7
Aliu, E., Aune, T., Beilicke, M., et al. 2011, ApJ, 742, 127
Amenomori, M., Ayabe, S., Cui, S. W., et al. 2003, ApJ, 598, 242
Archambault, S., Arlen, T., Aune, T., et al. 2013, ApJ, 776, 69
Archambault, S., Aune, T., Behera, B., et al. 2014, ApJL, 785, L16
Arlen, T., Aune, T., Beilicke, M., et al. 2013, ApJ, 762, 92
Arsioli, B., Chang, Y.-L., & Musiimenta, B. 2020, MNRAS, 493, 2438
Ball, N. M., & Brunner, R. J. 2010, IJMPD, 19, 1049
Ballet, J., Burnett, T. H., Digel, S. W., & Lott, B. 2020, arXiv:2005.11208
Baron, D. 2019, arXiv:1904.07248
Bartoli, B., Bernardini, P., Bi, X. J., et al. 2011, ApJ, 734, 110
Bartoli, B., Bernardini, P., Bi, X. J., et al. 2012, ApJ, 758, 2
Biteau, J., & Williams, D. A. 2015, ApJ, 812, 60
Błażejowski, M., Sikora, M., Moderski, R., & Madejski, G. M. 2000, ApJ, 545, 107
Bloom, S. D., & Marscher, A. P. 1996, ApJ, 461, 657
Buitinck, L., Louppe, G., Blondel, M., et al. 2013, arXiv:1309.0238
Cao, Z., Chen, M.-J., Chen, S.-Z., & Hu, H.-B. 2019, ChA&A, 43, 457
Chandra, P., Rannot, R. C., Yadav, K. K., et al. 2012, JPhG, 39, 045201
Chandra, P., Yadav, K. K., Rannot, R. C., et al. 2010, JPhG, 37, 125201
Chang, Y. L., Arsioli, B., Giommi, P., & Padovani, P. 2017, A&A, 598, A17
Chang, Y.-L., Arsioli, B., Giommi, P., Padovani, P., & Brandt, C. H. 2019, A&A, 632, A77
Chiaro, G., Meyer, M., Di Mauro, M., et al. 2019, ApJ, 887, 104
Costamante, L., Bonnoli, G., Tavecchio, F., et al. 2018, MNRAS, 477, 4257
Costamante, L., & Ghisellini, G. 2002, A&A, 384, 56
Costamante, L., Ghisellini, G., Giommi, P., et al. 2001, A&A, 371, 512
Daniel, M. K., Badran, H. M., Bond, I. H., et al. 2005, ApJ, 621, 181
Dermer, C. D., & Schlickeiser, R. 1993, ApJ, 416, 458
Dimitrakoudis, S., Mastichiadis, A., Protheroe, R. J., & Reimer, A. 2012, A&A, 546, A120
Domínguez, A., Wojtak, R., Finke, J., et al. 2019, ApJ, 885, 137
Fan, J.-H. 2002, PASJ, 54, L55
Fan, J.-H., Bastieri, D., Yang, J.-H., et al. 2014, RAA, 14, 1135
Fan, J. H., Kurtanidze, O., Liu, Y., et al. 2017, ApJ, 837, 45
Fan, J. H., Kurtanidze, S. O., Liu, Y., et al. 2021, ApJS, 253, 10
Fan, J. H., Yang, J. H., Liu, Y., et al. 2016, ApJS, 226, 20
Fawcett, T. 2006, PaReL, 27, 861
Fermi-LAT Collaboration, Abdollahi, S., Ackermann, M., et al. 2018, Sci, 362, 1031
Foffano, L., Prandini, E., Franceschini, A., & Paiano, S. 2019, MNRAS, 486, 1741
Ghisellini, G., Maraschi, L., & Treves, A. 1985, A&A, 146, 204
Gilmore, R. C., Somerville, R. S., Primack, J. R., & Domínguez, A. 2012, MNRAS, 422, 3189
Giommi, P., Capalbi, M., Fiocchi, M., et al. 2002, in Blazar Astrophysics with BeppoSAX and Other Observatories, ed. P. Giommi, E. Massaro, & G. Palumbo, 63
Godambe, S. V., Rannot, R. C., Chandra, P., et al. 2008, JPhG, 35, 065202
Gupta, A. C., Agarwal, A., Bhagwan, J., et al. 2016, MNRAS, 458, 1127
Gupta, A. C., Kalita, N., Gaur, H., & Duorah, K. 2016, MNRAS, 462, 1508
Hearst, M. A. 1998, IISys, 13, 18
HESS Collaboration, Abramowski, A., Acero, F., et al. 2010, A&A, 520, A83
HESS Collaboration, Abramowski, A., & Acero, F. 2013, MNRAS, 434, 1889
Kam Ho, T. 1995, in Proc. 3rd Int. Conf. Document Analysis and Recognition, Vol. 1 (Piscataway, NJ: IEEE), 278
Kang, S.-J., Li, E., Ou, W., et al. 2019, ApJ, 887, 134
Kapanadze, B., Dorner, D., Vercellone, S., et al. 2018, MNRAS, 473, 2542
Kneiske, T. M., Bretz, T., Mannheim, K., & Hartmann, D. H. 2004, A&A, 413, 807
Korsós, M. B., Erdélyi, R., Liu, J., & Morgan, H. 2021, FrASS, 7, 113
Lin, C., & Fan, J.-H. 2016, RAA, 16, 103
Lott, B., Gasparrini, D., & Ciprini, S. 2020, arXiv:2010.08406
MAGIC Collaboration 2000, in AIP Conf. Proc. 515, GeV-TeV Gamma Ray Astrophysics Workshop: Towards a Major Atmospheric, ed. B. L. Dingus, M. H. Salamon, & D. B. Kieda (Melville, NY: AIP), 510
Majumder, A., Mitra, K., Chatterjee, R., et al. 2019, MNRAS, 490, 124
Maraschi, L., Ghisellini, G., & Celotti, A. 1992, ApJL, 397, L5
Massaro, F., Paggi, A., Errando, M., et al. 2013, ApJS, 207, 16
McCulloch, W. S., & Pitts, W. 1943, The Bulletin of Mathematical Biophysics, 5, 115
Nievas Rosillo, M., Domínguez, A., Chiaro, G., et al. 2022, MNRAS, 512, 137
Osorio, M., Sacahui, R., González, M. M., Fraija, N., & García-González, J. A. 2019, ICRC (Madison, WI), 36, 686
Padovani, P., & Giommi, P. 1995, ApJ, 444, 567
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, JMLR, 12, 2825
Piner, B. G., Pant, N., & Edwards, P. G. 2010, ApJ, 723, 1150
Primack, J. R., Bullock, J. S., Somerville, R. S., & MacMinn, D. 1999, APh, 11, 93
Raschka, S. 2015, Python Machine Learning (Birmingham: Packt Publishing) https://www.packtpub.com/product/python-machine-learning/9781783555130
Sahu, S., López Fortín, C. E., & Nagataki, S. 2019, ApJL, 884, L17
Saldana-Lopez, A., Domínguez, A., Pérez-González, P. G., et al. 2021, MNRAS, 507, 5144
Sambruna, R. M., Aharonian, F. A., Krawczynski, H., et al. 2000, ApJ, 538, 127
Schroedter, M., Badran, H. M., Buckley, J. H., et al. 2005, ApJ, 634, 947
Sharma, M., Nayak, J., Koul, M. K., et al. 2015, NIMPA, 770, 42

Sikora, M., Begelman, M. C., & Rees, M. J. 1994, ApJ, 421, 153
Singh, K. K., Bhatt, H., Bhattacharyya, S., et al. 2019, AdSpR, 63, 766
Tavecchio, F. 2014, MNRAS, 438, 3255
Taylor, C. 2019, Applications Of Dynamic Programming To Agricultural Decision Problems (Boca Raton, FL: CRC Press)
The CTA Consortium, Actis, M., Agnetta, G., et al. 2011, ExA, 32, 193
Urry, C. M., & Padovani, P. 1995, PASP, 107, 803
van Eldik, C., Holler, M., Berge, D., et al. 2015, ICRC (The Hague), 34, 847
Varma, S., & Simon, R. 2006, BMC Bioinform., 7, 91
VERITAS Collaboration, Badran, H. M., Blaylock, G., et al. 2005, in AIP Conf. Proc. 745, High Energy Gamma-Ray Astronomy, ed. F. A. Aharonian, H. J. Völk, & D. Horns (Melville, NY: AIP), 633
Villata, M., Raiteri, C. M., Balonek, T. J., et al. 2006, A&A, 453, 817
Wakely, S. P., & Horan, D. 2008, ICRC (Mérida), 3, 1341
Wills, B. J., Wills, D., Breger, M., Antonucci, R. R. J., & Barvainis, R. 1992, ApJ, 398, 454
Xiao, H., Fan, J., Yang, J., et al. 2019, SCPMA, 62, 129811
Xiao, H., Zhu, J., Fu, L., Zhang, S., & Fan, J. 2022, PASJ, 74, 239
Xiao, H. B., Fan, J. H., Rando, R., Zhu, J. T., & Hu, L. J. 2020, AN, 341, 462
Xiao, H. B., Pei, Z. Y., Xie, H. J., et al. 2015, Ap&SS, 359, 39
Yang, J. H., Fan, J. H., Liu, Y., et al. 2022, ApJS, 262, 18
Yang, W.-X., Xiao, H.-B., Wang, H.-G., et al. 2022, RAA, 22, 085002
Youden, W. J. 1950, Cancer, 3, 32