



# Identification of 4FGL Uncertain Sources at Higher Resolutions with Inverse Discrete Wavelet Transform

Haitao Cao<sup>1</sup> , Hubing Xiao<sup>2</sup> , Zhijian Luo<sup>2</sup>, Xiangtao Zeng<sup>3,4,5</sup>, and Junhui Fan<sup>3,4,5</sup>

<sup>1</sup> School of Information Engineering, Guangzhou Panyu Polytechnic, Guangzhou, 511483, People's Republic of China

<sup>2</sup> Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai, 200234, People's Republic of China; [hubing.xiao@shnu.edu.cn](mailto:hubing.xiao@shnu.edu.cn)

<sup>3</sup> Center for Astrophysics, Guangzhou University, Guangzhou 510006, People's Republic of China

<sup>4</sup> Astronomy Science and Technology Research Laboratory of Department of Education of Guangdong Province, Guangzhou 510006, People's Republic of China

<sup>5</sup> Key Laboratory for Astronomical Observation and Technology of Guangzhou, Guangzhou 510006, People's Republic of China

Received 2023 July 31; revised 2023 November 1; accepted 2023 November 17; published 2024 January 17

## Abstract

In the forthcoming era of big astronomical data, it is a burden to find target sources from ground-based and space-based telescopes. Although machine-learning methods have been extensively utilized to address this issue, the incorporation of in-depth data analysis can significantly enhance the efficiency of identifying target sources when dealing with massive volumes of astronomical data. In this work, we focused on the task of finding active galactic nucleus (AGN) candidates and identifying BL Lacertae objects (BL Lac) or flat spectrum radio quasar (FSRQ) candidates from the 4FGL\_DR3 uncertain sources. We studied the correlations among the attributes of the 4FGL\_DR3 catalog and proposed a novel method, named fractal dimension–inverse discrete wavelet transform (FDIDWT), to transform the original data. The transformed data set is characterized as low-dimensional and feature-highlighted, with the estimation of correlation features by fractal dimension theory and the multi-resolution analysis by inverse discrete wavelet transform (IDWT). Combining the FDIDWT method with an improved lightweight MatchboxConv1D model, we accomplished two missions: (1) to distinguish the AGNs from others (non-AGNs) in the 4FGL\_DR3 uncertain sources with an accuracy of  $96.65\% \pm 1.32\%$ , namely Mission A; and (2) to classify blazar candidates of uncertain type into BL Lacs or FSRQs with an accuracy of  $92.03\% \pm 2.2\%$ , namely Mission B. There are 1354 AGN candidates in Mission A, and 482 BL Lacs candidates and 128 FSRQ candidates were found in Mission B. The results show a high consistency of greater than 98% with the results in previous works. In addition, our method has the advantage of finding less variable and relatively faint sources than ordinary methods.

*Unified Astronomy Thesaurus concepts:* [Blazars \(164\)](#); [Classification \(1907\)](#); [Active galactic nuclei \(16\)](#); [Wavelet analysis \(1918\)](#)

*Supporting material:* machine-readable table

## 1. Introduction

Active galactic nucleus (AGN) has been a hot topic for 60 yr in astronomy since its discovery in 1963 (Schmidt 1963). It is believed that AGNs centered a supermassive black hole, which is surrounded by an accretion disk, and the whole system provides energy for AGN radiation (Lynden-Bell 1969; Blandford & Znajek 1977; Blandford & Payne 1982). The emissions from AGNs are observed to span over the entire electromagnetic spectrum, and these emissions are found to be strong and variable. Based on the ratio of radio emission strength to optical emission strength, AGNs are divided into radio-loud and radio-quiet AGNs (Strittmatter et al. 1980; Kellermann et al. 1989). This method has recently been modified by a double-criterion method (Xiao et al. 2022a). Blazars, a subclass of radio-loud AGNs with jets pointing toward the observer, show multi-band high and fast variability, high and variable polarization, strong and variable  $\gamma$ -ray emissions, and apparent superluminal motion (Wills et al. 1992; Urry & Padovani 1995; Fan 2002; Villata et al. 2006; Fan et al. 2014, 2021; Gupta et al. 2016; Xiao et al. 2019, 2020a, 2022b; Abdollahi et al. 2020). Blazars consist

of BL Lacertae objects (BL Lacs) and flat spectrum radio quasars (FSRQs), the former shows strong emission lines (rest-frame equivalent width,  $EW > 5 \text{ \AA}$ ) and the latter demonstrates no or weak emission features ( $EW < 5 \text{ \AA}$ ) (Urry & Padovani 1995; Scarpa & Falomo 1997).

Until the launch of the Large Area Telescope on board the Fermi Gamma-Ray Space Observatory (Fermi-LAT) in 2008, the study of blazars has been severely limited due to the small sample size. The Large Area Telescope has an unprecedented performance, a better energy resolution, angular resolution, and a wider effective area in both low-energy and high-energy bands<sup>6</sup> than its predecessor EGRET (Thompson et al. 1993), at energy ranges from 20 MeV to 300 GeV. Fermi-LAT collaboration has released 5 main  $\gamma$ -ray sources catalogs, namely 0FGL, 1FGL, 2FGL, 3FGL, and 4FGL (Abdo et al. 2009, 2010; Nolan et al. 2012; Acero et al. 2015; Abdollahi et al. 2020). However, there are still significant Fermi sources that are unrelated to any known class, e.g., 1010 unassociated sources in 3FGL and 2291 uncertain sources (2157 unassociated sources + 134 unknown sources) in the latest 4FGL\_DR3, 573 blazar candidates of uncertain types (BCUs) in the 3FGL, and 1493 BCUs in the latest 4FGL\_DR3.



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.

<sup>6</sup> [https://www.slac.stanford.edu/exp/glast/groups/canda/lat\\_Performance.htm](https://www.slac.stanford.edu/exp/glast/groups/canda/lat_Performance.htm)

It is time consuming to verify these uncertain sources individually through optical observation, thus more efficient methods must be explored. Many machine-learning (ML)-based algorithms have been employed for this issue. For instance, Saz Parkinson et al. (2016) applied two different ML methods—Random Forest (RF) and Logistic Regression (LR)—to identify 1008 unassociated sources in 3FGL. Among them, 334 sources were predicted as pulsars (PSRs), and 559 sources were predicted as AGNs. Chiaro et al. (2016) utilized blazar flaring patterns (B-FlaP) as an identification approach on BCUs. Since variability is one of the characterizing properties of blazars (Paggi et al. 2011), the light curve of blazars was used by an artificial neural network (ANN) to identify 573 BCUs in 3FGL. These BCUs were associated with 342 BL Lacs and 154 FSRQs, and 77 sources remained uncertain. Xiao et al. (2020b) carried out an ensemble ML method, picked out 748 AGNs candidates from 1010 3FGL unassociated sources, and identified 573 BCUs to be 326 BL Lac candidates and 247 FSRQ candidates. Moreover, Kang et al. (2019) studied the classification of 1312 BCUs from 4FGL\_DR1 via three supervised ML methods and obtained 724 BL Lac and 332 FSRQ candidates.

The task of Fermi source classification can be seen as feature extraction with ML methods due to their ability to learn patterns from data and provide valuable insights, decisions, and predictions (Jordan & Mitchell 2015; Zhou et al. 2017). However, traditional ML methods are limited when dealing with the increasing volume of big astronomical data with the successful launch of more telescopes and detectors. In recent years, the popularity of graphics processing units had led to the research of deep learning (DL) to learn features from massive-scale data using deep neural networks (DNNs). This has become a major research focus in ML (Yu & Deng 2010; Liu et al. 2017). DNNs have proven to be successful in various real-world applications (Chen et al. 2018; Alemany et al. 2019; Jifara et al. 2019; Lam et al. 2019; Zyner et al. 2019). Furthermore, it has been shown that more complex problems require deeper networks (Bengio et al. 2009; He et al. 2016), which has led to the development of sophisticated networks such as VGG (Simonyan & Zisserman 2014), ResNet (He et al. 2016), and ChatGPT (Brown et al. 2020).

However, rather than solely focusing on the deep structure of networks that helps to learn intrinsic features, attribute analysis should also be emphasized. We propose that there exists a correlation feature among the attributes of raw data, and it will further improve the learning performance. In addition, numerous studies have shown that real-world data often contain highly redundant and unimportant attributes (Bakshi & Stephanopoulos 1993; Bengio et al. 2009; Glorot et al. 2011). This redundancy can lead to sparsity in high-dimensional space, where most samples in the data set are far away from each other. In classification tasks, this sparsity can result in less reliable predictions than in low dimensions because predictions are based on larger extrapolations (Géron 2017). Therefore, we believe that attribute analysis presents an opportunity for better results through additional correlation features and dimension reduction.

In this paper, we focus on two missions for 4FGL\_DR3, i.e., classifying 2291 uncertain sources into AGN or non-AGN and associating 1493 BCUs to BL Lac or FSRQ, which are named Mission A and Mission B, respectively. First, we study some popular attribute analysis methods and briefly examine the

attributes of 4FGL\_DR3 sources. We then find the core attributes based on the fractal dimension (FD) theory and step into the research of multi-attribute analysis from the perspective of the whole data set. Based on the results, we propose a novel method, called fractal dimension–inverse discrete wavelet transform (FDIDWT), which is a combination of FD and inverse discrete wavelet transform (IDWT), to extract correlation features at a higher resolution. With FDIDWT, the original data set is transformed into a low-dimensional and feature-highlighted set, which benefits the subsequent learning process. Finally, we combine the FDIDWT method with a lightweight convolutional neural network (CNN) model as our contribution to accomplish the classification missions.

This paper is organized as follows. Section 2 describes the data sets in two missions and presents some commonly used attribute analysis methods. Based on that, Section 3 interprets our proposed method in detail. The experiments and results are reported in Section 4. Further discussions and conclusions are presented in Sections 5 and 6, respectively.

## 2. Data Sets and Attribute Analysis

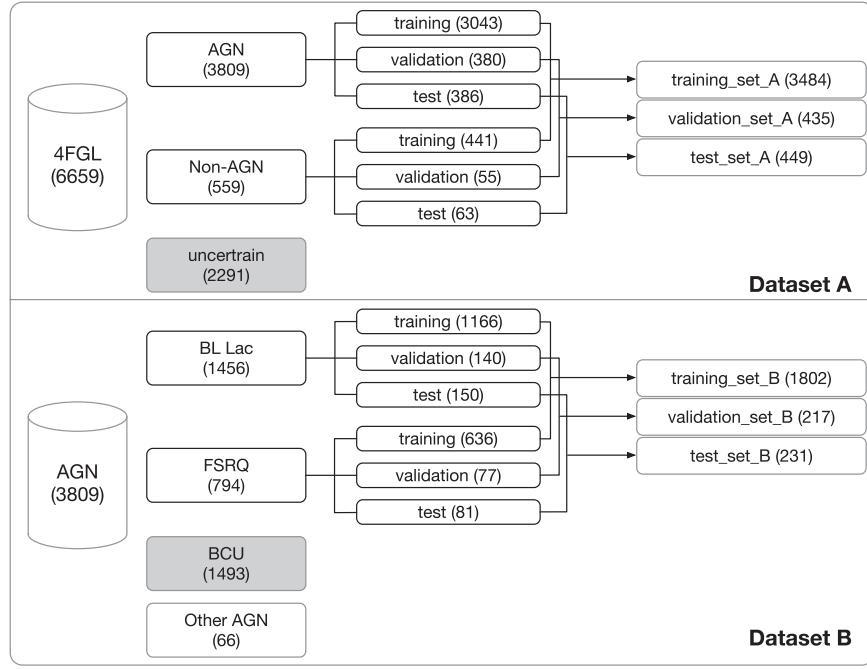
### 2.1. Samples of 4FGL\_DR3

The Fermi-LAT collaboration has recently released the incremental version of the 12 yr Fermi-LAT Gamma-ray Source Catalog (4FGL\_DR3; Abdollahi et al. 2022), which contains 6659 sources, among which 3809 sources are AGNs, 559 sources are associated with non-AGNs (including pulsars, high-mass binaries, and supernova remnants, etc.), and 2291 uncertain sources (134 sources are associated with counterparts of unknown nature and 2157 unassociated sources). Within the AGNs, there are 3743 sources confirmed with blazars, among which 1456 blazars are associated with BL Lacs, 794 blazars are associated with FSRQs, and the other 1493 Blazars are BCUs that have not been tagged as BL Lacs or FSRQ.

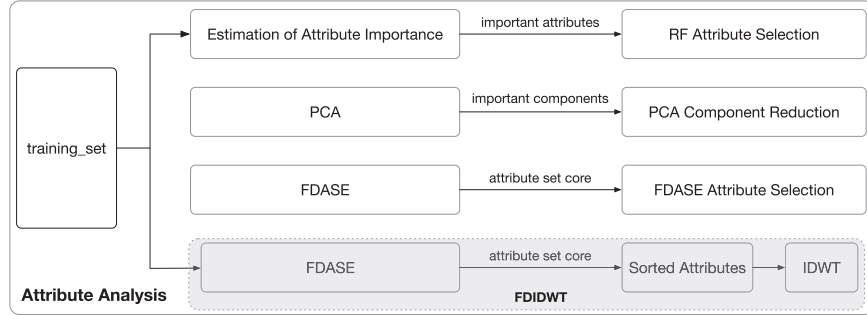
To accomplish Mission A and B, we need to select features that can distinguish one from another. Variability is one of the widely known characteristics of AGNs with respect to others that are detected by Fermi-LAT (Abdollahi et al. 2020, 2022). Therefore, those features that contain variability information (Flux1000, Flux\_Band, Variability\_Index, and Frac\_Variability) should be included for accomplishing Mission A. While FSRQs and BL Lacs demonstrate significant  $\gamma$ -ray spectra, the GeV  $\gamma$ -ray regime situates at the different regimes of the higher hump of blazar spectral energy distribution (Fan et al. 2016; Yang et al. 2022, 2023). Those features that contain spectral information (Flux\_Band, Pivot\_Energy, PL\_Index) should be included for Mission B.

In this case, we compile the data of 13 attributes from 4FGL\_DR3, as listed in Table 1.

For Mission A and Mission B, we randomly split the data into three subsets with a ratio of 8:1:1, as shown in Figure 1. The training set is used to fit the ML methods or DNN models, and the validation set aids in fine-tuning hyper-parameters for better performance. Finally, the generalization capability of the model is independently evaluated on the test set. Moreover, to explore the robustness of the model, this split policy was repeated 10 times and in the end 10 data sets were prepared for each mission. In this section, we analyze the attributes of the training set with the methods shown in Figure 2.



**Figure 1.** Data Set A (upper panel) and Data Set B (lower panel) for Mission A and Mission B, respectively.



**Figure 2.** The commonly used three attribute analysis methods and our proposed method (gray color).

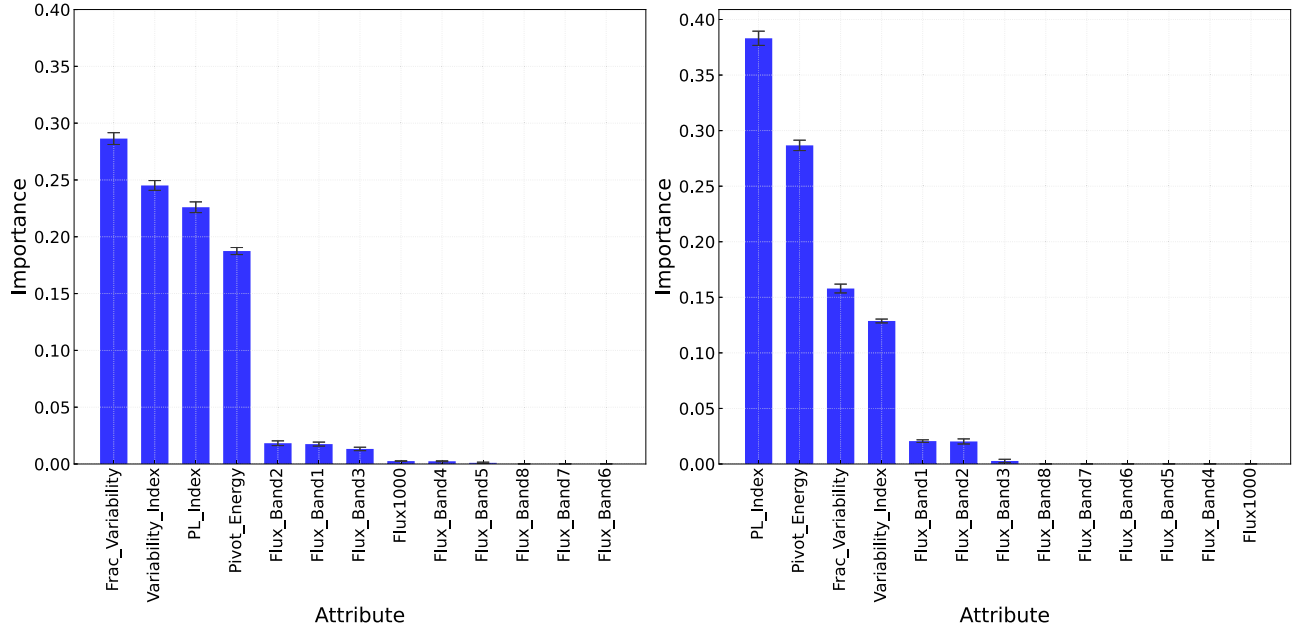
**Table 1**  
The 13 Selected Attributes from 4FGL\_DR3

Index	Attribute	Description
$a_1$	Pivot_Energy	Energy at which error on differential flux is minimal
$a_2$	Flux1000	Integral photon flux from 1 to 100 GeV
$a_3$	PL_Index	Best fit power-law index
$a_4$	Variability_Index	Sum of 2A-log (Likelihood) difference between the flux fitted in each time interval and the average flux over the full catalog interval
$a_5$	Frac_Variability	Fractional variability computed from the fluxes in each year
$a_6$	Flux_Band1	Integral photon flux in the spectral band 0.05–0.1 GeV
$a_7$	Flux_Band2	Integral photon flux in the spectral band 0.1–0.3 GeV
$a_8$	Flux_Band3	Integral photon flux in the spectral band 0.3–1 GeV
$a_9$	Flux_Band4	Integral photon flux in the spectral band 1–3 GeV
$a_{10}$	Flux_Band5	Integral photon flux in the spectral band 3–10 GeV
$a_{11}$	Flux_Band6	Integral photon flux in the spectral band 10–30 GeV
$a_{12}$	Flux_Band7	Integral photon flux in the spectral band 30–100 GeV
$a_{13}$	Flux_Band8	Integral photon flux in the spectral band 100–1000 GeV

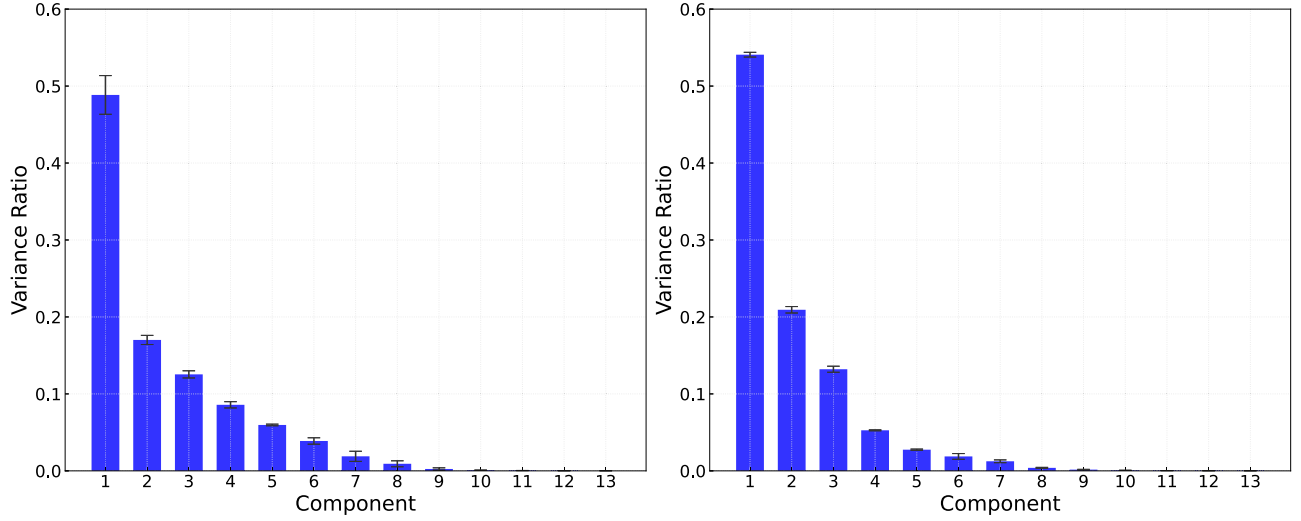
## 2.2. Attribute Importance

As mentioned earlier, it is found that real-world data often contain highly redundant and unimportant attributes. The decision tree (DT) is a popular technique to estimate the

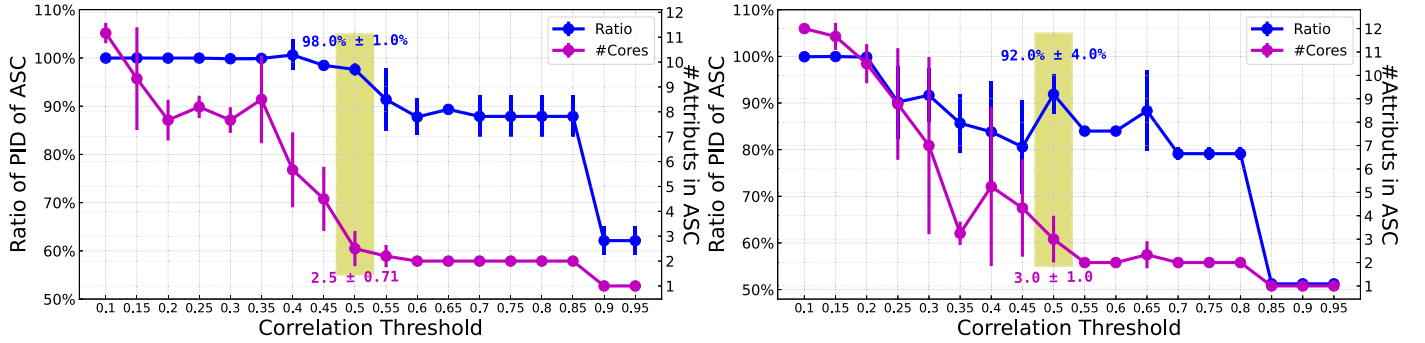
importance of attributes. The attributes that appear in the tree are considered important, with the frequency of their appearance being their attribute importance. An attribute that appears less frequently is assumed to be less important. RF is



**Figure 3.** The average attribute importance of the training sets in Data Set A (left) and Data Set B (right).



**Figure 4.** The average variance ratio of the training sets in Data Set A (left-hand panel) and Data Set B (right-hand panel).

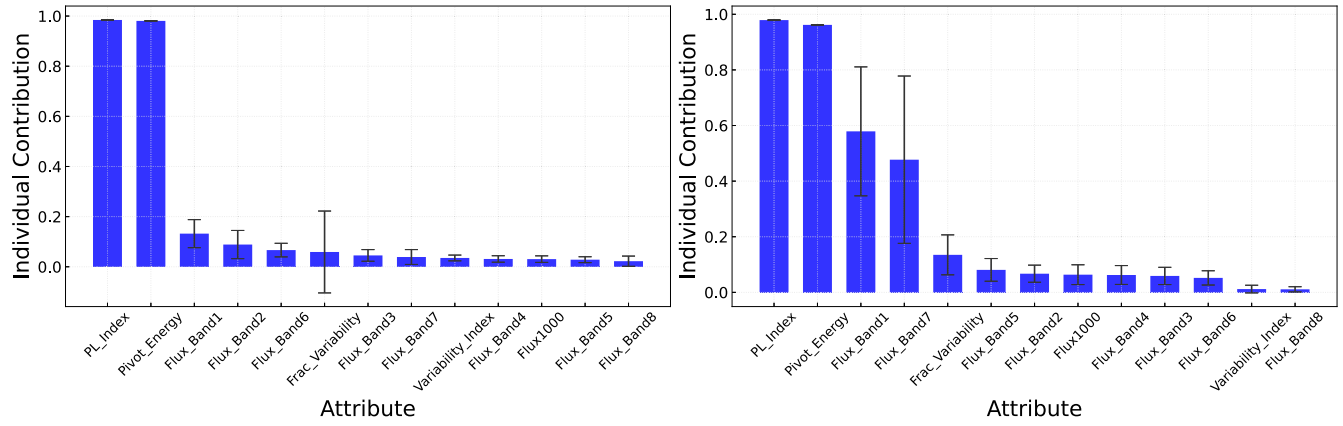


**Figure 5.** The average ratio of the PID of ASC to ID (magenta color) and the average number of ASC (blue color) with different correlation threshold  $\xi$  for the training sets of Data Set A (left-hand panel) and Data Set B (right-hand panel).

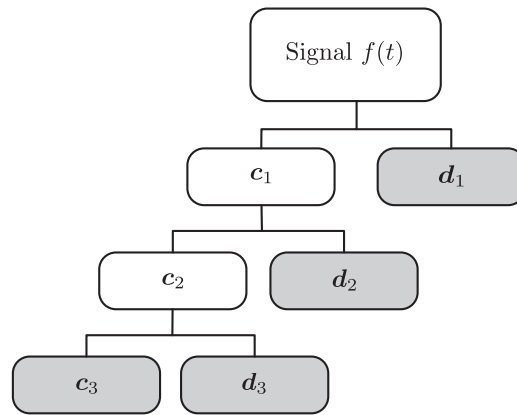
composed of multiple DTs and reduces the bias in estimating attribute importance (Breiman 2001). It is always the practice to remove the unimportant attributes and perform the so-called attribute selection (AS) for dimension reduction.

We employ 50,000 DTs to build the RF and take the entropy as the criterion. Other hyper-parameters align to the default values of scikit-learn (Pedregosa et al. 2011). The results of attribute importance on the training sets are averaged over 10

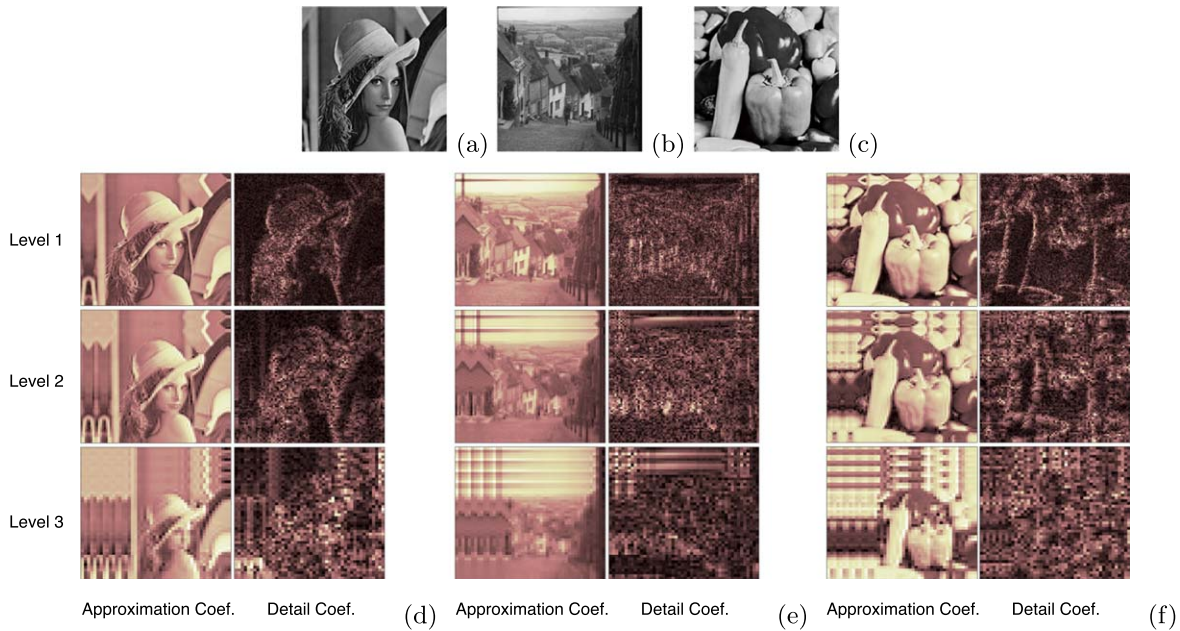




**Figure 6.** The IC for each attribute when  $\xi = 0.5$  for the training sets of Data Set A (left-hand panel) and Data Set B (right-hand panel).



**Figure 7.** Wavelet coefficients for a three-level pyramid transform.



**Figure 8.** Representations of images of Lena, Goldhill, and Peppers in (a)–(c) spatial domain, and (d)–(f) wavelet domain, respectively. For display performance, the images of coefficients have been mapped into pink color. The detail coefficients at a particular level drawn in the figures are the sum of the corresponding horizontal, vertical, and diagonal detail coefficients.

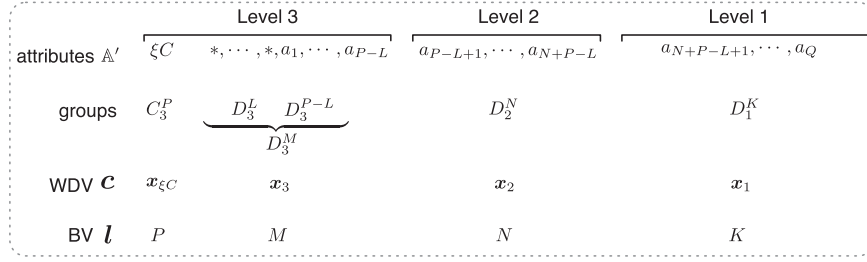


Figure 9. The groups divided from sorted attributes, taking decomposition level  $J = 3$  as an example.

data sets and are shown in Figure 3. Interestingly, from the results, we find that the four most important attributes are Pivot\_Energy, PL\_Index, Variability\_Index, and Frac\_Variability in both Data Set A and B.

### 2.3. Principal Component Analysis

Another method to study attributes is to project the data from the attribute space to a new space. In this new space, the direction with maximum variance is considered to contribute most to the data, while the directions with small variances can be removed without sacrificing crucial information. This method is called principal component analysis (PCA) (Jolliffe & Cadima 2016). For each split data set, we normalize the samples and perform PCA on training sets. The variance ratio is averaged and depicted in Figure 4. We find that most of the information of the training data is concentrated in the first three components (variance ratio larger than 0.1) in both Data Set A and B.

### 2.4. Attribute Significance Estimator based on the Fractal Dimension (FDASE)

Even though the RF model aggregates the attribute importance values from all the DTs (see Section 2.2), it may result in redundancy because it does not consider the correlations among important attributes. To address this issue, the FD theory is applied to estimate the potential contribution of each attribute to the data set and measure the correlations among attributes (Belussi & Faloutsos 1995, 1998). This method is based on the observation that independent attributes contribute more to the data set, while correlated attributes contribute less.

Mathematically, for a data set  $\mathbf{A}$  with  $E$  attributes:  $\mathbf{A} = \{a_1, a_2, \dots, a_E\}$ , the FD theory supposes that the potential existence of correlated attributes leads the set of points in the original  $E$ -dimensional space to describe one spatial object in a dimension that is lower than or equal to  $E$ . The dimension of the object represented by the data set is called the intrinsic dimension (ID), denoted by  $D$ ,  $D \in \mathbb{R}^+$ . The ceiling of the ID  $\lceil D \rceil$  is the minimum number of attributes that must be retained to keep the essential characteristics of a data set (de Sousa et al. 2007).

However, the ID of a data set is difficult to obtain. Alternatively, we consider the ID projecting the data set on a subset  $\mathbf{C} \subset \mathbf{A}$ , where  $\mathbf{C}$  is defined by an attribute subspace  $\mathbf{C} \subset \mathbf{A}$ . It is named partial intrinsic dimension (PID) on  $\mathbf{C}$ :  $pD(\mathbf{C})$ . Based on the definitions, the attribute  $a_i \in (\mathbf{A} - \mathbf{C})$  increases  $pD(\mathbf{C})$  by at most its individual contribution (IC) according to the degree of the correlation between  $a_i$  and the attributes in  $\mathbf{C}$ . The IC of  $a_i$ , i.e.,  $iC(a_i)$ , is the maximum potential contribution of the attribute  $a_i$  to  $pD(\mathbf{C})$ . The greater

Table 2  
The Proposed FDIDWT Method

<b>Input:</b>	original data set defined on attribute space $\mathbf{A}$ decomposition level $J$ correlation threshold $\xi$ scale range $n$
<b>Output:</b>	new data set defined on attribute space $\mathbb{B}$
1:	run FDASE algorithm to find out the ASC $\xi C$ , denote its length as $P$ ;
2:	calculate $I_a(a_i)$ , $a_i \in (\mathbf{A} - \xi C)$ , with Equation (6);
3:	arrange the attributes of $\mathbf{A}$ into $\mathbf{A}'$ according to the order of $I_a(a_i)$ as shown in Equation (7);
4:	based on $P$ , calculate the length of other groups, insert placeholder attributes if needed;
5:	construct the WDV $\mathbf{c}$ and BV $\mathbf{l}$ ;
6:	perform IDWT on each normalized sample with $\mathbf{c}$ and $\mathbf{l}$ ;
7:	output the new data set defined on $\mathbb{B}$ ;

the correlation between  $a_i$  and the attributes in  $\mathbf{C}$ , the lower its contribution to  $pD(\mathbf{C})$ .

In addition,  $iC(a_i)$  can be measured by  $pD(\{a_i\})$  and it ranges in  $[0, 1]$ . A more independent distribution of the values of  $a_i$  leads to  $iC(a_i)$  closer to one, while a more structured distribution brings  $iC(a_i)$  closer to zero (de Sousa et al. 2007). Thus, the  $E$ -dimensional data set can be seen as formed by adding the attributes with different contributions to the  $D$ -dimensional sub-data set.

Moreover, the degree of correlations among attributes can be measured by a threshold  $\xi$ : a sub-data set  $\mathbf{B} \subset \mathbf{A}$  is said to be  $\xi$ -correlated to another sub-data set  $\mathbf{C} \subset \mathbf{A}$  (their attribute spaces  $\mathbb{B} \cap \mathbb{C} = \emptyset$ ) if every attribute  $a_i \in \mathbb{B}$  does not contribute more than  $\xi * iC(a_i)$  to  $pD(\mathbf{C})$ . The threshold  $\xi \in [0, 1]$  tunes how strong the correlation between attributes in  $\mathbb{B}$  and attributes in  $\mathbf{C}$  should be to be detected.

A greedy algorithm FDASE was developed by de Sousa et al. (2007) to find a subset of attributes whose PID approaches the ID of the whole data set. The resulting subset is called the attribute set core  $\xi C$  (ASC), with the given correlation threshold  $\xi$  and scale range  $n$ . Therefore, the ratio  $pD(\xi C)/ID$  normalizes the contribution of ASC to the whole data set. Similarly, for each split data set, we utilize the FDASE algorithm on the training sets with scanning correlation threshold  $\xi$  and compare the value of  $pD(\xi C)/ID$  with the number of ASC in Figure 5. We fix the scale range  $n = 50$  and plot the IC for each attribute in Figure 6 under a suitable  $\xi$ . From the figures, we find that when  $\xi = 0.5$  two attributes (PL\_Index and Pivot\_Energy) contribute around 98% information to Data Set A. While for Data Set B, four attributes (PL\_Index, Pivot\_Energy, Flux\_Band1, and Flux\_Band7) contribute around 92% information when  $\xi = 0.5$ .

**Table 3**  
The WDV and BV for the IDWT in Two Missions

Decomposition Level	WDV	BV	Wavelet Name	Case
1	A: $(\{x_3, x_1, x_5, x_{13}, x_{10}, x_2, x_9\}, \{0, x_{12}, x_4, x_8, x_{11}, x_7, x_6\})$ B: $(\{x_3, x_1, x_5, x_{13}, x_4, x_{12}, x_{11}\}, \{0, x_7, x_8, x_6, x_9, x_2, x_{10}\})$	(7,7,13)	db1	#1
		(7,7,12)	db2	#2
		(7,7,11)	db2	#3
		(7,7,10)	db3	#4
		(7,7,9)	db3	#5
		(7,7,8)	db4	#6
		(7,7,7)	db4	#7
		(7,7,6)	db5	#8
		(7,7,5)	db5	#9
		(7,7,4)	db6	#10
		(7,7,3)	db6	#11
		(7,7,2)	db7	#12
		(7,7,1)	db7	#13
3	A: $(\{x_3, x_1, x_5\}, \{x_{13}, x_{10}, x_2\}, \{x_9, x_{12}, x_4\}, \{x_8, x_{11}, x_7, x_6\})$ B: $(\{x_3, x_1, x_5\}, \{x_{13}, x_4, x_{12}\}, \{x_{11}, x_7, x_8\}, \{x_6, x_9, x_2, x_{10}\})$	(3,3,3,4,5)	db2	#14
		(3,3,3,4,6)	db2	#15

**Note.** Value 0 in the WDV means one placeholder attribute is inserted.

**Table 4**  
The Hyper-parameters of Attribute Analysis Methods and Classifiers

Attribute Analysis Methods	Hyper-parameters	Classifiers	Hyper-parameters
RF Attribute Selection (AS)	#DTs: 50,000 criterion: entropy	RF	#DTs: 700 criterion: entropy
PCA Component Reduction (CR)	normalization: StandardScaler	SVM	normalization: StandardScaler C = 5.0 degree = 1
FDASE Attribute Selection (FDASE)	scale range: 50 correlation threshold: 0.5	AdaBoost	normalization: MinMaxScaler #estimators: 300 learning rate: 0.5
		MLP	hidden units: [32, 32, 64, 32, 32] dropout: 0.5 activation function: Leaky ReLU batch size: 64 epochs: 500 learning rate: 0.0005

### 3. Proposed Method

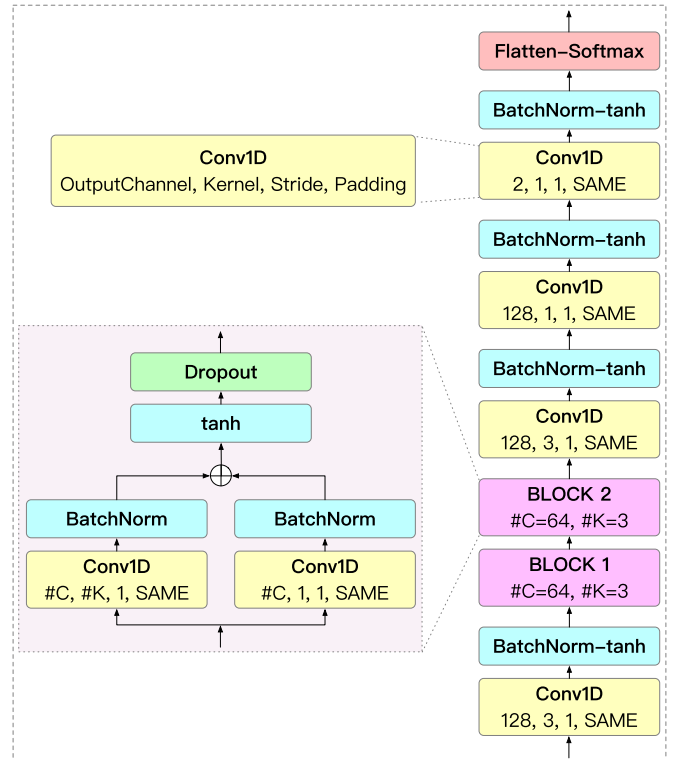
#### 3.1. Inverse Discrete Wavelet Transform

The attribute analysis methods introduced in Section 2 are used to find the important attributes or components that can be retained for the learning process, while other attributes or components are removed for dimension reduction. However, the crude removal may cause the loss of correlation features that degrade the performance.

We propose to retain all of the attributes but perform IDWT on the original samples for representations at higher resolutions. After IDWT, the correlation features are supposed to be highlighted, while the dimension is possibly reduced, which is important when dealing with big data.

The well-known discrete wavelet transform (DWT) is usually implemented with filtering operations for high efficiency. The filters are designed based on the standpoint of multiresolution: the difference of information between the approximation of a signal at the resolutions  $2^{m+1}$  and  $2^m$  (where  $m$  is an integer) can be extracted by decomposing this signal on an orthonormal basis of wavelets (Mallat 1989). The pyramidal structure of the wavelet filter bank makes it possible to infer information at a low resolution from information at a high resolution. IDWT is the converse process of DWT and provides representations at high resolutions for DL.

However, in practical applications, a finite signal should be considered. The length of the signals varies at different



**Figure 10.** The proposed lightweight MatchboxConv1D model.

**Table 5**  
The Hyper-parameters of the Proposed Method and Model

Proposed Method	Hyper-parameters	Proposed Model	Hyper-parameters
FDIDWT	wavelet type: Daubechies max decomposition level: 3	MatchboxConv1D	dropout: 0.9 activation function: tanh batch size: 64 epochs: 500 learning rate: 0.0005 more can be found in Figure 10

**Table 6**  
The Average Test Accuracy Results of Mission A with the ML Methods

Attribute Analysis Methods	#Dimension	RF	SVM	AdaBoost	MLP
RF Attribute Selection (AS)	1	86.08% $\pm$ 1.15%	87.02% $\pm$ 0.0%	87.95% $\pm$ 0.41%	87.02% $\pm$ 0.0%
	2	85.15% $\pm$ 1.34%	87.02% $\pm$ 0.0%	87.77% $\pm$ 0.61%	87.02% $\pm$ 0.0%
	3	87.65% $\pm$ 1.21%	87.02% $\pm$ 0.0%	88.22% $\pm$ 1.0%	87.02% $\pm$ 0.0%
	4	90.39% $\pm$ 1.18%	89.38% $\pm$ 0.49%	88.91% $\pm$ 1.58%	86.56% $\pm$ 0.0%
	5	90.73% $\pm$ 1.16%	91.12% $\pm$ 0.59%	90.77% $\pm$ 1.0%	89.98% $\pm$ 0.0%
	6	<b>90.82% <math>\pm</math> 1.18%</b>	90.93% $\pm$ 0.64%	90.87% $\pm$ 0.91%	90.2% $\pm$ 0.0%
	7	90.55% $\pm$ 0.98%	92.23% $\pm$ 0.88%	93.35% $\pm$ 0.93%	91.34% $\pm$ 0.0%
	8	90.0% $\pm$ 0.92%	93.14% $\pm$ 0.81%	94.4% $\pm$ 1.03%	92.03% $\pm$ 0.0%
	9	90.73% $\pm$ 1.03%	93.26% $\pm$ 0.78%	94.69% $\pm$ 1.02%	92.71% $\pm$ 0.0%
	10	90.64% $\pm$ 0.97%	93.3% $\pm$ 0.83%	94.81% $\pm$ 1.14%	<b>93.39% <math>\pm</math> 0.0%</b>
	11	90.27% $\pm$ 0.79%	93.35% $\pm$ 0.78%	95.33% $\pm$ 0.88%	91.57% $\pm$ 0.0%
	12	89.32% $\pm$ 0.83%	<b>93.42% <math>\pm</math> 0.76%</b>	95.4% $\pm$ 0.77%	91.57% $\pm$ 0.0%
	13	88.38% $\pm$ 0.59%	93.37% $\pm$ 1.04%	<b>95.47% <math>\pm</math> 0.86%</b>	90.43% $\pm$ 0.0%
PCA Component Reduction (CR)	1	80.75% $\pm$ 2.17%	87.15% $\pm$ 0.75%	86.97% $\pm$ 0.14%	87.93% $\pm$ 0.0%
	2	85.97% $\pm$ 1.5%	87.49% $\pm$ 0.81%	86.63% $\pm$ 0.82%	87.47% $\pm$ 0.0%
	3	91.85% $\pm$ 1.05%	90.84% $\pm$ 0.53%	91.5% $\pm$ 1.16%	88.61% $\pm$ 0.0%
	4	92.69% $\pm$ 0.75%	91.41% $\pm$ 0.61%	91.64% $\pm$ 1.0%	89.07% $\pm$ 0.0%
	5	92.78% $\pm$ 0.77%	91.75% $\pm$ 0.64%	92.71% $\pm$ 1.08%	90.43% $\pm$ 0.0%
	6	93.37% $\pm$ 0.61%	92.71% $\pm$ 0.6%	93.3% $\pm$ 0.88%	91.34% $\pm$ 0.0%
	7	93.33% $\pm$ 0.95%	92.78% $\pm$ 0.59%	93.39% $\pm$ 0.9%	<b>91.8% <math>\pm</math> 0.0%</b>
	8	94.03% $\pm$ 0.97%	93.12% $\pm$ 0.67%	93.9% $\pm$ 0.58%	91.8% $\pm$ 0.0%
	9	93.96% $\pm$ 0.89%	93.28% $\pm$ 0.78%	94.03% $\pm$ 0.52%	91.57% $\pm$ 0.0%
	10	94.03% $\pm$ 1.05%	93.35% $\pm$ 0.9%	<b>94.17% <math>\pm</math> 0.83%</b>	91.57% $\pm$ 0.0%
	11	94.37% $\pm$ 0.89%	<b>93.37% <math>\pm</math> 1.04%</b>	94.12% $\pm$ 0.74%	91.57% $\pm$ 0.0%
	12	94.76% $\pm$ 0.91%	93.37% $\pm$ 1.04%	94.15% $\pm$ 1.04%	91.57% $\pm$ 0.0%
	13	<b>94.87% <math>\pm</math> 0.93%</b>	93.37% $\pm$ 1.04%	94.17% $\pm$ 0.85%	91.34% $\pm$ 0.0%
FDASE Attribute Selection (FDASE)	2	87.88% $\pm$ 1.36%	88.88% $\pm$ 0.45%	87.38% $\pm$ 0.37%	87.7% $\pm$ 0.0%
	3	87.43% $\pm$ 0.45%	89.16% $\pm$ 0.49%	88.2% $\pm$ 0.76%	88.84% $\pm$ 0.0%
	4	89.18% $\pm$ 0.85%	89.86% $\pm$ 0.61%	89.43% $\pm$ 0.95%	88.84% $\pm$ 0.0%
	5	88.2% $\pm$ 0.34%	90.14% $\pm$ 0.79%	89.61% $\pm$ 0.8%	88.84% $\pm$ 0.0%
	6	90.73% $\pm$ 1.05%	91.28% $\pm$ 0.57%	90.93% $\pm$ 1.03%	89.75% $\pm$ 0.0%
	7	89.09% $\pm$ 0.61%	92.6% $\pm$ 0.8%	93.14% $\pm$ 1.12%	90.66% $\pm$ 0.0%
	8	87.93% $\pm$ 0.59%	92.57% $\pm$ 0.75%	93.39% $\pm$ 1.27%	91.12% $\pm$ 0.0%
	9	<b>90.91% <math>\pm</math> 0.87%</b>	92.32% $\pm$ 0.61%	93.83% $\pm$ 0.89%	91.34% $\pm$ 0.0%
	10	90.48% $\pm$ 0.88%	93.12% $\pm$ 0.69%	95.4% $\pm$ 1.02%	<b>91.8% <math>\pm</math> 0.0%</b>
	11	89.95% $\pm$ 0.93%	93.23% $\pm$ 0.79%	<b>95.49% <math>\pm</math> 1.05%</b>	91.34% $\pm$ 0.0%
	12	89.64% $\pm$ 0.89%	<b>93.42% <math>\pm</math> 0.76%</b>	95.4% $\pm$ 0.77%	91.8% $\pm$ 0.0%
	13	88.34% $\pm$ 0.69%	93.37% $\pm$ 1.04%	95.47% $\pm$ 0.86%	91.57% $\pm$ 0.0%

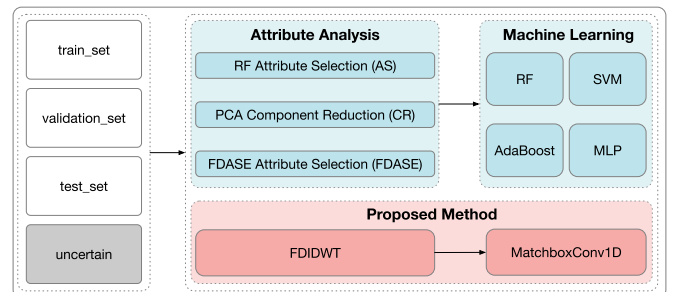
**Note.** The highest test accuracy for each combination of the attribute analysis method and the classifier is highlighted in bold face.

resolutions, which is due to the operations of downsampling, upsampling, and filtering. In the IDWT process, if we denote  $p$  and  $s$  as the length of the signal at low resolution and high resolution, respectively, then Rajmic & Prusa (2014) give

$$s = \begin{cases} 2p - u + 1, & \text{for even } s + u - 1, \\ 2p - u + 2, & \text{for odd } s + u - 1, \end{cases} \quad (1)$$

where  $u$  is the length of the filters.

To perform IDWT on the original data set, the attribute data of one sample will be seen as the signal at a lower resolution. As a result, the attributes should be rearranged in some order



**Figure 11.** The flowchart of attribute analysis and learning process.



**Table 7**  
The Average Test Accuracy Results of Mission B with the ML Methods

Attribute Analysis Methods	#Dimension	RF	SVM	AdaBoost	MLP
RF Attribute Selection (AS)	1	80.31% $\pm$ 3.08%	87.05% $\pm$ 1.85%	86.61% $\pm$ 1.69%	86.34% $\pm$ 0.0%
	2	87.97% $\pm$ 1.97%	88.9% $\pm$ 1.67%	87.89% $\pm$ 1.97%	88.55% $\pm$ 0.0%
	3	89.65% $\pm$ 2.12%	<b>90.04%</b> $\pm$ <b>0.89%</b>	<b>89.82%</b> $\pm$ <b>1.45%</b>	89.87% $\pm$ 0.0%
	4	89.69% $\pm$ 1.46%	89.69% $\pm$ 1.02%	89.74% $\pm$ 1.86%	89.43% $\pm$ 0.0%
	5	89.65% $\pm$ 1.79%	89.34% $\pm$ 1.05%	89.12% $\pm$ 2.08%	89.43% $\pm$ 0.0%
	6	89.82% $\pm$ 1.93%	89.38% $\pm$ 1.13%	89.52% $\pm$ 1.83%	89.87% $\pm$ 0.0%
	7	89.91% $\pm$ 1.55%	89.52% $\pm$ 1.03%	89.38% $\pm$ 2.19%	89.87% $\pm$ 0.0%
	8	<b>90.04%</b> $\pm$ <b>1.62%</b>	89.74% $\pm$ 0.95%	89.3% $\pm$ 2.04%	89.87% $\pm$ 0.0%
	9	89.96% $\pm$ 1.54%	89.74% $\pm$ 1.3%	89.07% $\pm$ 2.32%	89.87% $\pm$ 0.0%
	10	89.91% $\pm$ 1.69%	89.82% $\pm$ 1.35%	89.16% $\pm$ 1.82%	90.31% $\pm$ 0.0%
	11	89.87% $\pm$ 1.38%	90.0% $\pm$ 1.21%	89.12% $\pm$ 2.29%	90.75% $\pm$ 0.0%
	12	89.52% $\pm$ 1.39%	90.0% $\pm$ 1.14%	89.34% $\pm$ 1.81%	<b>91.19%</b> $\pm$ <b>0.0%</b>
	13	89.69% $\pm$ 1.78%	90.0% $\pm$ 1.3%	89.43% $\pm$ 1.83%	90.75% $\pm$ 0.0%
PCA Component Reduction (CR)	1	61.15% $\pm$ 2.41%	65.42% $\pm$ 1.85%	89.03% $\pm$ 1.16%	63.44% $\pm$ 0.0%
	2	89.07% $\pm$ 1.11%	90.4% $\pm$ 1.54%	88.81% $\pm$ 1.65%	89.87% $\pm$ 0.0%
	3	88.94% $\pm$ 1.4%	90.53% $\pm$ 1.44%	89.3% $\pm$ 1.54%	90.75% $\pm$ 0.0%
	4	88.9% $\pm$ 1.66%	<b>90.57%</b> $\pm$ <b>1.44%</b>	88.81% $\pm$ 2.08%	90.75% $\pm$ 0.0%
	5	89.65% $\pm$ 1.64%	90.13% $\pm$ 1.27%	88.94% $\pm$ 1.72%	90.75% $\pm$ 0.0%
	6	89.78% $\pm$ 1.47%	89.96% $\pm$ 1.33%	88.85% $\pm$ 1.64%	90.75% $\pm$ 0.0%
	7	90.09% $\pm$ 1.43%	90.09% $\pm$ 1.25%	88.85% $\pm$ 2.08%	90.75% $\pm$ 0.0%
	8	89.96% $\pm$ 1.58%	90.09% $\pm$ 1.32%	<b>89.52%</b> $\pm$ <b>1.73%</b>	90.75% $\pm$ 0.0%
	9	89.56% $\pm$ 1.47%	90.0% $\pm$ 1.18%	89.3% $\pm$ 1.48%	90.75% $\pm$ 0.0%
	10	89.82% $\pm$ 1.46%	89.87% $\pm$ 1.31%	88.99% $\pm$ 2.13%	90.75% $\pm$ 0.0%
	11	89.87% $\pm$ 1.66%	89.91% $\pm$ 1.38%	89.08% $\pm$ 2.02%	90.75% $\pm$ 0.0%
	12	90.04% $\pm$ 1.6%	90.0% $\pm$ 1.3%	88.94% $\pm$ 1.67%	<b>91.19%</b> $\pm$ <b>0.0%</b>
	13	<b>90.26%</b> $\pm$ <b>1.74%</b>	90.0% $\pm$ 1.3%	88.59% $\pm$ 1.56%	90.31% $\pm$ 0.0%
FDASE Attribute Selection (FDASE)	2	88.06% $\pm$ 2.15%	88.9% $\pm$ 1.67%	87.89% $\pm$ 1.97%	88.55% $\pm$ 0.0%
	3	88.63% $\pm$ 1.99%	88.63% $\pm$ 1.97%	87.71% $\pm$ 2.13%	86.78% $\pm$ 0.0%
	4	88.28% $\pm$ 1.78%	88.68% $\pm$ 2.0%	87.97% $\pm$ 2.13%	86.78% $\pm$ 0.0%
	5	89.69% $\pm$ 2.14%	89.74% $\pm$ 0.95%	89.12% $\pm$ 1.62%	90.31% $\pm$ 0.0%
	6	89.82% $\pm$ 2.08%	89.52% $\pm$ 0.74%	88.9% $\pm$ 1.41%	89.87% $\pm$ 0.0%
	7	90.0% $\pm$ 1.92%	89.52% $\pm$ 0.9%	89.3% $\pm$ 1.74%	90.75% $\pm$ 0.0%
	8	90.09% $\pm$ 2.16%	89.6% $\pm$ 0.86%	89.16% $\pm$ 1.64%	90.31% $\pm$ 0.0%
	9	89.96% $\pm$ 1.99%	89.56% $\pm$ 0.91%	<b>89.52%</b> $\pm$ <b>1.8%</b>	90.31% $\pm$ 0.0%
	10	<b>90.13%</b> $\pm$ <b>2.09%</b>	89.6% $\pm$ 1.06%	89.12% $\pm$ 1.9%	90.75% $\pm$ 0.0%
	11	90.13% $\pm$ 2.06%	89.74% $\pm$ 1.18%	89.03% $\pm$ 1.6%	90.75% $\pm$ 0.0%
	12	89.78% $\pm$ 1.61%	<b>90.0%</b> $\pm$ <b>1.14%</b>	89.34% $\pm$ 1.81%	90.31% $\pm$ 0.0%
	13	89.91% $\pm$ 1.37%	90.0% $\pm$ 1.3%	89.43% $\pm$ 1.83%	<b>91.19%</b> $\pm$ <b>0.0%</b>

**Note.** The highest test accuracy for each combination of the attribute analysis method and the classifier is highlighted in bold face.

based on wavelet theory. This can be achieved with the alignment of the information in the wavelet domain and attribute space, which will be discussed next.

### 3.2. Information in Wavelet Domain: $I_c$

The process of DWT is to analyze the original signal from a fine scale to a coarse scale. The representation of a finite signal  $f(t)$  in the wavelet domain after DWT is a collection of vectors:

$$\{c_j(k): k \in \mathbb{Z}\} \cup \{d_j(k): 1 \leq j \leq J; k \in \mathbb{Z}\}, \quad (2)$$

where  $J \in \mathbb{N}$  is called decomposition level;  $c_J$  contains the approximation coefficients at level  $J$ , i.e., the lowest resolution; and  $d_j$  contains the detail coefficients at level  $j$ , i.e., the high resolutions. An example of the three-level pyramid transform is illustrated in Figure 7. We find that the original signal  $f(t)$  can be seen as the approximation coefficients at the highest resolution, i.e.,  $c_0$ .

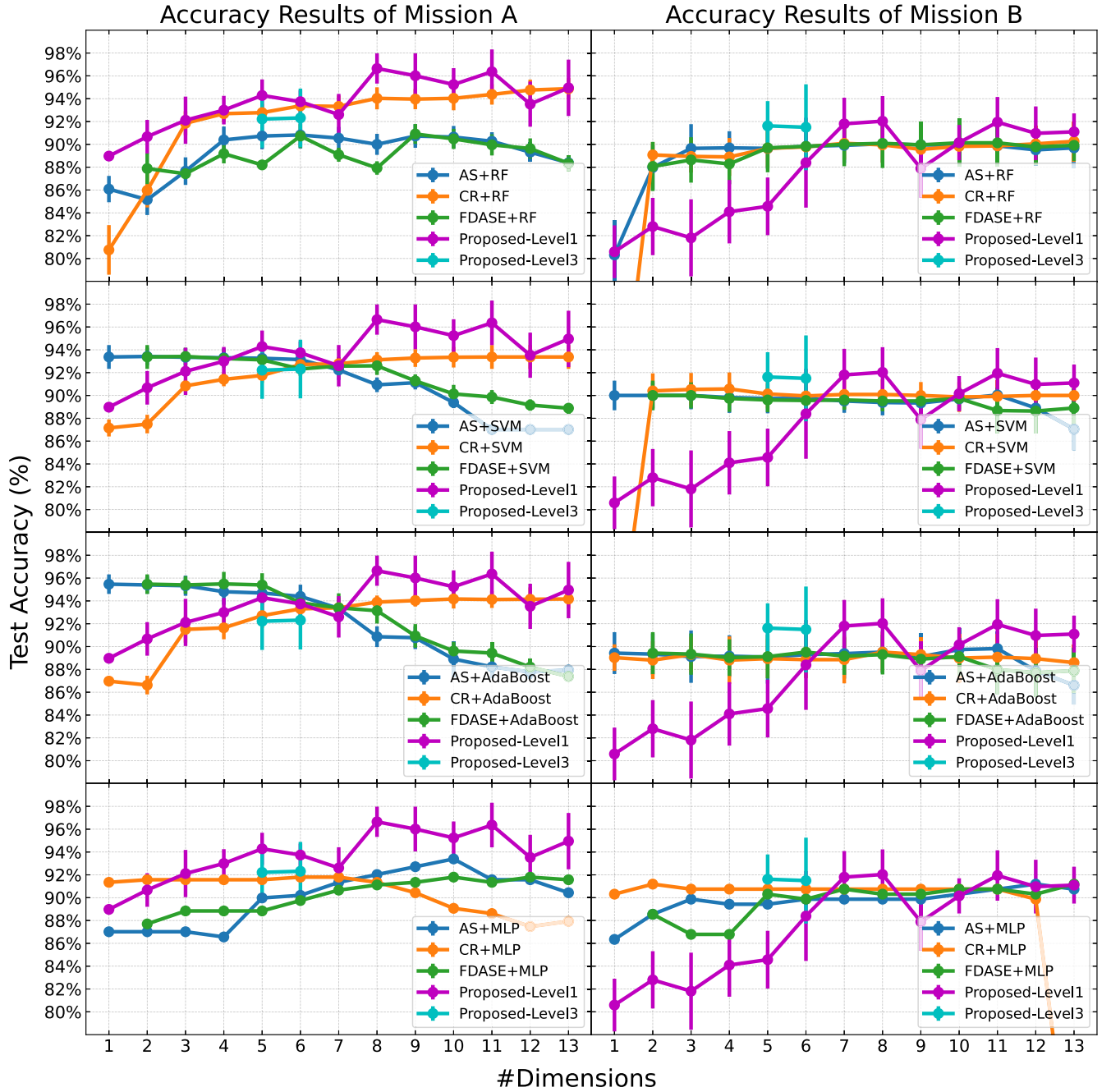
However, based on prior knowledge, it is often observed that at a specific level, the majority of information in a natural

signal is typically presented in the approximation coefficients. Additionally, the information is typically reduced by half during the decomposition of the signal from level  $j$  to  $j+1$  in DWT, due to the downsampling operation. Hence, it can be found that the information of coefficients at different levels roughly respect

$$I_c(d_j) < \dots < I_c(d_2) < I_c(d_1) < I_c(c_j). \quad (3)$$

As an example, the three-level DWT is performed on three images to intuitively illustrate the relationship, as shown in Figure 8. These figures suggest that the approximation coefficients at each level retain the image contour and contain a significant amount of information, while the objects in the image cannot be easily identified through the detail coefficients. However, to some extent, the detail coefficients at a low level appear to provide more information than those at a high level, i.e.,  $I_c(d_{j+1}) < I_c(d_j)$ .

To transform the coefficients from lower resolutions to higher resolutions in the IDWT process, the wavelet decomposition vector (WDV)  $c$  and bookkeeping vector (BV)  $l$  are



**Figure 12.** A comparison of average test accuracy results in two missions. The results of the proposed methods in each mission are depicted in every panel.

required according to multiresolution analysis. The WDV includes the coefficients shown in Equation (2):

$$\mathbf{c} = (c_j, d_j, d_{j-1}, \dots, d_1), \quad (4)$$

while the BV is made up of the number of coefficients in  $\mathbf{c}$ :

$$\mathbf{l} = (\text{len}(\mathbf{c}_j), \text{len}(\mathbf{d}_j), \text{len}(\mathbf{d}_{j-1}), \dots, \text{len}(\mathbf{d}_1), \text{len}(f(t))), \quad (5)$$

where  $\text{len}(\cdot)$  indicates the number of the elements in a vector.

### 3.3. Information in Attribute Space: $I_a$

In the field of big data, an attribute's information reflects its contribution to the data set. As explained in Section 2.4, the IC estimates the potential contribution of an individual attribute to the data set. However, the presence of correlations among

attributes means that the actual contribution of an attribute  $a_i$ , i.e., its information  $I_a(\{a_i\})$ , cannot be precisely determined if  $a_i$  involves other attributes.

Moreover, as discussed, ASC is the smallest subset of attributes that can fully characterize the entire data set. Given a correlation threshold  $\xi$ , the ASC  $\xi C$  can be found with the FDASE algorithm (de Sousa et al. 2007). The attributes in ASC are not  $\xi$ -correlated with each other and they contain the most information in the data set.

Based on the analysis, if  $\mathbb{A} = \{a_1, a_2, \dots, a_E\}$  is denoted as the universal attribute set of the data set that has  $E$  attributes, then the real contribution of attribute  $a_i \in (\mathbb{A} - \xi C)$  to the data set can be seen as the degree of correlation between  $a_i$  and the attributes in  $\xi C$ . The weaker the correlation between  $a_i$  and  $\xi C$ ,

**Table 8**

The Average Test Accuracy Results of Two Missions with the FDIDWT Method and MatchboxConv1D Model

#Dimension	Case	Test Accuracy	
		Mission A	Mission B
13	#1	94.95% $\pm$ 2.47%	91.1% $\pm$ 1.61%
12	#2	93.53% $\pm$ 1.98%	90.97% $\pm$ 2.35%
11	#3	96.36% $\pm$ 1.95%	91.94% $\pm$ 2.21%
10	#4	95.24% $\pm$ 1.44%	90.16% $\pm$ 1.54%
9	#5	96.01% $\pm$ 1.96%	87.89% $\pm$ 2.53%
8	#6	<b>96.65% <math>\pm</math> 1.32%</b>	<b>92.03% <math>\pm</math> 2.2%</b>
7	#7	92.61% $\pm$ 1.81%	91.8% $\pm$ 2.28%
6	#8	93.74% $\pm$ 0.99%	88.39% $\pm$ 3.94%
5	#9	94.29% $\pm$ 1.4%	84.57% $\pm$ 2.53%
4	#10	93.0% $\pm$ 1.25%	84.1% $\pm$ 2.78%
3	#11	92.11% $\pm$ 2.07%	81.81% $\pm$ 3.38%
2	#12	90.67% $\pm$ 1.47%	82.8% $\pm$ 2.51%
1	#13	88.97% $\pm$ 0.01%	80.59% $\pm$ 2.31%
5	#14	92.22% $\pm$ 2.51%	91.62% $\pm$ 2.17%
6	#15	92.31% $\pm$ 2.57%	91.50% $\pm$ 3.76%

**Note.** The highest test accuracy for each combination of the attribute analysis method and the classifier is highlighted in bold face.

**Table 9**

The Prediction Results of Uncertain Sources in Mission A and BCUs in Mission B

Mission	Source Name	Predict Probability	Predict Class
A	4FGL J0000.3-7355	0.9963 $\pm$ 0.0028	AGN
	4FGL J0000.5+0743	0.9948 $\pm$ 0.0048	AGN
	4FGL J0000.7+2530	0.9822 $\pm$ 0.0229	AGN
	4FGL J0001.6+3503	0.9805 $\pm$ 0.0153	AGN
	4FGL J0002.1+6721c	0.9698 $\pm$ 0.0111	non-AGN
B	4FGL J0001.2+4741	0.9315 $\pm$ 0.0404	BL Lac (BCU)
	4FGL J0001.6-4156	0.9887 $\pm$ 0.0159	BL Lac
	4FGL J0001.8-2153	0.8635 $\pm$ 0.0966	BL Lac (BCU)
	4FGL J0002.3-0815	0.9881 $\pm$ 0.0047	BL Lac
	4FGL J0002.4-5156	0.8595 $\pm$ 0.1072	BL Lac (BCU)

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

the higher the contribution of  $a_i$  to the data set, i.e., attribute  $a_i$  contains more information. Therefore, the information of attribute  $a_i$  can be estimated by

$$I_a(\{a_i\}) = |pD(\{a_i\} \cup \xi C) - pD(\xi C)|, a_i \in (\mathbb{A} - \xi C), \quad (6)$$

where  $pD(\cdot)$  is the PID of a sub-data set.

By estimating the information for each attribute  $a_i \in (\mathbb{A} - \xi C)$ , we can obtain a rough ordering of attributes based on their information. In this case, the data under the attributes can be treated as wavelet coefficients and arranged to form the WDV  $\mathbf{c}$  and BV  $\mathbf{l}$  for IDWT. The data set is then transformed into representations at a higher resolution where the correlation features are presented. This is the main idea of our method, which will be detailed next.

### 3.4. FDIDWT

The  $J$ -level IDWT aims to transform the original data into representations at the  $J$ -higher resolution. To achieve this, the WDV and BV must be constructed with the attribute data of

each sample. However, the positions of the data in the WDV are determined by the estimation of the attribute information with Equation (6). As a result, this approach is referred to as FDIDWT.

As discussed earlier, the attributes in  $\text{ASC } \xi C \subset \mathbb{A}$  contain the most information of the data set defined in  $\mathbb{A} = \{a_1, a_2, \dots, a_E\}$ . It is assumed that the information of attributes in  $\mathbb{A} - \xi C$  respect an ascending order:

$$I_a(\{a_1\}) < I_a(\{a_2\}) < \dots < I_a(\{a_i\}) < \dots < I_a(\{a_Q\}) < I_a(\xi C), \quad (7)$$

where  $a_i \in (\mathbb{A} - \xi C)$  and  $1 \leq i \leq Q$ . If the number of attributes in  $\xi C$  is  $P$ , then  $P + Q = E$ . Contrasting with the order of the information in the wavelet domain shown in Equation (3), we group the attributes in  $\mathbb{A} - \xi C$  and obtain a similar order of the information in the attribute space:

$$I_a(D_f^M) < I_a(D_{f-1}^N) < \dots < I_a(D_f^L) < \dots < I_a(D_1^K) < I_a(\xi C), \quad (8)$$

where  $D_f^L$  is the attribute group containing  $L$  attributes whose information respects to Equation (7), i.e.,

$$\begin{aligned} D_f^M &= \{a_1, a_2, \dots, a_M\}, \\ D_{f-1}^N &= \{a_{M+1}, a_{M+2}, \dots, a_{M+N}\}, \\ &\dots \\ D_1^K &= \{a_{Q-K+1}, a_{Q-K+2}, \dots, a_Q\}, \\ M + N + \dots + K &= Q. \end{aligned} \quad (9)$$

Thus, the data under the attributes of  $D_f^L$  and  $\xi C$  can be taken as the detail coefficients  $\mathbf{d}_f$  and approximation coefficients  $\mathbf{c}_J$ , respectively. Specifically, if denote  $\mathbf{x}_J$  and  $\mathbf{x}_{\xi C}$  as the data under the attributes in  $D_f^L$  and  $\xi C$ , respectively, for one sample of the original data set, then the WDV  $\mathbf{c}$  and BV  $\mathbf{l}$  for the IDWT process can be constructed as

$$\begin{aligned} \mathbf{c} &= (\mathbf{x}_{\xi C}, \mathbf{x}_J, \mathbf{x}_{J-1}, \dots, \mathbf{x}_1), \\ \mathbf{l} &= (P, M, N, \dots, K, O), \end{aligned} \quad (10)$$

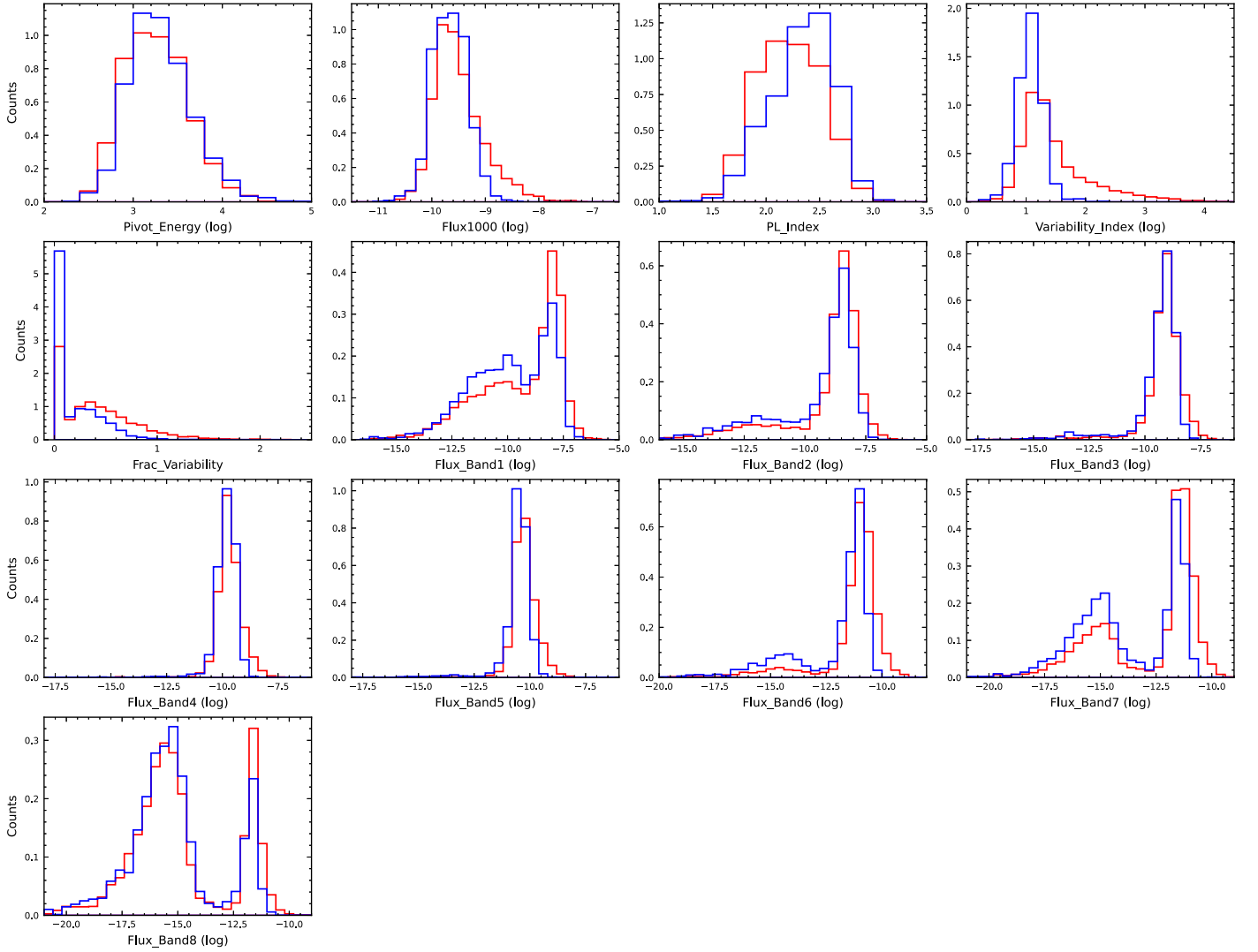
where  $O$  is the number of attributes in the transformed data set, i.e., the transformed dimension in the new attribute space, which can be derived with Equation (1). Table 2 shows the procedure of the proposed FDIDWT method. To better describe how the algorithm works, we fix the decomposition level  $J = 3$ . Moreover, as in the common case, the length of filters  $u$  for the pyramid transform is set to be even. Then:

- Steps 1–3: Given the correlation threshold  $\xi$  and scale range  $n$ , the *FDASE* algorithm (de Sousa et al. 2007) is used to find out the attribute set core  $\xi C$ . The information  $I_a(\{a_i\})$  of the remaining attributes  $a_i \in (\mathbb{A} - \xi C)$  is then calculated. The attributes of  $\mathbb{A}$  are arranged into a set  $\mathbb{A}'$  according to the ascending order of  $I_a(\{a_i\})$ .<sup>7</sup> For instance, if the relationship in Equation (7) holds, then

$$\mathbb{A}' = \xi C \cup \{a_1, a_2, \dots, a_i, \dots, a_Q\}. \quad (11)$$

In our experiment, we scanned the correlation threshold  $\xi$  in the range [0.05, 1.0] with step 0.05, and set scale range  $n = 50$ . We got the ASC  $\{a_3, a_1\}$  and  $\{a_3, a_1, a_5\}$  for Data Set A and B, respectively, as shown in Figures 5 and

<sup>7</sup> Here,  $\mathbb{A}'$  is a sequence since the information of its attributes respects an order.



**Figure 13.** A comparison of 13 attributes between 4FGL\_DR3 AGNs and predicted AGNs. The histogram of 4FGL\_DR3 AGNs is shown in red and the histogram of predicted AGNs is shown in blue; the area under the histogram integrates into one.

6. Then, the arranged attribute set  $\mathbb{A}'$  for the two data sets will be  $\{a_3, a_1, a_5, a_{13}, a_{10}, a_2, a_9, a_{12}, a_4, a_8, a_{11}, a_7, a_6\}$  and  $\{a_3, a_1, a_5, a_{13}, a_4, a_{12}, a_{11}, a_7, a_8, a_6, a_9, a_2, a_{10}\}$ .
2. Steps 4–5:  $P$  is the number of attributes in  $\xi C$ . In the proposed method, the length of the approximation coefficients at level  $J$  is assumed to be not shorter than  $P$ . If this length equals  $P$ , then Figure 9 shows other groups divided from  $\mathbb{A}'$ .  $C_3^P$  denotes the approximation coefficient group of attributes in  $\xi C$ . Note that the original dimension  $E$ , i.e., the number of attributes of the original data set, may be smaller than the length of WDV  $c$  due to the given decomposition level  $J$  and filter length  $u$ . In this case, some necessary placeholder attributes “\*” are inserted following the ASC group  $C_3^P$ . The role of the placeholder attributes is to ensure that IDWT works while not changing the information of the data. Therefore, the values under the placeholder attributes are set to zeros so that they contribute nothing to the data set. In the signal processing field, signals are always transformed into the wavelet domain with DWT for processing, and the lengths of the wavelet coefficients at each level are recorded for the reconstruction with IDWT. However,

there is no DWT in the proposed method. Hence, many cases exist for the lengths of the divided groups based on Equation (1):

$$\begin{aligned} M &= P, \\ N &= \begin{cases} 2M - u + 1, & \text{for even } N + u - 1 \\ 2M - u + 2, & \text{for odd } N + u - 1 \end{cases}, \\ K &= \begin{cases} 2N - u + 1, & \text{for even } K + u - 1 \\ 2N - u + 2, & \text{for odd } K + u - 1. \end{cases} \end{aligned} \quad (12)$$

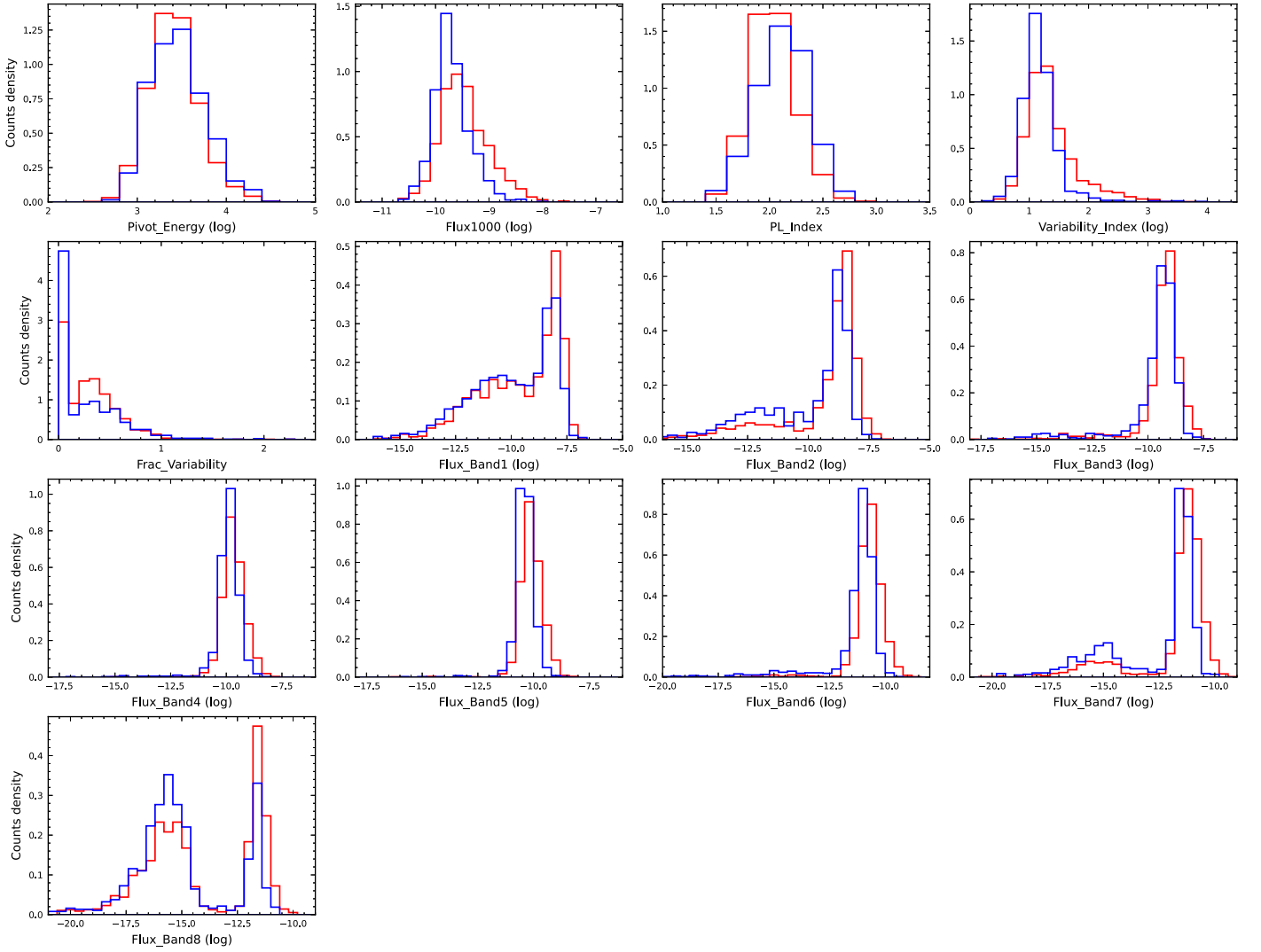
The length of the placeholder attributes  $L$  can be derived based on the equality of the length of WDV, i.e.,  $P + M + N + K = E + L$ . Then, the WDV and BV are constituted as follows:

$$\begin{aligned} c &= (x_{\xi C}, x_3, x_2, x_1), \\ l &= (P, M, N, K, O), \end{aligned} \quad (13)$$

where  $O$  is the transformed dimension computed with:

$$O = \begin{cases} 2K - u + 1, & \text{for even } O + u - 1 \\ 2K - u + 2, & \text{for odd } O + u - 1. \end{cases} \quad (14)$$

We used the Daubechies filter, which can be indicated by



**Figure 14.** A comparison of 13 attributes between 4FGL\_DR3 BL Lacs and predicted BL Lacs. The histogram of 4FGL\_DR3 BL Lacs is shown in red and the histogram of predicted BL Lacs is shown in blue; the area under the histogram integrates into one.

its length  $u$ . For example, the filter length  $u = 8$  indicates that the db4 wavelet filter was used. Under the assumption of the maximum decomposition level  $J = 3$ , we built the WDV and BV for two missions in Table 3.

3. Steps 6–7: Finally, IDWT is performed on each sample of the normalized data set, and a new data set defined on  $\mathbb{B} = \{b_1, b_2, \dots, b_i, \dots, b_O\}$  is generated.  $b_i$  is the  $i$ th new attribute.

### 3.5. A Lightweight CNN

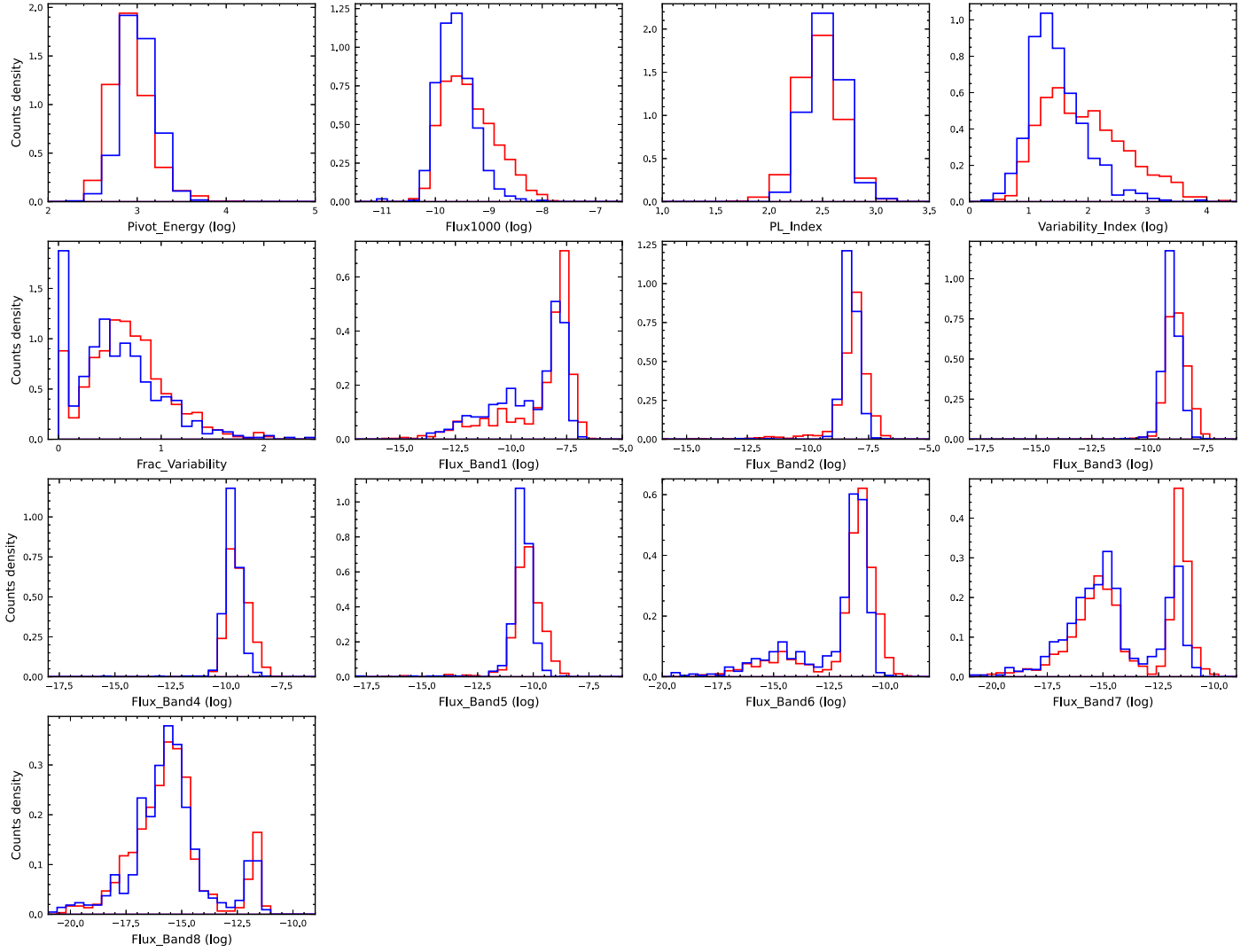
Since the new attribute space  $\mathbb{B}$  is considered to respect some natural order after FDIDWT, we believe that the convolution operation with a kernel will learn more features and achieve better performance. The kernel of CNN determines a visual field on the adjacent attributes in DL. Thus, we utilized a lightweight CNN model modified from the Matchbox net (Majumdar & Ginsburg 2020) but with a smaller size of only 94k parameters for high efficiency, which we refer to as MatchboxConv1D. Its structure is depicted in Figure 10.

## 4. Experiments and Results

With the three attribute analysis methods introduced in Section 2, we independently used four ML classifiers, i.e., RF, support vector machine (SVM), AdaBoost, and multilayer perceptron (MLP) to carry out Mission A and Mission B. The hyper-parameters are listed in Table 4. The FDIDWT method is designed to work with the MatchboxConv1D model proposed in Section 3.5 and the hyper-parameters are shown in Table 5. After training the models with the training set, fine-tuning with the validation set, and evaluating with the test set, the uncertain samples, i.e., the uncertain sources in Mission A and the BCUs in Mission B, are finally predicted with the model achieving the highest test accuracy. This flowchart is illustrated in Figure 11.

We performed the experiments on each split data set and the accuracy results were averaged over 10 test sets. In the experiments, we extensively evaluated the performance of an increasing number of the most important attributes or principal components. These attributes and components were sorted in descending order of importance and variance ratio, respectively, as shown in Figures 3, 4, and 6. The ML methods were implemented with scikit-learn (Pedregosa et al. 2011) and the





**Figure 15.** A comparison of 13 attributes between 4FGL\_DR3 FSRQs and predicted FSRQs. The histogram of 4FGL\_DR3 FSRQs is shown in red and the histogram of predicted FSRQs is shown in blue; the area under the histogram integrates to one.

training process of DNNs was accelerated by TensorFlow (Abadi et al. 2016) on NVIDIA GeForce RTX 3060. The code that was used in this experiment has been uploaded to GitHub.<sup>8</sup>

The average accuracy results of the commonly used attribute analysis methods and classifiers on test sets are compared in Tables 6 and 7 for two missions. These results are also illustrated in Figure 12. From the results, we find that the highest test accuracy is  $95.49\% \pm 1.05\%$  in mission A, which comes from the AdaBoost classifier working with the FDASE attribute selection method in the case of reducing two dimensions. In mission B, the highest test accuracy is  $91.19\% \pm 0.0\%$ . This results from the MLP classifier working with different attribute analysis methods but reduces at most one dimension.

Table 8 shows the average test accuracy results of the proposed method with different transformed dimensions (see Table 3) for two missions. We find that case #6 achieves the highest test accuracy and outperforms the results of Tables 6 and 7, while reducing five dimensions. Finally, the prediction results of uncertain sources and BCUs in two missions are listed in Table 9.

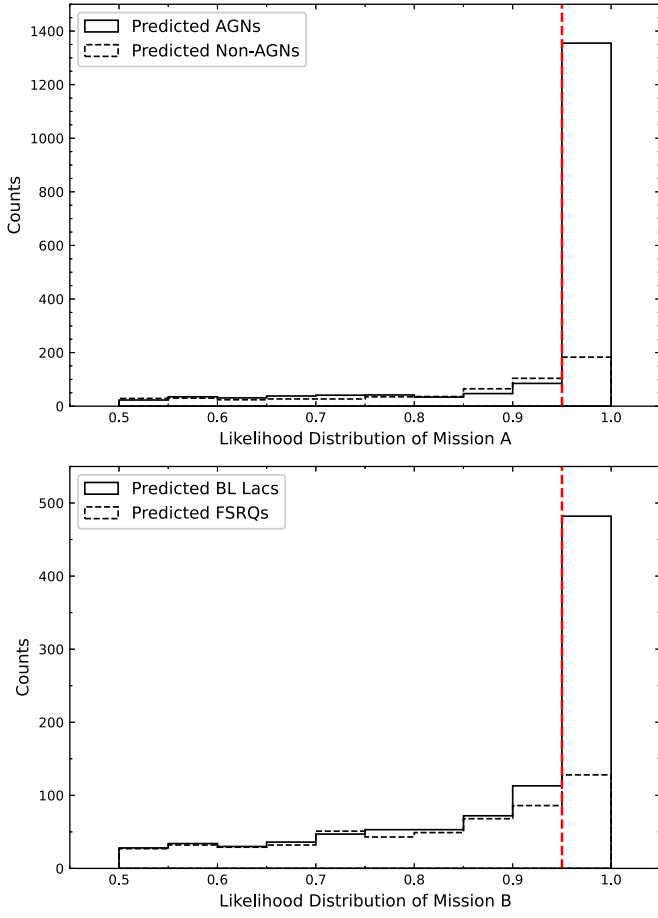
## 5. Discussion

### 5.1. The Proposed Method

Extracting correlation features among the attributes of Fermi sources is a promising research field. The commonly used attribute analysis methods, such as the estimation of attribute importance (see Section 2.2) or principal components (see Section 2.3), aim to find the significant attributes or components, and remove the unimportant ones for dimension reduction. However, the correlation features will be removed with these methods, which results in inferior classification performance (see the results of Tables 6 and 7).

The FDASE gives a global perspective of the whole data set and estimates the correlation features among all of the attributes based on FD theory. The ASC is the resulting attribute subset of the FDASE algorithm and its data contains most of the information of the data set under some correlation threshold and scale range. Nevertheless, the retention of ASC also suffers from the loss of correlation features (see the results of FDASE attribute selection in Tables 6 and 7). Thus, we propose to rearrange the original attributes to highlight correlation features at higher resolutions based on FDASE and IDWT, which is referred to as the proposed FDIDWT method.

<sup>8</sup> [https://github.com/chtfrank/astro\\_FDIDWT](https://github.com/chtfrank/astro_FDIDWT)



**Figure 16.** The likelihood probability distribution of Mission A (upper panel) and Mission B (lower panel). The AGN and non-AGN candidates are shown in the upper histogram with the solid line and the dashed line, respectively. The BL Lac and FSRQ candidates are shown in the lower histogram with the solid line and the dashed line, respectively. The dashed red lines are the set boundary of likelihood probability equals 95%.

The resulting attribute space is also considered to respect some natural order after FDIDWT. We believe that the convolution operation will extract more correlation features from the ordered attributes. Meanwhile, the structure of the CNN gives more potentiality for better classification performance. Therefore, by combining FDIDWT with the MatchboxConv1D model, we obtain the best test accuracy for both Mission A and B in case #6. In addition, the dimension is transferred to be 8, which is significant for reducing the computing burden in big astronomical data processing. It may also be concluded that most features of Data Set A and B exist at one higher resolution because case #6 corresponds to decomposition level 1 (see Table 3).

The prediction results have been carried out with an accuracy of  $96.65\% \pm 1.32\%$  and  $92.03\% \pm 2.2\%$  for Mission A and Mission B, respectively, using the proposed method. The results are listed in Table 9. A general comparison between the predicted AGNs of Mission A and the originally confirmed AGNs is shown in Figure 13. A comparison between the predicted BL Lacs of Mission B and the originally confirmed BL Lacs is shown in Figure 14. Finally, a comparison between the predicted FSRQs of Mission B and the originally confirmed FSRQs is shown in Figure 15. From the comparison, we can find that the distribution shapes of the 13 attributes of the predicted sources resemble the shapes of the original

4FGL\_DR3 sources with the corresponding classification, in general. The results indicate that our predicted sources are correctly classified.

We notice that the histograms of the attribute Variability\_Index show a longer high-variation-index tail for the original 4FGL\_DR3 sources than for the predicted sources (including predicted AGNs, predicted BL Lacs, and predicted FSRQs in Figures 13–15, respectively), and the predicted sources have more contribution on the high-variation-index head than the original 4FGL\_DR3 sources. This result suggests that our method has the advantage of finding less variable sources from the uncertain sources. We also notice that the original 4FGL\_DR3 sources give more contribution to the histogram tails than the predicted sources, while the predicted sources give more contribution to the histogram head for the attributes of multi-band intensities (Flux1000, Flux\_Band1, Flux\_Band2, Flux\_Band3, Flux\_Band4, Flux\_Band5, Flux\_Band6, Flux\_Band7, and Flux\_Band8). This suggests that our method has the advantage of finding relatively faint  $\gamma$ -ray sources from the uncertain sources. It also encourages us to believe that our method should be able to make efforts to identify less variable and faint sources in the era of survey telescopes, e.g., the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. 2009), China Space Station Telescope (CSST; Zhan 2011), etc.

## 5.2. The Classification Results of Mission A and Mission B

By employing the proposed method, we managed to predict 2291 uncertain sources into 1731 AGNs and 560 non-AGNs, and predict 1493 BCUs into 948 BL Lacs and 545 FSRQs. The likelihood probabilities of these predicted uncertain sources and BCUs are shown in Table 9 and the distribution is displayed in Figure 16.

We set a boundary of likelihood probability greater than 95%, which is shown as the dashed-red lines in Figure 16, to claim this source as a candidate of the corresponding class. In this case, we get further constrained results into 1354 AGN candidates in Mission A, 482 BL Lacs candidates and 128 FSRQ candidates in Mission B, as shown in the last column of Table 9.

Several methods have been employed to classify 4FGL BCUs as FSRQ candidates or BL Lac candidates, as Mission B of the present work, in the previous works. Kang et al. (2019) utilized three supervised ML methods (RF, SVM, and ANN) to make a classification for 1312 4FGL\_DR1 BCUs, and carried out a combined classification result of 724 BL Lac candidates and 332 FSRQ candidates. By crossing matching the results from Kang et al. (2019) and our results, we found that there are 419 overlapping BCUs in these two works. Among the overlapping 419 BCUs, 324 BCUs are predicted as BL Lac candidates and 90 BCUs are predicted as FSRQ candidates in both works, which gives our results a consistency of 98.8% with Kang et al.’s work.

Similarly, five different supervised ML algorithms (RF, LR, XGBoost, CatBoost, and neural network) were applied to the 4LAC\_DR3 BCUs in Agarwal (2023), and 610 BCUs were able to be classified as BL Lac candidates and 333 BCUs were classified as FSRQ candidates. A comparison between our work and Agarwal’s work suggests that there are 481 overlapping BCUs in the two works, 392 BCUs are classified as BL Lac candidates and 87 BCUs are classified as FSRQ candidates in both works, which gives our results a consistency of 99.6% with their work.

Fan et al. (2022) employed three physical parameters and built diagrams among them (i.e., the photon spectrum index against the photon flux diagram, the photon spectrum index against the variability index diagram, and the variability index against the photon flux) to separate the known BL Lacs from known FSRQs. They then used the boundary to divide BCUs into BL Lacs and FSRQs. In their work, 751 BCUs were classified as BL Lac candidates and 210 BCUs were classified as FSRQ candidates. There are 492 overlapping BCUs in the two works, 409 BCUs are classified as BL Lac candidates and 83 BCUs are classified as FSRQ candidates in both works, which gives our results a consistency of 100% with their work.

## 6. Conclusion

In this paper, the correlation features of attribute space of the 4FGL\_DR3 data set are highlighted by the proposed FDIDWT method and the intrinsic features hidden in data are further extracted by a lightweight MatchboxConv1D model. With the combination of the FDIDWT method and the MatchboxConv1D model, we have obtained results with an accuracy of  $96.65\% \pm 1.32\%$  for Mission A and an accuracy of  $92.03\% \pm 2.2\%$  for Mission B. For the likelihood probability boundary of 95%, we managed to classify 1354 AGN candidates in Mission A, and 482 BL Lacs candidates and 128 FSRQ candidates in Mission B. A high consistency of greater than 98% emerges by comparing our predicted candidates with those from previous works. More importantly, our method has the advantage of finding less variable and faint sources.

## Acknowledgments

H.B.X acknowledges the support from the National Natural Science Foundation of China (NSFC) under grant No.12203034, from the Shanghai Science and Technology Fund under grant No. 22YF1431500, and from the science research grants from the China Manned Space Project. Z.J.L acknowledges the support from NSFC grant 12141302, the Shanghai Science and Technology Fund under grant No. 20070502400, and from the science research grants from the China Manned Space Project. J.H.F acknowledges the support of the NSFC U2031201, NSFC 11733001, the Scientific and Technological Cooperation Projects (2020–2023) between the People's Republic of China and the Republic of Bulgaria, the science research grants from the China Manned Space Project with No. CMS-CSST-2021-A06, and the support for Astrophysics Key Subjects of Guangdong Province and Guangzhou City.

## ORCID iDs

Haitao Cao  <https://orcid.org/0000-0002-8626-8686>  
 Hubing Xiao  <https://orcid.org/0000-0001-8244-1229>  
 Junhui Fan  <https://orcid.org/0000-0002-5929-0968>

## References

- Abadi, M., Barham, P., Chen, J., et al. 2016, in 12th USENIX Symp. on Operating Systems Design and Implementation (USENIX Association), 265
- Abdo, A. A., Ackermann, M., Ajello, M., et al. 2009, *ApJS*, **183**, 46
- Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010, *ApJS*, **188**, 405
- 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
- Agarwal, A. 2023, *ApJ*, **946**, 109
- Aleman, S., Beltran, J., Perez, A., & Ganzfried, S. 2019, arXiv:1802.02548
- Bakshi, B. R., & Stephanopoulos, G. 1993, *AChE*, **39**, 57
- Belussi, A., & Faloutsos, C. 1995, in Proc. 21st Int. Conf. on Very Large Data Bases, ed. P. M. D. Dayal et al. (San Mateo, CA: Morgan Kaufmann Publishers), 299
- Belussi, A., & Faloutsos, C. 1998, *ACM Trans. Inf. Syst.*, **16**, 161
- Bengio, Y. 2009, *Found. Trends Mach. Learn.*, **2**, 1
- Blandford, R. D., & Payne, D. G. 1982, *MNRAS*, **199**, 883
- Blandford, R. D., & Znajek, R. L. 1977, *MNRAS*, **179**, 433
- Breiman, L. 2001, *Mach. Learn.*, **45**, 5
- Brown, T. B., Mann, B., Ryder, N., et al. 2020, arXiv:2005.14165
- Chen, X., Li, L.-J., Li, F.-F., & Gupta, A. 2018, in IEEE/CVF Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 7239
- Chiaro, G., Salvetti, D., La Mura, G., et al. 2016, *MNRAS*, **462**, 3180
- de Sousa, E. P. M., Traina, C., Jr., Traina, A. J. M., Wu, L., & Faloutsos, C. 2007, *Data Min. Knowl. Discovery*, **14**, 367
- 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
- Fan, J. H., Yang, J. H., Liu, Y., et al. 2016, *ApJS*, **226**, 20
- Géron, A. 2017, *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (Sebastopol, CA: O'Reilly Media, Inc.)
- Glorot, X., Bordes, A., & Bengio, Y. 2011, in Proc. 14th Int. Conf. on Artificial Intelligence and Statistics (PMLR), 315
- Gupta, A. C., Agarwal, A., Bhagwan, J., et al. 2016, *MNRAS*, **458**, 1127
- He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conf. on Computer Vision and Pattern Recognition (Piscataway, NJ: IEEE), 770
- Jifara, W., Jiang, F., Rho, S., Cheng, M., & Liu, S. 2019, *J. Supercomput.*, **75**, 704
- Jolliffe, I. T., & Cadima, J. 2016, *RSPTA*, **374**, 20150202
- Jordan, M. I., & Mitchell, T. M. 2015, *Sci*, **349**, 255
- Kang, S.-J., Li, E., Ou, W., et al. 2019, *ApJ*, **887**, 134
- Kellermann, K. I., Sramek, R., Schmidt, M., Shaffer, D. B., & Green, R. 1989, *AJ*, **98**, 1195
- Lam, M. W., Chen, X., & Hu, S. 2019, in 2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (Piscataway, NJ: IEEE), 7235
- Liu, W., Wang, Z., Liu, X., et al. 2017, *Neurocomputing*, **234**, 11
- LST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
- Lynden-Bell, D. 1969, *Natur*, **223**, 690
- Majumdar, S., & Ginsburg, B. 2020, arXiv:2004.08531
- Mallat, S. G. 1989, *ITPAM*, **11**, 674
- Nolan, P. L., Abdo, A. A., Ackermann, M., et al. 2012, *ApJS*, **199**, 31
- Paggi, A., Cavaliere, A., Vittorini, V., D'Ammando, F., & Tavani, M. 2011, *ApJ*, **736**, 128
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Rajmic, P., & Prusa, Z. 2014, *Int. J. Wavelets, Multiresolution and Inf. Process.*, **12**, 1450001
- Saz Parkinson, P. M., Xu, H., Yu, P. L. H., et al. 2016, *ApJ*, **820**, 8
- Scarpa, R., & Falomo, R. 1997, *A&A*, **325**, 109
- Schmidt, M. 1963, *Natur*, **197**, 1040
- Simonyan, K., & Zisserman, A. 2014, arXiv:1409.1556
- Strittmatter, P. A., Hill, P., Pauliny-Toth, I. I. K., Steppe, H., & Witzel, A. 1980, *A&A*, **88**, L12
- Thompson, D. J., Bertsch, D. L., Fichtel, C. E., et al. 1993, *ApJS*, **86**, 629
- Urry, C. M., & Padovani, P. 1995, *PASP*, **107**, 803
- Villata, M., Raiteri, C. M., Balonek, T. J., et al. 2006, *A&A*, **453**, 817
- Wills, B. J., Wills, D., Breger, M., Antonucci, R. R. J., & Barvainis, R. 1992, *ApJ*, **398**, 454
- Xiao, H., Fan, J., Rando, R., Zhu, J., & Hu, L. 2020a, *AN*, **341**, 462
- Xiao, H., Fan, J., Yang, J., et al. 2019, *SCPMA*, **62**, 129811
- Xiao, H., Zhu, J., Fu, L., Zhang, S., & Fan, J. 2022a, *PASJ*, **74**, 239
- Xiao, H. B., Cao, H. T., Fan, J. H., et al. 2020b, *A&C*, **32**, 100387
- Xiao, H. B., Zhu, J. T., Fan, J. H., et al. 2022b, *MNRAS*, **517**, 4202
- Yang, J., Fan, J., Liu, Y., et al. 2023, *SCPMA*, **66**, 249511
- Yang, W.-X., Xiao, H.-B., Wang, H.-G., et al. 2022, *RAA*, **22**, 085002
- Yu, D., & Deng, L. 2010, *ISPM*, **28**, 145
- Zhan, H. 2011, *SSPMA*, **41**, 1441
- Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. 2017, *Neurocomputing*, **237**, 350
- Zyner, A., Worrall, S., & Nebot, E. 2019, *IEEE Trans. Intell. Transp. Sys.*, **21**, 1584