The Astrophysical Journal, 782:41 (7pp), 2014 February 10 doi:10.1088/0004-637X/782/1/41

SEARCH FOR GAMMA-RAY-EMITTING ACTIVE GALACTIC NUCLEI IN THE *FERMI*-LAT UNASSOCIATED SAMPLE USING MACHINE LEARNING

M. Doert[1,3] and M. Errando[2]

[1] Fakultät Physik, Technische Universität Dortmund, D-44221 Dortmund, Germany; marlene.doert@tu-dortmund.de

[2] Department of Physics and Astronomy, Barnard College, Columbia University, NY 10027, USA; errando@astro.columbia.edu

Received 2013 November 12; accepted 2013 December 19; published 2014 January 24

ABSTRACT

The second *Fermi*-LAT source catalog (2FGL) is the deepest all-sky survey available in the gamma-ray band. It contains 1873 sources, of which 576 remain unassociated. Machine-learning algorithms can be trained on the gamma-ray properties of known active galactic nuclei (AGNs) to find objects with AGN-like properties in the unassociated sample. This analysis finds 231 high-confidence AGN candidates, with increased robustness provided by intersecting two complementary algorithms. A method to estimate the performance of the classification algorithm is also presented, that takes into account the differences between associated and unassociated gamma-ray sources. Follow-up observations targeting AGN candidates, or studies of multiwavelength archival data, will reduce the number of unassociated gamma-ray sources and contribute to a more complete characterization of the population of gamma-ray emitting AGNs.

Key words: catalogs – galaxies: active – gamma rays: galaxies – methods: statistical

Online-only material: color figures, machine-readable table

1. INTRODUCTION

The identification of astrophysical MeV and GeV sources has been a long-standing question in gamma-ray astronomy, mainly due to the limited angular resolution of gamma-ray detectors. The pioneer *SAS-2* and *COS-B* satellites reported detections of 26 sources with median location error of ~1° (Hartman et al. 1979; Swanenburg et al. 1981). However, only the emission from the Crab and Vela pulsars and the quasar 3C 273 could be firmly identified. The deeper survey by EGRET reported 271 gamma-ray sources with a median location error of 0.°65, but only 101 identifications were reported (Hartman et al. 1999).

The Large Area Telescope (LAT) on board the *Fermi Gamma-ray Space Telescope* started operations in 2008. The increased effective area, reduced dead time, and use of silicon tracker technology resulted in an order of magnitude improvement in source location compared to its predecessors. The second *Fermi*-LAT source catalog (2FGL; Nolan et al. 2012) characterizes 1873 gamma-ray sources between 0.1 and 100 GeV with a median location error of 0.°07. A total of 1297 sources in the 2FGL are either identified through variability or morphology, or reliably associated with counterparts from catalogs of candidate gamma-ray-emitting source classes. The remaining 576 sources for which no counterpart was identified are left unassociated.

Several studies have addressed the nature of unidentified gamma-ray sources. Searches for counterparts have been carried out through dedicated observations of the source error boxes (e.g., Ozel et al. 1988; Crawford et al. 2006) or cross-correlations with multiwavelength catalogs (Romero et al. 1999; Sowards-Emmerd et al. 2003). Others studied latitude and flux distributions to discriminate between different source populations in the unidentified sample (Lamb & Macomb 1997; Gehrels et al. 2000), or compared the properties of unidentified sources to those of candidate source populations (Mukherjee et al. 1995; Kaaret & Cottam 1996; Merck et al. 1996). A new possibility offered by the large number of sources reported by *Fermi*-LAT is to train machine-learning algorithms on populations of known gamma-ray sources to find similar candidates in the unassociated sample. Ackermann et al. (2012) characterized the gamma-ray properties of pulsars and active galactic nuclei (AGNs) in the LAT 11 month catalog (Abdo et al. 2010b), and listed unassociated sources with similar characteristics. Mirabal et al. (2012) followed a similar approach, finding candidate classifications for unassociated 2FGL sources at high Galactic latitudes ($|b| > 10°$), while Lee et al. (2012) used a Bayesian approach to find pulsar candidates.

In this work, machine-learning algorithms are used to identify unassociated sources in the 2FGL catalog with properties similar to gamma-ray-emitting AGN. Two different learning algorithms are trained on the gamma-ray properties of the known AGN in the 2FGL catalog. Only the sources selected by both algorithms independently are considered AGN candidates, adding robustness to the classification method. In addition, a realistic way of estimating the performance of classification methods is presented that takes into account the differences between the associated and unassociated source samples. Section 2 of this paper describes the properties of the 2FGL catalog. Section 3 shows how the data was prepared and which classification algorithms were tested, while Section 4 discusses how the algorithms were optimized. The method for performance estimation is discussed in Section 5, and the final results are presented in Section 6. Finally, Section 7 summarizes the main conclusions of this study.

2. SOURCE CLASSES IN THE 2FGL CATALOG

There are fourteen classes of gamma-ray sources represented in the 2FGL catalog (Table 1). The different types of AGNs add up to 60% of the population. The rest of the catalog is distributed among unassociated sources (31%), and source classes with smaller number counts.

In this study, the classification of unassociated 2FGL sources is approached as a two-class problem, where each source is either labeled as "AGN" or "non-AGN" (see Table 1). Of the

[3] Now at Department of Physics, Columbia University, New York, NY 10027, USA.

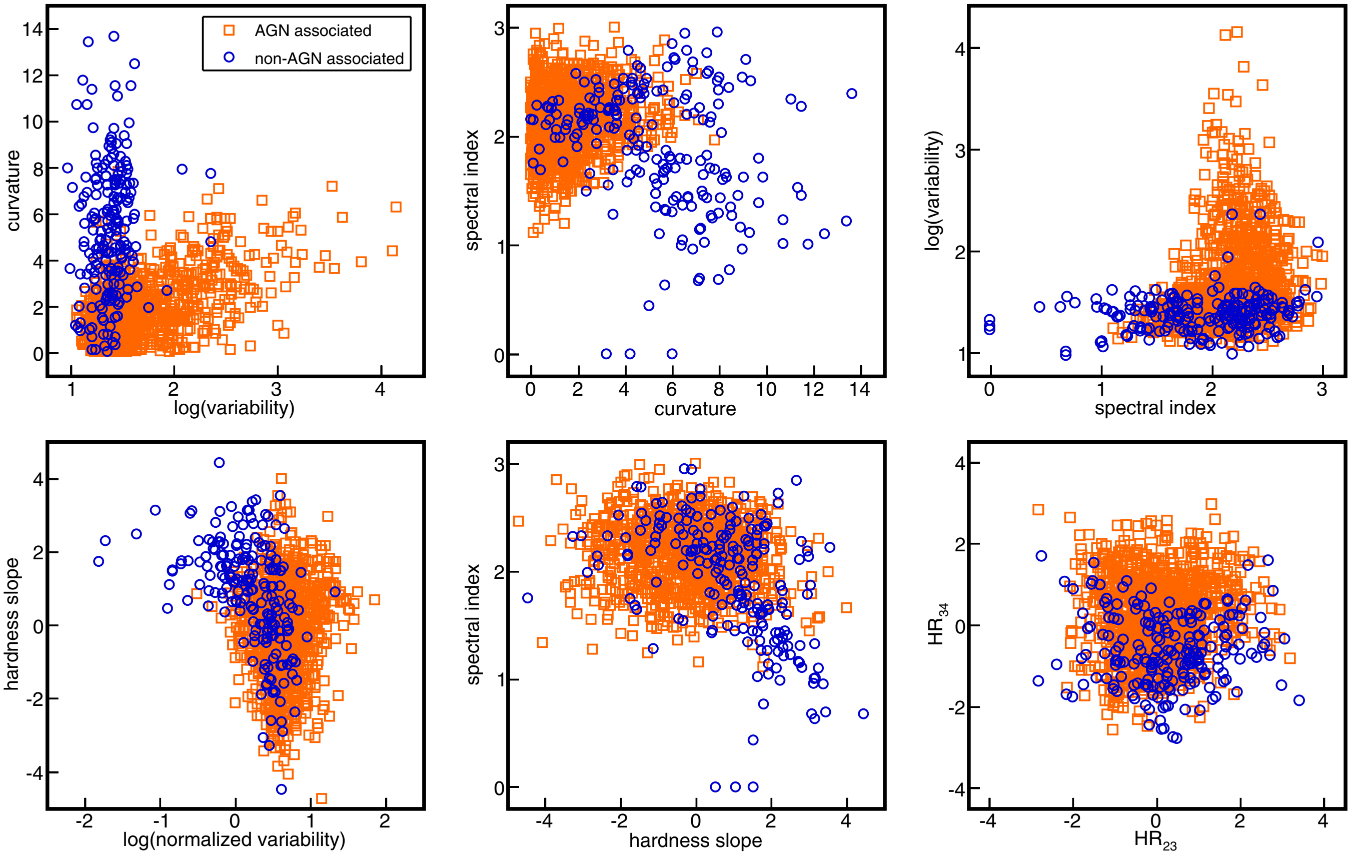

Figure 1. Scatter plots showing some gamma-ray properties listed in the 2FGL catalog (top panels) and parameters used by the machine-learning algorithms (bottom panels; defined in Section 3) for AGN and non-AGN sources.

(A color version of this figure is available in the online journal.)

Table 1
List of Source Classes in the 2FGL Catalog

Class	Description	Source count	Label
bzb	BL Lac-type blazar	436	AGN
bzq	FSRQ-type blazar	370	AGN
agu	AGN of uncertain type	257	AGN
agn	Non-blazar AGN	11	AGN
rdg	Radio galaxy	12	AGN
sey	Seyfert galaxy	6	AGN
psr	Pulsar	108	Non-AGN
glc	Globular cluster	11	Non-AGN
snr	Supernova remnant	10	Non-AGN
pwn	Pulsar wind nebula	3	Non-AGN
spp	SNR/PWN	58	Non-AGN
hmb	High-mass binary	4	Non-AGN
nov	Nova	1	Non-AGN
gal	Normal galaxy	6	Non-AGN
sbg	Starburst galaxy	4	Non-AGN
	Unassociated sources	576	

total number of associated sources, 1092 are labeled as AGNs and 205 as non-AGNs: mostly pulsars, pulsar wind nebulae and supernova remnants.

The gamma-ray properties of LAT-detected sources are discussed in detail in Nolan et al. (2012). Bright AGN exhibit significant flux variability, while pulsars show indication of spectral curvature (Ackermann et al. 2012). The top panels of Figure 1 show differences between AGNs and non-AGNs in some parameter distributions. Given these differences, machine-learning algorithms can be trained to recognize unassociated sources with AGN-like properties. Although pulsars also have distinct gamma-ray properties, they are not treated as a separate population to produce a list of pulsar candidates in this work. Detailed searches for pulsar candidates and multiwavelength counterparts have received much more attention than AGNs in the recent literature (see, e.g., Keith et al. 2011; Kerr et al. 2012; Lee et al. 2012; Abdo et al. 2013).

3. DATA PREPARATION AND CLASSIFICATION METHODS

Before starting the learning process, the sample of associated sources (1297 objects) was split into two subsamples: training (70% of the sources) and test (30%). Subsamples were selected using stratified sampling to avoid biasing the parameter distributions. The training sample was used to train the learning algorithms and optimize their performance, while the test sample was set aside to evaluate the performance of the classification methods once all the optimizations were made.

Two quantities characterize the performance of classification algorithms: recall and false-association rate. The recall is calculated in this study as the fraction of true AGNs that are correctly labeled as AGNs, and the false-association rate is defined as the fraction of non-AGN sources that are misclassified as AGNs.

In a first step, a variety of supervised machine-learning classification methods were investigated, covering random forest (Breiman 2001), support vector machines (Chang & Lin 2011), support vector networks (Cortes & Vapnik 1995), Bayesian classification (Berger 1985), logistic regression (Hosmer & Lemeshow 2000), nearest-neighbor pattern classification (Cover & Hart 1967), and multi-layer perceptrons, also known as neural networks (Rosenblatt 1962; Cybenko 1989). Algorithms were trained using the variables from Ackermann et al. (2012) and default settings (e.g., number of iterations). The performance parameters were estimated using tenfold cross-validation on the training sample, where the classifier is iteratively trained on 90% of the sample and tested on the remaining 10%, repeating the process ten times until the entire training sample has been tested.

Based on performance, random forest (RF) and neural networks (NN) were selected. The choice of two independent algorithms adds robustness to the overall classification scheme (RF & NN), which requires both RF and NN to label a source as AGN for it to be considered an AGN candidate. Combinations of three or more learning algorithms were also explored without showing any significant improvement of the performance.

The selection of RF and NN was done after a coarse test over several algorithms. It is not excluded that, after a better optimization, other algorithms could slightly improve the results presented here.

4. OPTIMIZATION OF THE LEARNING ALGORITHMS

Optimization of the RF and NN methods was done by selecting the set of parameters that optimizes the learning process, tuning the running parameters of the classification algorithms, and adjusting the confidence thresholds to select AGN candidates.

Different attributes from the 2FGL catalog were used during the learning process: spectral index, F_i (flux in the five reported energy bands), variability, curvature, and significance (square root of the test statistic value). The best separation power between the populations of AGN and non-AGN was found using spectral index and seven combinations of the abovementioned parameters (many already introduced in Ackermann et al. 2012): HR_{12}, HR_{23}, HR_{34}, HR_{45}, hardness slope, normalized variability, and normalized curvature. HR_{ij} describes the hardness ratio between the energy fluxes measured in two contiguous spectral bands:

$$\mathrm{HR}_{ij} = \frac{F_i E_i - F_j E_j}{F_i E_i + F_j E_j} \tag{1}$$

where F_i and E_i are, respectively, the flux and mean energy in the ith spectral energy band. A hardness slope parameter was also defined as

$$\text{hardness slope} = \mathrm{HR}_{23} - \mathrm{HR}_{34}, \tag{2}$$

which presents a powerful handle to separate possible AGN candidates from pulsar-like sources. Two additional parameters were also included

$$\text{normalized variability} = \frac{\text{variability}}{\text{significance}} \tag{3}$$

$$\text{normalized curvature} = \frac{\text{curvature}}{\text{significance}}. \tag{4}$$

Direct use of variables correlated with the overall flux of each source was avoided, and all parameter distributions were renormalized between 0 and 1 to minimize the influence of their very different ranges.

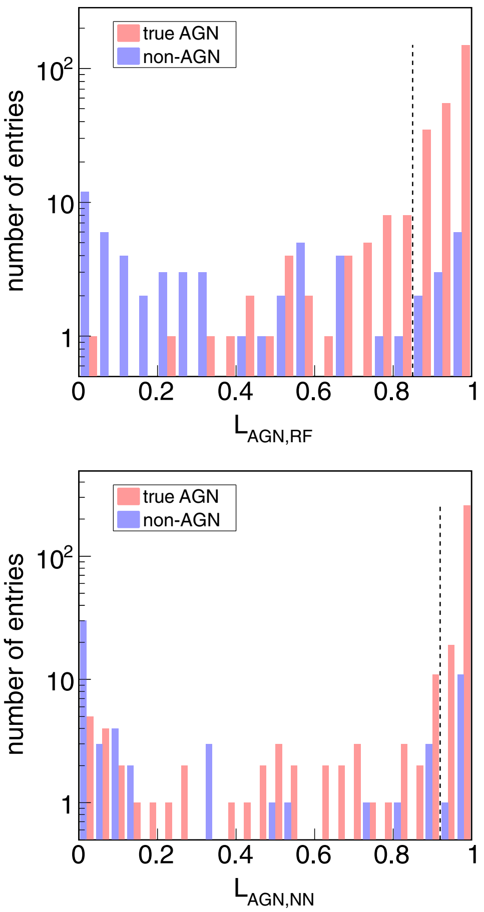

Figure 2. Distributions of the likelihood of an AGN classification for AGN and non-AGN sources in the test sample. The distributions are shown for random forest (top panel) and neural networks (bottom panel). Dashed black lines indicate the likelihood threshold of each algorithm to label a source as AGN.

(A color version of this figure is available in the online journal.)

The two selected learning algorithms have parameters that can be tuned to improve the performance of the method (see Breiman 2001; Cybenko 1989, for definitions). The RF parameters were adjusted to number of trees = 100 and depth of trees = 10. For NN, values of number of cycles = 1000, learning rate = 0.2, and momentum = 0.1 were found to be optimal.

After the learning process, RF and NN independently give a likelihood L_{AGN} of a tested source to be an AGN. Figure 2 shows likelihood distributions obtained with the RF & NN applied to the test sample. Thresholds of $L_{\mathrm{AGN,RF}} \geqslant 0.85$ and $L_{\mathrm{AGN,NN}} \geqslant 0.92$ were required for each method to label a source as an AGN. The thresholds were optimized targeting a false-association rate of ~10% for the combined classification method (RF & NN).

5. PERFORMANCE OF THE CLASSIFICATION METHOD

The performance of the classification algorithms is evaluated on the test sample, and is used to predict the completeness and number of spurious sources present in the final list of AGN candidates. However, the gamma-ray properties of associated (test) and unassociated sources differ in parameters that affect the performance of the classification methods. Figure 3 shows that unassociated sources appear to be more clustered at low

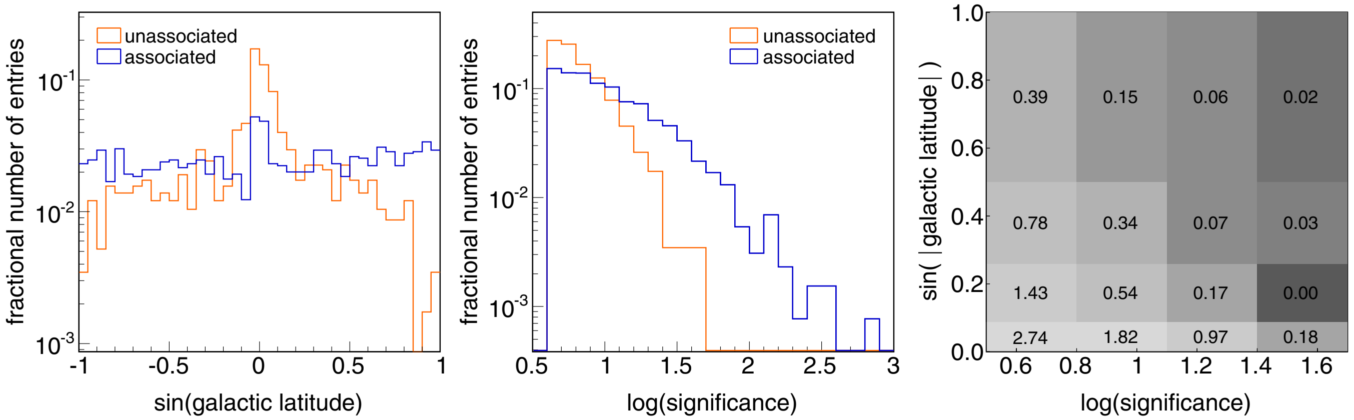

Figure 3. Galactic latitude (left panel) and significance (middle panel) distributions of associated and unassociated sources in the 2FGL catalog. Right panel: weights applied to the sources in the test sample to obtain a realistic performance estimate, as described in Section 5.

(A color version of this figure is available in the online journal.)

significances and low Galactic latitudes than sources in the associated sample. This is expected, as association probabilities are lower for sources with larger location errors, and counterpart catalogs tend to be incomplete near the Galactic plane (see Nolan et al. 2012, also discussion in Section 7).

Low-significance and low-latitude sources also challenge machine-learning classification algorithms. They are often too faint and/or influenced by the bright Galactic foreground to extract definitive information on their spectral shape and flux variability. As shown in the top panel of Figure 4, the number of sources incorrectly labeled as AGNs at low latitudes has a probability of $<1\%$ of arising from random sampling the test sample. It is also shown (Figure 4, bottom panel) that all misclassified sources have significance<12. This scenario has a chance probability of 1% if the performance of the classification algorithm would not depend on significance. These trends, and the differences between populations shown in Figure 3, imply that sources in the test sample are easier to classify (or less likely to be misclassified) than unassociated sources. Therefore, a false-association rate directly evaluated on the test sample will lead to an over-optimistic performance estimate.

To overcome this limitation, sources are binned in significance and Galactic latitude. Then, weights are calculated as $w_{ij} = N_{ij}^{\mathrm{ua}}/N_{ij}^{\mathrm{a}}$, where N_{ij}^{a} and N_{ij}^{ua} are, respectively, the number of associated and unassociated sources in the i, jth bin. The actual binning and weight values are shown in the right panel of Figure 3. The false-association rate is then estimated as

$$\text{false-association rate} = \frac{\sum\limits_{i,j} N_{ij}^{\mathrm{fa}} \cdot w_{ij}}{\sum\limits_{i,j} N_{ij}^{\mathrm{AGN}} \cdot w_{ij}} \qquad (5)$$

where N_{ij}^{fa} is the number of sources misclassified as AGNs on each bin, and N_{ij}^{AGN} the number of sources labeled as AGNs. The use of weights corrects the bias introduced by the differences between source populations, giving a realistic estimate of the false-association rate.

The performance of the classification algorithm (RF & NN), together with the individual performance of each learning method, is shown in Table 2. The algorithm is expected to recognize 80% of the AGNs present in the unassociated sample, with a false-association rate of 11%.

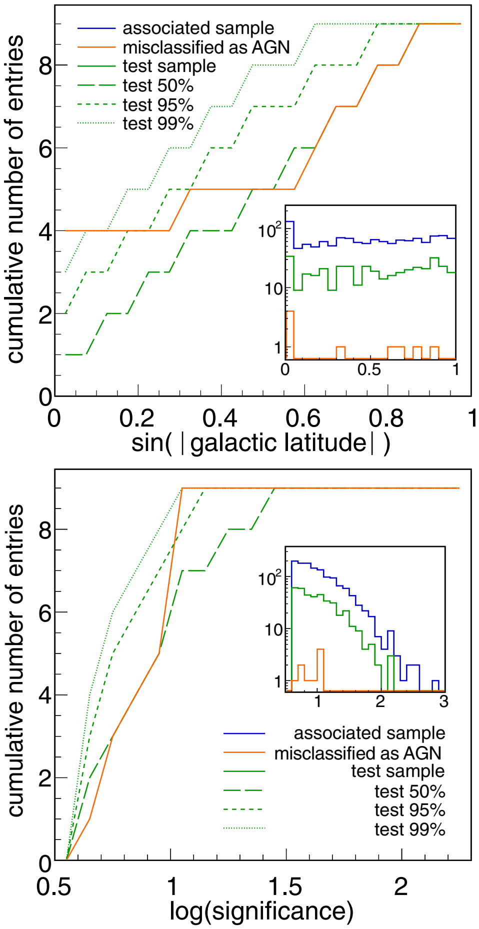

Figure 4. Cumulative distribution of sources in the test sample that were misclassified as AGNs (orange lines) with increasing Galactic latitude (top panel) and significance (bottom panel). To test the compatibility of the distributions of misclassified AGNs with a random sampling of the 389 sources on the test sample, 10000 histograms with nine entries randomly picked from the distribution of sources in the test sample were generated. The long-dashed, dashed, and dotted green lines show the cumulative distributions covering 50%, 95%, and 99% of the simulated histograms. Insets show the differential source distributions for the associated, test, and misclassified source samples.

(A color version of this figure is available in the online journal.)

Table 2
Performance of the Random Forest (RF), Neural Networks (NN), and Combined Algorithm (RF & NN) Evaluated on the Test Sample, Containing 389 Sources: 328 AGNs and 61 Non-AGNs

	AGN→AGN	Non-AGN→AGN	Recall	False-assoc. Rate
Random forest	289	12	88.1%	16.3%
Neural networks	278	12	84.7%	13.5%
RF & NN	261	9	79.6%	11.2%

Notes. Columns show the number of true AGNs correctly labeled as AGNs, non-AGNs misclassified as AGNs, and the recall and false-association rate.

Table 3
List of High-confidence AGN Candidates, Ordered by R.A.

2FGL Name	R.A. (°)	decl. (°)	L_{RF}	L_{NN}	(1)	(2)
J0004.2+2208	1.056	22.137	0.98	0.97	A	
J0014.3-0509	3.581	−5.153	1.00	1.00	A	
J0031.0+0724	7.775	7.414	0.99	1.00	A	b
J0032.7-5521	8.179	−55.356	1.00	1.00	A	
J0048.8-6347	12.218	−63.79	0.90	0.92	A	b
J0102.2+0943	15.553	9.726	1.00	1.00	A	b
J0103.8+1324	15.953	13.401	0.96	1.00	A	b
J0116.6-6153	19.174	−61.887	1.00	1.00	A	ab
J0133.4-4408	23.364	−44.142	0.99	1.00	A	ab
J0143.6-5844	25.917	−58.745	0.98	1.00	A	abc

Notes. (1) Class predicted by Mirabal et al. (2012); A: AGN, -: Uncertain. (2) Counterparts: a: infrared counterpart in Massaro et al. (2013a), b: X-ray counterpart in Paggi et al. (2013), c: AGN candidate in Acero et al. (2013).

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

6. LIST OF AGN CANDIDATES

The classification algorithm (RF & NN) was applied to the sample of unassociated 2FGL sources to produce the list of high-confidence AGN candidates shown in Table 3, which is the main result of this study. A total of 231 AGN candidates are found among the 576 unassociated sources that were studied. The sky distribution of the AGN candidates, together with the unassociated sources that were not conclusively labeled, is shown in Figure 5.

The 231 AGN candidates constitute 40% of the 2FGL unassociated sources. This is in line with estimates from similar works, that predicted 35%–55% of unassociated gamma-ray sources detected by *Fermi*-LAT to be AGNs (Ackermann et al. 2012; Ferrara et al. 2012). Out of 159 AGN candidates at $|b| \geqslant 10°$, 156 are also listed as likely AGNs according to a previous work by Mirabal et al. (2012) that focused only on 2FGL sources outside the galactic plane.

Sources potentially confused with interstellar emission (flagged with a “c” designator in the 2FGL catalog) were treated as regular sources throughout the analysis. Those constitute 22% of the unassociated sample, and have typically low detection significances. The list of 231 AGN candidates contains 22 sources with “c” designator (9%), showing that confused sources were less likely to be labeled as high-confidence AGNs, as expected for weak sources where the spectral and variability properties are less certain.

The classification algorithm finds 11 sources at $|b| \geqslant 10°$ with no significant similarities with known AGN ($L_{\mathrm{AGN,RF}} < 0.5$ & $L_{\mathrm{AGN,NN}} < 0.5$). These could potentially be interesting, as searches for dark matter annihilation or decay signals from dark subhalos target high-latitude unassociated sources with no obvious counterparts (Nieto et al. 2011; Zechlin & Horns 2012). However, all but 2FGL J0538.5-0534c are pulsar candidates (Ackermann et al. 2012; Mirabal et al. 2012; Lee et al. 2012) or have known X-ray counterparts (Takeuchi et al. 2013).

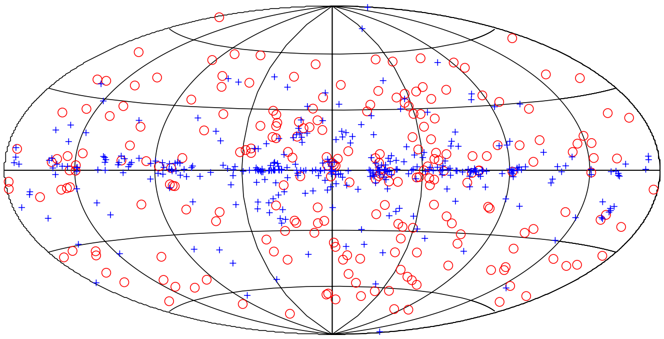

Figure 5. Sky distribution in Galactic coordinates of all unassociated 2FGL sources. AGN candidates are shown as red circles, while blue crosses indicate sources that were not labeled as AGNs. Adapted from Doert & Errando (2013).

(A color version of this figure is available in the online journal.)

7. DISCUSSION AND CONCLUSIONS

This work studied the sample of unassociated gamma-ray sources in the *Fermi*-LAT 2FGL catalog, finding 231 AGN candidates based on their gamma-ray properties. Two independent machine-learning algorithms (random forest and neural networks) were used to assess the likelihood of each source to be an AGN, and intersected to add robustness to the classification method and reduce the number of false associations.

The study includes for the first time an estimate of the false-association rate that takes into account the differences between associated and unassociated gamma-ray sources. By evaluating the performance using a test sample weighted in significance and Galactic latitude, the obtained 11% false-association rate can be considered a realistic estimate of the fraction of spurious sources present in the AGN-candidate list. Ackermann et al. (2012) obtained a lower false-association rate directly evaluated on the test sample, which is likely an optimistic performance estimate as discussed in Section 5. Similarly, Mirabal et al. (2012) used cross-validation on the training sample, which is known to give an optimistic performance, as the same sources used to optimize the classification method are used to calculate the false-association rate.

The list of AGN candidates (Table 3) covers the whole sky, studying for the first time the strip that covers the Galactic plane, where more than 50% of the unassociated 2FGL sources are located. About 210 AGNs are expected at $|b| \leqslant 10°$ extrapolating from high-latitude observations (Ackermann et al. 2011), while only 104 are listed in the 2FGL catalog. Even though low-latitude sources are harder to classify, 72 AGN candidates were found at $|b| \leqslant 10°$ (see Figure 6), which could be a significant fraction of the missing AGNs close to the Galactic plane. At $|b| \geqslant 10°$, the list of AGN candidates is in good agreement with previous work by Mirabal et al. (2012). Their study found 60 additional AGN candidates, which could be a combination of their method being more sensitive, as it was trained on a cleaner sample of high-latitude sources, and a lower confidence threshold to identify AGN candidates.

Close to the Galactic plane, AGNs are difficult to identify due to optical extinction and the bright foreground in radio and soft X-rays. Counterpart catalogs are usually incomplete at low

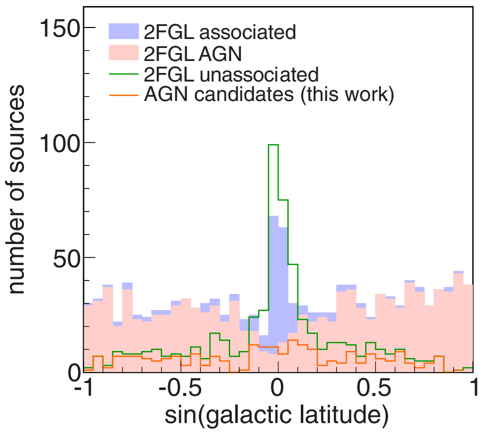

Figure 6. Galactic latitude distribution of the 2FGL sources. Shown separately are associated and unassociated sources, known AGNs, and AGN candidates identified in this work.

(A color version of this figure is available in the online journal.)

latitudes or skip the Galactic plane altogether (e.g., Healey et al. 2008; Massaro et al. 2009). Galactic absorption for gamma rays is negligible below 10 TeV (Moskalenko et al. 2006), making low-latitude AGNs detectable in the gamma-ray band but difficult to catalog at lower frequencies. In fact, numerous identifications of AGNs behind the Galactic plane have been triggered by gamma-ray detections at GeV (Mukherjee et al. 2000; Mirabal & Halpern 2009; Kara et al. 2012) and TeV energies (Abramowski et al. 2011; Archambault et al. 2013).

The list of 231 candidate AGNs presented here cannot be considered source associations, but only objects likely to be associated with an AGN. In case of gamma-ray-emitting AGNs, detectable levels of non-thermal emission in radio, optical, and X-ray frequencies are expected, and follow-up observations in those bands are needed to unambiguously identify the nature of the gamma-ray emission. Observations in the X-ray band (0.2–10 keV) have been successful in finding counterparts of unidentified gamma-ray sources (e.g., Mukherjee et al. 2000). The angular resolution of X-ray telescopes (∼0.′3 for *Swift*/XRT) can resolve individual sources inside the typical ∼5′ error box of unassociated 2FGL sources. Follow-up observations in radio and optical spectroscopy of candidate X-ray counterparts can then provide a solid AGN identification and spectral class (e.g., Halpern et al. 2001).

The *Swift* X-ray satellite has observed a good fraction of the 2FGL unassociated sources. A complete summary of these observations can be found in http://www.swift.psu.edu/unassociated/ (Stroh & Falcone 2013). So far, 135 out of the 231 candidate AGNs have at least 2 ks of *Swift*/XRT exposure. A good fraction of these have also been analyzed in Paggi et al. (2013), finding 85 sources with at least one point-like X-ray counterpart. Infrared counterparts with AGN-like spectra have been identified for 56 of the AGN candidates in the *WISE* all-sky survey (Massaro et al. 2013a). Recently, Acero et al. (2013) presented a multiwavelength study of seven unassociated sources where four AGN candidates were investigated and confirmed to have AGN-like properties in the radio and X-ray bands. These counterparts are listed in Table 3.

Counterpart catalogs of gamma-ray-emitting AGN candidates are mostly based on the population of AGNs detected by EGRET (Mukherjee et al. 1997), that contained a large number of flat-spectrum radio quasars (FSRQ; 71%) and fewer BL-Lac-type objects (27%). Similar FSRQ/BL Lac ratios are found in counterpart catalogs such as CGRaBS (84% FSRQ / 10% BL Lac; Healey et al. 2008) or BZCAT (54%/39%; Massaro et al. 2009). However, BL Lacs are more numerous than FSRQ in the 2FGL catalog (34%/40%; see Table 1). LAT-detected BL Lacs have a median radio flux density of 86 mJy (Ackermann et al. 2011), with a low-flux tail extending well below the completeness limit of the CRATES/CGRaBS catalog (65 mJy, Healey et al. 2007). The potential deficit of BL Lacs in counterpart catalogs suggests that a fraction of unassociated 2FGL sources might indeed be BL-Lac-type blazars that have not yet been cataloged. This could become relevant in searches for TeV-emitting AGN with present ground-based observatories (e.g., Massaro et al. 2013b), and prospects for future installations like CTA (Sol et al. 2013), as the harder gamma-ray spectra of BL Lacs favor their detection at TeV energies over FSRQ.

Follow-up studies on the AGN candidates presented here will reduce the number of unassociated gamma-ray sources and yield a more complete picture of the characteristics of gamma-ray-loud AGNs. Additionally, future observations could prove whether the population of gamma-ray-emitting BL Lacs extends to sources with low radio flux density. If confirmed, gamma-ray emission from BL Lacs with luminosities $\lesssim 10^{44}$ erg s^{-1} will give additional information on the low end of BL Lac luminosity function in the gamma-ray band (Ajello et al. 2013), which is a key ingredient to estimate their contribution to the isotropic diffuse gamma-ray background (Abdo et al. 2010a).

This work was supported by DFG (SFB 823, SFB 876), DAAD (PPP USA) and HAP. M.E. acknowledges support from NASA grant NNX12AJ30G. The authors thank Reshmi Mukherjee and Daniel Nieto for very useful discussions on the structure of the paper and the interpretation of the results, and Sabrina Einecke and Ann-Kristin Overkemping for proofreading the manuscript. This study used RapidMiner (Mierswa et al. 2006): an open-source data mining framework developed at Technische Universität Dortmund and maintained and distributed by Rapid-I.

REFERENCES

Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010a, PhRvL, 104, 101101
Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010b, ApJS, 188, 405
Abdo, A. A., Ajello, M., Allafort, A., et al. 2013, ApJS, 208, 17
Abramowski, A., Acero, F., Aharonian, F., et al. 2011, A&A, 529, A49
Acero, F., Donato, D., Ojha, R., et al. 2013, ApJ, 779, 133
Ackermann, M., Ajello, M., Allafort, A., et al. 2011, ApJ, 743, 171
Ackermann, M., Ajello, M., Allafort, A., et al. 2012, ApJ, 753, 83
Ajello, M., Romani, R. W., Gasparrini, D., et al. 2013, ApJ, 780, 73
Archambault, S., Arlen, T., Aune, T., et al. 2013, ApJ, 776, 69
Berger, J. O. 1985, Statistical Decision Theory and Bayesian Analysis (New York: Springer)
Breiman, L. 2001, Mach. Learn., 45, 5
Chang, C.-C., & Lin, C.-J. 2011, ACM Transactions on Intelligent Systems and Technology, 2, 27
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 273
Cover, T., & Hart, P. 1967, ITIT, 13, 21
Crawford, F., Roberts, M. S. E., Hessels, J. W. T., et al. 2006, ApJ, 652, 1499
Cybenko, G. 1989, Math. Control Signals Syst., 2, 303
Doert, M., & Errando, M. 2013, to appear in Proc. 33rd Int. Cosmic Ray Conf. (Rio de Janeiro), arXiv:1306.6529
Ferrara, E. C., Ojha, R., Monzani, M. E., & Omodei, N. 2012, in Proc. 2012 Fermi & Jansky - eConf C1111101 (arXiv:1206.2571)
Gehrels, N., Macomb, D. J., Bertsch, D. L., Thompson, D. J., & Hartman, R. C. 2000, Natur, 404, 363
Halpern, J. P., Eracleous, M., Mukherjee, R., & Gotthelf, E. V. 2001, ApJ, 551, 1016
Hartman, R. C., Bertsch, D. L., Bloom, S. D., et al. 1999, ApJS, 123, 79
Hartman, R. C., Kniffen, D. A., Thompson, D. J., et al. 1979, ApJ, 230, 597

Healey, S. E., Romani, R. W., Cotter, G., et al. 2008, ApJS, 175, 97
Healey, S. E., Romani, R. W., Taylor, G. B., et al. 2007, ApJS, 171, 61
Hosmer, D., & Lemeshow, S. 2000, Applied Logistic Regression (2nd ed; New York: Wiley)
Kaaret, P., & Cottam, J. 1996, ApJL, 462, L35
Kara, E., Errando, M., Max-Moerbeck, W., et al. 2012, ApJ, 746, 159
Keith, M. J., Johnston, S., Ray, P. S., et al. 2011, MNRAS, 414, 1292
Kerr, M., Camilo, F., Johnson, T. J., et al. 2012, ApJL, 748, L2
Lamb, R. C., & Macomb, D. J. 1997, ApJ, 488, 872
Lee, K. J., Guillemot, L., Yue, Y. L., Kramer, M., & Champion, D. J. 2012, MNRAS, 424, 2832
Massaro, E., Giommi, P., Leto, C., et al. 2009, A&A, 495, 691
Massaro, F., D'Abrusco, R., Paggi, A., et al. 2013a, ApJS, 206, 13
Massaro, F., Paggi, A., Errando, M., et al. 2013b, ApJS, 207, 16
Merck, M., Bertsch, D. L., Dingus, B. L., et al. 1996, A&AS, 120, 465
Mierswa, I., et al. 2006, in Proc. 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (New York: ACM Press), 935
Mirabal, N., Frías-Martinez, V., Hassan, T., & Frías-Martinez, E. 2012, MNRAS, 424, L64
Mirabal, N., & Halpern, J. P. 2009, ApJL, 701, L129
Moskalenko, I. V., Porter, T. A., & Strong, A. W. 2006, ApJL, 640, L155
Mukherjee, R., Bertsch, D. L., Bloom, S. D., et al. 1997, ApJ, 490, 116
Mukherjee, R., Bertsch, D. L., Dingus, B. L., et al. 1995, ApJL, 441, L61
Mukherjee, R., Gotthelf, E. V., Halpern, J., & Tavani, M. 2000, ApJ, 542, 740
Nieto, D., Aleksić, J., Barrio, J. A., et al. 2011, in Proc. 33rd Int. Cosmic Ray Conf., 5, 153
Nolan, P. L., Abdo, A. A., Ackermann, M., et al. 2012, ApJS, 199, 31
Ozel, M. E., Schlickeiser, R., Sieber, W., & Younis, S. 1988, A&A, 200, 195
Paggi, A., Massaro, F., D'Abrusco, R., et al. 2013, ApJS, 209, 9
Romero, G. E., Benaglia, P., & Torres, D. F. 1999, A&A, 348, 868
Rosenblatt, F. 1962, Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms (Washington, DC: Spartan Press)
Sol, H., Zech, A., Boisson, C., et al. 2013, APh, 43, 215
Sowards-Emmerd, D., Romani, R. W., & Michelson, P. F. 2003, ApJ, 590, 109
Stroh, M. C., & Falcone, A. D. 2013, ApJS, 207, 28
Swanenburg, B. N., Bennett, K., Bignami, G. F., et al. 1981, ApJL, 243, L69
Takeuchi, Y., Kataoka, J., Maeda, K., et al. 2013, ApJS, 208, 25
Zechlin, H.-S., & Horns, D. 2012, JCAP, 11, 050