Classifying Unidentified X-ray Sources in the Chandra Source Catalog Using a Multi-wavelength Machine Learning Approach

Hui Yang,[1] Jeremy Hare,[2,3] Oleg Kargaltsev,[1] Igor Volkov,[1] Steven Chen,[1] and Blagoy Rangelov[4]

[1]*Department of Physics, The George Washington University, 725 21st St. NW, Washington, DC 20052*
[2]*NASA Goddard Space Flight Center, Greenbelt, MD, 20771*
[3]*NASA Postdoctoral Program Fellow*
[4]*Department of Physics, Texas State University, 601 University Drive, San Marcos, TX 78666*

ABSTRACT

The rapid increase in serendipitous X-rays source detections requires the development of novel approaches to efficiently explore the nature of X-ray sources. If even a fraction of these sources could be reliably classified, it would enable population studies for various astrophysical source types on a much larger scale than currently possible. Classification of large numbers of sources from multiple classes characterized by multiple properties (features) must be done automatically and supervised machine learning (ML) seems to provide the only feasible approach. We perform classification of Chandra Source Catalog version 2 (CSCv2) sources to explore the potential of the ML approach and identify various biases, limitations, and bottlenecks that present themselves in these kinds of studies. We establish the framework and present a flexible and expandable Python pipeline, which can be used and improved by others. We also release the training dataset of 2,941 X-ray sources with confidently established classes. In addition to providing probabilistic classifications of 66,369 CSCv2 sources (21% of the entire CSCv2 catalog), we perform several narrower-focused case studies (high-mass X-ray binary candidates and X-ray sources within the extent of the H.E.S.S. TeV sources) to demonstrate some possible applications of our ML approach. We also discuss future possible modifications of the presented pipeline, which are expected to lead to substantial improvements in classification confidences.

1. INTRODUCTION

X-ray astrophysics is currently in an unprecedented era with observatories such as *Chandra*, *XMM-Newton*, and *Swift-XRT*, all viewing the sky. Data are continuously being produced in large quantities and this amount will continue to increase as more sensitive observatories begin functioning and/or reach their design sensitivity (e.g., *eROSITA* has already detected roughly one million sources in its first all-sky survey). This has resulted in a substantial growth in the number of detected X-ray sources, the vast majority of which have been detected serendipitously. Consequently, most of these sources have not been studied or classified. This implies that for every observation there are large amounts of data not being utilized to their fullest potentials. Developing and testing methods that facilitate the automated classification of these sources is important because it will enable population studies (e.g., evolution, spatial distribution) with much larger samples. It will also help find remarkable outliers, which may represent new classes of high-energy objects or rare cases of already known objects that push current models to their limits. Additionally, higher energy observatories (e.g., *Fermi*, H.E.S.S., HAWC) are also scanning the sky across many decades in energy and discovering tens of thousands of high-energy γ-ray sources. One strategy to understand the nature of these extreme

huiyang@gwmail.gwu.edu

particle accelerators is to explore the classifications of all X-ray sources in the extent of the γ-ray source to find a lower wavelength counterpart. This is often difficult as the extent of the γ-ray sources are often several arcminutes or more in size and can contain many potential X-ray counterparts.

Traditional multi-wavelength (MW) classification methods typically rely on examining various two-parameter plots in the MW parameter space (e.g., color-color diagrams, color-magnitude diagrams, X-ray hardness ratios) to conceive simple criteria (e.g., a single dividing line) for differentiating between source types (e.g., Kaplan et al. 2006; Misanovic et al. 2010). Besides being very time-consuming and tedious, the traditional classification approach does not provide any prescription on how to assign the confidence in these classifications. One solution to these problems is to develop an efficient, automated classification of a large number of astronomical sources using machine learning (ML) methods. In relation to X-ray sources, the ML approach was pioneered by McGlynn et al. (2004), who applied a supervised ML algorithm known as Oblique Decision Trees (Murthy et al. 1994) to classify $\sim$80,000 sources from the *ROSAT* all sky surveys (RASS) into six distinct classes, i.e., stars, white dwarfs, X-ray binaries (XRBs), galaxies, active galactic nuclei (AGNs), and galaxy clusters. In addition to the RASS catalog (Voges et al. 1999), the ML algorithm used optical (USNO-B; Monet et al. 2003) and radio data (SUMMS and NVSS; Murphy et al. 2007; Condon et al. 1998) to define nine source attributes (positions, X-ray fluxes, two hardness ratios, source extent, B and R magnitudes, and radio counterpart flags). However, the limited positional accuracy of *ROSAT* sources ($\sim 10''$–$30''$) resulted in a large degree of confusion with only $\sim$50% of optical associations being true counterparts in the Galactic plane.

Recently, Lo et al. (2014) and Farrell et al. (2015) used the Random Forest (RF; Breiman 2001) ML algorithm to classify 411 and 2,876 variable sources in the serendipitous *XMM-Newton* source catalogs (2XMMi-DR2; Watson et al. 2009 and 3XMM-DR4; Rosen et al. 2015) achieving an overall classification accuracy of 92%. The sources were classified into seven different classes including AGNs, cataclysmic variables (CVs), XRBs, stars, gamma-ray bursts (GRBs), super soft sources, and ultraluminous X-ray sources (ULXs), all known to be variable, using a training dataset (TD) consisting of $\sim$870 sources. The other recent large-scale study by Tranin et al. (2022), used a more interpretable ML algorithm, known as the naive Bayes classifier (Hand & Yu 2001), to classify 315,378 sources (55% of the whole 4XMM-DR10 catalog). The authors defined four classes (AGNs, stars, XRBs, and CVs), built a large ($\sim 25,000$ sources) TD, and achieved good performance with $> 90\%$ of AGNs, stars, and XRBs expected to be accurately classified, while only 34% of CVs were classified correctly. Tranin et al. (2022) also tried a RF algorithm and found that it performed equally well but they preferred the naive Bayes classifier as it provides a more straightforward interpretation of each classification. A similar (naive Bayes) approach was implemented earlier by Broos et al. (2011) in a narrower study to classify X-ray sources from the Chandra Carina Complex Project and to identify young stellar objects (YSOs) belonging to a starburst region. Zhang et al. (2021) used various ML methods to classify 4XMM-DR9 X-ray sources into three broad classes (quasars, galaxies and stars). ML methods have also been used to classify X-ray sources in a few interesting but small fields, using the *Chandra X-ray Observatory* (*CXO*) (Arnason et al. 2020) and *XMM-Newton* data (Sonbas et al. 2016; Hare et al. 2016, 2017; Klingler et al. 2020).

X-ray sources detected by the *CXO* during its first $\sim$15 years of the mission (i.e., before the end of 2014) are compiled in the Chandra Source Catalog (currently version 2.0, hereafter CSCv2; Evans et al. 2010, 2020). The CSCv2 contains $\sim$317,000 unique sources and covers $\sim$550 square degrees of the sky down to fluxes as low as $10^{-18}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ in the 0.5-7.0 keV band for the deepest fields. It also provides a large number of properties for each source (e.g., positions, fluxes in multiple bands, hardness ratios, variability properties). The sub-arcsecond *CXO* source localizations greatly reduce the degree of confusion with the MW counterparts in the crowded regions of the Galactic plane, compared to the several arcsecond localizations of *Swift-XRT* and *XMM-Newton* EPIC. To date there was no published attempt to classify all or a large fraction of CSCv2 sources.

In this paper we present the automated MUlti-Wavelength CLASSification pipeline (MUWCLASS) of the X-ray sources and a couple of its potential applications. Section 2 discusses our TD, consisting of eight source classes that are commonly detected in X-rays, as well as the X-ray and MW properties extracted for the X-ray sources. Section 3 describes the pipeline workflow, the data processing, and the choice of the algorithm. We also detail the approach we use to account for the uncertainties of various measured source properties. In Section 4 we discuss the optimization of the chosen ML algorithm and MUWCLASS's performance. We then apply the pipeline to classify 66,369 well-characterized CSCv2 sources in Section 5 and discuss the classification outcomes as well as several related consistency tests. In Section 6 we explore several high-mass X-ray binary (HMXB) candidates and a number of interesting X-ray sources within the extent of unidentified TeV sources from the H.E.S.S. Galactic plane survey. Finally, Section 7 discusses current limitations and future developments. We conclude with a summary in Section 8.

Table 1. Training Dataset Breakdown by Source Class

Source Type	Number of CSCv2 sources
Active galactic nucleus (AGN)	1390
Cataclysmic variable (CV)	44
High-mass star (HM-STAR)	118
High-mass X-ray binary (HMXB)	26
Low-mass star (LM-STAR)	207
Low-mass X-ray binary (LMXB)	65
Pulsar and isolated neutron star (NS)	87
Young stellar object (YSO)	1004
Total	2941

2. DATA

2.1. *Training Dataset (TD)*

Supervised ML relies on a TD to train (fit) a classification model. All sources in the TD must have known, confidently assigned labels (classes). To ensure the reliability of the classifications used in the TD, we only used sources from peer-reviewed publications that were confidently classified by detecting a feature (or a set of features) which is unique to a particular kind of source (e.g., the red-shift for an AGN, the pulsation period for a pulsar, the orbital period and the donor star type for an XRB, an associated star-forming region and the infrared excess for a YSO). We have compiled a TD of 2,941 literature-verified sources[1] belonging to eight classes: AGNs[2], CVs, high-mass stars (HM-STARs)[3], HMXBs, low-mass stars (LM-STARs), low-mass X-ray binaries (LMXBs)[4], pulsars and isolated neutron stars (NSs)[5], and YSOs of various kinds. Admittedly, the definitions of the classes are broad and some of the classes are rather heterogeneous (i.e., include sources with quite different properties). However, we have tried other class definitions and found these to be a reasonable compromise between having an astrophysically meaningful class and having a large enough number of confidently identified members that we could assign to each of the classes. Using more detailed classes would result in a very small number of sources in some classes, which would lead to poor performance for that class.

To construct the TD we first select catalogs for each source type and then cross-match sources from those catalogs with CSCv2 using a circular region with $3''$ radius (which we later reduce; see below). The selected catalogs are:

- AGNs from Veron Catalog of Quasars & AGN 13th Edition (Véron-Cetty & Véron 2010);

- CVs from the Cataclysmic Variables Catalog 2006 Edition (Downes et al. 2001);

- Stars from the Catalog of Stellar Spectral Classifications (Skiff 2014)[6] with O, B or W (e.g., WN, WR stars) types are labeled as HM-STARs and A, F, G, K, or M types are labeled as LM-STARs;

- LM-STARs from the APOGEE-2 data in SDSS Data Release 16 (Jönsson et al. 2020)[7];

- HM-STARs from the VIIth Catalog of Galactic Wolf-Rayet Stars (van der Hucht 2001) and its annex catalog (van der Hucht 2006);

---

[1] These are sources with well-established robust classifications based on traditional, non-ML based methods and have been studied extensively in the literature.

[2] These include quasars, AGNs and BL Lacs.

[3] These include Wolf-Rayet, O, B/Be stars.

[4] This class also includes non-accreting X-ray binaries, such as wide-orbit binaries with millisecond pulsars, as well as red-back and black widow systems.

[5] This class includes 11 magnetars.

[6] We remove faint sources with magnitudes >23, and Orion type stars to avoid mixing of LM-STARs and YSOs, and sources with ambiguous information in the SpType or Remark columns of this catalog.

[7] We also remove stars that lack reliable effective temperature or surface gravity measurements or show evidence of binarity with VSCATTER > 1 km s^{-1} and/or VSCATTER > 5 VERR_MED or are not flagged as a star based on Washington/DDO 51 photometry.

- HMXBs from the Catalog of HMXBs in the Galaxy 4th Edition (Liu et al. 2006);

- LMXBs from the Low Mass X-ray Binary Catalog (Liu et al. 2007);

- CVs and LMXBs from the Catalog of CVs, LMXBs and related objects (Seventh edition) (Ritter & Kolb 2003);

- NSs and non-accreting X-ray binaries from the ATNF Pulsar Catalog (Manchester et al. 2005);

- YSOs from multiple molecular clouds and open clusters (Megeath et al. 2012; Povich et al. 2011; Ozawa et al. 2005; Giardino et al. 2007; Rebull et al. 2011; Delgado et al. 2011);

- LMXBs, HMXBs and CVs from the INTEGRAL General Reference Catalog (IGRC) v43 where LMXBs and HMXBs have been confirmed by Sazonov et al. (2020) and Walter et al. (2015);

- HMXBs from the BeSS catalog with their SIMBAD (Wenger et al. 2000) types classified as HMXBs.

Sources from populous classes (AGNs, HM-STARs, LM-STARs and YSOs) are omitted if their class-specific catalog and X-ray combined 2-σ positional uncertainties[8] (PUs) are $> 1''$ or if the separations of the class-specific catalog and the CSCv2 coordinates exceed the 2-σ PUs. There are several cases within underpopulated classes (CVs, HMXBs, LMXBs, and NSs) where the CSCv2 positions of the sources are offset by $> 1''$ from their class-specific catalog coordinates, likely due to poor absolute astrometry, limited angular resolution of the instrument used in the class-specific catalog, or large proper motion. For some of these sources, we manually confirm the classifications/matches by reviewing the literature (besides the catalog itself) and/or by inspecting the X-ray and MW images. If the associations are deemed to be credible, we add them to our TD. Next, the CSCv2 coordinates of the sources remaining after the above-described vetting procedure are matched to SIMBAD (Wenger et al. 2000) and sources with classifications conflicting with the main SIMBAD class are omitted from the TD (unless a mistake in the SIMBAD class is obvious from looking at the original publications). Sources that are classified as candidates in the peer-reviewed publications and/or SIMBAD are also omitted. We also omit sources from our TD residing in some crowded environments such as globular clusters, the Large Magellanic Cloud (LMC), the Small Magellanic Cloud (SMC) and the Galactic center as well as sources strongly affected by complex diffuse emission around them, e.g., sources within bright pulsar wind nebulae (PWNe) or supernova remnants (SNRs). Finally, we omit stars (i.e., HM-STARs, LM-STARs and YSOs) if they have no cross-matched MW counterpart (see Section 2.3). The final content of the TD is summarized in Table 1. We note that our TD is not all-encompassing and its scope is limited by the time and efforts we could allocate for its creation. It is certainly possible to find reliable classifications for more CSCv2 sources in the published literature. We will continue updating our TD and we hope, in future, to turn it into a community driven effort with the web-based open database of classified X-ray sources.

For each source, our pipeline extracts and calculates up to a total of 29 MW features (i.e., attributes or parameters to be used by a ML algorithm) with X-ray features described in Section 2.2, and MW features described in Section 2.3. All of the features used in our TD can be found in Table 2. We provide the Python/Jupyter notebook in the GitHub reporitory[9] which has more details and can be used to reconstruct the TD from scratch. The TD is also available in the electronic (machine-readable) format and a subset of the whole TD is shown in Table 9 (see Appendix E).

We have also developed an interactive web-based plotting tool to visualize the TD's content through various two-dimensional (2D) slices of the multidimensional feature space (Yang et al. 2021)[10]. Two examples of such plots, shown in Figure 1, demonstrate a good degree of separation between sources of some classes for the choice of features shown in the plots. The top panel shows the plot of the X-ray hardness ratio (HR, see its definition in Section 2.2) vs. the X-ray to optical flux ratio, $\log(F_{\rm X}/F_{\rm O})$, which is often used in traditional classification methods. The X-ray to optical flux ratio is calculated by dividing the broadband X-ray flux $F_{\rm b}$ in the $0.5 - 7\,{\rm keV}$ energy range by the *Gaia G* band flux (see the conversion of the G band magnitude to energy flux in section 3.4). Note, however, that AGNs in the TD come from high-latitude surveys and, hence, are weakly extincted/absorbed by the intervening interstellar medium. If viewed through the Galactic plane, the AGNs would not show such a good degree of separation from other classes (e.g., YSOs).

---

[8] The X-ray positional uncertainties are approximated as circles with the radius equal to the semi-major axis of the 2-σ error ellipse in the CSCv2

[9] https://github.com/huiyang-astro/MUWCLASS_CSCv2

[10] The interactive plots are available at https://home.gwu.edu/~kargaltsev/XCLASS/

Figure 1. Examples of 2D slices of the multidimensional feature space showing the TD content (see also Yang et al. 2021). The clustering of sources belonging to certain classes (see the legend on top of the plots) makes these features a popular choice for non-ML classification. Note, however, that not all classes separate well with just these features and there is a substantial overlap between the sources from different classes.

2.2. X-ray features

After cross-matching the TD of literature-verified sources to the CSCv2 and applying a few cleaning steps (see Section 2.1), we extract the *per-observation* information from CSCv2 for all TD sources. This allows us to use sources with missing master-level[11] fluxes and to calculate an additional (to the CSCv2) inter-observation variability metric.

---

[11] In the CSCv2, master-level products combine information from multiple *CXO* observations, if they are available.

Table 2. List of MW Features Used for Classification.

Feature	Description
$F_{\rm s}$	Flux in the $0.5-1.2$ keV band
$F_{\rm m}$	Flux in the $1.2-2$ keV band
$F_{\rm h}$	Flux in the $2-7$ keV band
$F_{\rm b}$	Flux in the $0.5-7$ keV band
$\mathrm{HR_{ms}}$	Medium-Soft Hardness Ratio $(F_{\rm m} - F_{\rm s})/(F_{\rm m} + F_{\rm s})$
$\mathrm{HR_{hm}}$	Hard-Medium Hardness Ratio $(F_{\rm h} - F_{\rm m})/(F_{\rm h} + F_{\rm m})$
$\mathrm{HR_{h(ms)}}$	Combined Hardness Ratio $(F_{\rm h} - F_{\rm m} - F_{\rm s})/(F_{\rm h} + F_{\rm m} + F_{\rm s})$
$P_{\rm inter}$	inter-observation variability probability
$P_{\rm intra}$	intra-observation variability probability
G	Gaia EDR3 G band Magnitude
BP	Gaia EDR3 BP band Magnitude
RP	Gaia EDR3 RP band Magnitude
J	2MASS J band Magnitude
H	2MASS H band Magnitude
K	2MASS K band Magnitude
W1	WISE W1 Magnitude
W2	WISE W2 Magnitude
G-BP	Gmag − BPmag color
G-RP	Gmag − RPmag color
G-J	Gmag − Jmag color
G-H	Gmag − Hmag color
G-K	Gmag − Kmag color
BP-RP	BPmag − RPmag color
J-H	Jmag − Hmag color
J-K	Jmag − Kmag color
H-K	Hmag − Kmag color
W1-W2	W1mag − W2mag color
W1-W3	W1mag − W3mag color
W2-W3	W2mag − W3mag color

NOTE—The top three sections of the table are the X-ray properties based on the CSCv2 followed by MW properties from *Gaia* eDR3, 2MASS, and several WISE surveys (see text for details). W3 band magnitude is dropped after the feature selection (see Section 4.3 for details).

Detections with off-axis angles $> 10'$ are dropped to reduce the degree of confusion during MW cross-matching (see Section 2.3), since these sources have larger PUs due to the larger and asymmetric off-axis point spread function (PSF) of *CXO*. The source detections are filtered based on the per-observation source flags to avoid pileup[12], saturation, and readout streak contamination.

For each detection, the CSCv2 provides the mode ($F_{\rm mode}$), as well as the lower and upper limits at 1-σ confidence ($F_{\rm lo}$ and $F_{\rm hi}$) of the X-ray flux distributions at the soft ($0.5 - 1.2$ keV), medium ($1.2 - 2$ keV), and hard ($2 - 7$ keV) bands. We assume the flux distribution to be the Fechner distribution, also known as the split normal distribution,

---

[12] Detections with `pileup_warning` > 0.3 counts/frame/pixel are dropped.

consisting of two half-normal distributions with the same mode. We calculate the mean, and the variance of the Fechner distribution with the equations from Possolo et al. (2019):

$$\mu = F_{\text{mode}} + \sqrt{2/\pi}(F_{\text{hi}} + F_{\text{lo}} - 2F_{\text{mode}}) \tag{1a}$$

$$\sigma^2 = (1 - 2/\pi)(F_{\text{hi}} + F_{\text{lo}} - 2F_{\text{mode}})^2 + (F_{\text{hi}} - F_{\text{mode}})(F_{\text{mode}} - F_{\text{lo}}) \tag{1b}$$

for the soft, medium, and hard band, respectively. Next we calculate the mean and variance of the broadband $(0.5 - 7 \text{ keV})$ flux by combining the soft, medium, and hard bands:

$$\mu_{\text{b}} = \mu_{\text{s}} + \mu_{\text{m}} + \mu_{\text{h}} \tag{2a}$$

$$\sigma_{\text{b}}^2 = \sigma_{\text{s}}^2 + \sigma_{\text{m}}^2 + \sigma_{\text{h}}^2 \tag{2b}$$

where the subscripts b, s, m, and h indicate the broad, soft, medium, and hard band respectively.
The weighted average flux for multiple observations (whenever available) can be then expressed as

$$\overline{\mu} = \frac{\sum_{i=1}^{N}(\mu_i/\sigma_i^2)}{\sum_{i=1}^{N}(1/\sigma_i^2)} \tag{3a}$$

$$\overline{\sigma}^2 = \frac{1}{\sum_{i=1}^{N}(1/\sigma_i^2)} \tag{3b}$$

for the broad, soft, medium, and hard bands where i indexes multiple observations of the same source and N is the number of observations available. If there is only one observation for a source, then the weighted average of the mean flux and the variance of the mean will be equal to the mean and the variance of the flux distribution of that single observation. If the weighted average flux and its variance for a band are having null values, a mode of 0 and an upper limit of $10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ are used to replace them with a mean and a variance calculated from equations 1. Sources with all band fluxes having null values are dropped. From the fluxes we calculate three HRs which are HR_{ms}, HR_{hm} and $\text{HR}_{\text{h(ms)}}$ respectively with their definitions in Table 2.
We also calculate the inter-observation variability probability parameter from the cumulative probability distribution of the chi-square statistic by fitting a constant model to broadband fluxes from multiple detections of the same source:

$$\chi^2 = \sum_{i=1}^{N} \frac{(\mu_{\text{b},i} - \overline{\mu_{\text{b}}})^2}{\sigma_{\text{b},i}^2} \tag{4a}$$

$$P_{\text{inter}} = \int_0^{\chi^2} \frac{x^{\nu/2-1}\exp(-x/2)}{2^{\nu/2}\Gamma(\nu/2)}dx. \tag{4b}$$

Here $\nu = N - 1$ is the number of degrees of freedom, $\mu_{\text{b},i}$ is the per-observation broadband mean flux, $\overline{\mu_{\text{b}}}$ is the weighted average of the broadband mean flux, and $\sigma_{\text{b},i}^2$ is the variance of the broadband flux of each observation.
For the intra-observation variability parameter, we adopt the highest value of Kuiper's test probability of variability across all observations available in the CSCv2.
We also note that, with the help of an improved method for Monte-Carlo (MC) sampling for CSCv2 aperture photometry (CSC team member Rafael Martinez-Galarza, private communication), a substantial fraction of fluxes missing (i.e., having "Null" values) in the current CSCv2 release (due to the lack of convergence in the MC calculation) have been recovered.

2.3. MW Data

We cross-match the master-level CSCv2 source coordinates to the MW catalogs including *Gaia* EDR3 (Gaia Collaboration et al. 2018a), *Gaia* EDR3 Distances (Bailer-Jones et al. 2021), Two Micron All-Sky Survey (2MASS; Skrutskie et al. 2006), AllWISE (Cutri et al. 2021), CatWISE2020 (Marocco et al. 2021), and unWISE (Schlafly et al. 2019). Initially, we use a large search radius of $10''$ not to miss any possible counterparts (this also enables the estimation of local field source density, see Appendix C). If a source has more than one MW match located within the $10''$ radius, we consider only the nearest match. Next, for each remaining match (potential counterpart), we co-add (in quadrature) the PUs from the CSCv2 source and the MW counterpart[13]. Since the proper motion is not accounted for in the CSCv2, uncertainties of X-ray positions of fast moving sources with multiple observations in the CSCv2 may be underestimated. The combined 2-σ PU circle radius is used to filter out any MW matches that lie outside of it.

To search for the optical counterparts we use the *Gaia* EDR3 catalog. If the counterpart is found, we extract *Gaia*'s G, BP, and RP band magnitudes and add them to the MW features to be used in the X-ray source classification. The *Gaia* EDR3 catalog, complete down to $G = 21$, provides an all-sky coverage together with an excellent positional accuracy of around 0.5 mas at $G = 20$.

In the near infrared (NIR), the 2MASS's J, H, and K magnitudes are used. The corresponding limiting magnitudes are $J = 15.8$, $H = 15.1$, and $K = 14.3$ at 10-σ detection level, with slightly less sensitive limits (~ 1 magnitude) in the Galactic plane due to confusion between sources (Skrutskie et al. 2006).

In the infrared (IR) we use the W1, W2, and W3 bands from AllWISE, CatWISE2020 and unWISE catalogs. The 90% completeness depth is achieved at W1=17.7 and W2=17.5 for the CatWISE2020 catalog and AllWISE achieves S/N=5 with flux at 54, 71, 730 and 5000 mJy (16.9, 16.0, 11.5 and 8.0 mag) in W1, W2, W3 and W4, respectively.

We do not use the W4 band due to the shallower depth of this band along with the larger PUs (in comparison with the W1 band[14]), which could lead to the increased confusion between the IR and CSCv2 sources. For W1 and W2 band magnitudes, which are available from all three WISE catalogs, we only use the CatWISE2020 catalog when the magnitudes were missing from the AllWISE catalog and the unWISE when both AllWISE and CatWISE2020 are lacking the magnitude measurements. W3 band magnitudes are only available from the AllWISE catalog. We note that the W3 band magnitude is dropped after the feature selection while two colors involving the W3 band are selected as important features used for the classification pipeline (see Section 4.3 for details).

3. CLASSIFICATION METHODS AND PROCEDURES

3.1. *MUWCLASS Pipeline Workflow*

The MUlti-Wavelength CLASSification pipeline (MUWCLASS; see the workflow chart in Figure 2) is designed to handle simultaneously the TD and the unclassified data. This allows us to account for the measurement uncertainties of various features (see Table 2) by randomly sampling the feature values (MC sampling box in Figure 2) from the assumed probability distribution function (PDF) of each feature (see Section 3.2). In addition, for each unclassified source the direction-specific (to the source) absorption/extinction is applied to all AGNs from the TD. This is done because an AGN outside of the Galactic plane will appear very different from the same AGN viewed through the Galactic plane (see Section 3.3) and nearly all AGNs in the TD are located off the plane. MUWCLASS also calculates the derived features (including HRs and colors) after applying the reddening to AGNs, and then standardizes the data to provide a scaled distance metric across each feature (see Section 3.4). To mitigate the imbalance between different classes in the TD, an oversampling algorithm is applied (see Section 3.5). The missing data are replaced with a large negative flag value (see Section 3.6) before passing them to the Random Forest classifier (see Section 3.7).

3.2. *Monte-Carlo (MC) Sampling*

The MC method is used to account for the uncertainties of the feature measurements by repeatedly randomly sampling the feature values from their PDFs. This method has also been introduced recently in Shy et al. (2022) and we incorporate the suggested factor of $\sqrt{2}$ to account for the underestimated measurement uncertainties. This is done for all features that have their uncertainties available and both for TD sources and sources to be classified. With many samplings, we obtain a large number of classification results for each unclassified source, accounting for the uncertainties in both the TD and unclassified source's features. We then calculate the PDFs of the classification outcomes (i.e. vectors with probabilities for a source to belong to each of the pre-defined classes). Thus, the output of

---

[13] The semi-major axis is taken as the uncertainty if the PU is given as an error ellipse. It is converted to a 2-σ level under the assumption of Gaussian distribution.

[14] see Table 1 at http://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec2_5.html
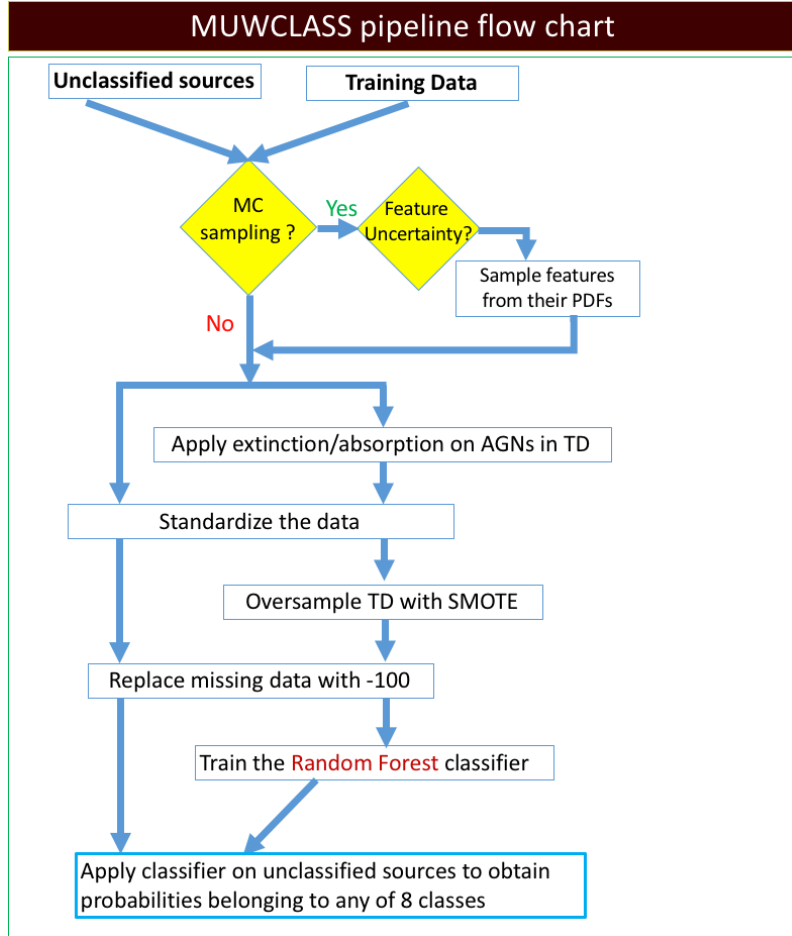
Figure 2. Workflow chart of MUWCLASS pipeline.

the classification is a PDF of probabilities of each class as opposed to a single probability obtained from traditional RF classification without accounting for uncertainties.

For each unclassified source, we run 1000 times MC samplings such that a reasonable convergence of the classification can be achieved (see Appendix B). It is extremely computational expensive, which we try to mitigate with the help of The George Washington University high-performance computing cluster (PEGASUS; Cluster 2022).

3.3. *Absorption/Extinction on AGNs*

Most ($> 99\%$) of the AGNs in our TD are located away from the Galactic plane ($|b| > 10°$), because surveys for AGNs are typically conducted outside of the plane, where the absorption and extinction are much lower than in the Galactic plane. This can potentially bias classification results because AGNs observed through the plane will look different from those in the TD, i.e., they will be dimmer, redder in optical/NIR and have harder X-ray spectra. In order to compensate for this bias, we redden all of the AGNs in our TD. The amount of reddening applied is determined by the location of the regions for which the classification is being performed. As far as we know, this step has not been performed in previous ML-based classification studies.

TD AGN X-ray fluxes are artificially absorbed using the X-ray photoelectric absorption cross-sections from Wilms et al. (2000) and the direction-specific hydrogen column density. The latter is calculated using the relation between the optical extinction and the hydrogen column density in the Galaxy $N_{\rm H}$ (cm^{-2}) $= (2.21 \pm 0.09) \times 10^{21} A_{\rm V}$ (mag) from Güver & Özel (2009) in the direction toward the unclassified source, where A_V is calculated from the SFD extinction map (Schlegel et al. 1998; Schlafly & Finkbeiner 2011), which has an angular resolution of 6.1′, using the standard

Table 3. Properties of photometric optical/NIR/IR surveys used for classification.

Band	Depth[a]	$f_{\mathrm{zp}}^{\mathrm{b}}$	$\lambda_{\mathrm{eff}}^{\mathrm{c}}$	$W_{\mathrm{eff}}^{\mathrm{d}}$
	Mag	$\mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}}$	Å	Å
G	21.5	2.5e-9	5822.39	4052.97
BP	21.5	4.08e-9	5035.75	2157.50
RP	21.0	1.27e-9	7619.96	2924.44
J	18.5	3.13e-10	12350.	1624.32
H	18.0	1.13e-10	16620.	2509.40
K	17.0	4.28e-11	21590.	2618.87
W1	18.5	8.18e-12	33526	6626.42
W2	17.5	2.42e-12	46028	10422.66
W3	14.5	6.52e-14	115608	55055.71

NOTE—[a]– corresponds to the faintest sources from the respective surveys found in our counterpart matches; [b]– zero-point spectral fluxes; [c]– the effective wavelength of the corresponding filter; [d]– the effective bandwidth of the corresponding filter. More details can be found at the VO Filter Profile Service website.

$A_{\mathrm{V}} = 3.1 \times E(B-V)$ (Fitzpatrick 1999). The absorption correction, C_{AGN}, is calculated by integrating the absorbed energy flux density within the CSCv2 energy bands:

$$C_{\mathrm{AGN}} = \frac{\int f(E) \exp(-N_{\mathrm{H}}\sigma(E))dE}{\int f(E)dE} \tag{5}$$

where the energy flux density $f(E)$ is assumed to be a power-law function $f(E) \propto E^{1-\Gamma}$ with the photon index $\Gamma = 2$, N_{H} is the estimated hydrogen column density, and $\sigma(E)$ is the photoelectric absorption cross-section. The X-ray fluxes of AGNs in the TD are multiplied by C_{AGN} corresponding to the direction toward the source to be classified prior to training the RF model (see Section 3.7). We are using N_{H} derived from the extinction maps instead of the HI maps because the latter may underestimate the absorption by not taking into account molecular hydrogen and they also have coarser angular resolution.

Similarly, the optical, NIR, and IR magnitudes are reddened using the same extinction map, and the amount of reddening is calculated at the effective wavelength of each band (see Table 3). This is achieved using the *extinction* python package[15]. In some places of the Galactic plane the absorption is very large and, if the AGN optical/NIR counterpart is faint, the reddening correction may push the AGN magnitudes beyond the survey detection limit. Therefore, we remove any magnitudes that are larger than the limiting magnitude of each survey (see Table 3).

To speed up the process when classifying many sources distributed across the sky, we split the sources up into bins within which the E(B-V) is assumed to be constant. The bin size is 25 mmag which is similar to the reddening uncertainty of the SFD extinction map Schlafly et al. (2014). Since E(B-V) ranges from 0 to ∼50, we have around 2,000 bins. We calculate the mean value of E(B-V) for all sources in each bin and apply it to redden the AGNs in the TD before classifying them.

3.4. *Pre-Processing and Standardization of the Data*

---

[15] https://github.com/kbarbary/extinction

After applying field-specific extinction/absorption to magnitudes/fluxes of AGNs from the TD, we calculate the colors and hardness ratios for both TD and unclassified sources. If the necessary magnitude is missing, the color is also considered to be missing (all calculated colors can be found in Table 2). Additionally, optical/IR magnitudes are converted to energy fluxes (in $\mathrm{erg\,s^{-1}\,cm^{-2}}$) using the conversion $F = f_{\mathrm{zp}}\exp(-mag/2.5)W_{\mathrm{eff}}$ to enable calculation of flux ratios involving the division by the broadband X-ray flux. The corresponding zero points (f_{zp}) and the effective band widths (W_{eff}) are taken from the VO Filter Profile Service[16] (Rodrigo et al. 2012) and are given in Table 3. We then divide the fluxes in all bands (except F_{b}) by the $0.5 - 7$ keV X-ray flux to help mitigate the impact of varying distances to the sources since we currently do not use in our classification. Finally, we take the base 10 logarithm of all flux quantities described in this paragraph, i.e., X-ray fluxes and optical/NIR/IR fluxes.

In order to allow our TD and unclassified data to be used with other ML algorithms (the provided Python/Jupyter notebooks allow for this flexibility) and to address the imbalance problem in our TD (see Section 3.5), we then standardize our data. The standardization is performed as:

$$X_i = \frac{x_i - \mu_i}{\sigma_i} \tag{6}$$

where x_i is the value of the ith feature for a particular source while, μ_i and σ_i are the corresponding feature's mean and standard deviation across the entire TD. The same standardization is applied to both the TD and any sources that we classify. The standardization allows the use of a distance metric for algorithms rely on clustering as a means of classification (e.g., K-Nearest Neighbors). Note that, the RF algorithm, primarily used in our study, does not rely on the distance metric.

3.5. *Imbalanced Data*

One major limitation of our TD is that it is heavily imbalanced. AGNs and YSOs substantially outnumber most other classes which can skew the performance of the classifier in favor of choosing the majority class when classifying unidentified sources. There are several ways to partly remedy this problem. The simplest one is to weigh the source classes to punish the algorithm more heavily for misclassifying sources as the most populous classes. There are two pre-defined weights in the scikit-learn package that can be used, but customized weights can also be implemented[17].

Another way to handle imbalanced data is to use oversampling. A popular implementation of this technique is the Synthetic Minority Over-sampling Technique (SMOTE; Chawla et al. 2011) written in Python[18]. This method creates a new synthetic source by choosing at random one out of 4 (a setting that can be changed) nearest neighbors from the same class for a given real source. Next, a synthetic source is created at a randomly selected point between the two real sources in the high-dimensional feature space (see Chawla et al. 2011 for details). Using this method, synthetic sources are added to all underpopulated classes until each class has as many objects as the most populous class (AGNs in our case). Figure 3 shows examples of applying SMOTE to different classes of sources in our TD for two selected 2-D slices of the multidimensional feature space. The features shown in Figure 3 have already been pre-processed and standardized as described in Section 3.4. As one can see the SMOTE procedure is not ideal as it leads to creation of artificial linear structures and also can have significant negative impact when one of the synthetic sources in the TD is an outlier with a wrong class assigned to it. Therefore, it is important to have a clean TD, for sources from classes with the smallest membership.

The SMOTE procedure has also influenced our choice of flag parameter for missing data (see Section 3.6). We set the flag value to -100 for missing data so that sources with missing data will be offset far away from those with existing data. This ensures that for any source the algorithm will not choose a neighboring source with missing data.

3.6. *Missing Data*

Our TD is fairly complete for X-ray features (all sources have flux values in at least one band, 99% in two bands, 94% in three bands) but the MW features can be missing for a large fraction of sources (optical magnitudes are missing for 24%, NIR magnitudes for 38%, IR magnitudes for 23%, and all of these magnitudes are missing for 8% of the TD sources). The data can be missing for several reasons, such as the insufficient survey depth, confusing environment (e.g., gas clouds or other diffuse emission), of the actual lack of emission in a particular band. In the latter case the

---

[16] http://svo2.cab.inta-csic.es/svo/theory/fps3/index.php?mode=voservice

[17] http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[18] https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html
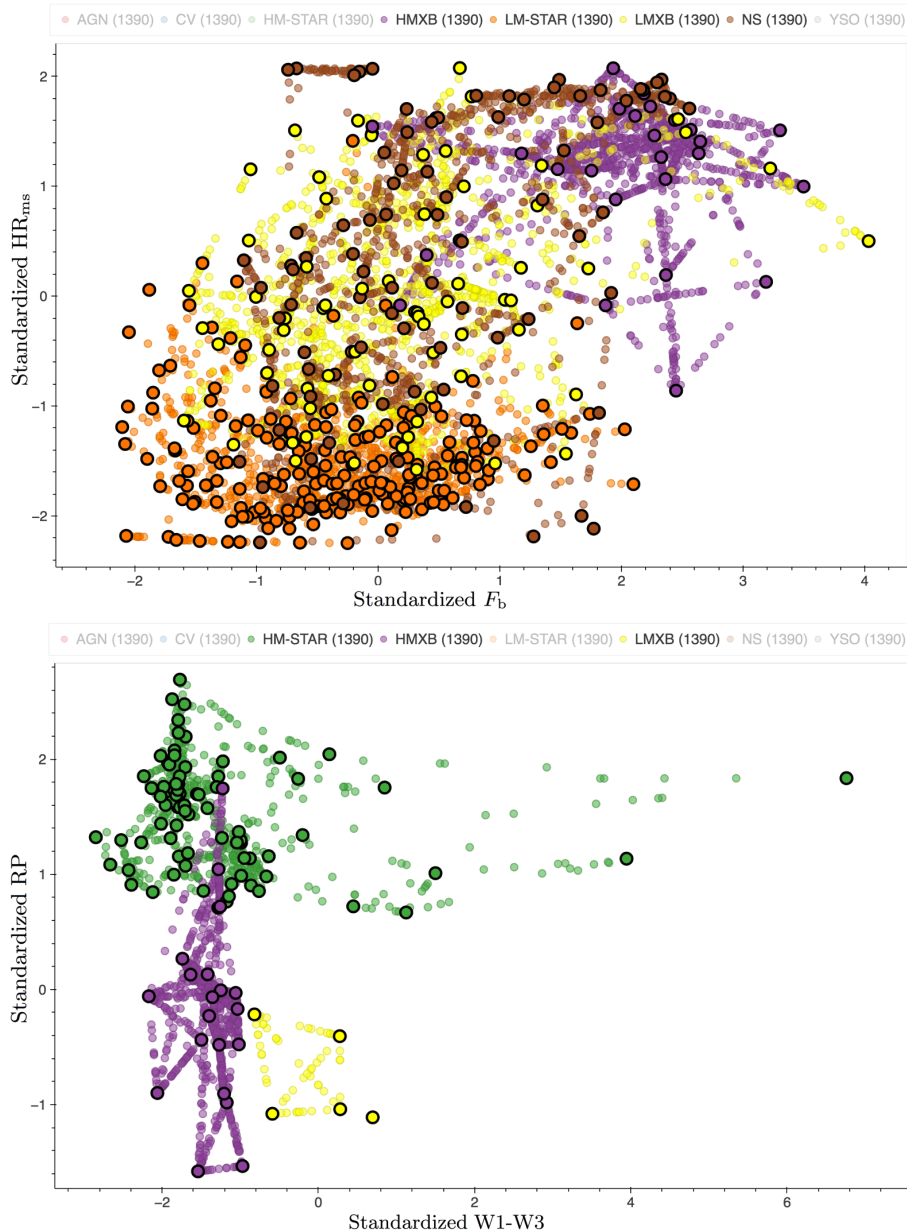
Figure 3. Examples of 2D slices of the multidimensional feature space (see Table 2) showing the TD content after applying SMOTE to create synthetic sources and balance the numbers of sources across all classes. The original (real) TD sources have black borders while the synthetic ones do not have them. Note that the SMOTE procedure results in artificial linear structures when the original (real) sources are far apart. The features have be pre-processed and standardized, as described in Section 3.4, prior to applying SMOTE.

lack of MW counterpart carries useful information (e.g., isolated NSs seen in X-rays are often identified by the lack of MW counterparts). There are multiple ways to deal with this missing data before classifying sources. One way, called imputation, replaces missing values with the mean values of each given feature from the training dataset. We disfavor this method as some sources, particularly solitary NSs, are extremely faint at optical/NIR/IR wavelengths (see e.g., Mignani 2011), and their optical magnitudes are typically much fainter than the limiting magnitudes of the surveys we use (see e.g., Shearer et al. 1997).

An alternative method is to replace all missing data with a large negative flag value. This approach is frequently used (e.g., Lo et al. 2014, Farrell et al. 2015) and it ensures that the sources with missing data will be offset far away

from those with no missing data. Hence, we adopt this approach and set all missing data in our TD to -100. In future, the use of more sensitive optical/NIR surveys should play the key role in the identification of X-ray sources from optical/NIR faint classes (e.g., isolated NSs) as long as accurate positions for X-ray sources can be obtained to combat the confusion which is expected to increase with increasing survey depth.

3.7. *Random Forest Algorithm*

Our MUWCLASS pipeline uses a supervised ensemble decision-tree algorithm, Random Forest (RF; Breiman 2001), which is implemented via the scikit-learn Python package (Pedregosa et al. 2012). In short, this classifier constructs an ensemble of decision trees from bootstrapped samples of the TD. It constructs the trees by using a randomly selected subset of features at each node, finding the optimal feature and the optimal splitting associated with this feature for this subset, and then repeating the process until all sources in a node are of the same class (at which point the node becomes a leaf). Optimal features (and splittings) are determined by minimizing the Gini impurity criterion, which is defined as:

$$I_{\mathrm{Gini}} = \sum_{\mathrm{class}} f_{\mathrm{class}}(1 - f_{\mathrm{class}})$$

where f_{class} is the fraction of sources belonging to a specific class which is separated by the selected feature and splitting. Unclassified sources are then fed through this ensemble of trees where each tree votes on the classification of these objects. RF also provides classification probabilities by counting the votes from each decision tree in the ensemble. While algorithms using a single decision tree (e.g., C4.5; Quinlan 1993) can be prone to overfitting, an ensemble of decision trees is more resistant to overfitting (Breiman 2001).

The RF algorithm can be easily replaced in our pipeline with some other algorithms from the scikit-learn Python package. This allows one to explore the whole suite of ML algorithms provided by the scikit-learn algorithm library. Also, since our data are already standardized (see Section 3.4), algorithms that rely on distance metrics (e.g., clustering-based algorithms such as K-Nearest Neighbors) can be readily used.

4. OPTIMIZATION AND PERFORMANCE EVALUATION

An important part of the ML approach to classification is the evaluation of how well the trained model (classifier) performs on data with known labels (classes) that were not used during the training. Below we describe several performance checks and explore the dependencies on various choices of parameters for the scikit-learn RF algorithm and MUWCLASS pipeline.

4.1. *Cross-Validation Method*

For validation we take a subset of the original (i.e. prior to SMOTE) TD to train the classifier and predict the labels (classes) for the rest of the TD sources. Then the predicted labels and the true labels are compared.

To validate the performance of our pipeline we use the leave-one-out cross-validation (LOOCV) procedure. At each iteration of the procedure, we remove one source, which is considered as the validation set, use the remaining (N-1) sources as the TD, and predict the classification of the left-out source. We iterate this procedure for all of the sources in the TD. This cross-validation procedure is an extreme case of k-fold cross-validation where $k = N$. It is computationally expensive to perform but it provides a reliable and unbiased estimate of model performance. Another reason we use LOOCV is that the numbers of sources for some classes are quite small (e.g., 26 for HMXBs) and, hence, setting aside a sizable fraction of these sources will substantially degrade the training of the classifier and the performance of the pipeline while the cross-validation will no longer reflect the true performance of the TD (with all of the sources included).

4.2. *Hyperparameter Tuning*

During the cross-validation process, we tune the hyperparameters[19] of the RF algorithm which could potentially improve the trained model performance. The RF algorithm has three main tunable hyperparameters, namely, the total number of trees (n_estimators), the maximum number of levels in each decision tree (max_depth), and the maximum number of features used at each node in the tree (max_features). We evaluate the dependence of the pipeline perfor-mance on the values of these hyperparameters using LOOCV on the TD while varying one hyperparameter and fixing

---

[19] see https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html for details.

the others at their default settings. We find that at the default settings of the hyperparameters, i.e., n_estimators = 100, max_depth=None[20], and max_features = $\sqrt{(\text{n\_features})}$[21], the pipeline performs well achieving an overall Accuracy of 88.6%, a Balanced Accuracy of 70.2%, a Macro F-1 Score of 68.1% and a Matthews Correlation Coefficient (MCC) of 82.9% (see Appendix A for definitions). The run time increases linearly with n_estimators while the performance remains nearly unchanged (within ∼0.5%) for n_estimators≥50. The run time and the performance are insensitive to the other hyperparameters around their default settings. Therefore, we choose to keep the hyperparameters at their default values.

4.3. *Feature Selection*

Feature selection is the process of keeping the most important features while dropping the less important ones to reduce the number of total input variables (features) used to fit (train) the model (classifier). It is desirable to reduce the number of the features to both reduce the computational costs and to avoid redundant (correlated) information.

To select the most important features, we train the RF classifier with all 54 features including 4 X-ray fluxes, 3 X-ray hardness ratios, 2 X-ray variability features, 3 optical band magnitudes, 3 NIR band magnitudes, 3 IR band magnitudes and 36 colors[22]. We run LOOCV on the TD 1,000 times with MC sampling of the feature uncertainties and calculate the feature importance by accumulating the impurity decrease within each tree for each classification. The importance values of all features are shown in Figure 4 where we calculate the mean and the standard deviation of each feature importance from the 1,000 runs. We also add a random feature, sampled from a uniform distribution, for which we find the importance to be 0.40±0.05%. Since any random feature carries no information, those real features that have similar (or smaller) importance values are regarded as completely non-informative. Thus, we conservatively use a threshold of 1% (2.5 times larger than the random feature) which leaves us with 29 features that are listed in Table 2. Unsurprisingly, all X-ray features are relatively important with importance >2%. This is primarily due to the fact that, by construction, all TD sources have at least one measured X-ray flux value, while MW features are missing for some fraction of the TD sources. The most important feature is HR_{ms} with ∼8% importance. There are other features that are also quite important, such as H and K magnitudes. Although the W3 magnitude by itself is relatively unimportant (dropped with our importance cut), the colors involving $W3$ mag ($W1 - W3$ and $W2 - W3$) turn out to be more informative so we keep those two colors.

4.4. *Algorithms Comparison*

We compare a number of supervised ML algorithms using the Receiver Operating Characteristic (ROC) curves (see Figure 5) and a number of other metrics (see Table 4 and Appendix A for the definitions of the performance metrics). Specifically, we compare the Random Forest (RF) classifier, the Gradient Boosting (GB) classifier[23], the K-Nearest Neighbors (KNN, with k=5) classifier, the Support Vector Classifier (SVC, with a radial basis function kernel and an auto Kernel coefficient 'auto' $\gamma= 1$ / n_features), the Bagging classifier, the Decision Tree classifier, and the ExtraTrees classifier. We choose to use a "Macro" averaged ROC curve (see Appendix A), which treats each class with the same weight, to measure the success of the classifier for underpopulated classes. The area under the curve (AUC) values (see Appendix A) for each algorithm are listed in the legend of Figure 5. The larger is the AUC value the better is the algorithm performance. Other metrics given in Table 4 include the Accuracy, the Balanced Accuracy, the "macro" F1 score, the MCC, and the average run time per LOOCV. Across the tested algorithms, the performance differences between the three ensemble-based decision tree algorithms (i.e., RF, Extratrees and GB algorithms) are marginal except that the GB algorithm is computationally expensive. Other algorithms perform worse by a few percentage points for each metric with the SVC method being the most computationally expensive. Therefore, we decide to adopt the RF algorithm because it is widely used (including the field of astronomy) and is also one of the best performing algorithms (see Figure 5). However, in our MUWCLASS pipeline any user can easily replace the RF algorithm with some other algorithm available from the scikit-learn package (including those mentioned above).

---

[20] The nodes expanded until all leaves are pure or until all leaves contain less than 2 samples.

[21] max_features = 6 for n_features=29 in our case.

[22] We note that the optical/NIR/IR band magnitudes are converted to energy fluxes, and all fluxes (including three X-ray fluxes at soft, medium, and hard bands and optical/NIR/IR band fluxes, except for the broadband X-ray fluxes) are then divided by the broadband fluxes, before taking the base 10 logarithm (see Section 3.4)

[23] Gradient boosting is an ensemble algorithm that fits boosted decision trees by minimizing a differentiable loss function. See https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html for details.

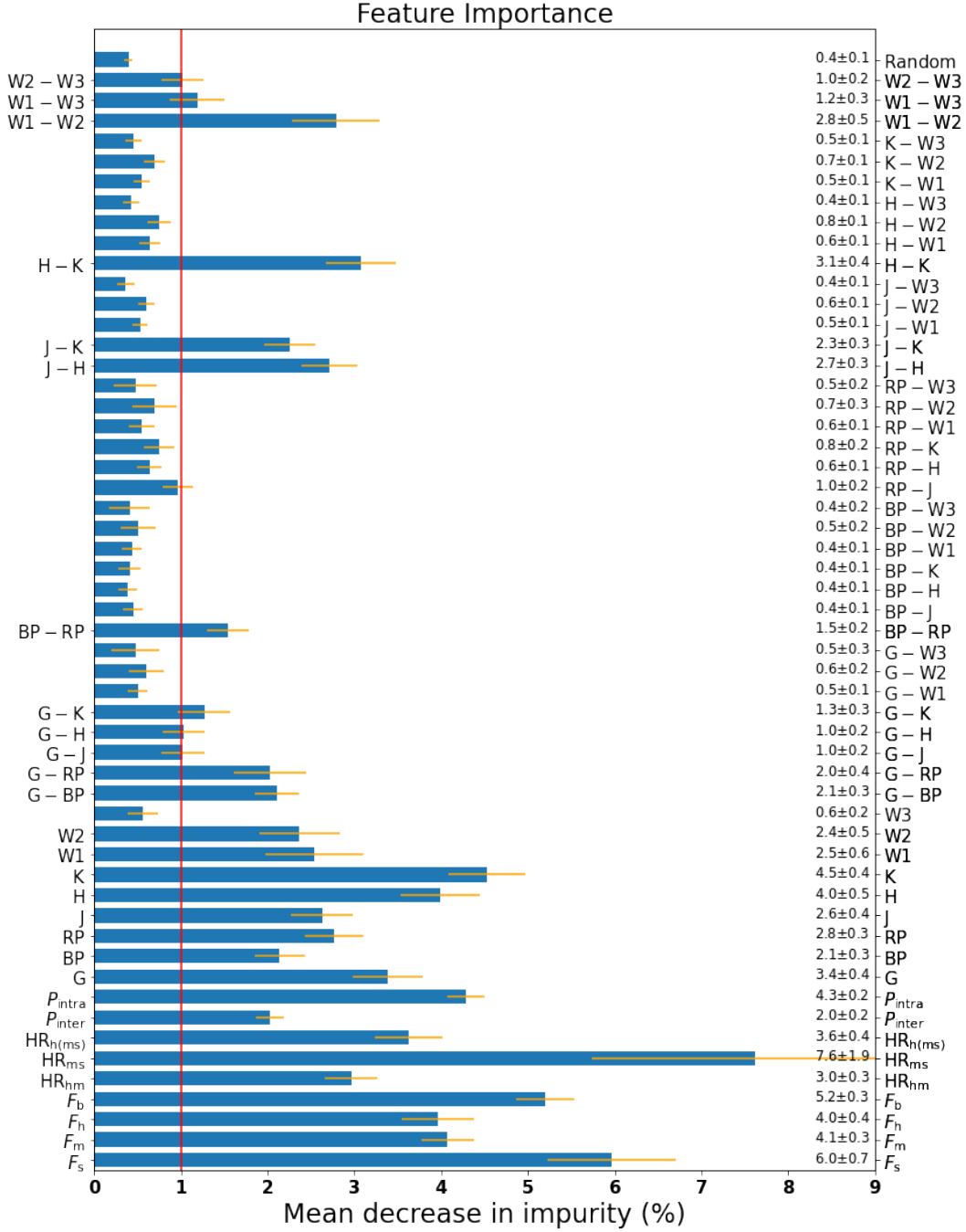**Figure 4.** Feature importance (blue bars) and their 1-σ uncertainties (orange errorbars) for all considered features with their names specified next to the vertical axis on the right. The values and their uncertainties are also shown next to the feature names. The subset of features used in the ML classification (see Table 2) has the importance exceeding that shown by the red line (1%) with their names shown next to the vertical axis on the left. We also added a dummy (random) feature and evaluated its importance (the top feature).

4.5. *The Final Pipeline Performance Evaluation*

We evaluate the pipeline performance by running LOOCV after performing the hyperparameter tuning and feature selection. We run 1,000 MC samplings for each TD source to account for feature uncertainties, which equals to training the model 1000×N times, where N is the number of sources in the TD. Besides the MC sampling of feature

Table 4. ML algorithm comparison.

Algorithms	RF	ExtraTrees	GB	Bagging	SVC	KNN	Decision Tree
Accuracy	0.886	0.882	0.874	0.867	0.856	0.835	0.882
Balanced Accuracy	0.702	0.680	0.720	0.668	0.651	0.651	0.680
Macro F1 Score	0.681	0.664	0.665	0.628	0.619	0.593	0.592
MCC	0.829	0.822	0.815	0.803	0.787	0.760	0.766
Run Time for LOOCV (s)	967	770	5080	841	6240	733	752
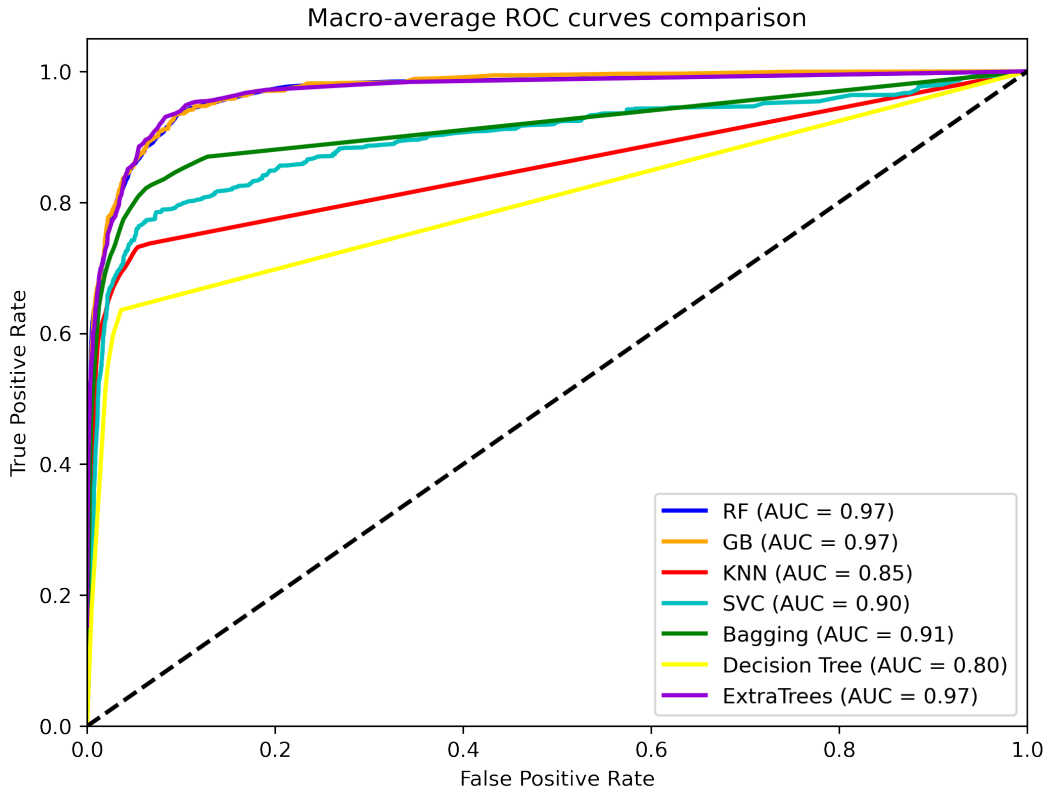
Figure 5. Macro-average ROC curves comparing different supervised machine learning algorithms: the RF classifier (blue), the GB classifier (orange), the KNN classifier (red), the SVC (cyan), the Bagging classifier (green), the Decision Trees classifier (yellow), and the ExtraTrees classifier (purple). The values listed in parentheses in the legend corresponds to the area under the ROC curve. The larger the area the better the performance of the algorithm.

uncertainties, there are other processes involving randomness. One is the SMOTE algorithm which creates synthetic sources and the other is the RF algorithm itself which randomly selects a subset of the TD for each tree and a subset of the features at each split. Each of these contributes to the variance in the classification (i.e. the classification uncertainty) but the main contribution comes from the feature measurement uncertainties (see Appendix B).

From 1,000 MC samplings we calculate the mean probability (P_{class}) for a source to belong to each of eight X-ray source classes and the standard deviation (ΔP_{class}; hereafter the classification probability uncertainty) which characterizes the width of the P_{class} distribution. The predicted class of the source is the class with the largest P_{class}. We define a classification confidence threshold (CT) as

$$\text{CT} = \min_{\text{class}}(\frac{P_{\text{predicted class}} - P_{\text{class}}}{\Delta P_{\text{predicted class}} + \Delta P_{\text{class}}}) \tag{7}$$
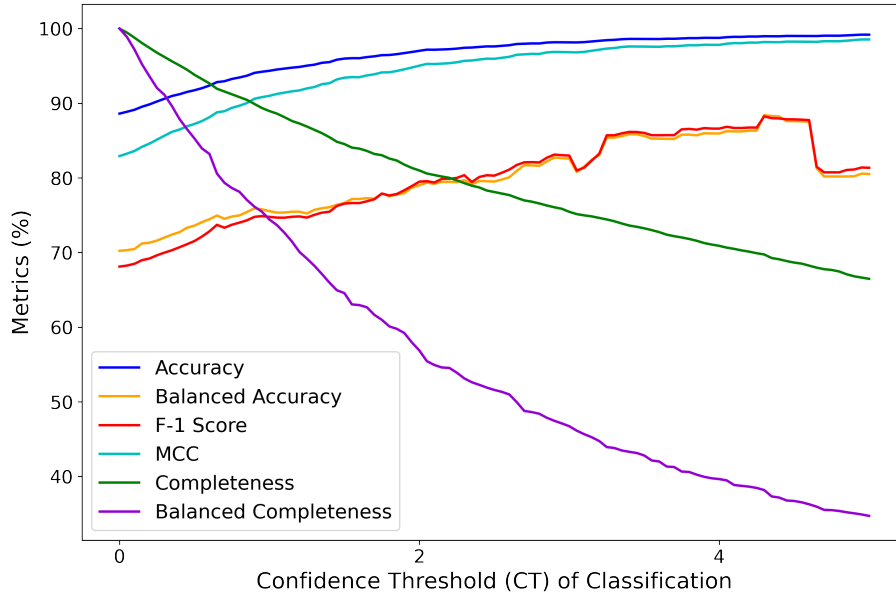
Figure 6. The metrics (including the Accuracy, the Balanced Accuracy, the F-1 Score and the MCC) as well as the Completeness and the Balanced Completeness plotted as a function classification confidence threshold. The values are computed with the TD using LOOCV method.

where class index runs through all 7 classes that are different from the the predicted class. We replace the classification uncertainty with 10^{-5} when the source is classified with zero uncertainty to avoid the zero division error.

To find the CT value which provides an optimal performance, we calculate the metrics including the Accuracy, the Balanced Accuracy, the Macro average F-1 Score and the MCC as well as the Completeness (defined as the fraction of sources that remain after the classification confidence threshold) and the Balanced Completeness (which is the average Completeness per class) as a function of CT in Figure 6. We find that metrics improve as we increase the CT while the Completeness drops. We choose CT = 2 since it achieves a good balance between Accuracy and Completeness, providing an Accuracy = 97.0%, Balanced Accuracy = 79.0%, F-1 Score= 79.5%, MCC=95.0% Completeness= 81.0% and Balanced Completeness = 56.9%. We note that there are some fluctuations of the F-1 Score and the Balanced Accuracy at higher CT values, which are caused by small number statistics in the minority classes. The users may choose a different confidence cut (e.g., choosing a different CT value or cutting $P_{\text{predicted class}}$ at a specific value). In such case, the performance of MUWCLASS should be re-evaluated using the TD provided.

For each classified source, we obtain the distribution of the classification probability for each of the eight classes. Examples of the classification probability distributions for several randomly selected sources are shown in Figure 7 for different levels of CT.

Figure 8 presents the normalized recall (upper row) and precision (lower row) confusion matrices (CMs, see Appendix A for definitions) for all sources in the TD (left panel) and confident classifications with CT $\geq$ 2 (right panel) based on 1,000 LOOCV runs. The source numbers (for each class) are shown below the class name on the vertical axis in the left panel, while the Completeness of each class after applying the confidence cut CT $\geq$ 2 is shown on the vertical axis in the right panel. The recall and precision CMs are obtained by normalizing a confusion matrix with raw counts of sources in each box (instead of percentages) by the total number of sources in a true class (recall CM), or by the total number of sources in a predicted class (precision CM). Below we summarize the performance metrics of MUWCLASS:

- The Accuracy for all classifications is 88.6% and it improves to 97.0% for confident classifications.

- The Balanced Accuracy for all classifications is 70.2% and it improves to 79.0% for confident classifications.
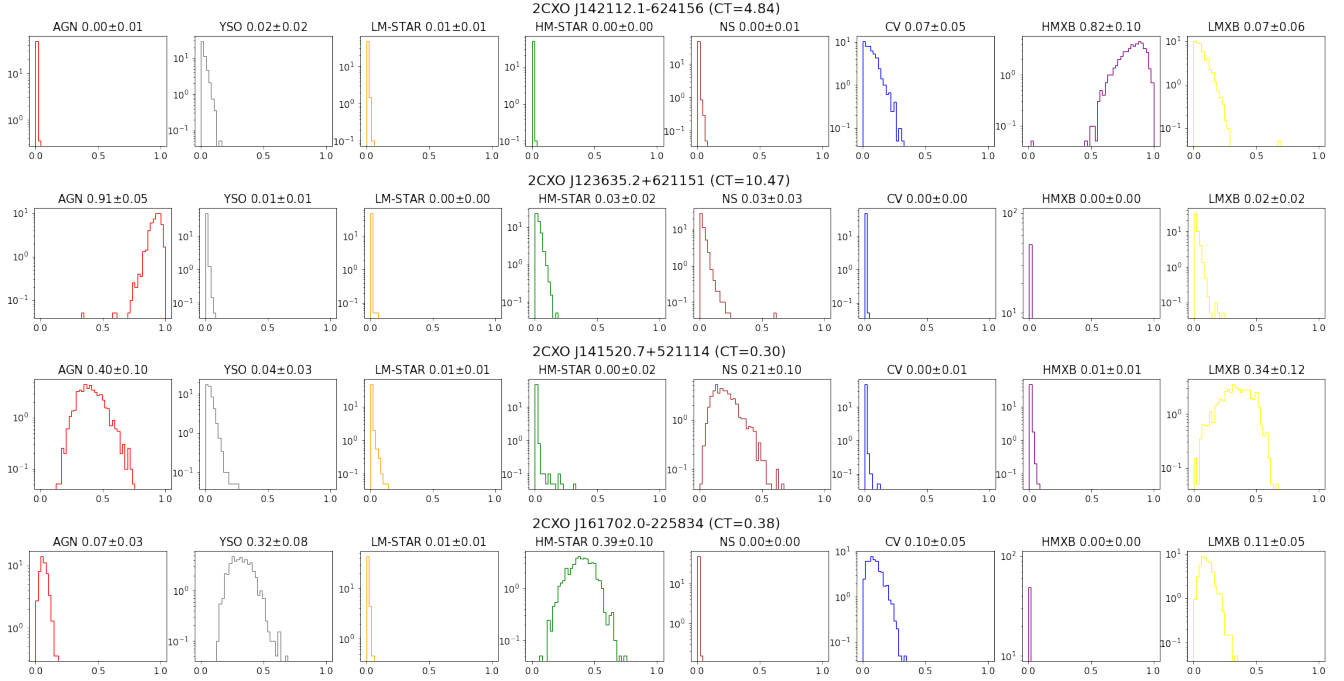
Figure 7. Examples of probability distributions per class (shown in the columns) for a few CSCv2 sources that we classify (one source per row). The name of the source and the confidence threshold are shown at the top of each row and the class and its probability with the $1 - \sigma$ uncertainties (see Sec. 4.5 for the definition) are shown at the top of each plot in the row.

- All classes improve their performance after the confident classification filter except for the recall rate for LMXBs.

- The overall Completeness (the fraction of classified sources after the confidence cut) drops from 100% to 81.0% and the Balanced Completeness drops from 100% to 56.9% after the confidence cut.

- The Completeness of confidently classified sources is high for populous classes like AGNs (87% for true AGNs, 93% for predicted AGNs) and YSOs (88%) and relatively high for LM-STARs with 70% while for other classes it is around or below 60%.

- The performance on AGNs and YSOs is extremely good with a recall rate and precision over 98% for confident classifications.

- The performance on LM-STARs is good with a recall rate of 96% and precision of 92% for confident classifications although they are sometimes confused with YSOs and HM-STARs.

- CVs, HM-STARs, HMXBs, and NSs perform reasonably well, with a recall rate and a precision of around 80% for confident classifications.

- LMXBs (including non interacting binaries) perform the worst and they are often confused with NSs, AGNs, and CVs.

We provide both recall and precision CMs for convenience because they can serve different purposes. For example, if one is interested in the fraction of a class that were retrieved (the ratio of the number of sources that are predicted as a particular class to the total number of sources of the same class) from the classification, it can be found on the diagonal of the recall CM (e.g., 86% of true LM-STARs will be classified correctly by MUWCLASS). To estimate the fraction of accurately classified sources among all confident classifications, one can look at the Confident precision CM (e.g., if there are 100 sources that are confidently classified as CVs, only 73 of them are expected to be true CVs). However, one has to keep in mind that the actual performance of MUWCLASS can be worse than indicated by these CMs if the TD does not represent well the sample of sources to be classified (i.e., there are selection biases). In reality,

Figure 8. The left panels shows the normalized confusion matrices (CMs) of all classifications while the right panels shows the normalized CMs for the confident classifications (CT $\geq$ 2). The upper panels show the normalized recall CMs and the lower panels show the normalized precision CMs. The value within each element of the CM is the percentage of sources in a predicted (true) class, shown on the horizontal axis, that are from the true (predicted) class, shown on the vertical axis, for the normalized recall (precision) CMs. The values under the class labels along the vertical axis in the left panels are the total numbers of the sources in the corresponding classes while in the right panels these values are the fractions of the sources surviving the confidence cut (CT$\geq$ 2) for each class. The darker the color the higher the percentage is.

these biases are often present and we discuss several of them in the paper. This difference between the expected and the actual performance is usually difficult to evaluate.

5. RESULTS

In order to select well characterized point sources with reliably measured X-ray features, we define a "good" CSCv2 sample (hereafter GCS) detected with significance S/N$\geq$3, PU $<1''$. Also we ensure that GCS sources do not have CSCv2 confused and extended source flags raised (i.e. *conf_flag* and *extent_flag* in CSCv2 source flags). Such filters leave 66,369 sources in GCS (constituting 21% of all CSCv2 sources). The classifications of GCS (with a source detection threshold at S/N=3) is presented by histograms in Fig. 9, with different CT cuts. A summary table of the classification numbers for GCS sources with 2 different source detection thresholds (S/N=3 and S/N=6) per class is

Table 5. Numbers and fractions of classifications per class

	AGN	YSO	LM-STAR	HM-STAR	NS	LMXB	CV	HMXB
GCS[a]	33,955	11,268	2,216	1,843	9,159	6,777	1,104	47
Confident fraction[b]	0.677	0.554	0.358	0.078	0.061	0.036	0.054	0.149
GCS (with S/N$\geq$ 6)[c]	14,529	4,988	1,137	690	3,484	2,556	487	24
Confident fraction (for S/N$\geq$ 6)[d]	0.775	0.585	0.383	0.075	0.122	0.088	0.119	0.250
MW counterpart fraction[e]	0.663	0.999	0.997	1.000	0.062	0.293	0.895	1.000

NOTE—[a]– # of GCS sources classified as the corresponding class; [b]– fractions of confident classifications (CT$\geq$ 2) for GCS sources per class; [c]– # of GCS sources with S/N$\geq$ 6 classified as the corresponding class; [d]– fractions of confident classifications (CT$\geq$ 2) of GCS sources with S/N$\geq$ 6 per class; [e]– fractions of classified GCS sources having at least one MW counterpart per class.

also shown in Table 5, together with the fractions of classifications after the confidence cut (CT$\geq$ 2) and the fractions of GCS sources having at least one MW counterpart. The classifications of GCS are also available in the electronic (machine-readable) format with a few sources shown in Table 8 as a subset of the entire table (see Appendix E).

In addition to classifying a large number of previously unclassified sources we apply several statistical tests to the confidently classified GCS (hereafter CCGCS) of 31,046 sources with CT$\geq$ 2 to compare them with those from the TD to look for possible biases. An investigation on some individual interesting sources/fields will be discussed in Sec 6.

As one can see, we find a large number of AGNs and YSOs. As the CT is increased, the number of classified sources in these classes does not drop as much (in %) as in other classes (cf., e.g., NSs). Also, recalling that the performance evaluation based on the TD (see Figure 8) shows that nearly all (99%) of AGNs and YSOs are correctly classified, we expect the confident AGN and YSO classifications to be reliable, unless there are strong biases making AGNs and YSOs from the TD poorly represent those that are classified, or the classified sources are in the crowded environments (see Section 7.1). The results for some other classes do not look nearly as good, e.g., for NSs and LMXBs only 80% and 33% of confident classifications are expected to be true. However, even after this correction the number of identified NSs is surprisingly large, suggesting that an unaccounted bias may be affecting the classifications for this class (see Section 7.1).

From our experience with *CXO* data and CSCv2, most of the S/N= 3 detections are very solid. However, it is still interesting to see what happens when the detection threshold is increased, i.e., how the numbers of classified sources per class change for brighter sources. Comparison of the first row and the third row in Table 5 (S/N$\geq$3 vs. S/N$\geq$6) shows that although the numbers of sources in each class drop by a factor of 2-3, there is no particular class whose classification efficiency would be disproportionately affected by the increase in the detection significance. This suggests that the bias associated with the TD being on average brighter than the GCS sources (see Section 5.2) may not be having a large impact on the classifications. Also, sources with higher detection significance are expected to have more accurate measurement for their X-ray features which should be translating to narrower distributions for their classification probabilities. Therefore, a larger fraction of classified sources would generally be expected to pass a particular CT threshold compared to that for the sources with lower detection significance. Indeed, as one can see from the second row and the fourth row in Table 5, this fraction increases for all sources but HM-STAR, with only marginal increases for LM-STAR and YSO and particularly large increases for NS, LMXB, and CV classes. This is understandable because most of sources classified as stars (including HM-STARs, LM-STARs and YSOs) have optical counterparts while for NS, LMXB, and CV classes the fraction of source with MW counterparts is much lower (see the fifth row in Table 5 and Section 5.2) and, hence, the accuracy of measurements of their X-ray features matters more.

5.1. *Comparison to catalogs with known source classes.*

We compare the classification results of CCGCS to several publicly available catalogs of classified sources (which we did not use in our TD for various reasons). To aid the comparison, we merged our LM-STAR and HM-STAR
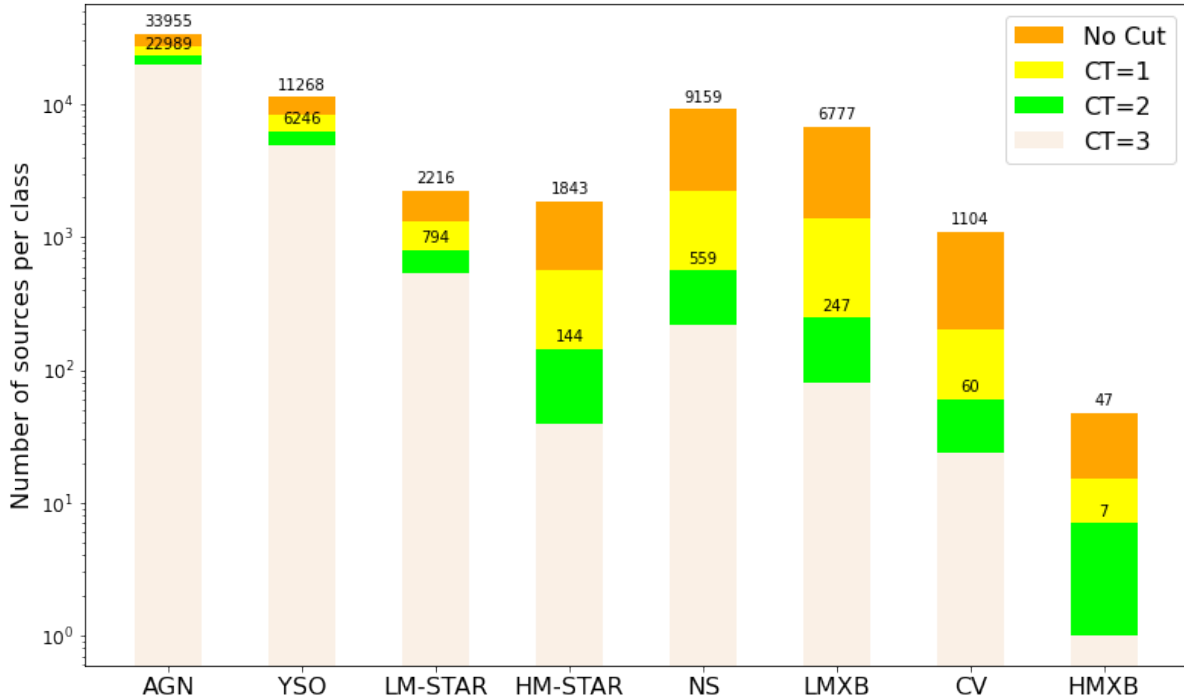
Figure 9. Summary of the classification outcomes for GCS with S/N≥3 shown as a function of classification confidence for three different values of CT. The values on top of the histogram bars show the numbers of sources for each class without (larger value) and with (smaller value) the CT=2 cut. Notice that for AGN, LM-STAR, and YSO classes (which are well represented in the TD) the relative change in the number of classified sources with the CT cut is relatively small compared to other source classes.

classes into a STAR class and LMXB and HMXB classes into an XRB class. The comparison of the classifications is summarized in Table 6 and briefly discussed below.

Cross-matching (by coordinates) our CCGCS with the FIRST-NVSS-SDSS AGN catalog (Smolčić 2009) gives 17 matches, all of which are classified by us as AGNs. A comparison to the combined WISE and Sloan Digital Sky Survey (SDSS) spectroscopic data catalog (Toba et al. 2014) resultes in 146 AGN matches and 7 YSO matches in our classification of GCS. The cross-matching to the two catalogs corresponds to the recall rates of 100% and 95.4% which are consistent with that estimated from the confusion matrix (99%) for confidently classified AGNs.

We also cross-match CCGCS to the TD of a recent large-scale ML-based classification study of 4XMM-DR10 (Tranin et al. 2022). The recall rates calculated from the cross-matching are 93.6%, 92.0%, 36.3% for AGNs, STARs, XRBs which are comparable to those estimated from the confident recall CM in Figure 8 while the recall rate for CVs is 25% with only 8 sources cross-matched. We check those sources that have discrepant classifications from our results and Tranin et al. (2022) and we find that most of them are from nearby, resolved galaxies or globular clusters, which are complicated and crowded environments where our MUWCLASS pipeline is not expected to work primarily due to the limitations of the MW surveys we are currently using (i.e., primarily confusion when cross-matching sources). There may also be some misclassificaitons in these complex environments in the TD compiled by Tranin et al. (2022). For instance, two sources (2CXO J031818.8–663230 and 2CXO J013647.4+154744) claimed to be XRBs by Tranin et al. (2022) are classified by MUWCLASS as LM-STARs and actually do appear to be foreground stars (based on their parallaxes and/or proper motions[24] in *Gaia* eDR3) which are coincident by chance with two resolved, nearby galaxies. We intentionally do not add the sources from nearby galaxies (or globular clusters) to our TD since cross-matching to MW counterparts in such crowded environments can easily result in false matches.

The SIMBAD catalog is overall accurate and has been used carefully to verify source classifications previously (e.g., Li et al. 2022; de Beurs et al. 2022). While constructing our TD, we find that some sources have wrong SIMBAD

---

[24] Note, we do not currently use this information for ML-classification with MUWCLASS and we checked it manually.

classifications (e.g., 2CXO J203213.1+412724 is a γ-ray binary with a known young pulsar, and belongs to our HMXB class (Lyne et al. 2015), but is labeled as a Be star in SIMBAD; 2CXO J112401.1–365319 is a black widow system, and belongs to our LMXB class (Gentile et al. 2014), but is labeled as a NS in SIMBAD. Although we cannot be sure that all SIMBAD classifications are accurate, we expect that most of them still are and proceed with the comparison under this assumption. By cross-matching CCGCS to the SIMBAD catalog, we find an estimated recall rate of 98.1% for AGNs, 97.3% for YSOs, 75% for NSs which are again comparable to those indicated by the confident recall CM in Figure 8. The recall for STARs is 29.7%, where the most confusion is from the YSOs (with 62.9% of STARS classified as YSOs). This is not very alarming since when YSOs are sufficiently evolved, they are not too different from STARs and the SIMBAD's definition of YSO class may differ from ours. The agreement for CVs and XRBs is rather poor. We check the CVs and XRBs that are classified differently (from SIMBAD) by MUWCLASS and find that most of them are located in globular clusters or resolved galaxies, where our classification results are unreliable due to the inability to identify the true MW counterpart in such dense environments with the MW surveys we currently use.

5.2. *X-ray Flux Distribution and MW Counterpart Fraction*

We compare the broadband X-ray flux distribution of CCGCS and the TD in Figure 10. Ideally, the distributions should closely resemble each other while the substantial difference would indicate a possible bias. We also compare the fraction of sources that have MW counterparts for each of the eight classes in both CCGCS and the TD. For an X-ray source that has at least one MW counterpart matched, the source will be added to the group of "with cp". Sources with no MW counterpart matched will be placed in the "no cp" category.

For populous classes, where a substantial fraction of sources is confidently classified (i.e., AGNs, YSOs and LM-STARs), the TD sources have systematically higher X-ray fluxes than those of CCGCS. The X-ray flux distribution of the TD sources tends to flatten below $10^{-13}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ until the detection limit, while that of CCGCS continuously grows below $10^{-13}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$. This probably reflects the fact that brighter X-ray sources are easier to study and classify using traditional approaches (e.g., spectroscopy, periodicity detection, characteristic variability patterns, etc.).

The discrepancy of the X-ray flux distributions represents a selection bias between CCGCS and the TD because the broadband flux is one of the features participating in our ML classification. However, all other fluxes are divided by the broadband flux before they are used as features which should help to mitigate the bias.

From Figure 10 one can also see that brighter sources have a higher chance of cross-matching to MW counterparts, resulting in a higher fraction of sources with MW matches for the TD. We also calculate the fractions of sources cross-matched to each MW survey at optical, NIR or IR band for CCGCS, which are 34%, 24% and 47%. This is also lower than those calculated for the TD, which are 76%, 62% and 77%. This makes sense as brighter sources are often more nearby, thus brighter at all wavelengths. This is another selection bias that may potentially skew the classification results (see discussion in Section 7.1). Confidently classified YSOs, STARs, and HMXBs from CCGCS all have counterparts, which is not surprising since the TD sources of these classes are required to be bright enough in the optical/IR bands so that the stellar spectral classification can be applied. On the other hand, since CCGCS sources are systematically fainter, we would expect that a fraction of them should be missing MW counterparts entirely as they are too faint to be detected by the surveys we are currently using. This aspect is currently not learned by our MUWCLASS pipeline, thus leading to a potential classification bias.

5.3. *Galactic Latitude Distribution*

The Galactic latitude parameter b has been used for classifications in several previous ML classification studies (e.g., Tranin et al. 2022; McGlynn et al. 2004). We do not use b as a feature in our work, since the distribution of b in the TD is dependent on the design of the surveys for some classes (e.g., AGNs) thus using b as a feature would cause an additional bias. Still it is useful to look at the distribution of b in the classified sources, as it provides an independent test of our classifications. Figure 11 shows the normalized histograms of $\sin(b)$ (the sin function transforms b into a uniform space of solid angles) of the TD (on the left panel) and CCGCS (on the right panel). We found that HMXBs, YSOs, and HM-STARs are highly concentrated on the Galactic Plane (in both the TD and CCGCS). This is in agreement with the fact that most of these types of sources are located in the Galactic plane. We notice that there is another concentration around $b = -20°$ or $\sin(b) = -0.34$ for YSOs in the TD which presents a selection bias associated with the fact that many TD YSOs come from of the Orion star-forming region (Megeath et al. 2012).
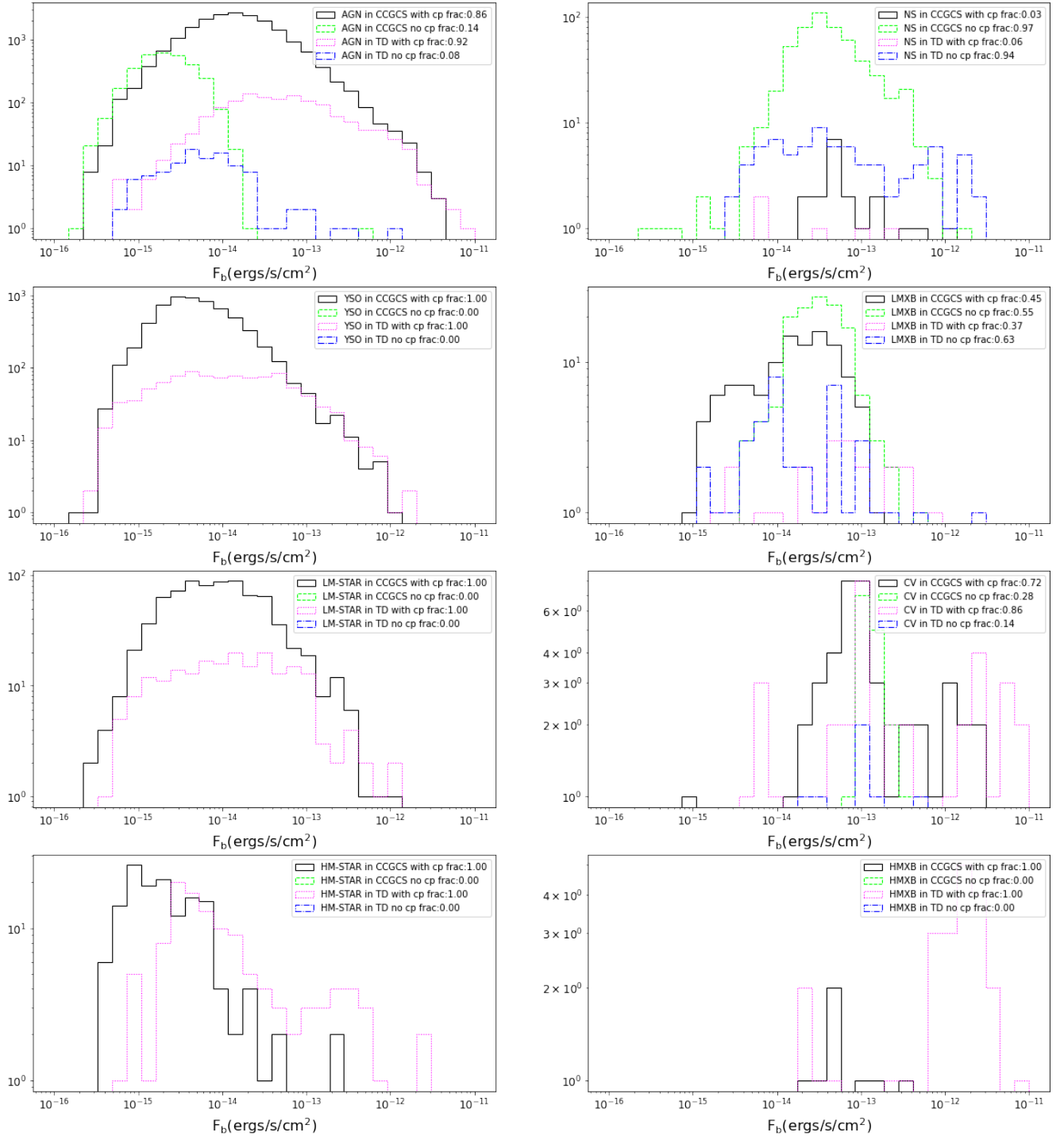
Figure 10. The distributions of the number of sources vs. broadband X-ray fluxes of CCGCS sources and sources from the TD. The solid black and dashed green histograms show the distributions of CCGCS with and without MW counterparts. The dotted magenta and dash-dotted blue histograms show the distributions of sources from the TD with and without MW counterparts. The fractions of sources with and without MW counterparts are also listed in the legends.

Table 6. Comparison of classifications to other catalogs.

Other Catalogs	MUWCLASS	Matches
FIRST-NVSS-SDSS AGN	AGN	17 (100%)
WISE-SDSS Galaxy	AGN	146 (95.4%)
	YSO	7 (4.6%)
4XMM-DR10 TD		
AGN	AGN	626 (93.6%)
	non-AGN	43 (6.4%)
STAR	STAR	172 (92.0%
	non-STAR	15 (8.0%)
XRB	XRB	101 (36.3%)
	NS	56 (20.1%)
	others	121 (43.5%)
CV	CV	2 (25%)
	non-CV	6 (75%)
SIMBAD		
Galaxy	AGN	5089 (98.1%)
	non-AGN	101 (1.9%)
YSO	YSO	2423 (97.3%)
	non-YSO	68 (2.7%)
STAR	STAR	479 29.7%
	YSO	1014 (62.9%)
	others	120 (7.4%)
XRB	XRB	30 (11.7%)
	non-XRB	226 (88.3%)
NS	NS	3 (75%)
	non-NS	1 (25%)
CV	CV	1 (5.3%)
	non-CV	18 (94.7%)

Note—STAR class consists of LM-STAR and HM-STAR classes while XRB class consists of LMXB and HMXB classes. We had to introduce merged classes in this table because of the limitations (or differences) of the catalogs used for comparison. The percentages in brackets in Matches column correspond to the fractions of sources classified as listed in the second (MUWCLASS) column among the sources belonging to the class listed in the first (Other Catalogs) column.

LM-STARs exhibit weaker concentrations toward the Galactic Plane for both the TD and CCGCS, which is consistent with the fact that LM-STARs are Galactic sources with soft X-ray spectra and have low X-ray luminosities so they can only be detected in X-rays when they are relatively nearby (unless there is a large coronal flare). For nearby LM-STARs whose distances are comparable to the local width of the Galactic disk, the distribution should not be strongly concentrated toward the Galactic Plane. As expected, AGNs lack any concentration toward the Galactic Plane for both the TD and CCGCS. In fact, the high extinction/absorption in the Galactic Plane makes many of the extragalactic sources too dim to be detected causing the apparent deficit of AGNs in the Plane. For CVs and LMXBs, we found that they also show reduced concentrations toward the Galactic Plane for CCGCS in comparison with the
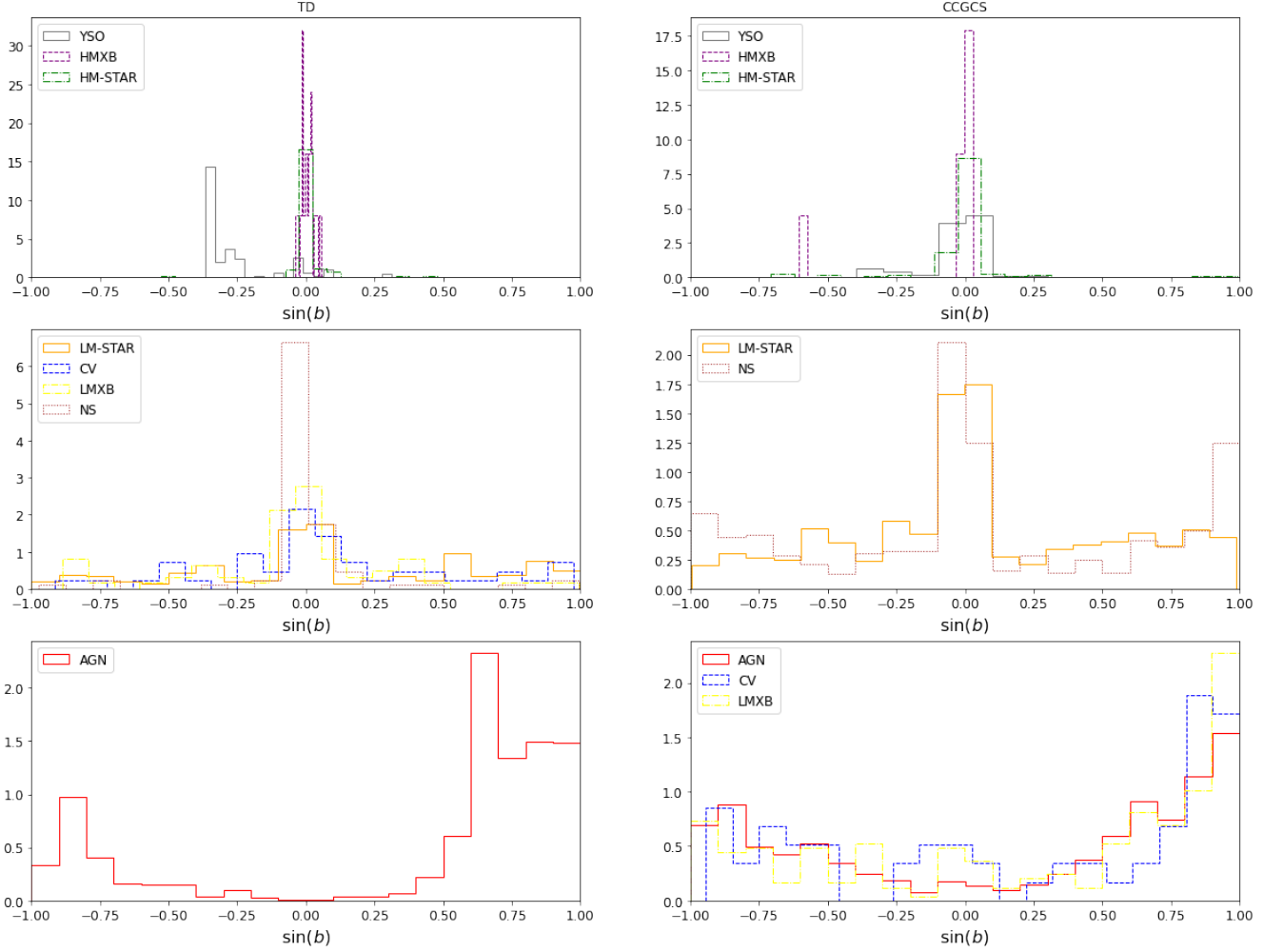
Figure 11. The distributions of the normalized number of sources vs. $\sin(b)$, where b is the Galactic latitude, for the TD (left panel) and CCGCS (right panel). Top panels show sources from YSO, HMXB, and HM-STAR classes whose distributions exhibit strong concentration toward the Galactic Plane. Middle panels show sources from those classes whose distributions exhibit intermediate concentration toward the Galactic Plane. Bottom panels show sources from those classes whose distributions exhibit weak or no concentration toward the Galactic Plane.

TD. We also note that the spatial distribution of both TD and confidently classified GSC sources are affected by the non-uniformity of the CSCv2's sky coverage and varying depth of *CXO* observations.

5.4. *Distance Distribution*

We also look at the distribution of the *rgeo* distances (from the *Gaia* EDR3 distance catalog; Bailer-Jones et al. 2021) for both CCGCS and TD. The parallax to parallax error ratio $\overline{\omega}/\sigma_{\overline{\omega}}$ is used to characterize the robustness of distance measurements. We define sources with $\overline{\omega}/\sigma_{\overline{\omega}} >= 3$ as those with good quality distance measurements. In Figure 12 the distributions of the distances for sources with good quality ($\overline{\omega}/\sigma_{\overline{\omega}} >= 3$) and bad quality ($\overline{\omega}/\sigma_{\overline{\omega}} < 3$) parallax measurements are plotted for both CCGCS and TD.

We see that very few AGNs have good quality distance measurements. Since *Gaia* can only reliably measure distances up to a few kpc, all extragalactic sources are expected to have unreliable distance measurements. HM-STARs and YSOs appear to have systematically higher values of *rgeo* compared to LM-STARs, which is consistent with the fact that LM-STARs in the TD and GCS have to be relatively nearby to be detected in X-rays and/or to be reliably classified via their optical spectra (for the inclusion into the TD).
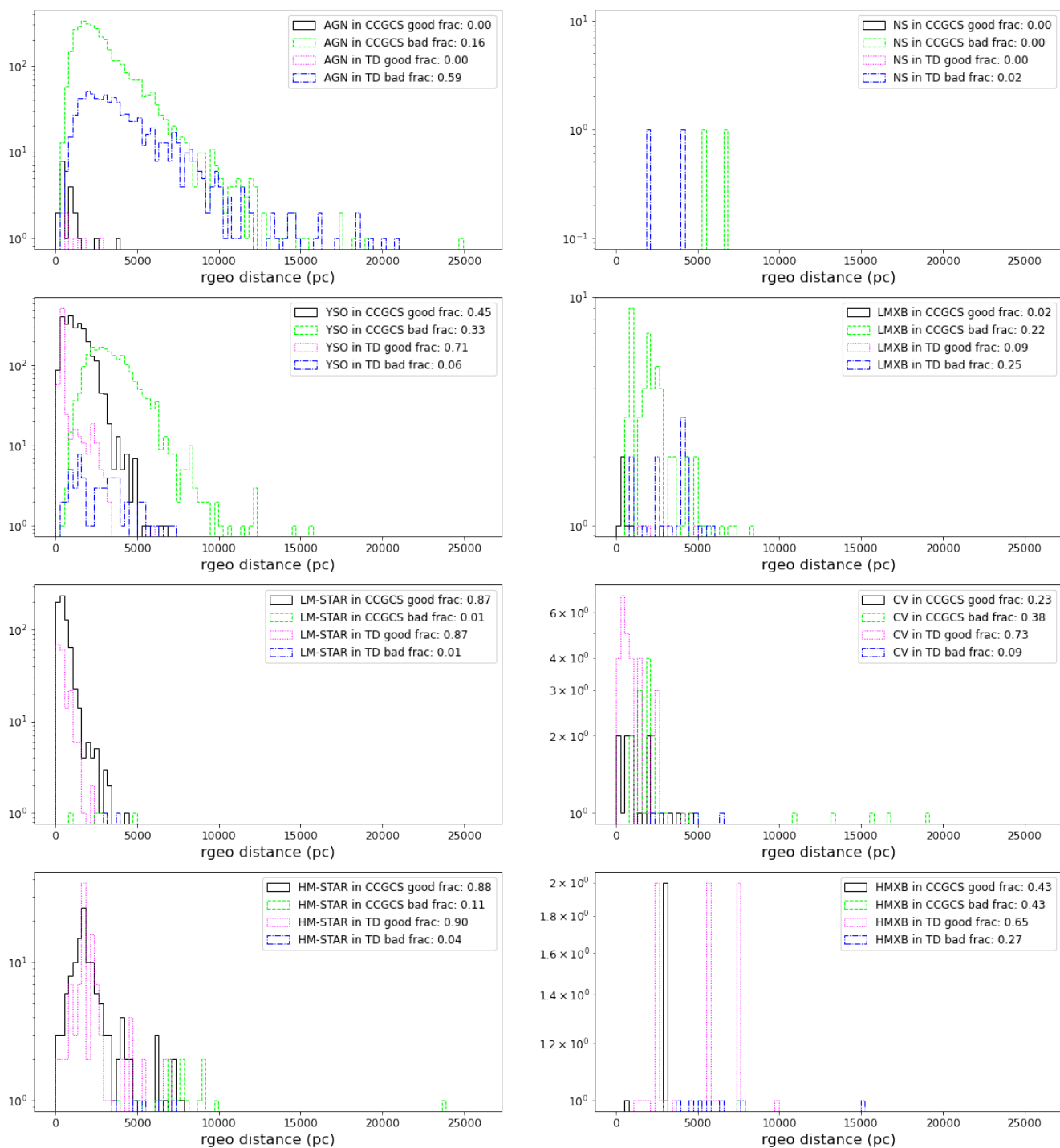
Figure 12. The distributions of the number of sources vs. *rgeo* distance for CCGCS and the TD. The black solid histogram shows the distribution for distance measurements of CCGCS with good quality parallax measurements while the dotted magenta histogram shows that for the TD. The histograms with bad quality parallax measurements are shown as green dashed and blue dot-dash lines for CCGCS and TD, respectively. The fractions of sources with good/bad distance measurements with respect to all sources of the same class in CCGCS or TD are listed in the legend.
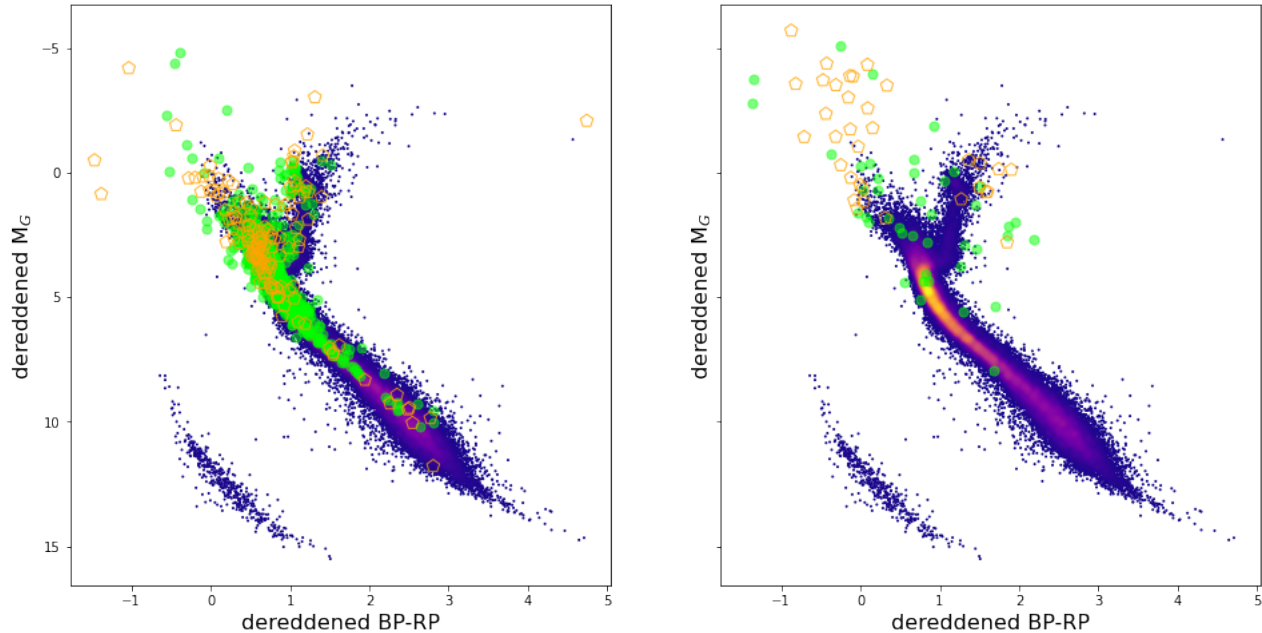
Figure 13. The left panel shows the extinction-corrected Color-Magnitude Diagram (CMD) for CCGCS of LM-STAR class as green circles and LM-STARs from the TD shown as open orange pentagons. The right panel shows a similar plot for the HM-STARs. Both plots are overlaid on the CMD of the Gaia DR2 low-extinction stars from Figure 5 in (Gaia Collaboration et al. 2018b). Color of the points indicates the local density.

5.5. *The Color-Magnitude Diagrams of Stars*

After applying the same data filtering as described in Section 2.1 of (Gaia Collaboration et al. 2018b) to ensure reliable astrometric solutions as well as accurate photometric measurements, 595 out of 794 LM-STARs and 66 out of 144 HM-STARs from CCGCS remained. We also apply the same data filtering for the TD leaving 157 out of 207 LM-STARs and 81 out of 118 HM-STARs. The distances are based on *rgeo* (Bailer-Jones et al. 2021) and the extinction correction (de-reddening) is applied using $E(B-V)$ inferred from the *combined19* model from the MWdust 3D extinction map (Bovy et al. 2016) using the extinction python package[25] at the effective wavelength of each band.

The de-reddened Color-Magnitude Diagrams (CMDs) for both LM-STARs and HM-STARs from CCGCS and the TD are shown in Figure 13 on top of the density plot of the CMD of a large number of low-extinction stars ($E(B-V) < 0.015$ mag) from Figure 5 in (Gaia Collaboration et al. 2018b). In these plots we exclude the sources from the Galactic Center (within 1 square degree) and a few known Galactic SFRs (Avedisova 2002) with poorly known (or/and strongly variable) extinction (matched within 1′).

Most X-ray sources classified as LM-STARs align well with the main sequence while the HM-STARs exhibit a larger scatter in the upper left part of the CMD with higher luminosities and bluer colors. One interesting source is 2CXO J084621.1+013755 which is the notable outlier to the far right in the CMD. It is a Mira Cet type variable star in SIMBAD, which is a pulsating star characterized by a very red color.

6. EXAMPLES OF CLASSIFICATION APPLICATIONS

The rapid classification of a large number of X-ray sources enables many different applications. For instance, these classifications could be used to systematically search for X-ray emitting compact objects in SNRs (see e.g., Pannuti et al. 2017). They can also be used to search for rarely found source types (e.g., quiescent HMXBs, LMXBs, NSs) or to search for potential X-ray counterparts to the numerous unidentified γ-ray (TeV and/or GeV) sources (see e.g., Hare et al. 2016, 2017). Below we use our classification results of CSCv2 sources to perform a more detailed exploration of 5 sources classified as HMXBs with high confidence (see Section 6.1). We also explore a sample of unidentified H.E.S.S.

---

[25] https://github.com/kbarbary/extinction

sources (see Table 7) with potential X-ray counterparts having interesting classifications (particularly, NS class) (see Section 6.2).

6.1. *HMXBs*

There are two known HMXBs that were not included in our TD and hence were classified by the pipeline. 2CXO J142112.1–624156 had the highest probability (among all classified sources) of being a HMXB (P_{HMXB} =82±10%. This source, also known as 2S 1417–642 is, in fact, a HMXB comprised of an accreting pulsar with a 17-s spin period orbiting a Be star with 42-day orbital period. It was observed by *Chandra* in quiescence (Tsygankov et al. 2017) and by Neutron Star Interior Composition Explorer (NICER) during an outburst (Mandal & Pal 2021). There was another known HMXB (2CXO J171556.4–385154) correctly classified as a HMXB with P_{HMXB} =57±16% and somewhat lower CT=1.6 than the threshold CT = 2 adopted above. This source is also known as XTE J1716–389 and is an obscured HMXB (see Ratti et al. 2010). Both sources were not included in our TD because their coordinates in the Liu et al. (2006) HMXB catalog were inaccurate and offset from their *CXO* positions by $> 5''$. These two sources will be added to the TD in future MUWCLASS pipeline releases.

2CXO J193309.5+185902 is classified as a HMXB with a probability of P_{HMXB} =69±14%. Interestingly, the source lies within the radio SNR G54.4–0.3, which is estimated to be at a distance of $d = 3 - 7\,\mathrm{kpc}$ (Junkes et al. 1992; Case & Bhattacharya 1998) although more recent estimates favor $d = 5 - 7\,\mathrm{kpc}$ (Ranasinghe & Leahy 2017; Lee et al. 2020). The radio-quiet PSR J1932+1916 is a middle-aged (spin-down age 35 kyrs) GeV pulsar which lies just outside of the northern rim of the SNR. It is still unclear whether or not the pulsar is related to the SNR (Pletsch et al. 2013; Karpova et al. 2017; Medvedev et al. 2019). The *Gaia* EDR3 counterpart of 2CXO J193309.5+185902 has Gmag=12.6 and a parallax distance of $d = 3.14^{+0.15}_{-0.12}\,\mathrm{kpc}$, with significant astrometric noise (> 6) and a Renormalized Unit Weight Error (RUWE) of 1.12, which suggests that the optical source is in a binary. Note that the distance estimate may be inaccurate given the large astrometric noise. Additionally, the optical counterpart is likely a known Hα emission line star, but no spectral type for the star has been published[26] (Kohoutek & Wehmeyer 1999). The 3D extinction maps of Bovy et al. (2016) give $E(B - V) = 1.26$ at the source's distance. Assuming the distance is correct, the dereddened *Gaia* absolute magnitude, $M_{\mathrm{G}} = -3.1$ and color $BP - RP = -0.1$ place this source in the O/B star region in the *Gaia* CMD, supporting the HMXB classification. At the parallax distance of 3.1 kpc, the source would have an observed X-ray luminosity of $L_X = 2 \times 10^{32}\,\mathrm{erg\,s^{-1}}$. An absorbed power-law model fit[27] to the source's spectrum gives $N_{\mathrm{H}} = (1.0 \pm 0.5) \times 10^{22}\,\mathrm{cm^{-2}}$ and a photon index $\Gamma = 1.7 \pm 0.4$. The spectral slope is compatible with those of a quiescent Be X-ray binary, high-mass γ-ray binary (HMGB; Dubus 2013), or perhaps a γ Cas type analog, which have comparable luminosities (see e.g., Smith et al. 2016). Future spectroscopic follow-up of this source can help to further elucidate its nature.

2CXO J085910.9–434343 has a HMXB classification probability $P_{\mathrm{HMXB}} = 64±9\%$, with its second highest classification probability being a YSO (P_{YSO} =30±8%). The source position on the sky overlaps with the young open cluster RCW 36, which has an age of ~ 1 Myr and is located at a distance of 700 pc (see, e.g., Ellerbroek et al. 2013). The source, which is highly variable (i.e., $P_{\mathrm{intra}} \approx 1.0$) has a *Gaia* EDR3 parallax distance of $d = 755^{+244}_{-139}$ pc, placing it in the RCW 36 cluster. Recently, Getman & Feigelson (2021) studied this source and found that it is a flaring pre-main-sequence star, which showed a "superflare" reaching a peak X-ray luminosity of $L_X \approx 10^{32}\,\mathrm{erg\,s^{-1}}$. The source is also strongly absorbed, with $A_V = 9.1$ (Getman & Feigelson 2021). It is likely that the large absorbing column density and strong variability led to the misclassification as a HMXB. According to the confident precision CM (Figure 8, right bottom panel) small fraction of sources confidently classified as HMXBs is expected to be actually YSOs.

2CXO J201641.4+370925 is classified as a HMXB with P_{HMXB} =61±11% and its second highest classification probability being a YSO with P_{YSO} =17±6%. The source has a counterpart in *Gaia* EDR3 with Gmag=18.4. The parallax measurement places it at a distance of about 3 kpc (Gaia Collaboration et al. 2018a; Bailer-Jones et al. 2021). At this distance, the source has an observed luminosity of $L_X = 4 \times 10^{31}\,\mathrm{erg\,s^{-1}}$. The 3D extinction maps of Bovy et al. (2016) give an absorption of $E(B - V) = 0.8$ at this distance. After correcting for absorption, the optical counterpart's absolute *Gaia* G-band magnitude is M_{G} =3.76 and and the color is $BP - RP = 0.75$. This places the source on the blue side of the main sequence in the *Gaia* CMD. We note that the parallax distance for this source has rather large uncertainties, spanning distances of $d \approx 2 - 5\,\mathrm{kpc}$. At the largest distance of 5 kpc, the absorption

---

[26] The coordinates of the Hα emitting star published in Kohoutek & Wehmeyer (1999) appear to be $\approx 8''$ offset from those of the *Gaia*'s optical counterpart to the X-ray source. However, since there are no other relatively bright stars within $10''$ of 2CXO J193309.5+185902, the position listed in Kohoutek & Wehmeyer (1999) is likely inaccurate.

[27] All spectra fit in this section were taken from the CSCv2 data products (Evans et al. 2020).

is larger ($E(B - V) = 1.8$), and the source is consistent with a high-mass star. Additionally, the X-ray luminosity becomes $L_X \approx 1 \times 10^{32}\,\mathrm{erg\,s^{-1}}$, which is consistent with those of HMXBs in quiescence, HMGBs, or γ Cas type Be binaries. However, the source is faint with very few counts having energies below $1.8\,\mathrm{keV}$. An absorbed power-law fit to the spectrum gives a photon index $\Gamma = 0.4 \pm 0.8$, suggesting a hard spectrum albeit with very large uncertainty. The source is relatively bright at NIR wavelengths having a 2MASS K-band mag of 14.3 (Skrutskie et al. 2006), thus spectroscopic observations of the source could help to better understand its nature.

6.2. *Unidentified TeV Sources from the H.E.S.S. Galactic Plane Survey*

Here we explore some interesting X-ray sources (detected with S/N≥ 6) within the extent of unidentified TeV sources from the H.E.S.S. Galactic plane survey (HGPS; H. E. S. S. Collaboration et al. 2018) that were confidently classified by MUWCLASS pipeline. Since many pulsars move fast, we classify all X-ray sources within two times the TeV source "size" value, corresponding to the actual source size for resolved sources, or two times the 68% positional uncertainty values for unresolved sources in HGPS (H. E. S. S. Collaboration et al. 2018). Unlike GCS, we do not perform filtering on the X-ray PUs and CSCv2 source flags to allow for extended sources to be also classified because pulsars surrounded by compact (and relatively faint) PWNe are often expected as counterparts to Galactic TeV sources (see, e.g., Kargaltsev et al. 2013). Individual X-ray source classifications (with all confidences and for all source detected with S/N≥ 3) are provided in the electronic (machine-readable) format (a few *CXO* sources discussed below are shown in Table 10 which represents a part of that larger table, see Appendix E). Additional data products (e.g., source class visualizations on top of X-ray images; see Figure 14 for examples) are available at the GitHub repository. The summary table of all *confident* (i.e. CT≥ 2) classifications for X-ray sources detected with S/N≥ 6 is given in Table 7. Below we discuss some interesting cases in more detail.

The field of the unidentified source HESS J1912+101 contains a relatively bright ($F_{\mathrm{b}} \approx 3 \times 10^{-13}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$) X-ray source 2CXO J191237.9+101044. It is confidently classified as a NS ($P_{\mathrm{NS}} = 78 \pm 15\%$). Two potential scenarios have been suggested for HESS J1912+1901. In the first scenario the TeV emission is produced by a PWN from the energetic pulsar PSR J1913+1011, however, no X-ray emission was detected from the pulsar or its putative PWN (Chang et al. 2008). The second scenario is that there is a low surface brightness supernova remnant (SNR) producing the TeV emission (see e.g., Puehlhofer et al. 2015), but no clear SNR has yet been detected (see e.g., Su et al. 2017; Reich & Sun 2019). 2CXO J191237.9+101044 is labelled as an extended source in the CSCv2, but its $10'$ off-axis location, makes it difficult to differentiate between extended emission or multiple point sources, as *Chandra*'s PSF degrades with increasing off-axis angle. The X-ray spectrum of the source is well fit by an absorbed power-law model with $N_{\mathrm{H}} = (0.7 \pm 0.4) \times 10^{22}\,\mathrm{cm^{-2}}$ and $\Gamma = 1.3 \pm 0.4$. Additionally, there are no NIR sources coincident with the *CXO* source in the deeper UKIDDS Galactic plane survey data (Lucas et al. 2008). If the source is truly extended, it may be a pulsar surrounded by a PWN, in which case it could be responsible for the TeV emission from HESS J1912+1901. A deeper on-axis *CXO* observation of this source could help to resolve the extended emission and better constrain the X-ray spectrum.

Another unidentified TeV source, HESS J1843−033, is resolved into three peaks in most recent HGPS images. The most separated peak was labelled as a new TeV source (HESS J1844−030), while HESS J1843−033 (J1843 hereafter) consists of two "hotspots", which we refer to as HGPSC 083 and HGPSC 084 (see Figure 28 in H. E. S. S. Collaboration et al. 2018). Given the presence of hotspots and the diffuse structure of J1843, it is likely that the TeV emission comes from multiple sources, including faint emission from SNR G28.6−0.1 (Helfand et al. 1989; Ueno et al. 2003; H. E. S. S. Collaboration et al. 2018). Interestingly, the HGPSC 084 region contains a source 2CXO J184335.8−034653, which is confidently classified as a NS, with $P_{\mathrm{NS}} = 87 \pm 16\%$. The source is also flagged by the CSCv2 as being extended, which may suggest that it is a pulsar surrounded by a PWN. However, the source is rather faint ($F_{\mathrm{b}} = 2.7 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$) and is imaged $\sim 7.6'$ off-axis. The source's spectrum is not well constrained, but appears to be fairly hard, having an absorbed power-law photon index $\Gamma = 1.1 \pm 0.9$, which is consistent with a PSR/PWN scenario. There are no MW counterparts to the source, even in deeper surveys which we checked. A deeper on-axis observation is needed to confirm the source classification and probe the extended nature of the source.

HESS J1813−126 is an unidentified TeV source spatially coincident with middle-aged GeV PSR J1813−1246 (included in our TD), which has been proposed as the X-ray counterpart to HESS J1813−126 (Abdo et al. 2009). However, Marelli et al. (2014) found that this fairly energetic pulsar does not have an X-ray PWN, which casts some doubt on a PWN scenario. 2CXO J181303.0−124907 lies within the extent of HESS J1813−126 and is offset from PSR J1813-1246 by $\sim 6'$. This source was confidently classified by our pipeline as an AGN with $P_{\mathrm{AGN}} = 95 \pm 4\%$. Based on the hardness

ratio of the source, Marelli et al. (2014) also concluded that this source was likely an AGN. The source's X-ray spectrum is well fit by an absorbed power-law model with $N_{\rm H} = (1.2 \pm 0.7) \times 10^{22}\,{\rm cm}^{-2}$ and $\Gamma = 1.0 \pm 0.4$, supporting the AGN classification. Additionally, the X-ray source position is coincident, within 2σ, with a bright ($181 \pm 6\,{\rm mJy}$ at 1.4 GHz) radio point source detected in the NVSS survey (Condon et al. 1998), further supporting its AGN nature. Future NIR/IR spectroscopic observations can help to constrain the AGNs redshift and determine whether or not it is plausible that the AGN contributes significantly to the TeV emission.

HESS J1018–589 is composed of two components (A and B), with A having a confirmed X-ray counterpart and B having a plausible X-ray counterpart. HESS J1018–589A has been shown to have TeV emission produced by the HMGB 1FGL J1018.6–5856 (H. E. S. S. Collaboration et al. 2015) (2CXO J101855.6–585645 in CSCv2). This source is not in our TD as it was not yet discovered when the HMXB catalog we use was produced. Therefore, it serves as a further test of the performance of our pipeline which correctly and confidently classified 1FGL J1018.6–5856 as a HMXB with a probability $P_{\rm HMXB} = 74 \pm 15\%$. The plausible counterpart to HESS J1018–589B is PSR J1016–5857 which has an X-ray PWN and is possibly associated with SNR G284.3–1.8 (H. E. S. S. Collaboration et al. 2012; Klingler et al. 2022). However, we also confidently classify an AGN, 2CXO J101812.9–585930, which lies in between 1FGL J1018.6–5856 and PSR J1016–5857. Interestingly, this source is also reported in the 40 month NuSTAR serendipitous source catalog (Lansbury et al. 2017), as it has a hard spectrum with a photon index $\Gamma = 1.3 \pm 0.2$. Unfortunately, the source is highly absorbed, thus no redshift was measured so the distance to the AGN is unconstrained. A NIR/IR spectrum should be taken to constrain the AGNs redshift to determine if it can contribute to the TeV emission in this complex region.

HESS J1614–518 is a very extended ($\approx 0.8°$ in diameter) TeV source with shell-like (SNR?) morphology which is mostly located outside of the field of view of the *CXO* ACIS observation. Out of 2 NS candidates identified in the *CXO* image only one (2CXO J161610.8–515545) overlaps with the eastern outskirts of the TeV source. The X-ray source lacks any optical or NIR counterpart and is confidently classified as a NS with $P_{\rm NS} = 85 \pm 16\%$. However, due to its off-center location and a very marginal coverage of HESS J1614–518 by *CXO*, the case for the actual association with HESS J1614–518 is rather weak. A deeper on-axis *CXO* observation of 2CXO J161610.8–515545 is needed to establish its nature reliably as the source is rather faint ($F_{\rm b} \approx 4.7 \times 10^{-14}\,{\rm erg\,s}^{-1}\,{\rm cm}^{-2}$) and is currently imaged $\approx 11'$ off-axis (which excludes the source from GCS). The other two NS candidates identified in the same *CXO* observation are located significantly farther from HESS J1614–518 and hence are even less likely to be related to it.

HESS J1841–055 is a similarly large source which is somewhat better covered by *CXO* thanks to the multiple ACIS observations, one of which (ObsID 7552) is both fairly close to the TeV source center and relatively deep (18.9 ks). In this observation, we classified 2CXO J184201.9–052823 as a confident NS candidate with $P_{\rm NS} = 84 \pm 17\%$. However, the source, which lacks any MW counterparts, is rather faint ($F_{\rm b} \approx 5.1 \times 10^{-14}\,{\rm erg\,s}^{-1}\,{\rm cm}^{-2}$) and is located about $7'$ off-axis which limits any further analysis with the existing *CXO* data.

HESS J1800–240 consists of hot spots A and B. The latter hotspot coincides with the massive SFR G5.89–0.39 (analyzed by Hampton et al. 2016). It was covered by *CXO* ACIS observation where an X-ray source 2CXO J180010.0–240129 is classified as a NS with $P_{\rm NS} = 73 \pm 11\%$. This X-ray source is offset by about $5'$ from HESS J1800–240B position, which would not be unusual for a NS born in the SFR, because NSs can get a substantial kick velocity at birth, and a relic PWN, responsible for the TeV emission, may be offset from the current pulsar position. The classified source is fairly faint in X-rays ($F_{\rm b} \approx 4 \times 10^{-14}\,{\rm erg\,s}^{-1}\,{\rm cm}^{-2}$) and also has a faint ($J{=}15$) NIR (2MASS) source located within the X-ray source's PU. However, given the large stellar density in this region, a chance coincidence probability is not negligible. We note that in our TD we have several NSs with accidental optical/NIR matches which we intentionally did not remove to allow sources with accidental matches still to be classified as NSs rather than another class. Of course, NIR emission from a real isolated NS is not expected to be detectable with the surveys we use. Deeper optical observations and spectroscopy can help to determine if 2CXO J180010.0–240129 really has a counterpart and verify its classification.

HESS J1857+026 has a fairly complex TeV morphology which has been modeled with two Gaussians in HGPS. The brightest part of the TeV source is likely associated with the known (included in our TD) PSR J1856+0245 which is likely responsible for the TeV emission. However, some contribution to the overall TeV emission of HESS J1857+026 can come from a different source. In the *CXO* ACIS ObsID 10513, which imaged the SW outskirts of the TeV source, we find 2CXO J185643.6+021921 confidently classified as a NS with $P_{\rm NS} = 83 \pm 17\%$. The source has no MW counterparts and is relatively bright ($F_{\rm b} \approx 3 \times 10^{-13}\,{\rm erg\,s}^{-1}\,{\rm cm}^{-2}$). However, the short 5 ks *CXO* exposure precludes any detailed analysis and a deeper observation may be warranted.

Table 7. Summary of confidently classified significant (S/N>6) X-ray sources within the fields of 20 unidentified HESS sources.

HESS Source Name	Summary of Confident Classifications	Possible Association
HESS J1912+101	2LM-STAR, 1NS	2CXO J191237.9+101044 (NS)
HESS J1843−033	1NS, 8YSO	2CXO J184335.8−034653 (NS)
HESS J1813−126	1AGN, 1YSO	PSR J1813-1246 / 2CXO J181303.0−124907 (AGN)
HESS J1018−589B	1AGN, 1HMXB, 1LM-STAR	PSR J1016-5857 / 2CXO J101812.9−585930 (AGN)
HESS J1614−518	2NS	unknown SNR? / 2CXO J161610.8−515545 (NS)?
HESS J1841−055	1NS, 2YSO	2CXO J184201.9−052823 (NS)
HESS J1800−240	1NS, 5YSO	2CXO J180010.0−240129 (NS)
HESS J1857+026	1LM-STAR, 1NS, 1YSO	PSR J1856+0245 / 2CXO J185643.6+021921 (NS)
HESS J1848−018	1NS, 1YSO	· · ·
HESS J1023−575	5HM-STAR, 1LM-STAR, 29YSO	· · ·
HESS J1458−608	1YSO	· · ·
HESS J1616−508	1LM-STAR, 4YSO	· · ·
HESS J1632−478	2YSO	· · ·
HESS J1634−472	3YSO	· · ·
HESS J1713−381	1YSO	· · ·
HESS J1718−385	4YSO	· · ·
HESS J1745−303	1LM-STAR	· · ·
HESS J1826−130	4YSO	· · ·
HESS J1844−030	1YSO	· · ·
HESS J1923+141	6YSO	· · ·

Note—In the Possible Association column we list a possible association based on our classifications and (in some cases) previously proposed association. Note, that for extended or multi-component TeV sources more than one association is possible. In this table we restrict ourselves to sources that pass the detection S/N$\geq$ 6 and CT$\geq$ 2 cuts. The supplementary products from the GitHub repository contain all classifications for S/N$\geq$ 6 as well as for S/N$\geq$ 3. Note that the fields of many TeV sources are only partly covered (see Figure 14). Also, the analysis is restricted to data existing in CSCv2 (which means we did not use any *CXO* observations after 2014). For 2CXO J161610.8−515545 (NS), the question mark is added because the X-ray source is very far from the TeV source center.

Finally, we would like to mention HESS 1848–018, which is located in the direction of the massive SFR W43. We did not find any NS candidates within the central part of this extended TeV source in *CXO* data represented in CSCv2. The only NS candidate (2CXO J184850.3–012509; classified with P_{NS} =83±15%) is located $\approx 32'$ north of the TeV source center in the vicinity of a massive old (likely globular) cluster GLIMPSE C01. Therefore, it is not likely to be related to HESS 1848–018. The NS candidate source is relatively faint ($F_{\mathrm{b}} \approx 4.7 \times 10^{-14}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$) and has no MW counterparts. More recent *CXO* data (not in CSCv2) that exist for this field must be analyzed to further explore the nature of this source.

In many of the HESS fields we have also confidently classified a number of YSOs, as well as low-mass and high-mass stars, which rules out a large number of X-ray sources as potential TeV counterparts, as these types of sources are not expected to produce TeV emission. We caution that currently our pipeline is not capable of identifying some classes of sources (such as NSs and HMXBs which are typically associated with Galactic TeV sources) with high certainty (see the validation results in Figure 8). For example, even after the confidence cut (CT$\geq$ 2; bottom right panel in Figure 8), 15% of sources identified as NSs are actually AGNs (some of AGNs may also be TeV sources). However, the pipeline classifications can be used to substantially reduce the number of sources that can be potential TeV emitters and to identify promising targets for future in-depth investigations. We also note that some HESS fields overlap with particularly crowded environments, such as HESS J1023–575 which is coincident with Westerlund 2. The classification results for these fields are less reliable because of the potential confusion in optical/NIR.
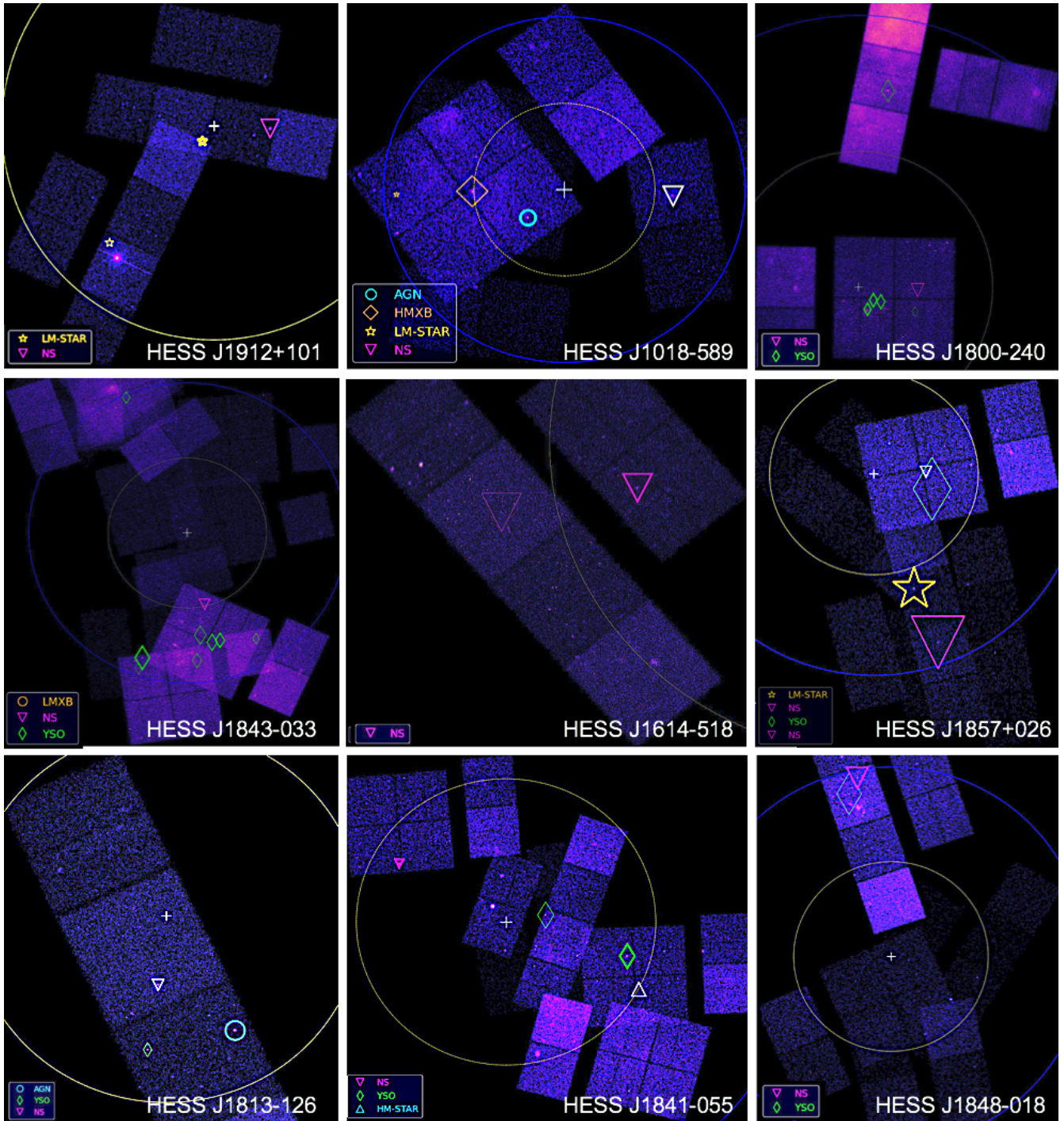
Figure 14.   The *CXO* images of nine HESS fields of that we investigated in Section 6.2. The names of the unidentified H.E.S.S. sources are shown in the lower right corner of each image. Each *CXO* image consists of ObsIDs made before the end of 2014 which were used for CSCv2. White crosses mark the center of the HESS sources, and yellow dashed circles indicate the "size" for resolved HESS sources or the 68% PUs for point HESS sources, and blue solid circles are twice the size of the yellow dashed circles which we used to search for potential X-ray counterparts of the TeV emission. Symbols show significant (S/N≥6) X-ray sources with confident classifications (CT> 2): gold stars − LM-STARs, magenta inverted triangles − NSs, cyan circles − AGNs, orange diamonds − HMXBs, green thin diamonds − YSOs, orange circles − LMXBs, and cyan triangles − HM-STARs. White inverted triangle and triangles mark NSs and HM-STARs from the TD. The widths of the symbol lines increase with the classification probability while the symbol sizes increase the X-ray detection significance. Similar pictures are available at the GitHub repository (see Section 6.2) for all unidentified TeV sources listed in Table 7, including less confident classifications.

7. LIMITATIONS AND FUTURE DEVELOPMENTS

MUWCLASS pipeline and the CSCv2 classification results presented here have a number of limitations which users must be aware of. Below we will discuss these limitations and outline possible solutions which we may implement in the future pipeline releases. We also encourage the users to clone the GitHub repository, independently run the pipeline, and try to improve it on their own.

7.1. *Limitations and Caveats*

One of the biggest limitations is the uneven performance of MUWCLASS for different source classes which is, to a large degree, caused by the unbalanced TD although it can also be due to the heterogeneous astrophysical nature of sources from certain classes as well as due to the strong interstellar absorption in many regions of the Galactic plane. As one can see from the CM plots in Figure 8, in most cases, we reliably classify AGNs, LM-STARs, and YSOs but the classification does not work so well for other classes with a particularly bad performance for the LMXB class which is strongly mixed with NSs. Aside from having few NS and LMXB sources in the TD, about 63% of LMXBs does not have any MW counterparts (likely due to large distances and reddening) making them appear similar to NSs. On the other hand, 5 NSs in the TD have spurious optical counterparts which we did not remove since some spurious counterparts are expected in the dense Galactic fields. With these limitations in mind, one can use the pipeline to "weed out" AGNs, LM-STARs, and YSOs if the goal is to look for other kinds of sources, or to create a combined class (e.g., NS+LMXB) to "improve" the performance if finding either of these is of interest. In the end, a manual investigation of a subset of interesting sources identified with the pipeline will likely still be needed (see Section 6).

Sources from the TD are systematically brighter than CCGCS (see Section 5.2) as a result of the selection effect when building the TD. The literature-verified sources from the TD are classified using reliable traditional classification methods which require the sources to be bright enough, e.g., NSs discovered by searching for the presence of a surface (pulsations, cyclotron resonance scattering feature, thermonuclear bursts etc.), stars and AGNs classified based on the optical/IR/X-ray spectroscopic data. This bias can lead to overestimation of the performance metrics that are evaluated using the TD. These selection effects can be mitigated by building a more complete TD by, e.g., creating synthetic fainter sources from the real ones by "placing" the latter at larger distances with appropriate extinction/reddening applied to them.

The eight classes of X-ray sources used in this paper (see Section 2.1) represent a compromise between the desire to have a separate class for astrophysically different sources and the need to maintain statistically meaningful number of sources for each class. As a result, some of the classes are astrophysically very heterogeneous. For example, we group magnetars together with rotation-powered pulsars in our NS class, and we group redback and black widow type systems with accreting LMXBs in our LMXB class. The HMXB class includes both accreting X-ray pulsars and gamma-ray binaries whose emission is likely powered by colliding winds. Moreover, some classes of source are simply missing from our current TD. For example, we do not include GRBs, ULXs, TDEs, or normal galaxies (not AGNs). Another way to deal with those rare type classes that are not included in the current TD is to introduce an outlier detection algorithm to search for sources that have distinctly different properties and/or do not belong to any of the classes in the TD, which has been done in some other works (Tranin et al. 2022).

Although *CXO* source localizations are excellent compared to other X-ray observatories, the PUs of some sources can still be as large as a few arcsec (e.g., large off-axis angles, faint sources). This can lead to a large degree of confusion when looking for MW counterparts in the crowded regions of the Galactic plane. For this reason we restrict PU$< 1''$ for both GCS and the populous classes in the TD. However, even $1''$ uncertainty may be too large in some dense Galactic environments. We remove sources from the crowded environments (e.g., Galactic Center, LMC/SMC and globular clusters, see Section 2.1) in the TD, but we do not apply this filter apply to GCS. Therefore, one has to be cautious about any classification results for those dense fields. In general, PUs also include a systematic error which is associated with the accuracy to which the telescope pointing is known. In CSCv2 this systematic error is included by adding $0.71''$ (95% confidence, derived by comparing CSCv2 source positions to the SDSS data) in quadrature to the statistically averaged PUs[28]. Currently, there is no attempt in CSCv2 to correct for the imperfect astrometry

---

[28] https://cxc.cfa.harvard.edu/csc/columns/positions.html#ra_dec

by cross-correlating X-ray source positions with their *Gaia* eDR3 counterparts positions. This can be done in the future but one has to be wary of parallaxes and proper motions which may need to be taken into account when such astrometric corrections are performed (it is easy to do on a per-observation level but then the X-ray position accuracy could be worse than that at the master level in CSCv2). Wherever the degree of confusion still remains large, a probabilistic cross-matching algorithm (e.g., Nway Salvato et al. 2018) can be used to take into account all possible counterparts within the positional error circle of an X-ray source.

The SFD extinction maps are less accurate at low Galactic latitudes (i.e., $|b| < 5°$; Schlegel et al. 1998), where many of GCS sources are located. The extinction maps from Sale et al. (2014) agree with the SFD maps within a factor of ∼2 (with most differences being within 20%, see Figure 11 from Sale et al. 2014) in most places of the Galactic plane ($|b| < 2°$). Due to the full sky coverage of the SFD extinction maps, we use them to select the E(B-V) in a given direction to adjust the fluxes of AGNs from the TD to the region where the *CXO* sources are being classified. However, the extinction/absorption correction can be inaccurate if the dust column in a given direction is extremely large and/or there is a significant gradient in the reddening on small angular scales.

7.2. *Future Developments*

This paper describes the first public release of MUWCLASS pipeline and provides some examples of its possible applications. We plan future releases of MUWCLASS pipeline as well as more detailed studies carried out with the help of the pipeline. By making the pipeline and the TD public and easily accessible, we hope to stimulate future progress in this direction, and to allow the broader community to contribute to the pipeline development. We discuss some the possible future improvements below.

To have a better balanced TD which is more adopted to the Galactic environment, it is important to increase the numbers of sources in currently underpopulated classes. We identified some resources that we plan to use in our future work, but we also hope that the community can contribute sources to an open database of classified X-ray sources which will have the TD presented here as the starting point. The work on this database is ongoing and a graphical user interface (GUI) that visualizes the TD content is already available online[29] (Yang et al. 2021). Future additions can include (but are not limited to) various types of X-ray sources from the INTEGRAL Reference Source Catalog [30] (Ebisawa et al. 2003), LMXBs from The Swift Bulge Survey, CVs from Suleimanov et al. (2022); Szkody et al. (2011); Drake et al. (2009), and some other online catalogs and projects (Kharchenko & Roeser 2009; Getman et al. 2005; Feigelson et al. 2007).

In addition to increasing and balancing the TD it is important to use deeper MW surveys and extend the MW coverage to the radio. The currently used 2MASS, WISE-based, and *Gaia* surveys are all-sky allowing for a maximum coverage. It may be beneficial to use deeper new surveys in the Galactic plane, but the varying depth of the surveys needs to be taken into account to avoid potential biases. In particular, replacing 2MASS with the high-resolution sensitive NIR survey should be the top priority. This will greatly improve MUWCLASS's ability to differentiate between CVs, LMXBs, HMXBs, and NS. The MW survey data that could be added in future include the Dark Energy Survey (Abbott et al. 2018), the Pan-STARRS Survey (Chambers et al. 2016), the VISTA surveys[31], the UKIRT Infrared Deep Sky Survey (Lawrence et al. 2007), and the VST Photometric Hα Survey of the Southern Galactic Plane and Bulge (Drew et al. 2014). In addition, the latest generation of radio surveys is able to reach a depth and resolution which warrants including radio domain information among the MW features. As soon as ASKAP (Hotan et al. 2021), MeerKAT (Jonas & MeerKAT Team 2016) and VLASS surveys (Lacy et al. 2020) release their source catalogs, we plan to add radio fluxes to our TD.

The time domain information in our current TD is limited to X-rays. The characterization of the variability time scale is also extremely coarse (i.e. intra and inter observation variability). In addition to determining how variable the source is, it would be beneficial to better to characterize the variability timescale within each observation using methods developed for the optical variability characterization, or to use per-observation data with extensive X-ray light curves. We also plan to add optical variability among the features in the future releases using such surveys as Transiting Exoplanet Survey Satellite (TESS) (Ricker et al. 2014), The Zwicky Transient Facility (ZTF) (Bellm et al. 2019), All-Sky Automated Survey for Supernovae (ASAS-SN) (Kochanek et al. 2017) which provide multiple observations and light curves, and *Gaia* DR3 with extensive variability analysis (Eyer et al. 2022).

---

[29] Our online visualization tool of the TD: https://home.gwu.edu/ kargaltsev/XCLASS/

[30] https://www.isdc.unige.ch/integral/science/catalogue

[31] https://www.eso.org/public/teles-instr/paranal-observatory/surveytelescopes/vista/surveys/#:∼:text=Six%20large%20public%20surveys%20conducted,wide%20range%20of%20scientific%20questions.

Thanks to a number of wide-area sensitive surveys, improved 3D extinction maps are becoming available (e.g., the 3D extinction map of the northern Galactic plane based on IPHAS photometry Sale et al. 2014). Besides the reddening/absorption affecting AGNs when viewed through the Galactic plane, other sources inhabiting the Galactic plane also exhibit different reddening/absorption depending on their position and distance. A better way to deal with the reddening bias between TD sources and unclassified sources in the Galactic plane is to use 3D extinction/absorption maps and distance information for the sources so that extinction/absorption corrections can be applied to all sources (not just AGNs) from the TD. A potentially better approach would be to de-redden (de-extinct) all sources in the TD and in the field to-be-classified before the classification. The reliability of this approach requires further investigation and testing because it may require filtering on the *Gaia* eDR3 distance uncertainties causing different kinds of biases.

Finally, MUWCLASS is designed in such a way that it can be easily adapted to incorporate X-ray source catalogs from other existing X-ray observatories such as *XMM-Newton* (in preparation), Swift-XRT, and eROSITA, as well as from future observatories like AXIS and Athena.

8. SUMMARY

We developed an automated multi-wavelength classification pipeline (MUWCLASS) of X-ray sources which relies on the supervised ML approach with the Random Forest algorithm. The pipeline is released together with a large TD of literature-verified sources that we have constructed over a period of a few years. It is the first supervised ML classifier to be applied to a significant fraction of sources from the CSCv2 which takes into account the measurement uncertainties and augments the X-ray properties with rich MW properties from the *Gaia* EDR3, 2MASS, and WISE catalogs. The main scientific products and the results are summarized below:

- A TD of 2,941 literature verified sources, consisting of 8 X-ray source classes including AGNs, CVs, HM-STARs, HMXBs, LM-STARs, LMXBs, NSs and YSOs, has been compiled and made publicly available. It was supplemented by a convenient web-based viewing tool which was released together with the TD[32].

- The pipeline combined several novel approaches that were not used previously at all or in combination. In particular, we included the Monte-Carlo method to account for the measurement uncertainties of the features and applied a field specific reddening on the AGNs from the TD to account for the additional large reddening for AGNs viewed through the Galactic Plane. We also introduced a novel confidence classification definition (the CT parameter) which relies on our ability to calculate distributions for the classification probabilities given the MC approach.

- We performed careful evaluation of the performance of MUWCLASS using LOOCV approach for the TD and discussed various possible biases, limitations, and improvements. We found that the Accuracy/Balanced Accuracy score increases from 88.6%/70.2% to 97.0%/79.0% after the confidence cut CT = 2 while the overall Completeness/Balanced Completeness drops from 100% to 81.0%/56.9%.

- We classified 66,369 "good" CSCv2 sources (∼21% of all CSCv2 sources) using MUWCLASS pipeline. We found ∼ 31,000 sources with reliable classification where most are classified as AGNs (74%) and YSOs (20%). However, the classification results for other classes are less reliable and need further verification with manual investigations and follow up studies. Although a large number of NSs have been identified, this may be the result of a selection bias related to the fact that the classified sources are generally fainter that the TD sources, thus are also more likely to be missing MW counterparts, similar to NSs.

- As an example of the scientific pipeline application, we studied a number of individual interesting sources, including several sources classified as HMXBs and X-ray sources within the extent of 20 unidentified H.E.S.S. Galactic Plane Survey sources. Several plausible NS candidates have been identified with some of them being possible particle accelerators associated with the TeV emission.

- MUWCLASS pipeline and the main results of this work are made publicly available at the GitHub repository[33]. The machine readable tables are also available for the properties and the classification results of GCS, TD and HESS field sources (see Appendix E).

---

[32] https://home.gwu.edu/~kargaltsev/XCLASS/
[33] https://github.com/huiyang-astro/MUWCLASS_CSCv2

Although the performance of the current MUWCLASS in many situations may be limited, it already can save substantial efforts by confidently identifying sources from some commonly occurring classes. Thus it can be used to efficiently shrink the sample of "more exotic" sources which need followup investigations or in-depth analysis. We hope that in future, MUWCLASS will become a powerful exploration tool capable of making accurate predictions of the astrophysical source nature rapidly for a large number of sources. In addition to the pipeline development, we strongly believe that the expansion of the TD must be a community driven effort with only limited mediation from the maintaining site. We expect that future development of the pipeline will make it useful for classification of eROSITA and XMM-Newton catalogs. We also envision that the tool could be useful for rapid classification of archival sources in large fields associated with gravitational wave sources, neutrino sources, and ultra-high energy cosmic ray sources. Although written with X-ray sources in mind the pipeline can be easily modified to classify any other sources (optical, NIR, IR, radio, etc.)

9. ACKNOWLEDGEMENT

REFERENCES

Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, ApJS, 239, 18, doi: 10.3847/1538-4365/aae9f0

Abdo, A. A., Ackermann, M., Ajello, M., et al. 2009, Science, 325, 840, doi: 10.1126/science.1175558

Arnason, R. M., Barmby, P., & Vulic, N. 2020, MNRAS, 492, 5075, doi: 10.1093/mnras/staa207

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Avedisova, V. S. 2002, Astronomy Reports, 46, 193, doi: 10.1134/1.1463097

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, AJ, 161, 147, doi: 10.3847/1538-3881/abd806

Ballio, F., & Guadagnini, A. 2004, Water Resources Research, 40, W04603, doi: 10.1029/2003WR002876

Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002, doi: 10.1088/1538-3873/aaecbe

Bovy, J., Rix, H.-W., Green, G. M., Schlafly, E. F., & Finkbeiner, D. P. 2016, ApJ, 818, 130, doi: 10.3847/0004-637X/818/2/130

Breiman, L. 2001, Machine Learning, 45, 5

Broos, P. S., Getman, K. V., Povich, M. S., et al. 2011, ApJS, 194, 4, doi: 10.1088/0067-0049/194/1/4

Case, G. L., & Bhattacharya, D. 1998, ApJ, 504, 761, doi: 10.1086/306089

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560. https://arxiv.org/abs/1612.05560

Chang, C., Konopelko, A., & Cui, W. 2008, ApJ, 682, 1177, doi: 10.1086/589225

---

[34] https://hvplot.holoviz.org/

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2011, arXiv e-prints, arXiv:1106.1813. https://arxiv.org/abs/1106.1813

Chicco, D., & Jurman, G. 2020, BMC Genomics, 21, 6, doi: 10.1186/s12864-019-6413-7

Cluster, T. G. W. U. H. P. C. 2022, Building a Shared Resource HPC Center Across University Schools and Institutes: A Case Study. https://arxiv.org/abs/2003.13629

Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693, doi: 10.1086/300337

Cutri, R. M., Wright, E. L., Conrow, T., et al. 2021, VizieR Online Data Catalog, II/328

de Beurs, Z. L., Islam, N., Gopalan, G., & Vrtilek, S. D. 2022, arXiv e-prints, arXiv:2204.00346. https://arxiv.org/abs/2204.00346

Delgado, A. J., Alfaro, E. J., & Yun, J. L. 2011, A&A, 531, A141, doi: 10.1051/0004-6361/201116491

Downes, R. A., Webbink, R. F., Shara, M. M., et al. 2001, PASP, 113, 764, doi: 10.1086/320802

Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, ApJ, 696, 870, doi: 10.1088/0004-637X/696/1/870

Drew, J. E., Gonzalez-Solares, E., Greimel, R., et al. 2014, MNRAS, 440, 2036, doi: 10.1093/mnras/stu394

Dubus, G. 2013, A&A Rv, 21, 64, doi: 10.1007/s00159-013-0064-5

Ebisawa, K., Bourban, G., Bodaghee, A., Mowlavi, N., & Courvoisier, T. J. L. 2003, A&A, 411, L59, doi: 10.1051/0004-6361:20031336

Ellerbroek, L. E., Bik, A., Kaper, L., et al. 2013, A&A, 558, A102, doi: 10.1051/0004-6361/201321752

Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, ApJS, 189, 37, doi: 10.1088/0067-0049/189/1/37

Evans, I. N., Primini, F. A., Miller, J. B., et al. 2020, in American Astronomical Society Meeting Abstracts, Vol. 235, American Astronomical Society Meeting Abstracts #235, 154.05

Eyer, L., Audard, M., Holl, B., et al. 2022, arXiv e-prints, arXiv:2206.06416. https://arxiv.org/abs/2206.06416

Farrell, S. A., Murphy, T., & Lo, K. K. 2015, ApJ, 813, 28, doi: 10.1088/0004-637X/813/1/28

Feigelson, E., Townsley, L., Güdel, M., & Stassun, K. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil, 313. https://arxiv.org/abs/astro-ph/0602603

Fitzpatrick, E. L. 1999, PASP, 111, 63, doi: 10.1086/316293

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018a, A&A, 616, A1, doi: 10.1051/0004-6361/201833051

Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018b, A&A, 616, A10, doi: 10.1051/0004-6361/201832843

Gentile, P. A., Roberts, M. S. E., McLaughlin, M. A., et al. 2014, ApJ, 783, 69, doi: 10.1088/0004-637X/783/2/69

Getman, K. V., & Feigelson, E. D. 2021, ApJ, 916, 32, doi: 10.3847/1538-4357/ac00be

Getman, K. V., Flaccomio, E., Broos, P. S., et al. 2005, ApJS, 160, 319, doi: 10.1086/432092

Giardino, G., Favata, F., Micela, G., Sciortino, S., & Winston, E. 2007, A&A, 463, 275, doi: 10.1051/0004-6361:20066424

Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98, doi: 10.3847/1538-3881/aafc33

Güver, T., & Özel, F. 2009, MNRAS, 400, 2050, doi: 10.1111/j.1365-2966.2009.15598.x

H. E. S. S. Collaboration, Abramowski, A., Acero, F., et al. 2012, A&A, 541, A5, doi: 10.1051/0004-6361/201218843

H. E. S. S. Collaboration, Abramowski, A., Aharonian, F., et al. 2015, A&A, 577, A131, doi: 10.1051/0004-6361/201525699

H. E. S. S. Collaboration, Abdalla, H., Abramowski, A., et al. 2018, A&A, 612, A1, doi: 10.1051/0004-6361/201732098

Hampton, E. J., Rowell, G., Hofmann, W., et al. 2016, Journal of High Energy Astrophysics, 11, 1, doi: 10.1016/j.jheap.2016.05.001

Hand, D. J., & Yu, K. 2001, Int. Stat. Rev.

Hare, J., Kargaltsev, O., Pavlov, G. G., Rangelov, B., & Volkov, I. 2017, ApJ, 841, 81, doi: 10.3847/1538-4357/aa6f07

Hare, J., Rangelov, B., Sonbas, E., Kargaltsev, O., & Volkov, I. 2016, ApJ, 816, 52, doi: 10.3847/0004-637X/816/2/52

Helfand, D. J., Velusamy, T., Becker, R. H., & Lockman, F. J. 1989, ApJ, 341, 151, doi: 10.1086/167479

Hotan, A. W., Bunton, J. D., Chippendale, A. P., et al. 2021, PASA, 38, e009, doi: 10.1017/pasa.2021.1

Jonas, J., & MeerKAT Team. 2016, in MeerKAT Science: On the Pathway to the SKA, 1

Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, AJ, 160, 120, doi: 10.3847/1538-3881/aba592

Junkes, N., Fuerst, E., & Reich, W. 1992, A&AS, 96, 1

Kaplan, D. L., Gaensler, B. M., Kulkarni, S. R., & Slane, P. O. 2006, ApJS, 163, 344, doi: 10.1086/501441

Kargaltsev, O., Rangelov, B., & Pavlov, G. 2013, in The Universe Evolution: Astrophysical and Nuclear Aspects. Edited by I. Strakovsky and L. Blokhintsev. Nova Science Publishers, 359–406

Karpova, A., Shternin, P., Zyuzin, D., Danilenko, A., & Shibanov, Y. 2017, MNRAS, 466, 1757, doi: 10.1093/mnras/stw3184

Kharchenko, N. V., & Roeser, S. 2009, VizieR Online Data Catalog, I/280B

Klingler, N., Kargaltsev, O., Pavlov, G. G., et al. 2022, arXiv e-prints, arXiv:2204.13167. https://arxiv.org/abs/2204.13167

Klingler, N., Yang, H., Hare, J., et al. 2020, ApJ, 901, 157, doi: 10.3847/1538-4357/abaf4b

Kochanek, C. S., Shappee, B. J., Stanek, K. Z., et al. 2017, PASP, 129, 104502, doi: 10.1088/1538-3873/aa80d9

Kohoutek, L., & Wehmeyer, R. 1999, A&AS, 134, 255, doi: 10.1051/aas:1999101

Lacy, M., Baum, S. A., Chandler, C. J., et al. 2020, PASP, 132, 035001, doi: 10.1088/1538-3873/ab63eb

Lansbury, G. B., Stern, D., Aird, J., et al. 2017, ApJ, 836, 99, doi: 10.3847/1538-4357/836/1/99

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599, doi: 10.1111/j.1365-2966.2007.12040.x

Lee, Y.-H., Koo, B.-C., & Lee, J.-J. 2020, AJ, 160, 263, doi: 10.3847/1538-3881/abc00e

Lemaître, G., Nogueira, F., & Aridas, C. K. 2017, Journal of Machine Learning Research, 18, 1. http://jmlr.org/papers/v18/16-365.html

Li, D., Starling, R. L. C., Saxton, R. D., Pan, H.-W., & Yuan, W. 2022, MNRAS, 512, 3858, doi: 10.1093/mnras/stac714

Liu, Q. Z., van Paradijs, J., & van den Heuvel, E. P. J. 2006, A&A, 455, 1165, doi: 10.1051/0004-6361:20064987

—. 2007, A&A, 469, 807, doi: 10.1051/0004-6361:20077303

Lo, K. K., Farrell, S., Murphy, T., & Gaensler, B. M. 2014, ApJ, 786, 20, doi: 10.1088/0004-637X/786/1/20

Lucas, P. W., Hoare, M. G., Longmore, A., et al. 2008, MNRAS, 391, 136, doi: 10.1111/j.1365-2966.2008.13924.x

Lyne, A. G., Stappers, B. W., Keith, M. J., et al. 2015, MNRAS, 451, 581, doi: 10.1093/mnras/stv236

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993, doi: 10.1086/428488

Mandal, M., & Pal, S. 2021, arXiv e-prints, arXiv:2103.13444. https://arxiv.org/abs/2103.13444

Marelli, M., Harding, A., Pizzocaro, D., et al. 2014, ApJ, 795, 168, doi: 10.1088/0004-637X/795/2/168

Marocco, F., Eisenhardt, P. R. M., Fowler, J. W., et al. 2021, ApJS, 253, 8, doi: 10.3847/1538-4365/abd805

McGlynn, T. A., Suchkov, A. A., Winter, E. L., et al. 2004, ApJ, 616, 1284, doi: 10.1086/424955

Medvedev, O. D., Karpova, A. V., Shibanov, Y. A., Zyuzin, D. A., & Pavlov, G. G. 2019, in Journal of Physics Conference Series, Vol. 1400, Journal of Physics Conference Series, 022018, doi: 10.1088/1742-6596/1400/2/022018

Megeath, S. T., Gutermuth, R., Muzerolle, J., et al. 2012, AJ, 144, 192, doi: 10.1088/0004-6256/144/6/192

Mignani, R. P. 2011, Advances in Space Research, 47, 1281, doi: 10.1016/j.asr.2009.12.011

Misanovic, Z., Kargaltsev, O., & Pavlov, G. G. 2010, ApJ, 725, 931, doi: 10.1088/0004-637X/725/1/931

Monet, D. G., Levine, S. E., Canzian, B., et al. 2003, AJ, 125, 984, doi: 10.1086/345888

Murphy, T., Mauch, T., Green, A., et al. 2007, MNRAS, 382, 382, doi: 10.1111/j.1365-2966.2007.12379.x

Murthy, S. K., Kasif, S., & Salzberg, S. 1994, arXiv e-prints, cs/9408103. https://arxiv.org/abs/cs/9408103

Ochsenbein, F., Bauer, P., & Marcout, J. 2000, A&AS, 143, 23, doi: 10.1051/aas:2000169

Ozawa, H., Grosso, N., & Montmerle, T. 2005, A&A, 429, 963, doi: 10.1051/0004-6361:20040480

Pannuti, T. G., Rho, J., Kargaltsev, O., et al. 2017, ApJ, 839, 59, doi: 10.3847/1538-4357/aa615c

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2012, arXiv e-prints, arXiv:1201.0490. https://arxiv.org/abs/1201.0490

Pletsch, H. J., Guillemot, L., Allen, B., et al. 2013, ApJL, 779, L11, doi: 10.1088/2041-8205/779/1/L11

Possolo, A., Merkatas, C., & Bodnar, O. 2019, Metrologia, 56, 045009, doi: 10.1088/1681-7575/ab2a8d

Povich, M. S., Smith, N., Majewski, S. R., et al. 2011, ApJS, 194, 14, doi: 10.1088/0067-0049/194/1/14

Puehlhofer, G., Brun, F., Capasso, M., et al. 2015, in International Cosmic Ray Conference, Vol. 34, 34th International Cosmic Ray Conference (ICRC2015), 886. https://arxiv.org/abs/1509.03872

Quinlan, J. R. 1993, C 4.5: Programs for machine learning

Ranasinghe, S., & Leahy, D. A. 2017, ApJ, 843, 119, doi: 10.3847/1538-4357/aa7894

Ratti, E. M., Bassa, C. G., Torres, M. A. P., et al. 2010, MNRAS, 408, 1866, doi: 10.1111/j.1365-2966.2010.17252.x

Rebull, L. M., Koenig, X. P., Padgett, D. L., et al. 2011, ApJS, 196, 4, doi: 10.1088/0067-0049/196/1/4

Reich, W., & Sun, X.-H. 2019, Research in Astronomy and Astrophysics, 19, 045, doi: 10.1088/1674-4527/19/3/45

Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, ed. J. Oschmann, Jacobus M., M. Clampin, G. G. Fazio, & H. A. MacEwen, 914320, doi: 10.1117/12.2063489

Ritter, H., & Kolb, U. 2003, A&A, 404, 301,
doi: 10.1051/0004-6361:20030330

Rodrigo, C., Solano, E., & Bayo, A. 2012, SVO Filter
Profile Service Version 1.0, IVOA Working Draft 15
October 2012,
doi: 10.5479/ADS/bib/2012ivoa.rept.1015R

Rosen, S., Watson, M., Pye, J., et al. 2015, in Astronomical
Society of the Pacific Conference Series, Vol. 495,
Astronomical Data Analysis Software an Systems XXIV
(ADASS XXIV), ed. A. R. Taylor & E. Rosolowsky, 319

Sale, S. E., Drew, J. E., Barentsen, G., et al. 2014,
MNRAS, 443, 2907, doi: 10.1093/mnras/stu1090

Salvato, M., Buchner, J., Budavári, T., et al. 2018,
MNRAS, 473, 4937, doi: 10.1093/mnras/stx2651

Sazonov, S., Paizis, A., Bazzano, A., et al. 2020, NewAR,
88, 101536, doi: 10.1016/j.newar.2020.101536

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103,
doi: 10.1088/0004-637X/737/2/103

Schlafly, E. F., Meisner, A. M., & Green, G. M. 2019,
ApJS, 240, 30, doi: 10.3847/1538-4365/aafbea

Schlafly, E. F., Green, G., Finkbeiner, D. P., et al. 2014,
ApJ, 786, 29, doi: 10.1088/0004-637X/786/1/29

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ,
500, 525, doi: 10.1086/305772

Shearer, A., Redfern, R. M., Gorman, G., et al. 1997,
ApJL, 487, L181, doi: 10.1086/310888

Shy, S., Tak, H., Feigelson, E. D., Timlin, J. D., & Babu,
G. J. 2022, AJ, 164, 6, doi: 10.3847/1538-3881/ac6e64

Skiff, B. A. 2014, VizieR Online Data Catalog, B/mk

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ,
131, 1163, doi: 10.1086/498708

Smith, M. A., Lopes de Oliveira, R., & Motch, C. 2016,
Advances in Space Research, 58, 782,
doi: 10.1016/j.asr.2015.12.032

Smolčić, V. 2009, ApJL, 699, L43,
doi: 10.1088/0004-637X/699/1/L43

Sonbas, E., Rangelov, B., Kargaltsev, O., et al. 2016, ApJ,
821, 54, doi: 10.3847/0004-637X/821/1/54

Su, Y., Zhou, X., Yang, J., et al. 2017, ApJ, 845, 48,
doi: 10.3847/1538-4357/aa7f2a

Suleimanov, V. F., Doroshenko, V., & Werner, K. 2022,
MNRAS, 511, 4937, doi: 10.1093/mnras/stac417

Szkody, P., Anderson, S. F., Brooks, K., et al. 2011, AJ,
142, 181, doi: 10.1088/0004-6256/142/6/181

Toba, Y., Oyabu, S., Matsuhara, H., et al. 2014, ApJ, 788,
45, doi: 10.1088/0004-637X/788/1/45

Tranin, H., Godet, O., Webb, N., & Primorac, D. 2022,
A&A, 657, A138, doi: 10.1051/0004-6361/202141259

Tsygankov, S. S., Wijnands, R., Lutovinov, A. A.,
Degenaar, N., & Poutanen, J. 2017, MNRAS, 470, 126,
doi: 10.1093/mnras/stx1255

Ueno, M., Bamba, A., Koyama, K., & Ebisawa, K. 2003,
ApJ, 588, 338, doi: 10.1086/368355

van der Hucht, K. A. 2001, NewAR, 45, 135,
doi: 10.1016/S1387-6473(00)00112-3

—. 2006, A&A, 458, 453, doi: 10.1051/0004-6361:20065819

Véron-Cetty, M. P., & Véron, P. 2010, A&A, 518, A10,
doi: 10.1051/0004-6361/201014188

Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A,
349, 389. https://arxiv.org/abs/astro-ph/9909315

Walter, R., Lutovinov, A. A., Bozzo, E., & Tsygankov, S. S.
2015, A&A Rv, 23, 2, doi: 10.1007/s00159-015-0082-6

Watson, M. G., Schröder, A. C., Fyfe, D., et al. 2009,
A&A, 493, 339, doi: 10.1051/0004-6361:200810534

Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS,
143, 9, doi: 10.1051/aas:2000332

Wilms, J., Allen, A., & McCray, R. 2000, ApJ, 542, 914,
doi: 10.1086/317016

Yang, H., Hare, J., Volkov, I., & Kargaltsev, O. 2021,
Research Notes of the American Astronomical Society, 5,
102, doi: 10.3847/2515-5172/abfcd4

Zhang, Y., Zhao, Y., & Wu, X.-B. 2021, MNRAS, 503,
5263, doi: 10.1093/mnras/stab744

APPENDIX

A. CLASSIFICATION PERFORMANCE METRICS

For binary classifications, a comparison of the predicted label and true label can lead to four outcomes: true positive (TP, a test result that correctly indicates the presence of a condition or attribute), true negative (TN, a test result that correctly indicates the absence of a condition or attribute), false positive (FP, a test result which wrongly indicates that a particular condition or attribute is present) and false negative (FN, a test result which wrongly indicates that a particular condition or attribute is absent). The metrics commonly used to evaluate the performance of a classifier include Accuracy, precision, recall, the F-1 Score, the Matthews correlation coefficient (MCC), the Receiver Operating Characteristic (ROC) curve, and the Confusion Matrix which are all built on the statistics of the four outcomes mentioned above.

Precision (also named as positive predictive value or efficiency) is the fraction of true positive predictions among all predicted positive outcomes; recall (also known as true positive rate or sensitivity) is the fraction of true positive predictions among all true outcomes; Specificity (also named as true negative rate) is the fraction of actual negative outcomes that have been predicted as negative; F-1 Score is the harmonic average of precision and recall:

$$\text{precision} \quad = \frac{TP}{TP+FP}, \tag{A1}$$

$$\text{recall} \quad = \frac{TP}{TP+FN}, \tag{A2}$$

$$\text{specificity} \quad = \frac{TN}{TN+FP}, \tag{A3}$$

$$\text{F-1 Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{A4}$$

In a multi-class classification problem, the metrics defined above for the binary classification can be utilized by treating one class as the positive class and all other classes as the negative class and repeating the calculations for each class. The so-called "Macro" average is used to calculate the arithmetic average of the metrics for each class. For the imbalanced TD that we have, the "Macro" average places the same weight on each class so that the metric will not be dominated by the contribution from the populous classes.

For a multi-class case, Accuracy is defined as the percentage of outcomes that are correctly predicted as their true classes with respect to the total number of outcomes. We also calculate the Balanced Accuracy which is the arithmetic average of the accuracy per class. The Balanced Accuracy is essentially a Macro average of recalls.

The MCC takes into account true and false positives and negatives and is designed to be a classification metric resistant to class imbalance (Chicco & Jurman 2020). The multi-class MCC is defined as[35]:

$$\text{MCC} = \frac{c \times s - \sum_k^K p_k \times t_k}{\sqrt{(s^2 - \sum_k^K p_k^2)(s^2 - \sum_k^K t_k^2)}} \tag{A5}$$

where t_k is the number of times class k occurred, p_k is the number of times class k predicted, c is the total number of samples correctly predicted and s is the total number of samples.

The performance of the pipeline for each source class can also be visualized using a Receiver Operator Characteristic (ROC) curve. A ROC curve shows the performance (recall vs. $1-$specificity) of a binary classifier as a function of classification confidence. A macro-averaged ROC curve can be calculated by averaging the ROC curves of each class which can be treated as the performance on the whole TD taking into account the underpopulated classes. For each ROC curve, the area under the curve (AUC) score represents an overall (averaged) performance of the classification pipeline. A perfect classifier has an AUC $= 1$ whereas a completely random classifier would have an AUC $= 0.5$.

The Confusion Matrix (CM), also known as the error matrix, is another way to visualize the performance of an algorithm using a matrix layout where each row represents the number of sources in a predicted (or actual) class while each column provides the number in an actual (or predicted) class. Since our TD is highly imbalanced, we care more

---

[35] https://scikit-learn.org/stable/modules/model_evaluation.html#matthews-corrcoef
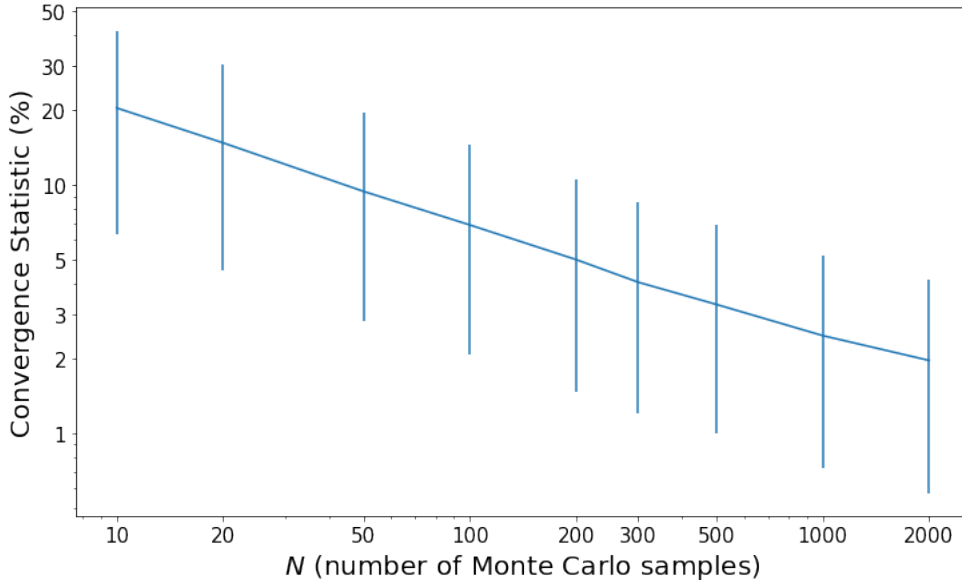
Figure 15. The median of the Convergence Statistic $\frac{|\overline{P_N} - \overline{P_{3000}}|}{\sqrt{\mathrm{Var}[P_{3000}]}}$ calculated for GCS, with the errorbars indicating the 16th and 84th percentiles (corresponding to 1-σ uncertainty), as a function of the number of MC samples N.

about the fraction of sources that are correctly predicted to each class rather than the absolute number. Therefore, we normalize the CM by the total number of sources in each row, such that each element represents the fraction of all true (predicted) sources of the row class, that are predicted (true) sources of the column class, to obtain the normalized recall (precision) CM. A more diagonal matrix represents a better performance.

B. CONVERGENCE AND UNCERTAINTY CONTRIBUTIONS IN CLASSIFICATION

While the Monte Carlo (MC) method is conceptually straightforward and very flexible, it is recognized as lacking well-established convergence criteria (Ballio & Guadagnini 2004). We implement a convergence analysis of the MC method that we use to account for the measurement uncertainties which are used to construct the classification probability distributions. Assuming the sample mean (arithmetic average) of the classification probability, $\overline{P_N} = \frac{1}{N} \sum_{i=1}^{N} P_i$, where N is the number of samples used to calculate the mean, will eventually converge to the true mean probability, we use a sample mean for a larger number ($N = 3,000$) of samples, $\overline{P_{3000}}$, as a pseudo true mean, for the convergence analysis. The sample variance of the classification probability with $N = 3,000$, $\mathrm{Var}[P_{3000}]$ is also calculated and used as a substitute for the true variance.

We define the Convergence Statistic as the absolute deviation of the sample mean of the classification probability from the pseudo true mean $\overline{P_{3000}}$, divided by the square root of the pseudo true variance, $\sqrt{\mathrm{Var}[P_{3000}]}$, calculated as a function of the number of samples N, i.e., $\frac{|\overline{P_N} - \overline{P_{3000}}|}{\sqrt{\mathrm{Var}[P_{3000}]}}$. Since we are most interested in the reliability of the predicted class, only its classification probability (the highest classification probability among 8 classes) is used in the calculation. Figure 15 shows the median of the Convergence Statistics calculated for GCS, with errorbars indicating the 16th and 84th percentiles, as a function of the number of MC samples N. Given our computational resource limitations we set the number of MC samples $N = 1000$ such that 84% of the Convergence Statistics values (for GCS) are less than 5%. We note that even if N is reduced by a factor of 10 ($N = 100$), 84% of the Convergence Statistics values are still less than 14%.

We evaluate the contributions to the uncertainties of the classification probabilities from three processes that involve random sampling: 1) MC sampling to account for the feature measurement uncertainties; 2) oversampling with the SMOTE algorithm; and 3) the Random Forest algorithm. To evaluate the relative contributions to the classification uncertainties from each process, we enable randomness in one process but not in the others and repeatedly run the classification pipeline 1000 times for each source in GCS. We then also run the classification pipeline with the randomness enabled for all three processes to be used as the benchmark. For each of these three sampling processes,
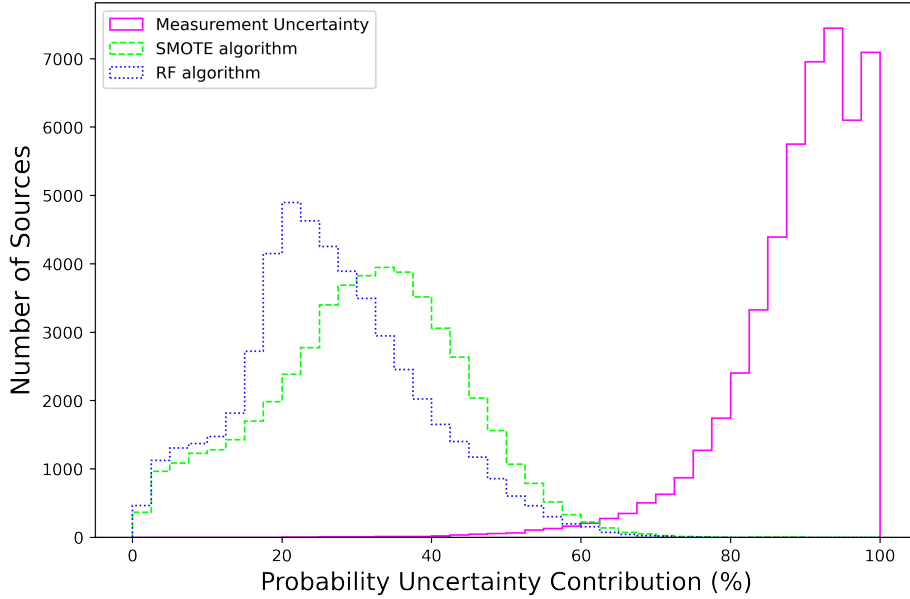
Figure 16. The distributions of the probability uncertainty contributions for GCS demonstrating the relative importance of various uncertainty factors.

we calculate the sample mean and the sample error (the square root of the sample variance) of the classification probability for the predicted class. For each source we evaluate three classification probability uncertainties (defined as the standard deviation of the classification probabilities characterizing the width of the distributions; see Figure 7) for the predicted class, which represent contributions to the total probability uncertainty from the measurement uncertainties, SMOTE algorithm, and RF algorithm, respectively. The individual contribution is calculated by dividing the probability uncertainty from one process by the combined probability uncertainty from three processes which is calculated by adding the three probability uncertainties in quadrature. The distributions of the probability uncertainty contributions are shown in Figure 16. The mean and the standard deviation of the distribution for the uncertainty contributions are 89±9%, 31±13%, and 27±12% for the measurement uncertainties, SMOTE algorithm and the RF algorithm respectively. The main uncertainty contribution is the measurement uncertainty. The SMOTE method and RF algorithm contribute comparable uncertainties which are less than the contribution from the measurement uncertainty. This underscores the importance of including the measurement uncertainty in the classification procedure.

It is likely that increasing n_estimators can decrease the uncertainty contribution from the RF algorithm. Thus, we recommend using larger n_estimators (e.g., 1000) when a smaller number of MC samplings (≤ 10) is chosen, e.g., to reduce the computing time.

C. CHANCE COINCIDENCE ESTIMATION FOR POSITIONAL CROSS-MATCHING

In this section, we estimate the potential rates of accidental matches of the MW counterparts to GCS. The X-ray sources are matched to MW catalogs (including *Gaia* EDR3, 2MASS, AllWISE, CatWISE2020, unWISE) using the combined X-ray and MW PUs. The chance coincidence probability for a field with an average MW catalog source density, ρ (arcsec^{-2}), and 2-σ combined PU uncertainty radius, r ($''$), is given by $P_{\mathrm{chance}} = 1 - \exp(-\rho \pi r^2)$, which is simply the probability of having one or more MW catalog source within a randomly placed circle of the chosen radius. We count all MW catalog sources within a larger ($R = 10''$) radius to calculate the field source density, $\rho = \frac{N}{A}$ where N is the number of MW catalog sources within the $R = 10''$ circle and $A = \pi R^2$ is the area. We then calculate the chance coincidence probabilities for each source in GCS and plot their distributions in Figure 17 with the average (per MW catalog) chance coincidence probability shown in the legend. The calculation presented here only offers a rough estimation and a variable radius (or other methods), for measuring the local source density, may be needed for more accurate estimates.
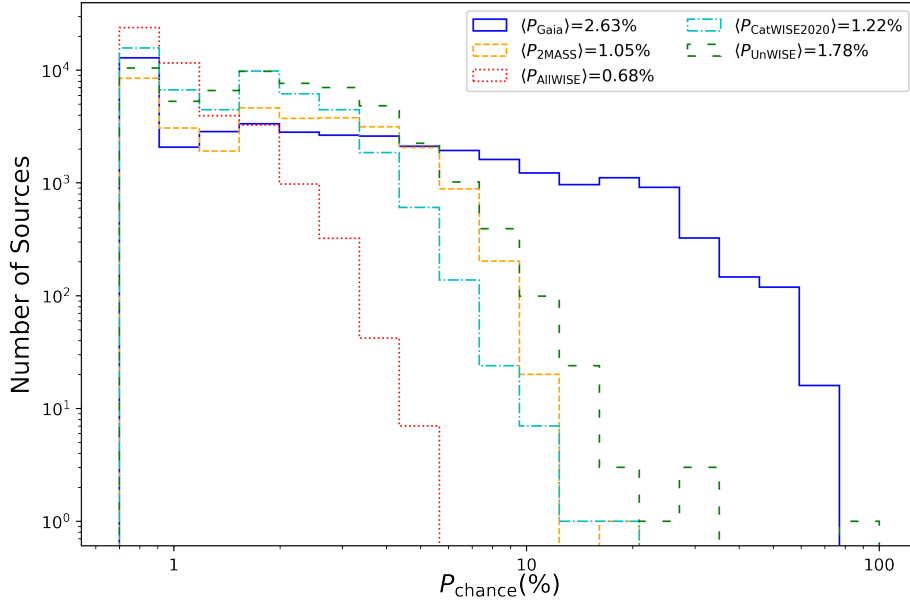
Figure 17. The distribution of the chance coincidence probability for GCS. The average chance coincidence probability calculated for each MW catalog is listed in the legend.

D. COMPUTATIONAL EFFICIENCY

A single classification within one extinction bin, requires ∼30 s with one core on The George Washington University Pegasus cluster. Classification of multiple sources sharing the same extinction bin takes almost the same time as for a single source. To classify all CSCv2 sources with ∼2000 extinction/absorption bins 1,000 times using 20 Nano nodes (each node with a Dual 20-Core 3.70GHz Intel Xeon Gold 6148 processor) on the Pegasus cluster running in parallel takes about one day. For comparison, a typical laptop (e.g., Acer Swift 3 with AMD Ryzen 7 4700U Processor, and an 8GB LPDDR4 memory) takes about 13 days to produce the same classification results for ∼66,000 GCS sources with 1,000 MC samplings. Therefore, although it is very much possible to run the classification pipeline on the laptop, we recommend to reduce the number of MC samplings by a factor of 10 if the number of source to be classified is large.

E. EXAMPLES OF CLASSIFIED SOURCES WITH THEIR PROPERTIES AND COLUMN DESCRIPTIONS FOR THE MACHINE READABLE TABLE

The properties and the classification results of GCS, TD, and of the X-ray sources within the unidentified HESS sources are available in the form of three machine readable tables (MRTs). A subset of each MRT is shown Table 8, Table 9 and Table 10. We also provide the list and description of columns used in the three MRTs.

CSCv2_name: CSCv2 source name in the form "2CXO Jhhmmss.s{+|−}ddmmss". *Column 1.* is the shortened version of **CSCv2_name** in Table 8, Table 9 and Table 10.

CSCv2_RA, CSCv2_DEC: CSCv2 master source J2000 Right Ascension and Declination, in degrees.

CSCv2_PU: CSCv2 master source err_ellipse_r0, the major radius of the 95% confidence level position error ellipse, in arcsecs, *Column 2.* in Table 8, Table 9 and Table 10.

CSCv2_S/N: CSCv2 master source X-ray significance, *Column 3.* in Table 8, Table 9 and Table 10.

Fb, Fs, Fm, Fh: Average broad, soft, medium, and hard band fluxes (see their calculations in Section 2.2), in unit of $\mathrm{erg\,s^{-1}\,cm^{-2}}$. *Column 4.* is **Fb** in unit of $10^{-13}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ in Table 8, Table 9 and Table 10.

P_intra: Highest value of Kuiper's test variability probability across all observations available in CSCv2, *Column 5.* in Table 8, Table 9 and Table 10.

P_inter: Inter-observation variability probability (see its calculation in Section 2.2), *Column 6.* in Table 8, Table 9 and Table 10.

Gmag, BPmag, RPmag: *Gaia* EDR3 G, BP, and RP band magnitudes. *Column 7.* is **Gmag** in Table 8, Table 9 and Table 10.

Jmag, Hmag, Kmag: 2MASS J, H, and K band magnitudes. *Column 8.* is **Jmag** in Table 8, Table 9 and Table 10.
W1mag, W2mag, W3mag: W1, W2, and W3 band magnitudes from AllWISE, CatWISE2020 and unWISE catalogs (see their definitions in Section 2.3). *Column 9.* is **W1mag** in Table 8, Table 9 and Table 10.
Plx: *Gaia* EDR3 absolute stellar parallax, in mas.
PM: *Gaia* EDR3 total proper motion, in $\mathrm{mas\,yr}^{-1}$.
rgeo, b_rgeo, B_rgeo: Median, 16th percentile, and 84th percentile of the geometric distance from *Gaia* EDR3 Distances catalog, in pc.
P_AGN, P_CV, P_HM-STAR, P_HMXB, P_LM-STAR, P_LMXB, P_NS, P_YSO: Classification probability of the X-ray source to be classified as each of the eight classes in our TD.
Class: The predicted class of the source with the highest classification probability among eight classes, *Column 10.* in Table 8, Table 9 and Table 10.
Class_prob: The classification probability of the predicted class calculated from MUWCLASS, *Column 11.* in Table 8, Table 9 and Table 10.
CT: The classification confidence threshold (see its definition in Section 4.5), *Column 12.* in Table 8, Table 9 and Table 10.
CSCv2_flags: The compilation of CSCv2 master source flags including conf_flag ("conf"), extent_flag ("extent"), and pileup_flag ("pileup"), jointed by a "|", *Column 14.* in Table 9.
Catalog_name: The name of the source from the literature verified catalogs for the classification of TD sources. This column is only available for the TD MRT and the MRT of the HESS field sources.
True_Class: The true class of the source from the TD, *Column 13.* in Table 9. This column is only available for the TD MRT and the MRT of the HESS field sources.
Class_ref: The reference of the classifications of the TD sources. This column is only available for the TD MRT and the MRT of the HESS field sources.
HESS_name: The HESS source name that the CSCv2 source resides in. *Column 13.* is the shortened version of **HESS_name** in Table 10. This column is only available for the MRT of the HESS field sources.
Columns of 1-σ uncertainty are are available for **Fb, Fs, Fm, Fh, Gmag, BPmag, RPmag, Jmag, Hmag, Kmag, W1mag, W2mag, W3mag, Plx, P_AGN, P_CV, P_HM-STAR, P_HMXB, P_LM-STAR, P_LMXB, P_NS, P_YSO, Class_prob** columns which are **e_Fb, e_Fs, e_Fm, e_Fh, e_Gmag, e_BPmag, e_RPmag, e_Jmag, e_Hmag, e_Kmag, e_W1mag, e_W2mag, e_W3mag, e_Plx, e_P_AGN, e_P_CV, e_P_HM-STAR, e_P_HMXB, e_P_LM-STAR, e_P_LMXB, e_P_NS, e_P_YSO, e_Class_prob** columns, respectively.

Table 8. A subset of GCS MRT.

2CXO name[a]	PU[b]	S/N[c]	F_b	P_{intra}	P_{inter}	G	J	W1	Class	Class Prob	CT
	arcsec		10^{-13} cgs[d]			mag	mag	mag		%	
J123635.2+621151	0.79	4.53	0.0038 ± 0.0006	...	0.224	...	...	...	AGN	91 ± 5	10.475
J141520.7+521114	0.86	3.33	0.02 ± 0.005	...	...	...	...	17.19 ± 0.09	AGN	40 ± 10	0.296
J031818.8−663230	0.72	7.87	0.12 ± 0.01	0.98	0.498	13.862 ± 0.007	12.22 ± 0.02	11.62 ± 0.02	LM-STAR	81 ± 15	2.543
J013647.4+154744	0.71	13.19	0.107 ± 0.006	1	0.996	...	10.93 ± 0.02	9.95 ± 0.02	LM-STAR	91 ± 6	9.562
J161702.0−225834	0.72	11.03	0.25 ± 0.03	0.924	...	14.796 ± 0.008	...	...	HM-STAR	39 ± 10	0.381
J201641.4+370925	0.93	9.16	0.33 ± 0.04	...	...	18.45 ± 0.003	15.59 ± 0.07	13.39 ± 0.02	HMXB	61 ± 11	2.587
J142112.1−624156	0.72	7.4	3.1 ± 0.5	0.979	...	16.192 ± 0.003	13.3 ± 0.02	11.75 ± 0.01	HMXB	82 ± 10	4.836
J085910.9−434343	0.73	16.12	1.2 ± 0.08	1	...	19.38 ± 0.03	14.48 ± 0.03	11.09 ± 0.03	HMXB	64 ± 9	2.106
J193309.5+185902	0.72	14.94	1.8 ± 0.1	0.906	...	12.616 ± 0.003	10.32 ± 0.02	9.11 ± 0.02	HMXB	69 ± 14	2.785
J171556.4−385154	0.93	3.06	3 ± 1	...	...	...	15.06	...	HMXB	57 ± 16	1.557

NOTE—The 1-σ uncertainties are shown whenever available. [a] – the shortened 2CXO name from the CSCv2 by removing the leading "2CXO" string; [b] – the 2-σ X-ray positional uncertainty; [c] – the CSCv2 master source X-ray significance; [d] – the energy flux in cgs unit of erg s^{-1} cm^{-2}.

Table 9. A subset of TD MRT.

2CXO name	PU	S/N	F_b	P_{intra}	P_{inter}	G	J	W1	Class	Class Prob	CT	True_Class	CSC_flags
	arcsec		10^{-13} cgs			mag	mag	mag		%			
J084621.1+013755	0.71	21.31	5.3 ± 0.3	0.9	$\cdots$	6.27 ± 0.007	1.5 ± 0.3	-1.8	YSO	36 ± 10	0.233	LM-STAR	$\cdots$
J112401.1−365319	0.73	10.27	0.58 ± 0.07	0.973	$\cdots$	$\cdots$	$\cdots$	$\cdots$	AGN	57 ± 8	2.867	LMXB	$\cdots$
J203213.1+412724	0.72	8.11	0.23 ± 0.02	0.943	1	11.28 ± 0.01	9.52 ± 0.02	8.48 ± 0.02	HM-STAR	57 ± 8	2.636	HMXB	extent|conf

Note—The 1-σ uncertainties are shown whenever available. The first 12 columns are the same as in Table 8.

Table 10. A subset of HESS field source MRT.

2CXO Name	PU	S/N	F_{b}	P_{intra}	P_{inter}	G	J	W1	Class	Class Prob	CT	HESS name[a]
	arcsec		10^{-13} cgs			mag	mag	mag		%		
J191237.9+101044*	1.58	14.66	2.7 ± 0.2	0.762	⋯	⋯	⋯	⋯	NS	78 ± 15	2.585	J1912+101
J161610.8−515545	2.02	7.15	0.47 ± 0.07	0.136	⋯	⋯	⋯	⋯	NS	85 ± 16	2.656	J1614−518
J184201.9−052823	1.36	6.22	0.51 ± 0.09	0.629	⋯	⋯	⋯	⋯	NS	84 ± 17	2.275	J1841−055
J180010.0−240129	0.78	11.39	0.4 ± 0.04	0.515	⋯	⋯	15.01 ± 0.09	⋯	NS	73 ± 11	3.254	J1800−240
J185643.6+021921	0.72	7.21	3.0 ± 0.4	0.026	⋯	⋯	⋯	⋯	NS	83 ± 17	2.185	J1857+026
J184850.3−012509	0.92	11.51	0.47 ± 0.05	0.052	⋯	⋯	⋯	⋯	NS	83 ± 15	2.523	J1848−018
J184335.8−034653*	1.23	8.44	0.27 ± 0.03	0.62	⋯	⋯	⋯	⋯	NS	87 ± 16	2.636	J1843−033
J181303.0−124907*	0.77	16.68	1.39 ± 0.09	0.913	⋯	⋯	16.12 ± 0.09	13.68 ± 0.04	AGN	95 ± 4	12.802	J1813−126
J101812.9−585930	0.72	35.62	3.8 ± 0.1	0.98	1	⋯	15.96	13.191 ± 0.009	AGN	90 ± 4	13.824	J1018−589
J101855.6−585645*	0.71	58.64	9.6 ± 0.2	0.872	1	12.269 ± 0.003	10.44 ± 0.02	9.9 ± 0.02	HMXB	74 ± 15	2.116	J1018−589

Note—The 1-σ uncertainties are shown whenever available. The first 12 columns are the same as in Table 8. * – Those sources are extended sources according to CSC_flags. [a] – the shortened HESS source from the HGPS by removing the leading "HESS" string.