Identifying the Physical Origin of Gamma-Ray Bursts with Supervised Machine Learning

Jia-Wei Luo1,2,3, Fei-Fei Wang4, Jia-Ming Zhu-Ge2,3, Ye Li5, Yuan-Chuan Zou6, and Bing Zhang2,3

1 College of Physics and Hebei Key Laboratory of Photophysics Research and Application, Hebei Normal University, Shijiazhuang, Hebei 050024, People’s Republic of China; ljw@hebtu.edu.cn
2 Nevada Center for Astrophysics, University of Nevada, Las Vegas, NV 89154, USA
3 Department of Physics and Astronomy, University of Nevada, Las Vegas, NV 89154, USA
4 School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, People’s Republic of China
5 Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 100012, People’s Republic of China
6 Department of Astronomy, School of Physics, Huazhong University of Science and Technology, Wuhan 430074, People’s Republic of China

Received 2023 July 28; revised 2023 September 23; accepted 2023 October 14; published 2023 December 5

Abstract

The empirical classification of gamma-ray bursts (GRBs) into long and short GRBs based on their durations is already firmly established. This empirical classification is generally linked to the physical classification of GRBs originating from compact binary mergers and GRBs originating from massive star collapses, or Type I and II GRBs, with the majority of short GRBs belonging to Type I and the majority of long GRBs belonging to Type II. However, there is a significant overlap in the duration distributions of long and short GRBs. Furthermore, some intermingled GRBs, i.e., short-duration Type II and long-duration Type I GRBs, have been reported. A multiparameter classification scheme of GRBs is evidently needed. In this paper, we seek to build such a classification scheme with supervised machine-learning methods, chiefly XGBOOST. We utilize the GRB Big Table and Greiner’s GRB catalog and divide the input features into three subgroups: prompt emission, afterglow, and host galaxy. We find that the prompt emission subgroup performs the best in distinguishing between Type I and II GRBs. We also find the most important distinguishing features in prompt emission to be $T_{90}$, the hardness ratio, and fluence. After building the machine-learning model, we apply it to the currently unclassified GRBs to predict their probabilities of being either GRB class, and we assign the most probable class of each GRB to be its possible physical class.

Unified Astronomy Thesaurus concepts: Gamma-ray bursts (629); Astronomy data analysis (1858)

Supporting material: machine-readable table

1. Introduction

Dating from the early days of gamma-ray burst (GRB) study, a clear bimodal distribution has been identified in their durations (Kouveliotou et al. 1993). Two classes of GRBs are then proposed, based on their durations, namely long GRBs (LGRBs) and short GRBs (SGRBs). The commonly used criterion is based on $T_{90}$, the time within which 90% of the fluence of the GRB is observed, with the dividing point set to be $T_{90} = 2 \text{s}$. LGRBs are thought to be produced by the core collapse of massive stars (Woosley 1993), and this theory has subsequently been supported by direct observational evidence of the association of some LGRBs with Type Ic supernovae (SNe; Galama et al. 1998; Woosley & Bloom 2006). SGRBs are thought to originate from compact star mergers (Eichler et al. 1989), and this theory has been supported by the multimessenger observations of the binary neutron-star merger event GW170817/GRB 170817A (Abbott et al. 2017a, 2017b, 2017c; Goldstein et al. 2017; Zhang et al. 2018).

However, this dichotomy is far from perfect. A significant overlap is present in the duration distributions of LGRBs and SGRBs, and the duration itself is dependent on the energy band in which it is measured (Mukherjee et al. 1998; Hakki et al. 2003; Horváth et al. 2006; Zhang & Choi 2008; Veres et al. 2010; Qin et al. 2012; Bromberg et al. 2013; Zhang et al. 2016). Moreover, there are some short-duration GRBs thought to be possibly produced by core-collapse massive stars (Antonelli et al. 2009; Greiner et al. 2009; Salvaterra et al. 2009; Tanvir et al. 2009; Zhang et al. 2009, 2021; Guelbenzu et al. 2011; Ahumada et al. 2021; Rossi et al. 2022), as well as some long-duration GRBs thought to possibly originate from compact star mergers (Della Valle et al. 2006; Fynbo et al. 2006; Gal-Yam et al. 2006; Gehrels et al. 2006; Zhang et al. 2007; Rastinejad et al. 2022; Troja et al. 2022; Yang et al. 2022b; Sun et al. 2023).

The existence of these “intermingled” GRBs challenges the practice of classifying GRBs solely based on duration, as well as the names of “long” and “short” GRBs. It is then apparent that more sophisticated classification criteria involving multiple observational parameters are needed. Throughout this study, we refer to the GRB classes based on their physical origins: namely, Type I for compact merger GRBs and Type II for collapsar GRBs, following the classification scheme of Zhang (2006) and Zhang et al. (2009). Many other schemes for GRB classification have also been put forward (e.g., Zhang et al. 2009, 2012; Lü et al. 2010, 2014; Bromberg et al. 2013; Li et al. 2016, 2020; Yang et al. 2016; Kulkarni & Desai 2017; Minaev & Pozanenko 2020), yet the classification of LGRBs and SGRBs is still largely based on community consensus, and there is a lack of objective classification models with minimal human interference.
In this case, machine learning comes in handy. Capable of automatically generating results without human input after training, machine learning can help us to fathom the differences between Type I and II GRBs, as well as aid us in the classification of newly discovered GRBs. Machine learning has already been widely adopted in the study of GRBs (e.g., Horváth et al. 2006, 2019; Rípa et al. 2012; Huertas-Company et al. 2015; Tarnopolski 2015; Modak et al. 2018; Jespersen et al. 2020; Modak 2021; Bhave et al. 2022; Salmon et al. 2022a, 2022b; Tarnopolski 2022; Steinhardt et al. 2023). However, the abovementioned studies predominantly use machine-learning methods of the unsupervised type, where only the observed features of the GRBs are inputted into the models, but not the labels (the GRBs’ physical classes being Type I or II). On the other hand, the other type of machine-learning methods, supervised methods, are also commonly employed by astronomy researchers in the classification of other astronomical objects (e.g., Luo et al. 2023; Zhu-Ge et al. 2023; Connor & van Leeuwen 2018; Butter et al. 2022; Coronado-Bláñquez 2022; de Beurs et al. 2022; Fan et al. 2022; Villa-Ortega et al. 2022; Yang et al. 2022a; Kaur et al. 2023), although studies on the application of supervised methods on GRB are scarce. Since supervised methods take both features and labels as input, and can produce deterministic predictions of the class of new GRBs, they can be helpful in identifying the true physical origin of intermingled GRBs.

In this study, we apply supervised machine-learning methods to the classification of Type I and II GRBs. The machine-learning model we use is the eXtreme Gradient Boosting (XGBOOST) classifier (Chen & Guestrin 2016). We employ XGBOOST, as it is one of the most popular and successful machine-learning frameworks to date and can handle missing input values natively. The GRB catalog we use contains many missing values, so the ability to handle them is vital. In Section 2, we introduce the GRB catalogs we utilize and the machine-learning methods we use. In Section 3, we present the classification results and feature the importance of the machine-learning models. In Section 4, we attempt to predict the classes of the unclassified GRBs. Finally, in Section 5, we put forward our conclusions and discuss the classifications of some recently discovered possible intermingled GRBs.

2. Data and Methods

We use an updated version of the GRB Big Table (Wang et al. 2020; F.-F. Wang et al. 2023, in preparation), which contains 7179 GRBs ranging from 1991 April 21 to 2021 July 8. Greiner’s GRB catalog, on the other hand, has 2261 GRBs in the same time range. We match the two catalogs, requiring $T_{90}$ of the selected GRBs in the Big Table to be known, and we label the GRBs based on their labels in Greiner’s catalog. GRBs with “S” at the end of their names are marked as Type I GRBs, while the others are marked as Type II GRBs. We also adopt the consensus classification of some intermingled GRBs: Type II GRB 090426 (Antonelli et al. 2009; Guelbenzu et al. 2011), Type I GRB 060505 (Fynbo et al. 2006), and Type I GRB 060614 (Fynbo et al. 2006; Gal-Yam et al. 2006; Gehrels et al. 2006; Zhang et al. 2007). This leaves us with 144 Type I and 1761 Type II GRBs. We acknowledge that this matching method substantially reduces the size of our sample, but the unmatched GRBs do not have many known features to begin with. Therefore, we did not discard too much information.

The classification inputs of our model are based on Greiner’s catalog, which collected the community consensus based on both $T_{90}$ and afterglow/host galaxy information as presented in the literature. It is possible that a small fraction of bursts have been misclassified, but the very strength of our machine-learning model is that it considers all the classifications of the input training sample. If there are a few GRBs that are wrongly classified, they would not have a significant impact on the overall accuracy.

In this study, we pay special interest to the intermingled GRBs. We define intermingled GRBs as GRBs classified as Type I in Greiner’s catalog but with $T_{90}$ values $>2$ s in the Big Table or as GRBs classified as Type II in Greiner’s catalog but with $T_{90} < 2$ s. There are 21 intermingled Type I GRBs and 59 intermingled Type II GRBs in our sample.

We also scrutinize the possible third intermediate GRB type proposed by some studies (e.g., Horvah 1998; Mukherjee et al. 1998; Hakkila et al. 2000, 2003; Balaguer et al. 2001; Horváth et al. 2006, 2008, 2010, 2018; Chattopadhyay et al. 2007; Rípa et al. 2009, 2012; Huja et al. 2009; Veres et al. 2010; Koen & Bere 2012; Zitouni et al. 2015; Kulkarni & Desai 2017) based on the $T_{90}$ distributions of GRBs, by creating an intermediate sample consisting of GRBs with $T_{90}$ within 1–4 s. There are 31 intermediate Type I and 100 intermediate Type II GRBs in our sample.

We then divide the features in the Big Table into three subgroups: prompt emission, afterglow, and host galaxy. The features in the three subgroups are listed in Tables 1–3. Three subsamples are subsequently created by requiring each GRB in the subsamples to have at least one feature other than $T_{90}$ in the corresponding feature group to be known. We also divide each subsample into training sets and test sets with a 7:3 ratio, while keeping the ratio of Type I to Type II GRBs the same in the training sets and test sets. The training sets are used to train the machine-learning model, while the test sets are used to test the performance of the model after it is trained.

While it is common practice to impute the missing values in the data with some type of algorithm, we find that imputation introduces false information in the feature importance we later calculate, which is also suggested by some other studies (e.g., Seijo-Pardo et al. 2019; Yu et al. 2022). Since the XGBOOST classifier (Chen & Guestrin 2016) can automatically handle missing values, we simply input our data without imputation.

Then, we note that the Type I and Type II GRBs in our sample are significantly imbalanced by a ratio of $\sim 1: 10$. Because this apparent ratio could be caused by selection effects, we should not introduce this ratio to our training data. However, the commonly used synthetic minority oversampling technique (SMOTE; Chawla et al. 2002) cannot be applied to data with missing values. Instead, we assign different sample weights for the two classes calculated with a balanced sample weight implemented in SCIKIT-LEARN (Pedregosa et al. 2011):

$$w_i = \frac{N}{kn_i},$$

where $w_i$ is the sample weight of the $i$th class, $N$ is the total number of data points, $k$ is the number of classes (in this study, two), and $n_i$ is the number of data points in the $i$th class.

Finally, we input the training sets into the XGBOOST classifier to train the machine-learning model. After training,
Spectral alpha

For features with multiple descriptions, we choose the one with the most known data. Directly measured features are listed above the horizontal line, while the derived features are listed below the line.

Table 1
List of Features Used in the Prompt Emission Subgroup

| Feature Name | Unit | Description | Log |
|--------------|------|-------------|-----|
| T90          | s    | Time within which 90% of the fluence of the GRB is observed | Y |
| F_g          | 10^{-6} erg cm^{-2} | Fluence in the 20–2000 keV energy band | Y |
| HR           | 1000–2000 keV | HR between 1000–2000 keV and 20–100 keV | Y |
| F_{pk1}      | 10^{-6} erg cm^{-2} | Peak flux in the 1 s time bin in the rest-frame 1–10 \times 10^{4} keV energy band | Y |
| F_{pk4}      | photon cm^{-2} | Peak photon flux in the 1 s time bin of 10–1000 keV | Y |
| alpha Band   | ...  | Low-energy spectrum index of the Band model | N |
| beta Band    | ...  | High-energy spectrum index of the Band model | N |
| E_P Band     | keV  | Spectral peak energy of the Band model | Y |
| alpha CPL    | ...  | Spectrum index of the cutoff power-law (CPL) model | N |
| E_P CPL      | keV  | Spectral peak energy of the cutoff-power-law (CPL) model | Y |
| alpha SPL    | ...  | Spectrum index of the simple power-law (SPL) model | N |
| spectral lag | ms MeV^{-1} | Spectral time lag | N |

Note. For features with multiple descriptions (e.g., variability and F_{pk}), we choose the one with the most known data. Directly measured features are listed above the horizontal line, while the derived features are listed below the line.

Table 2
List of Features Used in the Afterglow Subgroup

| Feature Name | Unit | Description | Log |
|--------------|------|-------------|-----|
| theta_j      | rad  | Jet opening angle | Y |
| Gamma0       | ...  | Initial Lorentz factor | Y |
| log_t_burst  | s    | Duration of the GRB central engine | Y |
| t_b          | d    | Jet break time | Y |
| F_{X11hr}    | s    | Flux density in the X-ray band 11 hr after the trigger time of the burst | Y |
| beta_X11hr   | ...  | Index in the X-ray band 11 hr after the trigger time of the burst | N |
| F_{Opt11hr}  | s    | Flux density in the optical band 11 hr after the trigger time of the burst | Y |
| T_{ai}       | s    | Rest-frame time at the end of the plateau phase in log t in X-ray | Y |
| L_a          | erg s^{-1} | Isotropc X-ray luminosity at the time Ta | Y |

Table 3
List of Features Used in the Host Galaxy Subgroup

| Feature Name | Unit | Description | Log |
|--------------|------|-------------|-----|
| offset       | kpc  | Distance from the burst location to the center of the host galaxy | Y |
| metallicity  | ...  | Metallicity of the host; the value is 12 + log[O/H] | N |
| A_{V}        | ...  | Dust extinction | N |
| SFR          | M_{\odot} yr^{-1} | SFR | Y |
| SSFR         | Gyr^{-1} | Specific SFR | Y |
| Age          | Myr  | The age of the GRB host galaxy | Y |
| Mass         | M_{\odot} | Stellar mass | Y |

The F_1 score is then calculated as the harmonic mean of the two metrics:

\[
F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)
\]

The resulting F_1 score is a value between 0 and 1, with 0 meaning total failure in predicting the correct labels for the test set and 1 meaning 100% accuracy in predicting the labels. Since the F_1 score is a stricter metric than accuracy, F_1 scores are usually significantly smaller than the accuracy scores calculated on the same classification results.

To test which input feature has the best capability in distinguishing between Type I and II GRBs, we use SHapley Additive exPlanations (SHAP; Lundberg & Lee 2017; Lundberg et al. 2020) to calculate the feature importance of the input features. For each data point, SHAP estimates the contribution toward the output result from each input feature in the form of SHAP values. Readers can refer to the abovementioned references for more detailed and mathematical descriptions of SHAP. In contrast to the also commonly used permutation feature importance (Breiman 2001; Altmann et al. 2010; Fisher et al. 2019), which generates a single feature importance value for a feature across all data points, SHAP can analyze the prediction contribution of features on individual data points. When the SHAP values from all the data points are combined,
SHAP can show not only the importance of the input features but also in which direction the feature values of each input data point draw the final output.

Figures 1(d) and 1(e) show an example of the results from SHAP. In Figure 1(d), the SHAP values from each individual data point are taken to be absolute values and averaged across

Figure 1. Examples of confusion matrices and SHAP feature importance values of the prompt emission subgroup.
different features. The length of each bar in the figure shows how important is each feature to the prediction result in general. Figure 1(e), on the other hand, shows the individual SHAP values of each feature in each data point. In this beeswarm plot, the X-axis shows the SHAP values, with higher SHAP values leaning toward Type I and lower SHAP values leaning toward Type II. The feature values of each data point are also shown with the color of the points so that readers can know in which direction a higher or lower value in one feature draws the prediction to.

3. Results

3.1. Prompt Emission

Many studies have suggested adding the hardness ratio (HR) to the $T_{90}$ classification criterion to form two-dimensional criteria will yield better results (e.g., Horváth et al. 2006, 2010, 2018; Ripa et al. 2012; Zhang et al. 2012, 2022; Bhat et al. 2016; Yang et al. 2016; Tarnopolski 2019). Similarly, the power-law index or peak energy $E_p$ of the spectrum of prompt emission can also take the place of the HR (Goldstein et al. 2010; Nava et al. 2011; Zhang et al. 2012).

In general, Type I GRBs have harder spectra compared with Type II GRBs. Goldstein et al. (2010) further propose classification on the $E_p$-fluence plane. Since fluence is highly related to duration, this scheme also follows the HR–$T_{90}$ scheme.

Some other studies (e.g., Zhang et al. 2009, 2012; Qin & Chen 2013; Tsutsui et al. 2013; Mineev & Pozanenko 2020) suggest that the famous Amati relation (Amati et al. 2002, 2009; Kumar & Zhang 2015) of the peak energy $E_p$ and the isotropic energy $E_{iso}$ of GRB prompt emission are different for Type I and II GRBs, and thus the $E_p$-$E_{iso}$ plane can be used to distinguish between Type I and II GRBs.

In addition, Gehrels et al. (2006), Norris & Bonnell (2006), Yi et al. (2006), Zhang (2006), Ukwatta et al. (2010), Mineev et al. (2014), Bernardini et al. (2015), and Shao et al. (2017) propose to classify Type I and II GRBs based on the spectral lag $\tau$ and the $\tau$-peak luminosity $L_{\tau}$ plane, where Type I GRBs have smaller spectral lags and peak luminosities.

Association with SNe is also a very important distinguishing factor between Type I and II GRBs, as SN associations provide smoking-gun evidence of the GRB progenitor. However, the Big Table only contains SN association information for 22 GRBs. For those GRBs without SN association information, it is unknown whether there truly was no SN associated with the GRB, whether there simply was no observation, or, most likely, whether there was an optical observation, but the SN was outshone by the bright optical afterglow. We find that including SN association in our model results in significantly lower $F_1$ scores. While the model correctly classifies GRBs with SN detection as Type II GRBs, including SN in our model also makes it more likely to classify GRBs without SN detection as Type I. Since most GRBs do not have SN detection in our data, because they are too far away for SN detection, and because there are more Type II GRBs than Type I, including SN will yield worse results. Furthermore, the model without SN can correctly classify almost all SN-associated GRBs as Type IIs. Therefore, we do not include the SN information in our model.

With the prompt emission subgroup, we are able to obtain an $F_1$ score of 0.758 on the test set, 0.667 on the intermingled GRBs, and 0.821 on the intermediate GRBs. The corresponding confusion matrices and feature importance are shown in Figure 1. Our model can predict most GRBs correctly based on prompt emission data, and $T_{90}$ is the most prominent feature, with feature importance much higher than the other features, as shorter $T_{90}$ pulls the predictions toward Type I. Since the intermingled GRBs are the ones that defy the classification based on $T_{90}$, the major features that cause their classifications to be different will be the features that have high feature importance other than $T_{90}$. The same is true for intermediate GRBs; if $T_{90}$ cannot classify them clearly, then they will be classified based on other important features.

However, when we remove $T_{90}$ from the prompt emission subgroup and carry out the same analysis, while we get a lower $F_1$ score of 0.581 on the test set, we also get a higher $F_1$ score of 0.833 on the intermingled GRBs. The $F_1$ score for intermediate GRBs is at a similar value of 0.888. The corresponding confusion matrices and feature importance are shown in Figure 2. This shows that $T_{90}$ can be misleading to the machine-learning model for intermingled GRBs, and multiple observational parameters are needed for more accurate classifications of GRBs.

We also find the fluence $F_{\gamma}$ and HR to be the most important features after $T_{90}$. A lower fluence and a higher HR pull the predictions toward Type I. Since fluence is directly related to the duration, our results confirm the findings of other studies.

In order to measure the importance of other features, we further exclude the fluence and HR from our feature group and carry out the same machine-learning analysis. We obtain an $F_1$ score of 0.485 on the test set, 0.815 on the intermingled GRBs, and 0.780 on the intermediate GRBs. The corresponding confusion matrices and feature importance are shown in Figure 3. While the general $F_1$ score drops again, the $F_1$ scores for intermingled and intermediate samples remain high. The most important features are again related to the spectral shape, such as $E_{iso}^{p1}$, $E_{iso}^{a1}$, $\alpha_{p1}$, and $\alpha_{a1}$. The flux-related features of $L_{pk}$, $F_{pk1}$, and $E_{pk4}$ are also important, as well as redshift and spectral lag. Generally, a harder spectrum, a lower flux, a shorter spectral lag, and a lower redshift pull the predictions toward Type I.

3.2. Afterglow

Gehrels et al. (2008), Nysewander et al. (2009), D’Avanzo et al. (2012), and Margutti et al. (2013) have pointed out that the afterglows of Type I GRBs mostly have lower X-ray luminosities and energies. The X-ray luminosities and energies of Type I GRBs also decay faster. There are also correlations among the afterglow X-ray energy, X-ray afterglow luminosity, prompt emission isotropic energy $E_{iso}^{\tau}$, peak luminosity $L_{\tau}$, and peak energy $E_{p1}$. Combined with the findings mentioned in Section 3.1, the X-ray afterglow luminosity can also be employed for GRB classification.

Kann et al. (2011) found that similar to X-ray, optical afterglows of Type I GRBs are significantly fainter than those of Type II GRBs, and similar afterglow–prompt emission correlations also exist in the optical band.

With the afterglow subgroup, we are able to obtain an $F_1$ score of 0.353 on the test set, 0.857 on the intermingled GRBs, and 0.75 on the intermediate GRBs. The corresponding confusion matrices and feature importance are shown in Figure 4. The model clearly classifies Type I and II GRBs based on the $F_1$ score, $E_{p1}$, and $L_{pk}$.
Figure 4. We found the most important feature to be the 11 hr beta index in X-rays. The 11 hr fluxes in the X-ray and optical bands are also important. A higher beta index and lower X-ray and optical fluxes pull the predictions toward Type I, consistent with other studies. In general, we find that afterglow features perform poorly in GRB classification.
Figure 3. Examples of confusion matrices and SHAP feature importance values of the prompt emission subgroup without $T_{90}$, $F_{iso}$, or HR.

(a) Confusion matrix on all GRBs in the test set

(b) Confusion matrix on intermingled GRBs in the test set

(c) Confusion matrix on intermediate GRBs in the test set

(d) Average SHAP values of each feature on the training set

(e) SHAP value beeswarm plot on the training set
Figure 4. Examples of confusion matrices and SHAP feature importance values of the afterglow subgroup.
3.3. Host Galaxy

The different progenitors of Type I and II GRBs also have a substantial correlation with the properties of their host galaxies. The short lifetimes of Type II GRB progenitors (Woosley et al. 2002) make their event rate generally follow the star formation rate (SFR) of the host galaxies, and Type II GRB host galaxies generally have higher SFR. (Bloom et al. 2002; Chary et al. 2007; Savaglio et al. 2009; Levesque et al. 2010a; Levesque 2014; Robertson & Ellis 2011; Wei et al. 2014; Cucchiara et al. 2015; Trenti et al. 2015; Lan et al. 2022). The redshift distribution of Type I GRBs is found to be delayed with respect to the star formation history, and thus host galaxies of Type I GRBs generally have lower SFR respectively (Piran 1992; Nakar et al. 2006; Zheng & Ramirez-Ruiz 2007; Virgili et al. 2011; Wanderman & Piran 2015; Luo et al. 2022).

Type II GRB hosts also have low metallicity, which is required to form high-mass progenitors (Fynbo et al. 2003; Prochaska et al. 2004; Fruchter et al. 2006; Levesque et al. 2010b; Campisi et al. 2011; Kocevski & West 2011; Mannucci et al. 2011; Graham & Fruchter 2017; Lesniakowska et al. 2022). Type I GRB hosts, on the other hand, are found to have higher metallicity (e.g., Berger 2014).

Type II GRBs usually occur in regions with active star formation and are, therefore, closer to the center of the galaxy and in brighter regions. Type I GRBs, however, have larger offsets from the galactic center, as the evolution of compact binary mergers requires SN events that “kick off” the binary system away from the location where they are formed (Bloom et al. 2002; Fruchter et al. 2006; Fong et al. 2013, 2022; Blanchard et al. 2016; Wang et al. 2018; Li et al. 2020; O’Connor et al. 2022).

With the host galaxy subgroup, we are able to obtain an $F_1$ score of 0.615 on the test set, 0.909 on the intermingled GRBs, and 0.933 on the intermediate GRBs. The corresponding confusion matrices and feature importance are shown in Figure 5. We found the most important feature to be offset, with higher offsets pulling the predictions toward Type I.

In order to find other important features, we also carry out the same analysis on the host galaxy subgroup without offset. We get an $F_1$ score of 0.56 on the test set, 0.909 on the intermingled GRBs, and 0.857 on the intermediate GRBs with the host galaxy subgroup without offset. The corresponding confusion matrices and feature importance are shown in Figure 6. $A_\lambda$, $V$, stellar mass, and SFR are fairly important. Stronger dust extinction, higher stellar mass, and lower SFR pull the predictions toward Type I.

3.4. All

We also combine all the feature subgroups to form an “all” group. We then train and test our machine-learning model with this group containing all the features. With all the features, we obtain an $F_1$ score of 0.8 on the test set, 0.649 on the intermingled GRBs, and 0.870 on the intermediate GRBs.

The corresponding confusion matrices and feature importance are shown in Figure 7. The most important features all come from the prompt emission subgroup, which shows that prompt emission data are most important in GRB classification.

3.5. Comparing the Feature Subgroups

Because the training and test set splitting process introduces randomness to the results, the $F_1$ scores from a single trial may not be able to fully reflect the abilities in distinguishing Type I and II GRBs for different feature subgroups. Therefore, we repeat the random splitting and training process 1000 times and record the $F_1$ scores of each feature subgroup on the entire test set and intermingled and intermediate GRBs.

We report the average $F_1$ scores, along with standard deviations based on the 1000 trials, for each feature subgroup of the entire test set and intermingled GRBs in Table 4. We found that the prompt emission subgroup performs the best in predicting Type I and II GRBs, while the average $F_1$ score of the afterglow subgroup is significantly lower. Using all features only marginally improves the performance of the model. The host galaxy comes in between the two subgroups. However, prompt emission including $T_{90}$ performs the worst on the intermingled GRBs. Also, all the feature subgroups perform reasonably well on the intermediate GRBs, which indirectly rejects the existence of a third intermediate GRB type.

Note that the intermingled and intermediate GRBs form a smaller sample that usually get more attention from the scientific community compared with all the GRBs. Among the 32 features we use in this study, requiring the GRBs to have $T_{90}$ and one other feature to be known, the GRBs in the general sample on average have 2.5 known features, while the GRBs in the intermingled and intermediate samples on average have 9.3 and 9.4 known features, respectively.

We also compare the performance of our model with the traditional way of classifying GRBs on the $T_{90}$–HR plane by building a decision tree (e.g., Breiman et al. 1984; Timofeev 2004; Loh 2011, 2014), with $T_{90}$ and HR as inputs. We use the implementation in SCIKIT-LEARN and set the maximum depth of the decision tree to be 3. We use this model instead of the more sophisticated XGBOOST model we use in the other parts of this study because we think the decision tree model better reflects the classification ability of a human scientist.

With this decision tree model, we are able to achieve an $F_1$ score of 0.761 on all GRBs, 0.125 on intermingled GRBs, and 0.667 on intermediate GRBs. Examples of the confusion matrices are shown in Figure 8. When comparing the average $F_1$ scores from the multiple trials listed in Table 4, we find that while the performances of our XGBOOST multiparameter model and the simple decision tree model are comparable on all GRBs, the multiparameter model performs significantly better on the intermingled and intermediate GRBs. This shows that our new classification method is an improvement over the traditional one, especially on the intermingled and intermediate GRBs.

4. Predicting Unclassified GRBs

After building the models, we then move on to predict the classes of the unclassified GRBs in the Big Table. Since the all-feature subgroup achieved the best performance, we use all the features to train our model and predict the classes.

We train the model using the same method described in Section 2, with all the classified GRBs with at least one feature we intend to use and $T_{90}$ known, and use the trained model to predict the probabilities of the unclassified GRBs being either class. We also require the unclassified GRBs to have at least one feature and $T_{90}$ known. 1455 GRBs are used for training, and the class probabilities of 2809 unclassified GRBs are predicted. For each unclassified GRB, the class in which they
are predicted with the highest probability is assigned as their class. 2181 GRBs are predicted as Type II, while 628 GRBs are predicted as Type I. The prediction results are listed in Table 5. We graph the probability distribution of the unclassified GRBs being Type II in Figure 9. To compare our results with the traditional method of classifying GRBs on the $T_{90}$–HR plane, we also plot our prediction results on the $T_{90}$–HR plane in Figure 10.
5. Conclusions and Discussions

In this paper, we have applied supervised machine-learning methods, mainly XGBOOST, to the classification of Type I and II GRBs. We come up with the following conclusions:

1. Classifying GRBs solely based on $T_{90}$ can yield unsatisfactory results, especially on intermingled GRBs. Criteria based on multiple observational parameters are needed.

2. Compared with traditional GRB classification methods, the machine-learning method can effectively classify GRBs, especially intermingled and intermediate ones.

3. The fact that a supervised machine-learning model with two classes of GRBs can effectively classify intermediate...
Figure 7. Examples of confusion matrices and SHAP feature importance values of the all features subgroup.
Figure 8. Examples of confusion matrices of the $T_{90}$-HR decision tree.

(a) Confusion matrix on all GRBs in the test set

(b) Confusion matrix on intermingled GRBs in the test set

(c) Confusion matrix on intermediate GRBs in the test set

Table 4: List of Average $F_1$ Scores and 16th/84th Percentile Values Obtained with Different Feature Subgroups and GRB Samples

| Feature Subgroup                  | All GRBs $F_1$ | Intermingled $F_1$ | Intermediate $F_1$ |
|-----------------------------------|----------------|-------------------|-------------------|
| Prompt emission                   | 0.755$^{+0.042}_{-0.046}$ | 0.633$^{+0.089}_{-0.087}$ | 0.848$^{+0.059}_{-0.057}$ |
| Prompt emission, without $T_{90}$ | 0.665$^{+0.057}_{-0.059}$ | 0.755$^{+0.070}_{-0.072}$ | 0.869$^{+0.040}_{-0.040}$ |
| Prompt emission, without $T_{90}$, HR, $F_g$ | 0.497$^{+0.062}_{-0.063}$ | 0.773$^{+0.084}_{-0.080}$ | 0.855$^{+0.040}_{-0.055}$ |
| Afterglow                         | 0.425$^{+0.108}_{-0.103}$ | 0.798$^{+0.135}_{-0.131}$ | 0.831$^{+0.110}_{-0.081}$ |
| Host galaxy                       | 0.641$^{+0.091}_{-0.081}$ | 0.909$^{+0.076}_{-0.066}$ | 0.941$^{+0.062}_{-0.056}$ |
| Host galaxy, without offset       | 0.490$^{+0.119}_{-0.113}$ | 0.803$^{+0.106}_{-0.136}$ | 0.878$^{+0.122}_{-0.109}$ |
| All                               | 0.767$^{+0.045}_{-0.043}$ | 0.638$^{+0.090}_{-0.086}$ | 0.859$^{+0.046}_{-0.044}$ |
| $T_{90}$-HR decision tree         | 0.757$^{+0.049}_{-0.044}$ | 0.144$^{+0.052}_{-0.024}$ | 0.636$^{+0.062}_{-0.066}$ |

Table 5: Prediction Results of the Unclassified GRBs

| GRB     | $p_l$ | $p_H$ | Type |
|---------|-------|-------|------|
| 170325B | 0.0   | 1.0   | II   |
| 170315A | 0.0   | 1.0   | II   |
| 170309A | 0.12  | 0.88  | II   |
| 170228B | 0.0   | 1.0   | II   |
| 170208C | 0.0   | 1.0   | II   |
| 170207A | 0.0   | 1.0   | II   |
| 170206C | 0.003 | 0.997 | II   |
| 170130A | 0.0   | 1.0   | II   |
| 170121A | 0.381 | 0.619 | II   |
| 170120A | 0.0   | 1.0   | II   |

Note. The probabilities of them being Type I or II are shown as $p_l$ and $p_H$, respectively. This is an example of the first 10 rows of the table. (This table is available in its entirety in machine-readable form.)
GRBs with $T_{90}$ between 1 and 4 s indirectly rejects the existence of a third intermediate GRB class proposed based on duration distribution.

4. We found that the best feature group in predicting Type I or II GRBs is prompt emission. Among the features on prompt emission, we found that $T_{90}$ still separates Type I and II GRBs the best. Besides $T_{90}$, fluence and HR are also important features. Since fluence is correlated with $T_{90}$, this is consistent with the traditional way of classifying GRBs on the $T_{90}$–HR plane.

5. We predict the class of some of the GRBs not present in Greiner’s catalog. Their predicted class and their probabilities of being either are shown in Table 5.

6. The methods employed in this study can be applied to future newly discovered GRBs to identify potentially peculiar GRBs in their early stages and help allocate resources for follow-up observations. The analysis code used in this study is available.8

Recently, three possible intermingled GRBs have gained a lot of attention from the scientific community. GRB 200826A

is thought to be an intermingled Type II GRB (Ahumada et al. 2021; Zhang et al. 2021; Rossi et al. 2022), while GRB 211221A and GRB 230307A are thought to be intermingled Type I GRBs (Rastinejad et al. 2022; Yang et al. 2022b; Sun et al. 2023). We apply our trained model to these three GRBs to examine their observational proprieties. The features we

---

8 http://github.com/Rigel7/grb-ml

---

Table 6
Observational Properties of GRB 200826A

| Name       | Value | Unit | Reference       |
|------------|-------|------|-----------------|
| $T_{90}$   | 1.14  | s    | GCN 28287       |
| $F_g$      | 4.85  | 10^{-6} erg cm^{-2} | Zhang et al. (2021) |
| HR         | 0.803 | ...  | Zhang et al. (2021) |
| $F_{pk1}$  | 9.11  | 10^{-6} erg cm^{-2} s^{-1} | Zhang et al. (2021) |
| $p_{pk4}$  | 39.06 | photon cm^{-2} s^{-1} | GCN 28287 |
| alpha_band | −0.41 | ...  | GCN 28287       |
| beta_band  | −2.4  | ...  | GCN 28287       |
| E_P_band   | 89.8  | keV  | GCN 28287       |
| alpha_cpl  | −0.68 | ...  | Zhang et al. (2021) |
| E_P_cpl    | 120.29| keV  | Zhang et al. (2021) |
| spectral_lag | 5233 | ms MeV^{-1} | Zhang et al. (2021) |
| z          | 0.714 | ...  | GCN 28301       |
| D_L        | 1.21  | 10^{28} cm | GCN 31209 |
| E_iso      | 0.709 | 10^{52} erg | Zhang et al. (2021) |
| L_{pk}     | 1.41  | 10^{52} erg s^{-1} | Zhang et al. (2021) |
| offset     | 2.6   | kpc  | Zhang et al. (2021) |
| metallicity| −0.37 | ...  | Gupta et al. (2022) |
| A_V        | 0.19  | ...  | Gupta et al. (2022) |
| SFR        | 3.49  | $M_\odot$ yr^{-1} | Gupta et al. (2022) |
| Age        | 4.74  | Myr  | Gupta et al. (2022) |
| Mass       | 8.32×10^9 | $M_\odot$ | Gupta et al. (2022) |

Table 7
Observational Properties of GRB 211221A

| Name       | Value | Unit | Reference       |
|------------|-------|------|-----------------|
| $T_{90}$   | 34.3  | s    | GCN 31210       |
| $F_g$      | 540   | 10^{-6} erg cm^{-2} | GCN 31223 |
| HR         | 3.6   | ...  | GCN 31209, GCN 31223 |
| $p_{pk4}$  | 324.9 | photon cm^{-2} s^{-1} | GCN 31210 |
| alpha_band | 1.3   | ...  | GCN 31210       |
| beta_band  | 2.4   | ...  | GCN 31210       |
| E_P_band   | 646.8 | keV  | GCN 31210       |
| alpha_cpl  | 1.56  | ...  | GCN 31209       |
| spectral_lag | 107 | ms MeV^{-1} | Rastinejad et al. (2022) |
| z          | 0.0763| ...  | GCN 31221       |
| D_L        | 0.105 | 10^{28} cm | GCN 31221 |
| E_iso      | 0.69  | 10^{52} erg | GCN 31223 |
| L_{pk}     | 0.194 | 10^{52} erg s^{-1} | Yang et al. (2022b) |
| offset     | 7.91  | kpc  | Rastinejad et al. (2022) |
| metallicity| −0.69 | ...  | Rastinejad et al. (2022) |
| A_V        | 0.05  | ...  | Rastinejad et al. (2022) |
| Age        | 4.00  | Myr  | Rastinejad et al. (2022) |
| Mass       | 6.92×10^8 | $M_\odot$ | Rastinejad et al. (2022) |
Since \( T_{90} \) can be misleading for the classification of intermingled GRBs, we use all the features except \( T_{90} \) to classify these three bursts. For GRB 200826A, the model predicts it to have 13% probability of being Type I and 87% probability of being Type II. For GRB 211211A, the model predicts it to have 2% probability of being Type I and 98% probability of being Type II. For GRB 230307A, the model predicts it to have 98% probability of being Type I and 2% probability of being Type II. The SHAP values explaining how the individual features affect the prediction results are shown in Figure 11.

The prediction results of GRB 200826A and GRB 230307A from our model match the proposed classifications of the two GRBs in other literature, while the prediction result of GRB 211211A matches the classification done solely based on its duration. This shows that these GRBs, particularly GRB 211211A, are truly peculiar and more study should be done on their observational properties and physical origins. Indeed, Yang et al. (2022b) propose that GRB 211211A could originate from a white dwarf–neutron star merger, while Barnes & Metzger (2023) argue that GRB 211211A could still be explained by the normal Type II collapsar GRB model.

With its ability to not only predict the physical types of GRBs but also explain the importance of each parameter in individual classifications, our model can provide independent opinions on the classifications of possible peculiar GRBs and help guide future observations and studies.

**Acknowledgments**

This work is partially supported by the Top Tier Doctoral Graduate Research Assistantship (TTDDGRA) and Nevada Center for Astrophysics at the University of Nevada, Las Vegas.

**ORCID iDs**

Jia-Wei Luo @ https://orcid.org/0000-0002-9642-9682

Yuan-Chuan Zou @ https://orcid.org/0000-0002-5400-3261

Bing Zhang @ https://orcid.org/0000-0002-9725-2524

**References**

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, *PhRvL*, 119, 161101

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, *ApJL*, 848, L12

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017c, *ApJL*, 848, L13

Ahumada, T., Singer, L. P., Anand, S., et al. 2021, *NatAs*, 5, 917

Altmann, A., Toloza, L., Sander, O., & Lengauer, T. 2010, *Bioinformatics*, 26, 1340

Amati, L., Frontera, F., & Guidorzi, C. 2009, *A&A*, 508, 173

Amati, L., Frontera, F., Tavani, M., et al. 2002, *A&A*, 390, 81

Antonelli, L. A., D’Avanzo, P., Perna, R., et al. 2009, *A&A*, 507, L45

Balashev, A., Ruiz-Lapuente, P., Canal, R., et al. 2001, *MNRAS*, 328, 283

Barnes, J., & Metzger, B. D. 2023, *ApJ*, 947, 55

Berger, E. 2014, *ARA&A*, 52, 43

Bernardini, M. G., Ghirlanda, G., Campana, S., et al. 2015, *MNRAS*, 446, 1129

---

**Table 8**

| Name       | Value   | Unit    | Reference          |
|------------|---------|---------|--------------------|
| T90        | 41.52   | s       | Sun et al. (2023)  |
| F_g        | 4050    | 10^{-6} erg cm^{-2} | GCN 33579 |
| F_pk1      | 448     | 10^{-6} erg cm^{-2} s^{-1} | Sun et al. (2023) |
| P_pk4      | 791     | photon cm^{-2} s^{-1} | GCN 33411 |
| alpha_band | −0.43   | ...     | GCN 33444 |
| beta_band  | −5.1    | ...     | GCN 33444 |
| E_P_band   | 1260    | keV     | GCN 33444 |
| alpha_cpl  | −1.198  | ...     | Sun et al. (2023) |
| E_P_cpl    | 1254.68 | keV     | Sun et al. (2023) |
| spectral_lag | 32     | ms MeV^{-1} | Sun et al. (2023) |
| z          | 0.065   | ...     | Sun et al. (2023) |
| D_L        | 0.0907  | 10^{28} cm | Sun et al. (2023) |
| E_liso     | 3.08    | 10^{52} erg | Sun et al. (2023) |
| L_pk       | 48.9    | 10^{52} erg s^{-1} | Sun et al. (2023) |
| offset     | 36.6    | kpc     | Sun et al. (2023) |
| metallicity| 8.2     | ...     | Levan et al. (2023) |
| SFR        | 0.547   | \( M_{\odot} \) yr^{-1} | Levan et al. (2023) |
| Age        | 1130    | Myr     | Levan et al. (2023) |
| Mass       | 1.65 \times 10^9 | \( M_{\odot} \) | Levan et al. (2023) |
Zhang, B. 2006, Nat, 444, 1010
Zhang, B., Lu, H.-J., & Liang, E.-W. 2016, SSRv, 202, 3
Zhang, B., Zhang, B.-B., Liang, E.-W., et al. 2007, ApJL, 655, L25
Zhang, B., Zhang, B.-B., Virgili, F. I., et al. 2009, ApJ, 703, 1696
Zhang, B.-B., Liu, Z.-K., Peng, Z.-K., et al. 2021, NatAs, 5, 911
Zhang, B.-B., Zhang, B., Sun, H., et al. 2018, NatCo, 9, 447
Zhang, F.-W., Shao, L., Yan, J.-Z., & Wei, D.-M. 2012, ApJ, 750, 88
Zhang, L., Luo, J.-J., Huang, Y.-F., Gong, Y.-J., & Wu, S. 2022, MNRAS, 517, 5770
Zhang, Z.-B., & Choi, C.-S. 2008, A&A, 484, 293
Zheng, Z., & Ramirez-Ruiz, E. 2007, ApJ, 665, 1220
Zhu-Ge, J.-M., Luo, J.-W., & Zhang, B. 2023, MNRAS, 519, 1823
Zitouni, H., Guessoum, N., Azzam, W. J., & Mochkovitch, R. 2015, Ap&SS, 357, 7