Scientific Research Publishing

International Journal of Astronomy and Astrophysics, 2026, 16(3), 202-213
https://www.scirp.org/journal/ijaa
ISSN Online: 2161-4725
ISSN Print: 2161-4717

Fermi Gamma-Ray Burst Classification Based on a Convolutional Autoencoder

Baocheng Qin

College of Physics and Electronic Information Engineering, Guilin University of Technology, Guilin, China
Email: qinbaocheng2019@163.com

How to cite this paper: Qin, B.C. (2026) Fermi Gamma-Ray Burst Classification Based on a Convolutional Autoencoder. *International Journal of Astronomy and Astrophysics*, **16**, 202-213.
https://doi.org/10.4236/ijaa.2026.163013

Received: April 11, 2026
Accepted: July 25, 2026
Published: July 28, 2026

Abstract

Traditionally, gamma-ray bursts (GRBs) are classified into long and short bursts based on the bimodal distribution of their duration. However, this phenomenological dichotomy has been challenged by recent observations, and current classification methods often rely on manual pre-analysis of light curves. In this paper, we propose an end-to-end feature learning and classification method based on a convolutional autoencoder (CAE), which enables automatic feature extraction and unsupervised classification directly from raw GRB count images. Using 3819 GRB samples observed by the Fermi Gamma-ray Burst Monitor (GBM) from July 2008 to January 2024, we construct count images with a resolution of 512 × 512 pixels that integrate temporal and spectral information. The CAE model is built upon a pre-trained ResNet-18 encoder with a convolutional block attention module (CBAM), learning high-dimensional feature vectors, which are then visualized using the UMAP algorithm. The results show that GRBs do not exhibit a clear subclass separation structure and are essentially continuously distributed. Key physical parameters such as T_{90}, E_p, S_γ, and F_p show clear evolutionary gradients. Short bursts tend to cluster in specific regions, supporting the idea that GRBs may indeed differ in their origins, but without a clear boundary in their physical properties.

Keywords

Gamma-Ray Bursts, Convolutional Autoencoder, Unsupervised Learning, UMAP

1. Introduction

Gamma-ray bursts (GRBs) are the most powerful explosions in the Universe. Traditionally, GRBs are classified into two categories based on the bimodal distribu-

tion of durations: long GRBs (LGRBs) with a duration longer than 2 s ($T_{90} > 2$ s), and short GRBs (SGRBs) with $T_{90} < 2$ s . Previous theoretical and observational studies support that LGRBs and SGRBs originate from the collapsars and compact binary mergers, respectively. However, the conventional classification schemes have been challenged by recent observations, as the phenonmenological dichotomy does not necessarily correspond to the two distinct physical origins. The classification of GRBs remains a central research focus and a persistent hotspot in astronomy [1]-[5]. It is widely recognized that current GRB classification methods often rely on manual pre-analysis of light curves, introducing a certain degree of subjectivity. Therefore, there is an urgent need for feature extraction and analysis techniques that can operate directly on raw data.

Currently, most GRB research based on Convolutional Neural Networks is conducted within a supervised learning framework [6], while the application of unsupervised learning to GRB feature mining and classification remains relatively underexplored. To address this gap, we propose an end-to-end feature learning and classification model for GRBs based on a Convolutional Autoencoder (CAE), aiming to advance classification methodologies. Using high-quality observational data from the Fermi Gamma-ray Burst Monitor (GBM), our CAE model automatically extracts physically discriminative feature vectors from raw count images, followed by dimensionality reduction and visualization. By combining the strengths of convolutional neural networks in image processing with the capability of unsupervised learning to autonomously capture data similarities and differences, we aim to reveal the distribution of GRBs in a deep feature space. This approach enables a more accurate classification of GRBs, facilitates investigation into their diversity and complexity, and offers new insights into their classification and evolution.

Section 2 describes the dataset and the data preprocessing steps. Section 3 details the methodology, as well as the architecture and fine-tuning of the convolutional autoencoder model. Section 4 presents the experimental results and corresponding analysis. Section 5 briefly discusses the model's performance and the practical implications of the proposed method.

2. Data and Samples

The core data for this study originates from the Gamma-ray Burst Monitor aboard NASA's Fermi Gamma-ray Space Telescope. Fermi GBM consists of 12 sodium iodide (NaI) crystal detectors and 2 bismuth germanate (BGO) crystal detectors. The 12 NaI detectors, with a detection energy range from 8 keV to 1 MeV, cover the primary energy band of GRB prompt emission and serve as the main data source for this study. Each NaI detector possesses 128 independent energy channels, providing fine spectral resolution.

Fermi GBM offers various data products, among which the Time-Tagged Event (TTE) data contains the highest time and energy resolution raw photon arrival information, making it ideal for deep feature analysis. This study selects available

GRB events from the publicly available catalog released by Fermi GBM, spanning from July 2008 to January 2024, resulting in a total of 3819 samples.

For each selected GRB event, we perform the following data processing procedure: Centered on the official T_{90} start time (T_{start}) and end time (T_{end}), we extend the interval by 20 seconds forward and 30 seconds backward, constructing a total analysis time range of $T_{90}+50$ seconds. This time window fully captures the main burst phase and significant pre- and post-burst activities. From the TTE data, we extract the arrival times and energy channel information for all photons within this interval. Subsequently, by counting the number of photons falling into each time bin and energy channel cell and concatenating the raw light curves, we construct a 512 × 512 pixel raw count image for each GRB, as shown in **Figure 1**. This count image intuitively displays the evolution of GRB radiation intensity with time and energy, serving as a comprehensive representation that integrates temporal and spectral information. To mitigate the impact of large variations in absolute flux between different GRB events on model training, we divide all pixel values of each count image by its global maximum, thereby scaling all pixel values to the interval $[0,1]$. This normalization enables the model to focus on learning morphological characteristics and relative spectral features of GRBs rather than their absolute brightness. We performed a statistical analysis of the T_{90} values for the 3819 samples, plotting their logarithmic distribution histogram and fitting

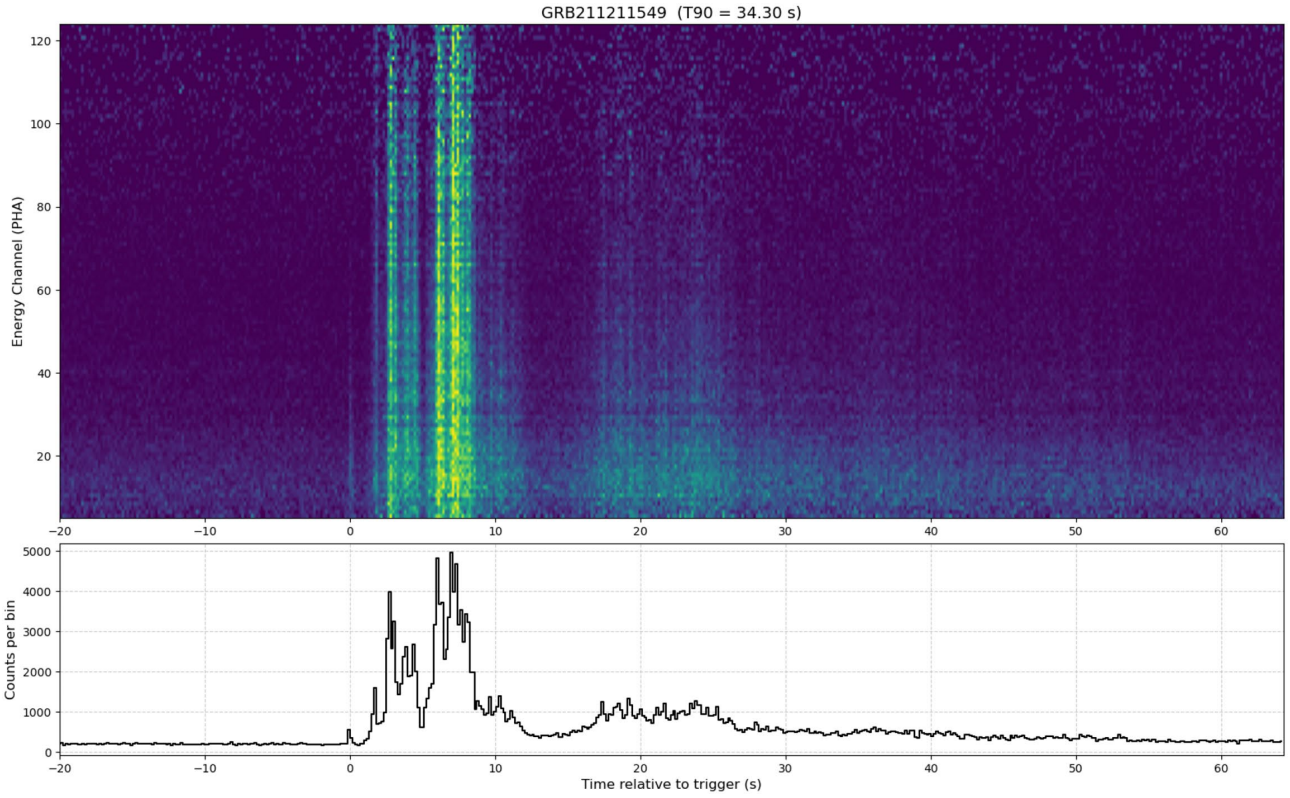

Figure 1. Original count map of GRB 211211A, constructed over the $T_{90}+50$ s window. The bright, multi-pulse main emission phase lasts approximately 10 s, followed by fainter extended emission.

it using a Gaussian Mixture Model, as shown in Figure 2. The results clearly reproduce the classic bimodal structure: one peak corresponds to the short burst group near $\log_{10} T_{90} \sim -0.01$ seconds, and the other peak corresponds to the long burst group near $\log_{10} T_{90} \sim 1.4$ seconds. Between the two peaks, within the range $\log_{10} T_{90} = 1 \pm 1$ seconds, there is a distinct valley, although a considerable number of events, approximately 5.7% of the total, are distributed in this overlapping region.

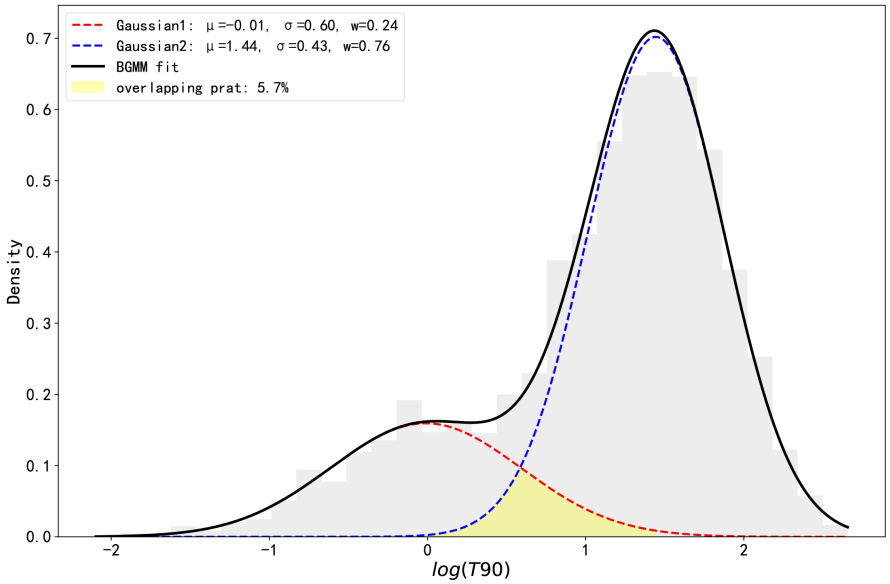

Figure 2. Gaussian Mixture Model fit to the logarithmic T_{90} distribution of all 3819 GRBs. The classic bimodal structure is reproduced, with short and long burst peaks separated by a valley, though about 5.7% of events fall in the overlapping region.

3. Methodology

3.1. Transfer Learning

Pan *et al.* [7] systematically compared the mechanisms of traditional machine learning and transfer learning. Their results indicated that the core idea of transfer learning is to transfer knowledge learned from a source task to a related target task, thereby improving learning efficiency and generalization ability. Subsequently, A. Sharif Razavian *et al.* [8] demonstrated through experiments that the Overfeat model pre-trained on ImageNet can serve as a generic feature extractor, achieving performance superior to many complex models in classification tasks. Yosinski *et al.* pointed out that if the features learned by a base network are generalizable, transferring them to a target network and employing training strategies such as freezing and fine-tuning can significantly improve the model's performance on the target task [9]. These works collectively suggest that features extracted by deep convolutional networks possess good generalizability for most vision tasks, and transfer learning helps accelerate model training and enhance generalization ability on new tasks.

However, the aforementioned studies mainly focus on supervised learning sce-

narios. For unsupervised classification tasks, J. Guérin *et al.* [10] extended the application of transfer learning. They found that combining deep convolutional networks pre-trained on ImageNet (e.g., VGG16, VGG19, InceptionV3, Xception, and ResNet50) with classical clustering algorithms (e.g., K-means, hierarchical clustering) could achieve performance comparable to specially optimized image clustering methods.

Based on these achievements, in the context of unsupervised GRB classification research, our convolutional autoencoder can adopt a similar paradigm. We first select a deep convolutional network model, ResNet-18, pre-trained on a large dataset whose data distribution is similar to GRB images. We utilize this as the encoder part of the convolutional autoencoder, subsequently combining it with a decoder and dimensionality reduction algorithms to achieve unsupervised classification. This approach not only fully leverages the representation capabilities of existing pre-trained models but also significantly reduces reliance on large amounts of labeled samples, thereby enhancing classification accuracy and efficiency.

3.2. Convolutional Block Attention Module

In our CAE model, processing image data generates a large number of feature maps, some of which contain crucial information for the current task while others are redundant or irrelevant noise. To identify and enhance important features while suppressing unimportant ones, we introduce a key module into our model: the Convolutional Block Attention Module (CBAM). CBAM is a simple yet effective attention module for feed-forward convolutional neural networks [11]. Given an intermediate feature map $\boldsymbol{F} \in \boldsymbol{R}^{C\times H\times W}$ as input, CBAM sequentially infers a 1D channel attention map $\boldsymbol{M}_c \in \boldsymbol{R}^{C\times 1\times 1}$ and a 2D spatial attention map $\boldsymbol{M}_s \in \boldsymbol{R}^{1\times H\times W}$. The overall attention process can be summarized as follows:

$$\begin{aligned} \boldsymbol{F}' &= \boldsymbol{M}_c\left(\boldsymbol{F}\right)\otimes \boldsymbol{F} \\ \boldsymbol{F}'' &= \boldsymbol{M}_s\left(\boldsymbol{F}'\right)\otimes \boldsymbol{F}' \end{aligned} \tag{1}$$

During the multiplication process, attention values are broadcast accordingly: channel attention values are broadcast along the spatial dimension, and vice versa. $\boldsymbol{F}''$ is the final output.

In the channel attention module, the primary focus is on generating a channel attention map by exploiting the inter-channel relationships of the feature map. Unlike previous methods that solely use average pooling, CBAM utilizes both average pooling and max pooling. First, spatial information of the feature map is aggregated using average pooling and max pooling to generate two different spatial context descriptors: $\boldsymbol{F}_{\text{avg}}^{c}$ and $\boldsymbol{F}_{\max}^{c}$, representing average-pooled and max-pooled features, respectively. These two descriptors are then forwarded to a shared network, composed of a Multi-Layer Perceptron (MLP), to generate the channel attention map $\boldsymbol{M}_c \in \boldsymbol{R}^{C\times 1\times 1}$. After applying the shared network to both descriptors, the output feature vectors are merged using element-wise summation. The channel attention is computed as follows:

$$\begin{aligned} \boldsymbol{M}_c(\boldsymbol{F}) &= \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}\left(\boldsymbol{F}\right)\right)+\mathrm{MLP}\left(\mathrm{MaxPool}\left(\boldsymbol{F}\right)\right)\right) \\ &= \sigma\left(\boldsymbol{W}_1\left(\boldsymbol{W}_0\left(F^c_{\mathrm{avg}}\right)\right)+\boldsymbol{W}_1\left(\boldsymbol{W}_0\left(F^c_{\mathrm{max}}\right)\right)\right) \end{aligned} \tag{2}$$

where σ denotes the sigmoid function, $\boldsymbol{W}_0 \in \boldsymbol{R}^{(C/r)\times C}$ and $\boldsymbol{W}_1 \in \boldsymbol{R}^{C\times(C/r)}$. Note that the MLP weights $\boldsymbol{W}_0$ and $\boldsymbol{W}_1$ are shared for both inputs, and the ReLU activation function follows $\boldsymbol{W}_0$.

In the spatial attention module, the primary focus is on generating a spatial attention map by exploiting the inter-spatial relationships of the feature map. Unlike channel attention, spatial attention focuses on which part is informative, complementing channel attention. To compute spatial attention, we first apply average pooling and max pooling operations along the channel axis and concatenate them to generate a feature descriptor. A convolution layer is then applied to this concatenated feature descriptor to generate the spatial attention map $\boldsymbol{M}_s(\boldsymbol{F}) \in \boldsymbol{R}^{H\times W}$, which encodes the locations to emphasize or suppress. Specifically, by using two pooling operations to aggregate the channel information of the feature map, we generate two 2D maps: $\boldsymbol{F}^s_{\mathrm{avg}} \in \boldsymbol{R}^{1\times H\times W}$ and $\boldsymbol{F}^s_{\mathrm{max}} \in \boldsymbol{R}^{1\times H\times W}$, representing average-pooled and max-pooled features across channels, respectively. These are then concatenated and convolved using a standard convolutional layer, producing the 2D spatial attention map. In short, the spatial attention is computed as follows:

$$\boldsymbol{M}_s(\boldsymbol{F}) = \sigma\left(f^{7\times 7}\left(\left[\mathrm{AvgPool}\left(\boldsymbol{F}\right);\mathrm{MaxPool}\left(\boldsymbol{F}\right)\right]\right)\right) = \sigma\left(f^{7\times 7}\left(\left[F^s_{\mathrm{avg}};F^s_{\mathrm{max}}\right]\right)\right) \tag{3}$$

where σ denotes the sigmoid function, and $f^{7\times 7}$ denotes a convolution operation with a filter size of 7×7.

Regarding the arrangement of the two attention modules, as they compute complementary attention, the two modules can be placed in parallel or sequential order. Experiments have shown that the sequential arrangement yields better results than the parallel arrangement. For the sequential process, the experimental results indicate that the channel-first order is slightly superior to the spatial-first order [11]. Therefore, the final CBAM module adopts the channel-spatial sequential order. CBAM is a lightweight, general-purpose module that can be seamlessly integrated into any CNN architecture; its parameter and computational overhead is negligible in most cases, and it can be trained jointly with the base CNN.

3.3. Feature Learning Framework Based on Convolutional Autoencoder

The feature learning framework proposed in this study is an end-to-end convolutional autoencoder. Its encoder core is a pre-trained ResNet-18 network. ResNet [12] addresses the vanishing gradient problem in deep network training by introducing residual connections. The encoder also integrates a Convolutional Block Attention Module (CBAM) to enhance feature representation capability. The overall architecture of this convolutional autoencoder is shown in **Figure 3**.

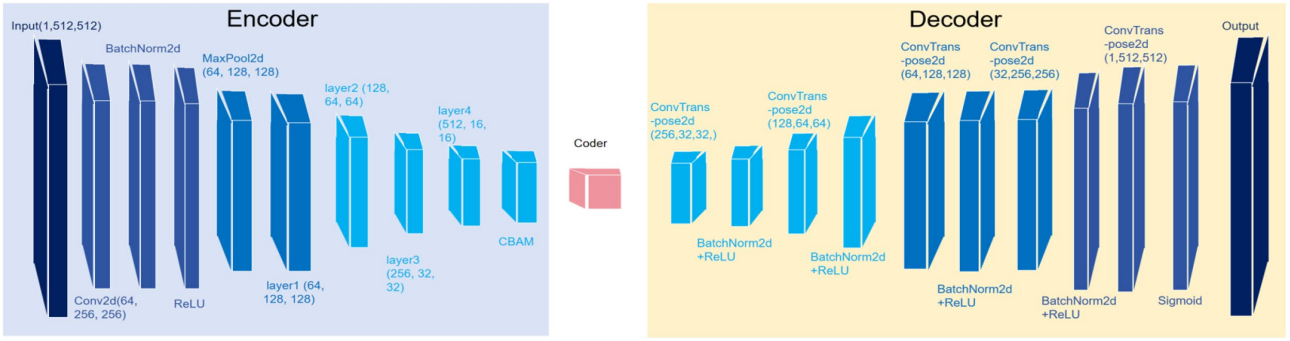

Figure 3. Architecture of the convolutional autoencoder. The encoder is a pre-trained ResNet-18 with a CBAM attention module; the decoder uses five transposed convolutional layers to reconstruct the 512 × 512 input image.

We input a single-channel grayscale image x of size 512 × 512. The model outputs a reconstructed image $\hat{x}$ of the same dimensions. All input images are uniformly resized to 512 × 512 pixels during preprocessing and normalized using ImageNet grayscale statistics (mean 0.485, standard deviation 0.229) to adapt to the input distribution of the pre-trained backbone network.

The design of the encoder is crucial to our model. We employ a ResNet-18 network pre-trained on the ImageNet dataset as the core network, modifying the input channel number of its first convolutional layer from 3 to 1 to accommodate single-channel astronomical images. The encoder retains the complete downsampling path of ResNet-18, including conv1, maxpool, and the four residual blocks layer1 through layer4. A CBAM module is embedded after the final residual block layer4 to enhance feature representation capability. To balance transferability and task adaptability, we freeze the parameters of layer1 and shallower layers during training, allowing only layer2, layer3, layer4, and the CBAM module to be fine-tuned.

The decoder aims to progressively upsample the high-dimensional feature maps output by the encoder, ultimately reconstructing an image of size 512 × 512. We design a symmetric structure consisting of five transposed convolutional layers. Each layer comprises a transposed convolution operation with stride 2, batch normalization, and a ReLU activation function. The final layer uses a Sigmoid activation function to ensure the output pixel values lie within the $[0,1]$ interval, matching the range of the normalized input data. The decoder performs five 2 × upsampling operations to precisely restore the original spatial resolution.

The model is trained using a single-stage, end-to-end strategy. The loss function is the pixel-wise Mean Squared Error (MSE):

$$\mathcal{L} = \left\| x - \hat{x} \right\|_2^2 \tag{4}$$

The optimizer is AdamW[13]. Initial learning rates are set to 4×10^{-4} for both the encoder and decoder, coupled with a cosine annealing learning rate scheduler ($T_{\max} = 20$). The entire training process runs for 100 epochs on an NVIDIA 3060 GPU.

Upon completion of training, the encoder is used as a feature extractor. For any input image x_i, its deep feature representation z_i is defined as the encoder's output feature map of size 512 × 16 × 16, which is then flattened to form a 131,072-

dimensional vector. This vector preserves spatial structure information and is suitable for subsequent unsupervised analysis.

After obtaining the feature vectors for all GRB data, we perform dimensionality reduction and visualization. We use the Uniform Manifold Approximation and Projection (UMAP) algorithm to reduce the feature vectors to a 2-dimensional plane. UMAP is effective at preserving both the global structure and local neighborhood relationships of the data, making it well-suited for visualizing high-dimensional clustering structures. Through the resulting scatter plot, we can intuitively assess whether GRBs naturally form distinct clusters.

4. Results

We input the 3819 raw GRB count images into the convolutional autoencoder, extract the deep feature vector for each event, and then employ the UMAP algorithm for dimensionality reduction and visualization. After extensive parameter tuning, we set n_neighbors = 45. This value balances the preservation of global topological structure and local clustering features. Concurrently, we mark the positions of special GRBs, those confirmed to be associated with kilonovae or supernovae, on the UMAP map, resulting in Figure 4. From Figure 4, it is obvious that the 3819 Fermi GRBs do not exhibit a clear separation in the 2D projection space. No separation based on progenitor type is observed; bursts associated with kilonovae and those associated with supernovae are scattered across different regions of the map without forming distinct clusters. This preliminary result suggests that the distribution of GRBs based on deep features may be more complex than traditionally recognized, and a simple dichotomy may be insufficient to fully describe their intrinsic structure.

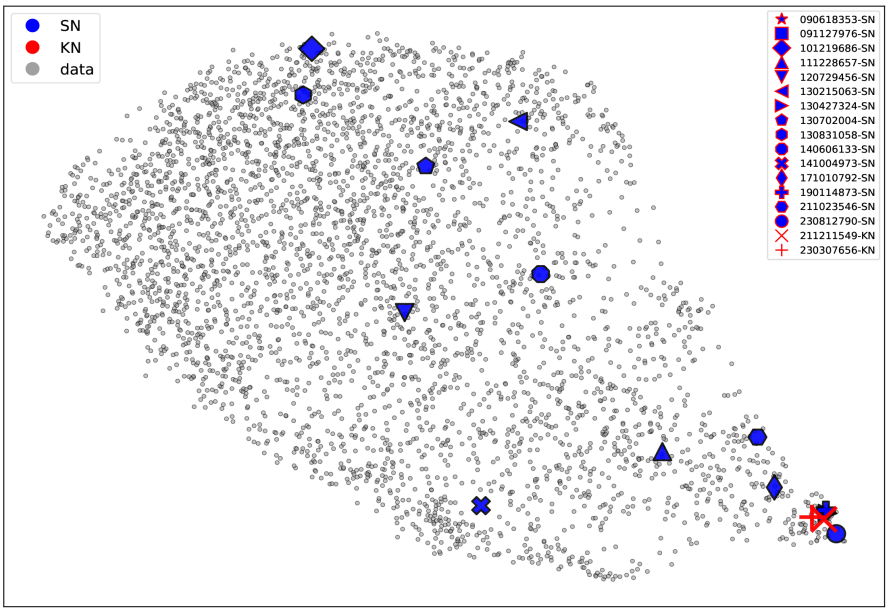

Figure 4. UMAP embedding of the 3819 GRBs based on CAE features. Blue and red symbols mark GRBs associated with supernovae and kilonovae, respectively. Neither population forms an isolated cluster, suggesting no sharp progenitor-based dichotomy in the deep feature space.

To further investigate the relationship between the deep feature space and physical parameters, we obtain the duration T_{90}, peak energy E_p, peak flux F_p, and fluence S_γ for each GRB from the dataset. We color the data points in the UMAP map based on the values of these parameters, shown in Figure 5. The E_p values are uniformly fitted using the Band function to ensure comparability across different bursts. Several notable features can be observed in Figure 5. First, the values of T_{90}, F_p, and S_γ exhibit clear gradients across the UMAP-projected space. Moving, for instance, from the lower-left to the upper-right region of the projection plane, the magnitudes of these parameters gradually increase, with colors transitioning from dark to bright. This continuous transition suggests that the overall parameter distribution of GRBs may tend towards a continuous rather than the traditional two distinct classes of short and long bursts. That is, the deep feature space captures a gradually evolving structure rather than a sudden classification boundary.

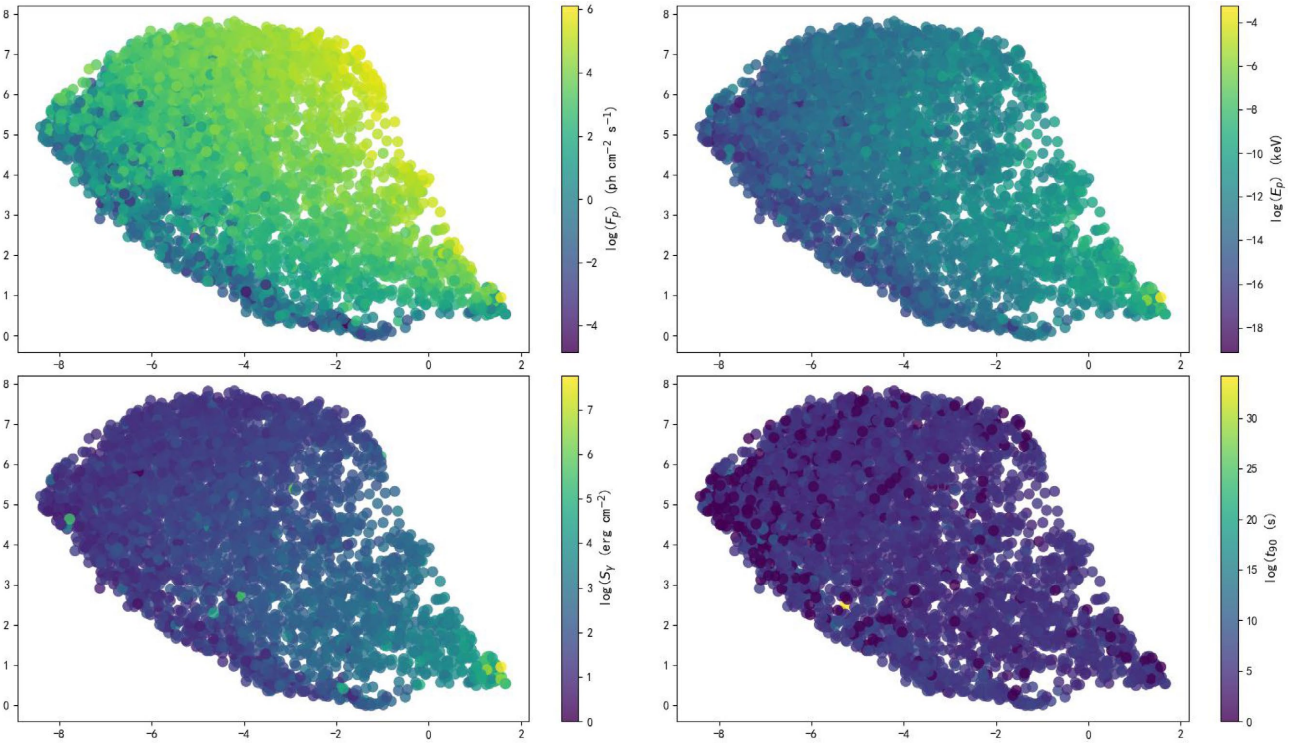

Figure 5. UMAP projection colored by T_{90}, E_p, S_γ, and F_p. All four parameters exhibit smooth gradients across the embedded space, indicating a continuous distribution of GRB properties rather than discrete classes.

To quantitatively evaluate whether the feature space supports the traditional long/short dichotomy, we applied K-Means ($k = 2$) clustering to the CAE feature vectors of all 3819 GRBs and computed the Silhouette Score and Davies-Bouldin Index (DBI). The results are summarized in Table 1.

The Silhouette Score of 0.076 is far below the commonly accepted threshold for weak clustering (~0.25), and the DBI of 5.277 is substantially larger than the criterion for good separation (<1.0). These values indicate that the two forced classes overlap severely in the high-dimensional feature space, and no meaningful binary

Table 1. Quantitative clustering assessment based on K-Means ($k = 2$).

Metric	Value	Reference Threshold
Silhouette Score	0.076	<0.25
Davies-Bouldin Index	5.277	<1.0
Permutation Test Mean Silhouette	−0.0001 ± 0.0013	-
Permutation Test p-value	<0.001	<0.05

boundary exists to support the traditional GRB dichotomy.

To assess the statistical significance of the observed weak structure, we performed a permutation test by randomly shuffling the K-Means labels 100 times. The mean Silhouette Score of the randomized labels was −0.0001 ± 0.0013. The observed score of 0.076 is significantly higher than the random baseline ($p < 0.001$). This statistical significance confirms the existence of a weak but non-random ordered structure in the feature space. However, the extremely low Silhouette Score and high DBI indicate that this structure does not correspond to discrete, well-separated clusters. Rather, as evidenced by the smooth color gradients of T_{90}, E_p, F_p, and S_γ in the UMAP projection (Figure 5), this structure is best interpreted as a continuous evolutionary trend of physical parameters, rather than distinct progenitor classes.

To better visualize the distribution of long and short bursts (defined by T_{90}) in the UMAP feature space, we color the UMAP map based on T_{90}: data points with $T_{90} > 2$ seconds are colored red, while those with $T_{90} < 2$ seconds are colored blue. The resulting map is shown in Figure 6. The figure clearly shows that, although

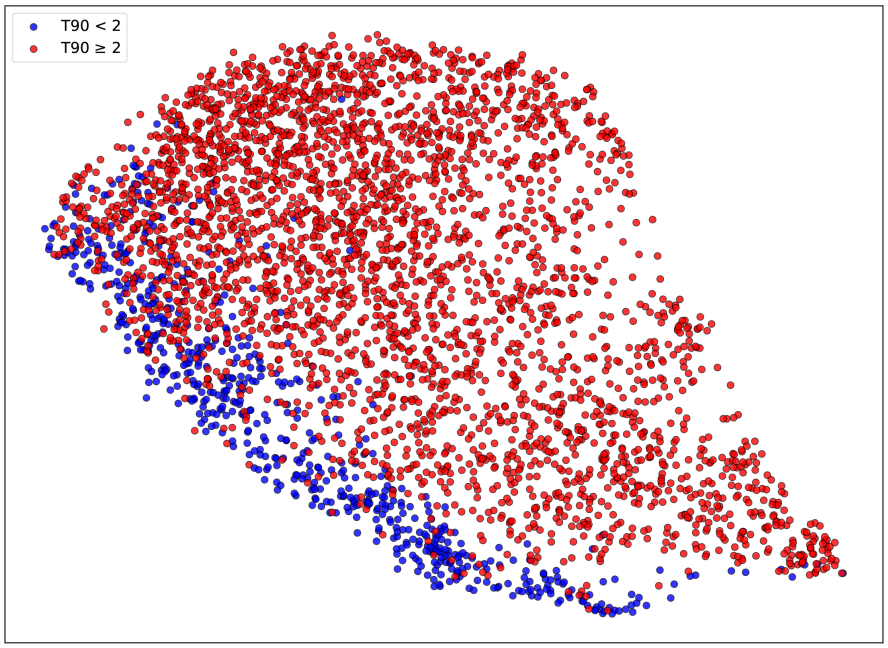

Figure 6. UMAP map color-coded by traditional T_{90} classification (blue: $T_{90} < 2$ s ; red: $T_{90} > 2$ s). Short bursts cluster in a confined bottom region, while long bursts spread more broadly, showing that duration information is partially retained in the deep feature space.

the overall distribution appears continuous, the blue points (short bursts) tend to cluster in a specific region towards the bottom of the projection plane. In contrast, long bursts are spread over a wider area but also show higher density in certain regions. This result indicates that duration, a traditional classification parameter, retains significant classification weight in the deep feature space.

5. Conclusions

By applying a convolutional autoencoder combined with the UMAP method, we analyze 3819 GRBs observed by the Fermi satellite. Our results show that the entire sample does not form distinct, separate clusters but instead exhibits a continuous transition in the parameter space. The color-coded distributions of key physical parameters— T_{90}, F_p, and S_γ —further illustrate the continuous evolutionary characteristics of all bursts. Additionally, we find that short and long bursts tend to cluster in different regions, confirming that short and long GRBs may indeed have differences in their physical origins. Our findings support the idea that while GRBs may differ in physical origin, their physical properties likely form a continuous distribution, consistent with the existence of mixed bursts that challenge the traditional dichotomy. It is worth noting that due to the lack of clear separation based on progenitor type, GRBs with different redshifts, luminosities, and environmental conditions are mixed together, which may obscure the features of certain subclasses. This also suggests that subsequent analyses should focus on more refined studies using specific subsamples.

The CAE model is capable of extracting effective features from raw count images and, through unsupervised learning, successfully captures the continuous evolutionary structure related to physical parameters while preserving the classification information inherent in traditional parameters. This feature learning approach provides a new perspective for a deeper understanding of the diversity and intrinsic structure of GRBs.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Zhang, B. (2011) Open Questions in GRB Physics. *Comptes Rendus Physique*, **12**, 206-225. https://doi.org/10.1016/j.crhy.2011.03.004

[2] Kumar, P. and Zhang, B. (2015) The Physics of Gamma-Ray Bursts & Relativistic Jets. *Physics Reports*, **561**, 1-109. https://doi.org/10.1016/j.physrep.2014.09.008

[3] Mészáros, P. (2019) Gamma-Ray Bursts: Theoretical Issues and Developments. *Memorie della Società Astronomica Italiana*, **90**, Article 57.

[4] Sun, H., Wang, C.W., Yang, J., *et al.* (2025) Magnetar Emergence in a Peculiar Gamma-Ray Burst from a Compact Star Merger. *National Science Review*, **12**, nwae401. https://doi.org/10.1093/nsr/nwae401

[5] Wang, C.W., Tan, W.J., Xiong, S.L., *et al.* (2025) A Subclass of Gamma-Ray Burst Originating from Compact Binary Merger. *The Astrophysical Journal*, **979**, 73.

https://doi.org/10.3847/1538-4357/ad98ec

[6] Zhang, P., Li, B., Gui, R., *et al.* (2024) Application of Deep-Learning Methods for Distinguishing Gamma-Ray Bursts from Fermi/GBM Time-Tagged Event Data. *The Astrophysical Journal Supplement Series*, **272**, 4.

[7] Pan, S.J. and Yang, Q. (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**, 1345-1359. https://doi.org/10.1109/tkde.2009.191

[8] Razavian, A.S., Azizpour, H., Sullivan, J. and Carlsson, S. (2014) CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. 2014 *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, 23-28 June 2014, 512-519. https://doi.org/10.1109/cvprw.2014.131

[9] Yosinski, J., Clune, J., Bengio, Y., *et al.* (2014) How Transferable Are Features in Deep Neural Networks? *Advances in Neural Information Processing Systems*, **27**, 3320-3328.

[10] Guérin, J., Gibaru, O., Thiery, S., *et al.* (2018) CNN Features Are Also Great at Unsupervised Classification. *Computer Science & Information Technology*, Singapore, 22-23 October 2018, 83-95. https://doi.org/10.5121/csit.2018.80308

[11] Woo, S., Park, J., Lee, J.Y., *et al.* (2018) CBAM: Convolutional Block Attention Module. In: *Lecture Notes in Computer Science*, Springer, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[12] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[13] Loshchilov, I. and Hutter, F. (2019) Decoupled Weight Decay Regularization. arXiv: 1711.05101. https://doi.org/10.48550/arXiv.1711.05101