

A machine learning approach for GRB detection in AstroSat CZTI data

Sheelu Abraham^{1,2}, Nikhil Mukund^{3,4,2}, Ajay Vibhute^{5,2}, Vidushi Sharma², Shabnam Iyyani², Dipankar Bhattacharya², A. R. Rao⁶, Santosh Vadawale⁷, and Varun Bhalerao⁸

¹ Marthoma College, Chungathara, Nilambur, Kerala

² Inter-University Center for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune, India

³ Max Planck Institute for Gravitational Physics (Albert Einstein Institute), D-30167 Hannover, Germany

⁴ Leibniz Universität Hannover, D-30167 Hannover, Germany

⁵ Savitribai Phule Pune University, Pune, Maharashtra, India.

⁶ Tata Institute of Fundamental Research, Mumbai, India

⁷ Physical Research Laboratory, Ahmedabad, Gujarat, India

⁸ Indian Institute of Technology, Bombay, India

June 25, 2019

ABSTRACT

We present a machine learning (ML) based method for automated detection of Gamma-Ray Bursts (GRBs) from the AstroSat CZTI data. We make use of density-based spatial clustering to detect excess power and carry out an unsupervised hierarchical clustering across all such events to identify the different categories of light curves present in the data. This representation helps in understanding the sensitivity of the instrument to the various GRB populations and identifies the major non-astrophysical noise artifacts present in the data. We make use of dynamic time wrapping (DTW) to carry out template matching to ensure the morphological similarity of the detected events with that of known typical GRB light curve. DTW alleviates the need for a dense template repository often required in matched filtering like searches, and the use of a similarity metric facilitates outlier detection suitable for capturing previously unmodeled events. Using the pipeline, we detect 35 new GRB events and briefly report their characteristics in this paper. Augmenting the existing data analysis pipeline with ML capabilities enables the instrument to quickly respond to alerts received from observatories such as the gravitational wave detectors and carry out robust follow up studies without the need for an onboard classification facility.

Key words. methods: data analysis, statistical; gamma rays: general; X-rays: bursts

1. Introduction

GRBs are one among the most energetic explosions known to occur in the Universe. A typical event can release energy of 10^{46} - 10^{52} erg/s lasting for a period of few milliseconds to minutes. Based on their duration, they are classified as short GRBs when the T90 is less than 2 s, and long GRBs when it is otherwise (Kouveliotou et al. 1993). These two types are also believed to originate from two different progenitors. Long GRBs are associated with the death of massive stars (Woosley 1993; Iwamoto et al. 1998; MacFadyen & Woosley 1999) and has been confirmed with the coincident detection of the supernova with the long GRB030329A (Stanek et al. 2003). On the other hand, short GRBs are believed to be produced as a result of the merger of compact objects like neutron star-neutron star or a neutron star-black hole (Eichler et al. 1989; Narayan et al. 1992). The recent discovery of gravitational waves from binary neutron star merger (GW170817) by advanced LIGO and advanced Virgo observatories (Abbott et al. 2017a) together with the detection of short GRB (GRB170817A) by various telescopes such as *Fermi*, SWIFT and others (Goldstein et al. 2017), have confirmed these proposed mechanisms, thus opening a new era in multi-messenger astronomy.

A GRB event can be divided into two main epochs: prompt emission phase in gamma rays which occurs immediately after the burst and a subsequent afterglow phase which is observed in

multiple wavelengths from gamma rays to radio extending over a period lasting from days to months. Timely identification of prompt emission is necessary to carry out follow-up observation in multiple wavelengths by ground and space-based telescopes. This can lead to the detection of afterglows, which is crucial in determining the red-shift as well as the overall properties of the GRB.

The onboard alert systems of Burst Alert Telescope (BAT) on *Neil Gehrels Swift* mission (Gehrels et al. 2004) and of Gamma-Ray Burst Monitor (GBM) on *Fermi* satellite (Meegan et al. 2009), have led to an increased number of detections along with more afterglow observations. They do not include any morphology recognition and hence need to be cross-checked to rule out the false triggers. Cadmium Zinc Telluride Imager (CZTI) onboard AstroSat, becomes a wide field hard X-ray detector above 100 keV because of the increasing transparency of collimators and surrounding supporting structures thus making it sensitive to GRBs (Rao et al. 2016). In this paper, we present a fully automated machine learning (ML) algorithm that enables GRB detection in CZTI data with a latency of ten seconds per single orbit thus overcoming the absence of an onboard detector.

The paper is organized as follows: in Section 2, the various pre-processing steps involved in generating light curves from CZTI data are presented. Section 3 briefly overviews the three machine learning algorithms used in this work and the proposed detection pipeline. Section 4 talks about the results from the

blind search, while conclusions and prospects are presented in Section 5.

2. AstroSat CZTI: Data and Prepossessing

AstroSat (Agrawal 2006; Singh et al. 2014) is India's first multi-wavelength space observatory capable of observing the X-ray wavelength band from 0.3 to 100 keV and the UV bands from 1300 - 3000 Å. It carries the following five science instruments for simultaneous observations of the source of interest: Ultra-Violet Imaging Telescope (UVIT; Tandon et al. 2017), Large Area X-ray Proportional Counters (LAXPC; Yadav et al. 2016), Soft X-ray Telescope (SXT; Singh et al. 2017), Cadmium Zinc Telluride Imager (CZTI; Rao et al. 2017) and Scanning Sky Monitor (SSM; Ramadevi et al. 2017). In particular, CZTI consists of an array of Cadmium Zinc Telluride (CZT) detectors, which are pixelated such that each pixel acts as an independent photon counting detector. CZTI has a detector area of 976 cm² build using CZT modules and makes use of Coded Aperture Mask (CAM) for imaging (Bhalerao et al. 2017). The total detection area is achieved by the use of 64 CZT modules of area 15.25 cm² each. These 64 modules are arranged in four identical and independent quadrants. The collimator walls separate the modules, and collimators above each detector module restrict the field of view to 4.6° x 4.6° (Full Width at Half Maximum) at photon energies below 100 keV. At energies above that, the collimator slats and the coded mask become progressively transparent, and the instrument behaves like an all-sky open detector enabling detection of GRBs. It also carries a Caesium Iodide (TI) based scintillator detector operating as anti-coincidence with the main CZT detector and is called a veto detector. The coded aperture telescope is sensitive to hard X-ray polarization and was recently used to measure the polarized hard X-ray emission from Crab nebula (Vadawale et al. 2018).

CZTI can be configured in 16 different modes. The default mode of operation is event mode, which is denoted as Mode M0 (Normal Mode). CZTI also records accumulated spectral and housekeeping information once in every 100 seconds and stores the recorded information when the mode is changed to Secondary Spectral Mode (Mode SS). Whenever the spacecraft passes through the South Atlantic Anomaly (SAA), High Voltage (HV) in the CZTI and Veto detectors are switched off, and the detector is in Mode M9 (SAA mode), during which only housekeeping information is recorded once every second. During the normal mode, whenever a photon hits a detector, CZT records the time of arrival of the photon, the position of the photon on the detector plane and energy of the photon. The time-tagged event list is stored in an event file. The events from four quadrants are stored as four different extensions of the event file. The recorded events also contain events generated due to the interaction of charged particles with the instrument or spacecraft body. The X-ray photons, consequently generated, can also deposit their energies in the CZTI detectors. Because of the pixelated nature of CZT, one charged particle can produce events in many pixels of CZT at the same time, and are referred to as 'bunches.' During preprocessing, these bunches, which do not belong to any astronomical source, are removed from the event file. Time intervals where data is not present due to SAA passage, data transmission error, etc. are ignored, and a Good Time Interval (GTI) file is produced. The events belonging to good time intervals are filtered and passed for further processing. During the data cleaning process, events from noisy or flickering pixels are removed. The on-board calibration source, Am-241 emits X-ray of energy 60 keV and an alpha particle simultaneously.

The alpha particle is absorbed in the CsI (TI) crystal whereas the X-ray gets detected in the CZT pixel and the alpha flag is set to 1. The events having the alpha flag equal to 1 are thereby removed from the event list. The cleaned event list that is finally obtained is used as the input to the GRB detection algorithm.

One event file from CZTI usually consists of multi-orbit data which may span 6000 - 30,000 seconds. We have divided the data into small chunks of 500 seconds each to check for any triggers present. Each event file consists of data from all the four quadrants as well as the four veto channels. In the preprocessing steps, we have considered only events which have energy higher than 60 keV because, at lower energies, the events belong to either source or background as the CZTI collimator blocks events from other directions. Although we perform preprocessing to clean the time series before feeding it to the analysis pipeline, it is never perfect, and traces of SAA are still visible in certain segments of the data. Our pipeline takes care of such noises present in the data, and in most cases, successfully vetoes the noise signals.

3. Detection Pipeline

This section describes the framework that is used to detect GRBs and issue subsequent Gamma-ray Coordination Network (GCN) alerts to the wider astronomy community. In particular, we discuss the different machine learning algorithms that are deployed to detect real GRBs from false triggers that arise from instrumental noise artifacts, cosmic rays, and random fluctuations. We start with the creation of a template bank for GRB light curves. The key idea is to minimize the number of templates while still achieving maximal coverage of the inherent morphologies present in the light curves. We obtain these templates from already identified events in CZTI data using the GCN trigger information published by the contemporary space observatories.

For each event, we perform generalized cross-correlation of the light curves observed within the four quadrants and correct for the delays among them. Light curves seen in all the quadrants are then stacked together to get a temporal sequence with higher SNR. These sequences are then clustered hierarchically, see Figure 1, to reveal the different clusters. Mean profile within each such cluster is then used to generate the GRB template bank. We use a bottom-up agglomerative clustering version of the algorithm where the objects start as individual clusters which are hierarchically combined to form a dendrogram. The technique provides freedom to the user to choose any valid distance metric to compare the similarity between the objects. By maximizing the sum of similarities among the adjacent clusters, we can achieve optimal leaf ordering within the dendrogram (Bar-Joseph et al. 2001). This ordering lets one observe the progressively changing morphology within the given data samples. Hierarchical clustering has previously been successful in identifying the dominant groups among the short duration transients such as those observed in gravitational wave observatories (Mukund et al. 2017).

We perform one second binning for the data chunks independently for each of the four quadrants. Based on the knowledge of detector noise fluctuations, we set a threshold and perform temporal clustering using the DBSCAN algorithm on samples which lie above this level. Absence of such data points either terminates the search or starts a new search with a lower threshold. DBSCAN which stands for Density-based spatial clustering of applications with noise (DBSCAN; Ester et al. 1996), groups together datasets which have similar features and identifies outliers in an automated fashion. As compared to K-Means like clustering algorithms (Hartigan & Wong 1979) there is no need to

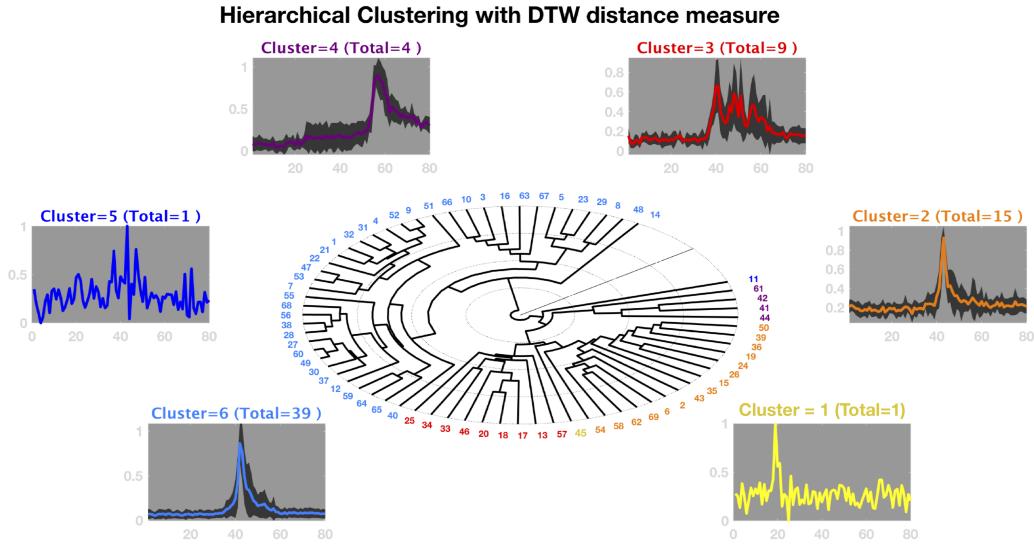


Fig. 1. Hierarchical clustering of known GRB light curves using dynamic time wrapping as the distance measure. Dominant templates identified through this unsupervised technique is further used for the detection pipeline.

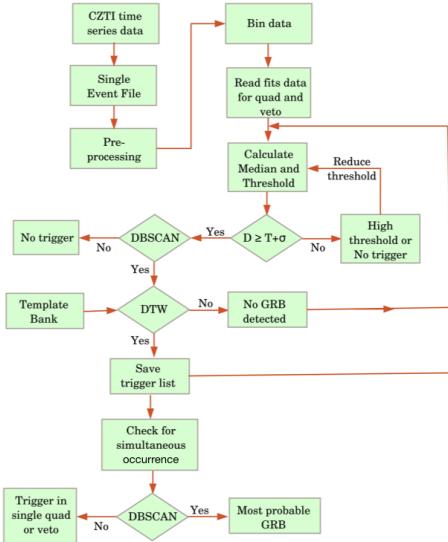


Fig. 2. Schematic showing the pipeline of our GRB detection algorithm.

specify the number of clusters present in the data. The only required parameters are the minimum number of points in each cluster and the maximum separation between samples for them to be considered as part of the same cluster. Once the triggers are identified, they are checked for the similarity to known astrophysical signals using Dynamic Time Warping (DTW) technique. DTW is a general method developed for time series alignment for speech and handwriting recognition which can be applied to detect similar temporal sequences which are relatively stretched or squeezed with the template (Sakoe & Chiba 1978). The method eliminates the need for feature extraction and can be easily extended to carry out the similarity search using a template bank of known sequences. DTW finds the optimal alignment between time series data, which allows a non-linear mapping of one signal to another, thereby minimizing the distance between the two. Pursuing alternative methods like cross-correlation or

matched filtering in this scenario would require a dense template bank, making them computationally challenging for rapid detection. Let X and Y be two vectors of lengths M and N , respectively. To create a mapping between the two vectors, we need to define a path which should start at $(0,0)$ and end at (M, N) . The aim is to find the path of minimum distance,

The optimal path starts from $(0,0)$, ends at (M, N) , and in between map the vectors on a common set of indices ix & iy such that the total sum of distances, d

$$d = \sum_{m \in ix} \sum_{n \in iy} d_{m,n}(X, Y) \quad (1)$$

is minimized where the distance $d_{m,n}$ is expressed in terms of symmetric Kullback-Leibler metric (Kullback & Leibler 1951),

$$d_{m,n}(X, Y) = (x_m - y_n)(\log x_m - \log y_n) . \quad (2)$$

The DTW path is constrained to move close to the diagonal by specifying a window around the main diagonal to minimize the effect of outliers. Additionally to ensure alignment of the complete signal and not just segments as well to prevent sample skipping, only the following transitions are permitted while the path proceeds from $(0,0)$ to (M, N) ,

$$\begin{aligned} (m, n) &\rightarrow (m + 1, n) \\ (m, n) &\rightarrow (m, n + 1) \\ (m, n) &\rightarrow (m + 1, n + 1) . \end{aligned}$$

Figure 3 shows one such instance of DTW based alignment of two GRB light curves. If the trigger matches any of the GRB models within the template bank, then checks for simultaneous occurrence in different quadrants and veto channels are carried out. The number of simultaneous occurrences required to a claim detection can be lowered to capture faint low SNR events but comes at the price of an increased number of false positives. High confidence detections are either reported as GCN alerts or sent to subject experts for further analysis. The final pipeline has

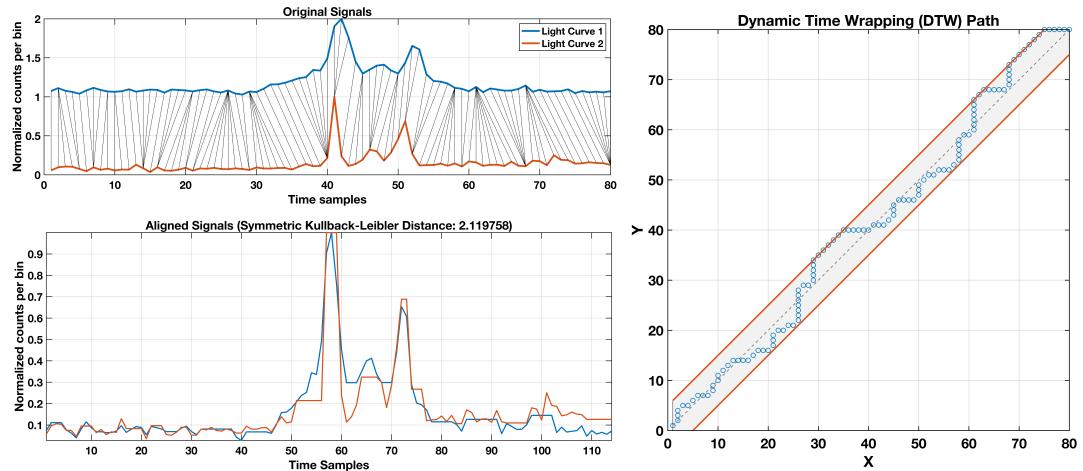


Fig. 3. Distance between two GRB light curves estimated using dynamic time wrapping with symmetric Kullback-Leibler metric and an adjustment window size of 5 samples.

a basic structure given in Figure 2. Initial prototyping of the algorithm was carried out in MATLAB while the final version of the pipeline is written in Python, and takes about 10 seconds for the execution of a single orbit data.

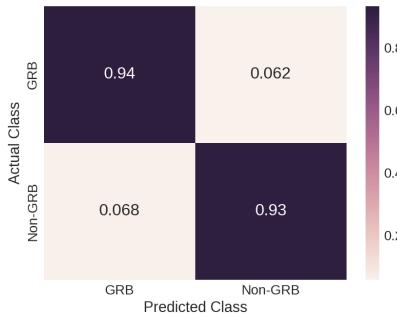


Fig. 4. Confusion matrix showing the prediction accuracy of our pipeline. The false alarm rate is high due to the noise associated with the event files.

4. Results

Before online deployment, we tested the performance of the scheme using a known set of GRBs which were previously reported by other similar missions. Our dataset included 106 samples consisting of 35 GRBs along with other instances of SAA and cosmic rays. We were able to recover all but two and the missed ones were successfully detected by lowering the threshold parameter from its base value of 3σ . Overall accuracy of the pipeline is given by the confusion matrix shown in Figure 4. Some of the detection scenarios involving true detections and false positives are shown in Figure 5. The upper panels shows the correctly identified events while the lower panels depict the instances when the events picked up where due to noise artefacts. We also carried out a blind search on the CZTI data leading to the discovery of 35 new GRB triggers. Table 1 lists these new events along with their trigger time, peak count rate, and T_{90} which is the duration over which 5 % to 95 % of the entire burst fluence (or total photon counts) is observed (Koshut et al. 1995). GRB180526A (Sharma et al. 2018a) and GRB180603A (Sharma

et al. 2018b) are two such events which were reported as GCN alerts based on the new GRB pipeline.

Observing short duration GRBs in conjunction with a GW trigger helps to better understand the kilonovae mechanisms, measure differences in speed of light and gravity, better scrutinize alternative theories of gravity, etc. thus is of great astrophysical significance (Abbott et al. 2017b). Simultaneous operation of multiple detectors capable of GRB detection would lead to improved sky coverage and help in constraining the time of occurrence of the event to a greater degree of precision. We tested the robustness of the developed pipeline with six available short GRBs and were able to recover all of them but had to lower the detection threshold from its default 3σ value. This led to an increased number of false positives, thus reducing the significance of true detection. One such a short GRB trigger event is shown in Figure 6. Better feature extraction from the temporal sequence coupled with a diverse template bank that includes more noise models could reduce the background, but this would be part of a future study. Nonetheless, the developed infrastructure would enable us to be part of the global network of detectors aiming for coincident detection of such short duration events.

5. Conclusions

We demonstrated the resourcefulness of various machine learning algorithms for robust GRB detection using AstroSat CZTI data. Automating such tasks can bring down the response time leading to efficient follow-up studies related to multi-messenger astronomy. Compared to conventional peak detection algorithms, incorporating morphology is seen to significantly decrease the false detection rate arising from instrumental artifacts and non-GRB phenomena. The newly developed scheme is now part of the data analysis pipeline and has led to the discovery of 35 new GRBs. The techniques presented in this work has the potential to be used in similar astronomical datasets like stellar spectra or gravitational wave transients. In the future, we would like to focus more on improving the detection efficiency for short GRBs through better localization of the trigger in time as well as frequency domain. The feasibility of embedding ML algorithms in FPGA based hardware for low latency on-board trigger detection is also worth exploring in the context of next-generation detectors.

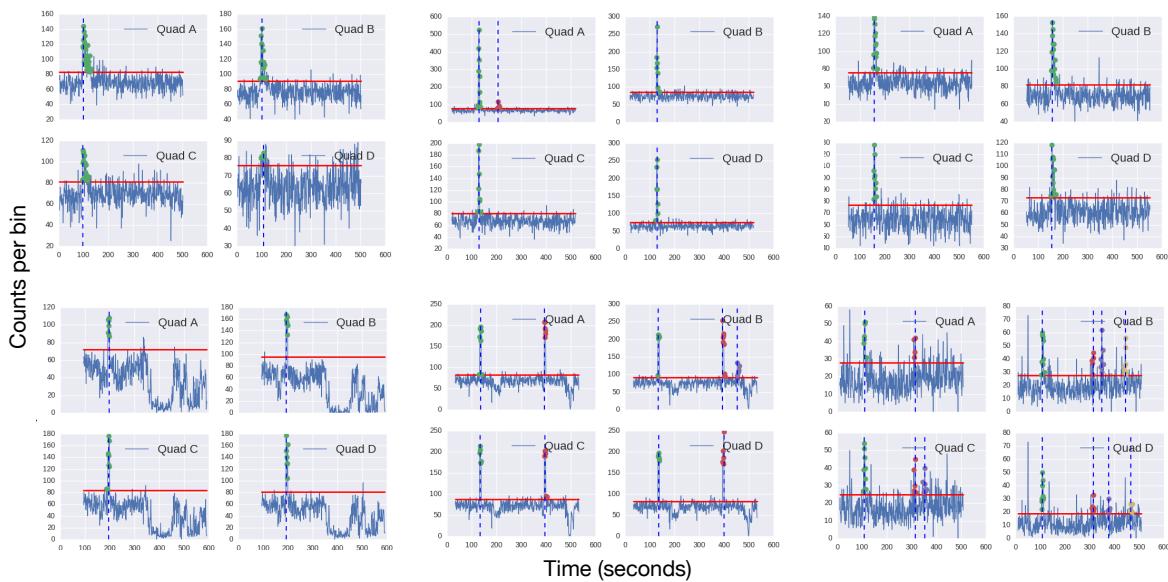


Fig. 5. Few detection scenarios encountered while testing the pipeline. The upper panel shows the correctly identified GRBs while the lower panel displays some of the false positives. The horizontal red line gives the detection while the dashed vertical line provides the cluster center and thus the trigger time.

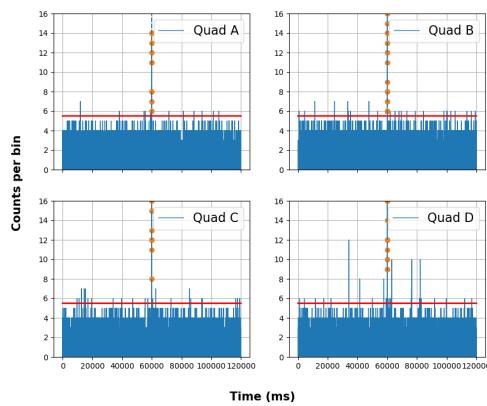


Fig. 6. Detection of short GRB (GRB170127C) in CZTI data.

6. Acknowledgments.

This publication uses the data from the AstroSat mission of the Indian Space Research Organisation (ISRO), archived at the Indian Space Science Data Centre (ISSDC). CZT-Imager is built by a consortium of Institutes across India including Tata Institute of Fundamental Research, Mumbai, Vikram Sarabhai Space Centre, Thiruvananthapuram, ISRO Satellite Centre, Bengaluru, Inter University Centre for Astronomy and Astrophysics, Pune, Physical Research Laboratory, Ahmedabad, Space Application Centre, Ahmedabad: contributions from the vast technical team from all these institutes are gratefully acknowledged. Authors express thanks to the CZTI AstroSat support cell at IUCAA for their help in data curation and preprocessing. NM acknowledges Council for Scientific and Industrial Research (CSIR), India, for providing financial support as Senior Research Fellow. Authors also express thanks to Ninan Sajeeth Philip for his valuable comments and suggestions.

References

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, *Physical Review Letters*, 119, 161101
- Abbott, B. P. et al. 2017b, *The Astrophysical Journal*, 848, L13
- Agrawal, P. 2006, *Advances in Space Research*, 38, 2989
- Bar-Joseph, Z., Gifford, D. K., & Jaakkola, T. S. 2001, *Bioinformatics*, 17, S22
- Bhalerao, V., Bhattacharya, D., Vibhute, A., et al. 2017, *Journal of Astrophysics and Astronomy*, 38, 31
- Eichler, D., Livio, M., Piran, T., & Schramm, D. N. 1989, *Nature*, 340, 126
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96* (AAAI Press), 226–231
- Gehrels, N., Chincarini, G., Giommi, P., et al. 2004, *ApJ*, 611, 1005
- Goldstein, A., Veres, P., Burns, E., et al. 2017, *ApJ*, 848, L14
- Hartigan, J. A. & Wong, M. A. 1979, *JSTOR: Applied Statistics*, 28, 100
- Iwamoto, K., Mazzali, P. A., Nomoto, K., et al. 1998, *Nature*, 395, 672
- Koshut, T. M., Paciesas, W. S., Kouveliotou, C., et al. 1995, in *Bulletin of the American Astronomical Society*, Vol. 27, *American Astronomical Society Meeting Abstracts #186*, 886
- Kouveliotou, C., Meegan, C. A., Fishman, G. J., et al. 1993, *ApJ*, 413, L101
- Kullback, S. & Leibler, R. A. 1951, *Ann. Math. Statist.*, 22, 79
- MacFadyen, A. I. & Woosley, S. E. 1999, *ApJ*, 524, 262
- Meegan, C., Lichti, G., Bhat, P. N., et al. 2009, *ApJ*, 702, 791
- Mukund, N., Abraham, S., Kandhasamy, S., Mitra, S., & Philip, N. S. 2017, *Phys. Rev. D*, 95, 104059
- Narayan, R., Paczynski, B., & Piran, T. 1992, *ApJ*, 395, L83
- Ramadevi, M. C., Seetha, S., Bhattacharya, D., et al. 2017, *Experimental Astronomy*, 44, 11
- Rao, A. R., Bhattacharya, D., Bhalerao, V. B., Vadawale, S. V., & Sreekumar, S. 2017, ArXiv e-prints [arXiv:1710.10773]
- Rao, A. R., Chand, V., Hingar, M. K., et al. 2016, *ApJ*, 833, 86
- Sakoe, H. & Chiba, S. 1978, *IEEE transactions on acoustics, speech, and signal processing*, 26, 43
- Sharma, V., Vibhute, A., & Bhattacharya, D. 2018a, 22742, 1
- Sharma, V., Vibhute, A., & Bhattacharya, D. 2018b, 22762, 1
- Singh, K. P., Stewart, G. C., Westergaard, N. J., et al. 2017, *Journal of Astrophysics and Astronomy*, 38, 29
- Singh, K. P., Tandon, S. N., Agrawal, P. C., et al. 2014, in *Proc. SPIE*, Vol. 9144, *Space Telescopes and Instrumentation 2014: Ultraviolet to Gamma Ray*, 91441S
- Stanek, K. Z., Matheson, T., Garnavich, P. M., et al. 2003, *ApJ*, 591, L17
- Tandon, S. N., Subramaniam, A., Girish, V., et al. 2017, *AJ*, 154, 128
- Vadawale, S. V., Chattopadhyay, T., Mithun, N. P. S., et al. 2018, *Nature Astronomy*, 2, 50
- Woosley, S. E. 1993, *ApJ*, 405, 273
- Yadav, J. S., Agrawal, P. C., Antia, H. M., et al. 2016, in *Proc. SPIE*, Vol. 9905, *Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray*, 99051D

New GRBs Detected			
GRB ID	Trigger Time (UT)	Peak Count Rate (s ⁻¹)	T ₉₀ (s)
GRB151019	182937927.159	1371	5.4
GRB160310	195323235.005	1478	16.2
GRB151219	188212280.016	276	3.9
GRB151224	188619996.81	200	13.8
GRB160214	193137447.586	531	27.1
GRB160223	193917545.015	238	8.8
GRB160325	196585165.043	1479	13.3
GRB160720	206735125.015	412	14.3
GRB160805	208131979.903	138	21
GRB160824B	209742690.004	212	18.7
GRB160829	210176329.007	1641	5.8
GRB160418	198698926.319	536	31.7
GRB160128	191681984.314	714	22.3
GRB160221	193777005.024	379	11.2
GRB151217	188024302.683	214	13.0
GRB160119	190868908.237	204	30.5
GRB170915	243143491.009	259	10.6
GRB170901	241963200.064	311	20.7
GRB170825	241358409.010	498	5.9
GRB170808	239927270.013	1291	15.8
GRB170614	235136404.011	655	25.2
GRB170423	230676925.038	435	17.9
GRB170316	227379745.004	258	14.3
GRB170311	226935913.014	481	10.9
GRB170228	226004584.004	732	14.2
GRB170216	224959176.014	294	22.7
GRB170210	224390896.045	999	34.3
GRB180504	263099760.020	246	42.3
GRB180426	262444262.028	439	12.4
GRB180416	261562255.020	416	9.1
GRB180411	261145715.007	461	28.1
GRB180403	260458375.024	181	25.1
GRB180401	260309858.072	798	20.9
GRB180526A	264942249.0	661	23.8
GRB180603A	265738980.0	609	31.1

Table 1. New GRBs detected with the machine learning algorithm described in this paper.