Serb. Astron. J. № 212 (2026), 107 - 116 **UDC 523.98 + 551.510.535 + 004.8**
DOI: https://doi.org/10.2298/SAJ2612107S Original scientific paper

DETECTION OF SHORT-TERM PEAKS IN VLF/LF SIGNALS AROUND THE TIME OF GAMMA-RAY BURST DETECTIONS USING DEEP LEARNING

M. Sarnovský[1], A. Nina[2], L. Č. Popović[3], V. Srećković[2], A. Khudevych[1], V. Krešňáková[1], P. Butka[1] and P. Bednár[1]

[1] *Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, 04200 Košice, Slovakia*

E–mail: *martin.sarnovsky@tuke.sk*

[2] *Institute of Physics Belgrade, National Institute of the Republic of Serbia, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia*

[3] *Astronomical Observatory, Volgina 7, 11060 Belgrade 38, Serbia*

(Received: December 27, 2025; Accepted: March 18, 2026)

SUMMARY: Gamma-ray bursts (GRBs) are recognized as the most energetic and luminous astrophysical phenomena whose high-energy photons can ionize atmospheric particles and thus affect electrical conductivity. Consequently, electromagnetic waves such as Very Low and Low Frequency (VLF/LF) radio signals used for the lower ionosphere monitoring can be used to detect disturbances, i.e. the impact of GRBs on the area in which these signals propagate. This paper addresses application of deep learning for automatic detection the short-term peaks in VLF/LF signals around the time of satellite registration of GRBs. We utilized a refined hybrid architecture combining Convolutional Neural Networks (CNN) for automatic feature extraction and Long Short-Term Memory network (LSTM) for modeling sequential dependencies in the multi-channel time-series data. We used derived, peak-based representations of the raw ionospheric waveforms and trained the proposed neural network architecture on such pre-processed data representations. The experiments with the sample of 54 short-lasting GRBs revealed limitations due to extreme class imbalance; however, a refined approach utilizing manual adjustment of positive-class weights and optimization of the classification threshold achieved the best detection performance. The final CNN–LSTM model demonstrated strong performance metrics, confirming the viability of CNN–LSTM approaches in this task.

Key words. Gamma-ray bursts – X-rays: bursts – Techniques: spectroscopic – Techniques: data analysis – Methods: statistical – Methods: numerical – Techniques: miscellaneous

1. INTRODUCTION

Gamma-ray bursts (GRBs) are among the most energetic and luminous events, releasing an immense amount of energy, capable of significantly affecting Earth's atmospheric layers, particularly the ionosphere Gehrels et al. (2009). There is a considerable interest in understanding the extent to which a GRB event can disturb the Earth's atmosphere. It is anticipated that high-energy photons can influence the ionosphere's electrical conductivity. In previous studies, it is shown that its disturbances can be manifested as changes in local electron density in the domain of low ionospheric heights and below, where Very Low Frequency (VLF, 3 kHz–30 kHz) and Low Frequency (LF, 30 kHz–300 kHz) radio signals propagate within the Earth-ionosphere waveg-

© 2026 The Author(s). Published by Astronomical Observatory of Belgrade and Faculty of Mathematics, University of Belgrade. This open access article is distributed under CC BY-NC-ND 4.0 International licence.

uide. The resulting perturbation in electron density causes time variations in the recorded amplitude and phase of these VLF/LF signals. Accurate detection and timely analysis of GRBs are crucial not only for astrophysical research but also for understanding their potential impacts on terrestrial systems such as communication networks.

While rare, the intensive GRB events have been reported to cause significant disturbances in the lower ionosphere that can persist for several minutes or over an hour (Fishman and Inan 1988, Huang et al. 2008, Hudec et al. 2010, Inan et al. 2007, Kerrache et al. 2024, Maeda et al. 2005, Pal et al. 2023, Tanaka et al. 2010, Slosiar et al. 2011). However, detecting low-intensity ionospheric perturbations is a very complex task, as the ionosphere is continually perturbed by numerous phenomena, including lightnings, solar activity, seismic processes, and electron precipitation (Arnaut et al. 2023, Hayakawa 2025, Kolarski et al. 2022a,b, Nina et al. 2021, Nina 2024a,b, Price et al. 2007, Silber and Price 2017). It is difficult to isolate a subtle perturbation and conclusively relate it to a specific GRB event. Nina et al. (2015) addressed this challenge by applying statistical analysis to the VLF/LF radio signals from six worldwide transmitters, successfully confirming detectability of short-term ionospheric responses to a sample of 54 short-lasting GRBs. This study showed that statistically significant perturbations could be observed immediately after the beginning of the GRB event, or with a delay of 60 to 90 seconds.

Despite the successful confirmation of the physical phenomenon through statistical techniques, the signals associated with these less intense reactions are often subtle and hover near the noise floor. A detailed, automated method is needed to reliably extract these faint, rare signatures from the multi-channel radio time-series data. Accurate detection and timely analysis are crucial for astrophysical research and for understanding potential impacts on terrestrial systems, such as communication networks.

The main motivation of this paper is to advance traditional statistical and threshold-based methods by investigating modern machine learning approaches for automatic detection of GRBs based on anomalies in the VLF/LF radio signals. This paper explores the application of deep learning, specifically leveraging a hybrid architecture combining Convolutional Neural Networks (CNNs) Zhao et al. (2024) with Long Short-Term Memory (LSTM) networks Speranza (2022). The CNN component is utilized for the automatic extraction of local patterns, such as transient oscillations or spectral spikes in the radio data, while the LSTM component processes these extracted features as a time series, enabling the network to capture crucial long-term sequential dependencies. Through the development and implementation of a refined CNN–LSTM model employing rolling-window preprocessing, this work demonstrates the viability of deep learning approaches for detection of changes in peak numbers before and after beginnings of the GRB events.

There are two main practical implications of this research study. First, the validation of GRB detectability via ground-based VLF/LF monitoring adds a valuable tool to the astrophysicists' repertoire, enabling synchronization between ground and space-based measurements. Second, the timing analysis of ionospheric responses supports radio engineers and forecasters, especially under intense space weather conditions when traditional ionospheric models may lack sufficient accuracy. Also, the study proved that a simple threshold-based technique is capable of detecting both long- and short-term ionospheric responses to GRBs.

The foundational success achieved through statistical peak counting established the basis for investigating modern machine learning methods to automatically extract these faint, rare signatures from noisy, multi-channel radio time-series data. The complexity of transient signal analysis in noisy environments has driven methodological convergence across related fields toward hybrid deep learning architectures. The CNN hierarchically extracts invariant spatial features or local patterns from individual inputs (such as transient oscillations or spectral spikes), which are then flattened and reshaped through a bridge mechanism to be compatible with recurrent processing. The LSTM unit subsequently processes these features to capture long-term temporal dynamics and complex sequential relationships, passing the information to a classification head—typically consisting of fully connected layers and a softmax function—to determine the final class. High-level architecture of such solution is depicted in Fig. 1.

This approach is methodologically aligned with recent research addressing transient events in related radio and gravitational time-series data. For example, the paper (Czech et al. 2018) employed a CNN–LSTM backbone in combination with sliding windows and explicit countermeasures against class imbalance (such as loss weighting) to perform eight-class identification of human-made radio-frequency interference (RFI) recorded by autonomous monitoring stations. Similarly, in (Lin and Wu 2021) authors utilized a complex Bayesian Convolutional Long-Short Term Memory Deep Neural Network (CLDNN) structure, also within a sliding-window framework, for the real-time detection and temporal segmentation of gravitational-wave (GW) events in interferometer strain data. These analogous studies validate the adoption of a hybrid CNN–LSTM architecture for processing raw one-dimensional astrophysical time-series data, providing a compelling motivation to apply these advanced sequential modeling capabilities to the challenging task of automatically detecting short-term GRB-induced VLF/LF anomalies. Both studies process raw one-dimensional astrophysical time-series data using a hybrid CNN–LSTM back-bone within a sliding-window framework, and

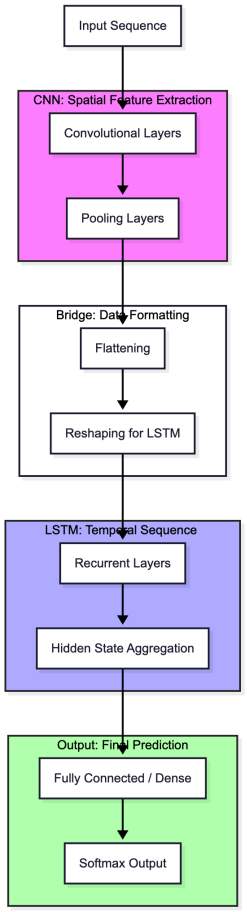

Fig. 1: Overall architecture of the CNN-LSTM network.

both explicitly address class imbalance during training. The key differences lie in the Bayesian versus deterministic treatment of weights. Also, both papers are methodologically related while targeting distinct transient phenomena under different experimental constraints.

Both investigations employ raw one-dimensional radio time-series, atwo-stage CNN-LSTM backbone, sliding windows for sample generation, and explicit countermeasures against class imbalance.Accordingly, Czech et al. (2018) and the present study are methodologically aligned, yet they remain distinct in scope: the former performs eight-class identification of human-made RFI, whereas the latter executes binary detection of astrophysical γ-ray bursts across multiple monitoring stations.

In this work, the CNN and LSTM networks were applied to 6 VLF and LF signals broadcast in Europe, the USA, and Australia and recorded in Serbia with a time resolution of 0.02 s. The periods of about 54 GRBs recorded by the Swift satellite were observed.

2. SIGNAL PROCESSING AND MODEL TRAINING PIPELINE FOR GAMMA-BURST DETECTION

This chapter presents an end-to-end pipeline for processing the VLF/LF radio-wave recordings and preparing them for training a neural network that detects GRBs. It begins by outlining the data set, its structure, and inherent limitations, proceeds through the full pre-processing workflow – from CSV conversion and cleaning to peak detection and adaptive binning at 5 s and 30 s resolutions and validates the implementation against the reference results. The chapter closes with three weighting strategies for an embedded Conv1D-LSTM network and a quantitative assessment of each on a held-out test set.

2.1. Dataset description

The data for gamma-ray burst (GRB) detection were provided by authors Nina *et al.* (2015) based on their publication. The dataset contains a list of all considered GRB events with the following information:

1. Burst date in the format `YYMMDD` (e.g., `090831` stands for 31 August 2009),
2. Start hour (UTC),
3. Start minute,
4. Start second,
5. Burst order within the day (1, 2, etc.).

The dataset contains amplitude values of 6 signals recorded by the AWESOME (Atmospheric Weather Educational System for Observation and Modeling of Electromagnetics) receiver located in Belgrade, Serbia in periods around 54 GRB events, each stored in a separate file. Each file includes data from 5 minutes before to 5 minutes after the burst, sampled at 0.02 s intervals. The columns in each file are:

1. Measurement index,
2. Signal amplitude from NWC transmitter (Australia),
3. Signal amplitude from NAA transmitter (USA),
4. Signal amplitude from NRK transmitter (Iceland),
5. Signal amplitude from ICV transmitter (Italy),
6. Signal amplitude from GQD transmitter (UK),
7. Signal amplitude from DHO transmitter (Germany).

The main limitations and potential sources of bias are:

- **Limited data volume.** Only 54 bursts, which may reduce the statistical significance of analyses on rare effects.
- **Geographic station distribution.** Uneven placement creates differences in sensitivity and signal delay across stations.

2.2. Data preprocessing

Methods and algorithms described in this section, as well as all peak detection formulas, are adopted from Nina et al. (2015) and adapted to reproduce the results of the original study.

2.2.1. Format conversion and initial cleaning

Raw VLF recordings were supplied as plain-text tables whose first eight columns correspond to transmitter amplitudes and auxiliary fields. All files were ingested and converted to comma–separated values (CSV) and stored in a dedicated repository to ensure a uniform downstream interface.

Six radio signals are included in the data — NWC (Australia), NAA (USA), NRK (Iceland), ICV (Italy), GQD (UK) and DHO (Germany). Because the GRB response is expected within a few minutes of the trigger, each record was clipped to the interval $[-120, +120]$ s. Given the native sampling rate $\Delta t = 0.02$ s this corresponds to exactly 12 000 samples. After slicing, the index was reset and converted to a physical time axis via $t_i = i\Delta t\,(i = 1, \ldots, 12\,000)$.

2.2.2. Signal-peak detection

For each of the six transmitting stations (NWC, NAA, NRK, ICV, GQD, DHO), the measured amplitude $A(t)$ was converted into a binary peak indicator. First, a 240 second segment was divided into consecutive 20 s windows. The mean amplitude within each window yielded a discrete series $A_{\mathrm{base}}(t_k)$. By applying linear interpolation between the centers of these windows, we obtained a continuous baseline curve $A_{\mathrm{base}}(t)$, which effectively removed any slow drift in the signal.

The deviation from this baseline was defined as

$$\Delta A(t) \;=\; A(t) \;-\; A_{\mathrm{base}}(t)\,. \tag{1}$$

To normalize $\Delta A(t)$, we computed a "robust" noise estimate A_{noise} as the 98th percentile of $|\Delta A|$ in other words, A_{noise} is the value exceeded by only the top 2% of absolute deviations. A data point at time t was classified as a peak if

$$\frac{A(t) \;-\; A_{\mathrm{base}}(t)}{A_{\mathrm{noise}}} \;\geq\; 2,3\,. \tag{2}$$

Whenever this condition was satisfied, the corresponding output column (`peak_australia`, `peak_italy`, etc.) was set to 1; otherwise, it remained 0. In this way, each continuous radio-amplitude trace is compressed into a sparse sequence of "peak" events, which serve as features for subsequent machine-learning classification of GRBs.

2.2.3. Temporal binning and adaptive thresholding

Single, isolated peaks often arise from random fluctuations. To suppress this noise and retain both rapid and slower accumulations of events, the binary peak indicators were aggregated on two timescales:

1. **Thirty-second bins.** Divide the 240 s interval into eight non-overlapping bins of duration 30 s each. For station s and bin b, let P_{bs} be the count of peaks in that bin, and let

$$\overline{P}_s \;=\; \frac{1}{8}\sum_{b=1}^{8} P_{bs} \tag{3}$$

be the mean number of peaks per bin for station s. We then mark bin b as "active" if

$$P_{bs} \;\geq\; 1.15\,\overline{P}_s\,. \tag{4}$$

This threshold (1.15 times the mean) identifies moderately elevated activity over a 30 s timescale.

2. **Five-second bins.** Subdivide the same 240 s interval into forty-eight consecutive 5 s bins. For each station s and 5 s bin b, compute P_{bs} as the number of peaks in that bin; then let

$$\overline{P}_s \;=\; \frac{1}{48}\sum_{b=1}^{48} P_{bs} \tag{5}$$

be the mean peak count across all forty-eight 5 s bins. A bin is flagged "active" if

$$P_{bs} \;\geq\; 2\,\overline{P}_s\,. \tag{6}$$

The stricter multiplier of 2 ensures that only significant short-term spikes survive.

Each station's results-stored as `bin_flag_australia`, `bin_flag_usa`, etc. By combining these two binning scales, we reduce random-noise artifacts while preserving both fast (5 s) and slower (30 s) accumulations of events.

Finally, all per-station tables are merged into a united table, which serves as the input feature set for the ANN classifier of gamma-ray bursts.

2.2.4. Verification of preprocessing results

To verify the correctness of the implemented preprocessing pipeline, the results were compared against the original data from Nina et al. (2015). The figures below show comparisons of the detected peak counts for individual stations obtained in this study with the corresponding results from original data.

The plots demonstrate that the distribution and quantitative characteristics of peaks in the analyzed time intervals agree closely with those reported in original study, confirming the validity of the pre-processing procedure and allowing us to proceed to model training.

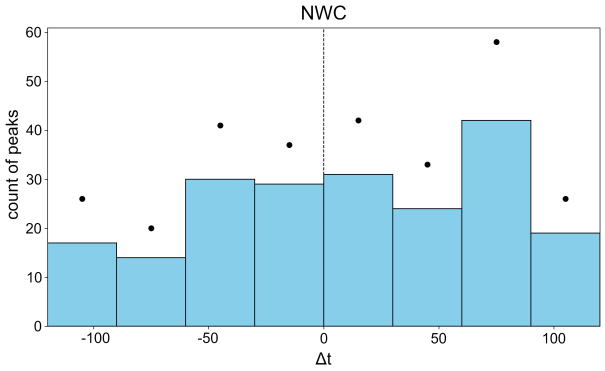

Fig. 2: Comparison of the number of peaks in NWC signal for 30-second bins obtained in this study (histogram) and in Nina et al. (2015) (dots). The value 0 at the x axis indicates the time of the beginning of GRB detection by the Swift satellite.

Example comparisons of peak-count plots

- Figure 2: Comparison for the NWC signal emitted in Australia (`peak_australia`) for 30-second bins.
- Figure 3: Comparison for the DHO signal emitted in Germany (`peak_germany`) for 5-second bins.

After this verification, we can proceed directly to the neural network model training stage.

Note: We trained and evaluated the model only on 30-second bins, because the 5-second-bin results were empirically indistinguishable.

2.3. Model training and evaluation

At the outset, the two directories – each containing 27 processed CSV files for individual GRB events - were concatenated into two master tables: one containing the 5 s binned data and the other containing the 30 s binned data. These two tables were identical in structure and differed only in the temporal resolution of the binned data. They formed the input datasets for all subsequent model preparation and training steps.

2.3.1. Experiment 1: distinct weighting assigned to each time bin

In this experiment we used a labelled time–series data set in CSV format as previously described in the earlier chapter that contains signals recorded by six geographic stations together with pre-computed binary flags for every thirty-second interval ("time bin"). Each row corresponds to a particular time stamp (seconds) and a specific station; the columns `bin_flag_peak_<station>` indicate whether a peak (1) or no peak (0) was present in that time bin. Samples belonging to the same case share the same identifier stored in the `data` column.

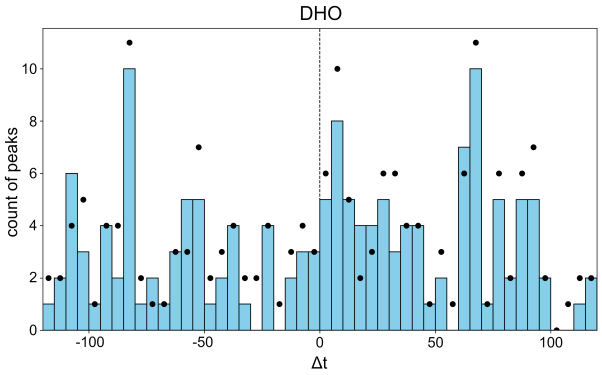

Fig. 3: Comparison of the number of peaks in DHO signal for 5-second bins obtained in this study (histogram) and in Nina et al. (2015) (dots). The value 0 at the x axis indicates the time of the beginning of GRB detection by the Swift satellite.

Rows containing missing values in any of the required columns (`data`, `Time`, the six station signals and the six pre-computed bin flags) were discarded. The unique case identifiers were then split at random (fixed seed) into training and test subsets in 80 %/20 % ratio. For each training case the time axis was divided into eight non-overlapping windows $[0, 30), [30, 60), \ldots, [210, 240)$ s. Within every window, all available signal values of each station were collected; if fewer than 1500 samples were present, the last value was repeated until the length reached 1500. Station–case pairs with zero samples in a window were removed. Thus, eight vectors of length 1500 were obtained per station, yielding a 1500×8 input matrix. The corresponding output label is the eight-element binary vector of pre-computed flags for that station. After processing all stations for all cases, the resulting matrices were stacked into a training tensor X_{train} of shape $(n_{\text{train}}, 1500, 8)$ and the binary labels into y_{train} of shape $(n_{\text{train}}, 8)$. The test set was built in exactly the same way, giving $X_{\text{test}} \in \mathbb{R}^{n_{\text{test}} \times 1500 \times 8}$ and $y_{\text{test}} \in \{0, 1\}^{n_{\text{test}} \times 8}$.

Because each of the eight output channels is strongly imbalanced, a time-bin-specific positive weight was introduced. For bin $i = 0, \ldots, 7$ let "pos"$_i$ and "neg"$_i$ be the numbers of ones and zeros in y_{train}; both counts were floored to 1 and

$$\text{pos_weight}_i = \frac{\text{neg}_i}{\text{pos}_i}$$

These eight weights were supplied to a custom weighted binary-cross-entropy loss implementation. If the true label equals 1 the loss term of that bin is scaled by pos_weight_i; if the label equals 0 the term is left unweighted.

The per-bin losses were averaged to obtain a scalar loss per sample. The architecture of the neural network used is presented in Table 1.

Table 1: Neural-network architecture.

Layer	Configuration	Output
Input	–	(1500, 8)
Conv1D	128 filters, kernel=5, padding=same, act=relu	(1500,128)
Batch norm	–	(1500,128)
Max pooling	pool=2	(750,128)
Dropout	rate=0.2	(750,128)
Conv1D	256 filters, kernel=3, padding=same, act=relu	(750,256)
Batch norm	–	(750,256)
Max pooling	pool=2	(375,256)
Dropout	rate=0.2	(375,256)
LSTM	64 units, return_seq=false	(64)
Dropout	rate=0.2	(64)
Dense	8 units, linear output	(8)

The model was compiled with the Adam optimiser (learning rate 0.001) and the custom loss. Training used a batch size of 16 for up to 100 epochs; 10 % of the training data were reserved for validation. Early stopping (patience = 10, best-weight restoration) and model checkpointing (lowest validation loss) were applied Hussein and Shareef (2024).

After training the logits of shape $(n_{\text{test}}, 8)$ were passed through a sigmoid and binarised at 0.5. Three metrics were reported:

1. exact-match accuracy: fraction of test samples whose eight predicted bits match the ground truth simultaneously,
2. bitwise accuracy: fraction of individual bits predicted correctly,
3. Matthews correlation coefficient (MCC) computed on the flattened bit vectors Obi (2023).

To optimise thresholding a grid search over $T \in [0, 1]$ in steps of 0.01 maximised the flattened MCC. The classification report (precision, recall, F_1-score, support) was then recomputed at the optimal threshold.

Table 2 summarises the global metrics on the test set before and after threshold tuning ($T = 0.51$) and a detailed flattened-bit classification report is given in Table 3.

Table 2: Performance before and after threshold tuning ($T = 0.70$) in Experiment 1.

Metric	Before	After
Exact-match acc.	0.0455	0.0152
Bitwise acc.	0.7083	0.6913
Overall MCC	0.0501	0.2646

Before threshold tuning the model achieved a bitwise accuracy of approximately 70.8 % but only 4.6 % exact matches; the MCC was close to 0, indicating difficulty in detecting the rare positive class. After tuning the threshold the MCC increased to 0.2646; precision for class (1) rose markedly while recall declined, reflecting a common precision–recall trade-off in imbalanced problems.

Table 3: Classification report for flattened bits in Experiment 1.

Class	Before tuning			After tuning ($T = 0.70$)		
	Precision	Recall	F_1	Precision	Recall	F_1
0	0.82	0.82	0.82	0.83	0.99	0.90
1	0.23	0.23	0.23	0.80	0.12	0.21
Macro avg.	0.53	0.52	0.53	0.81	0.56	0.55
Weighted avg.	0.71	0.71	0.71	0.82	0.83	0.77

2.3.2. *Experiment 2: universal weighting across all time bins*

In Experiment 2 we repeated the entire data-processing pipeline, network architecture and training protocol of Experiment 1, changing only the loss-weighting scheme. Instead of a separate positive weight for every time bin, all training labels were flattened into a single vector and one global positive weight was computed as

$$\text{pos_weight} = \frac{\text{total_negatives}}{\text{total_positives}},$$

which was applied uniformly to every bit.

Table 4 summarises the global metrics on the test set before and after threshold tuning ($T = 0.51$). Despite using a single weight, all metrics improve slightly compared with Experiment 1.

Table 4: Performance before and after threshold tuning ($T = 0.51$) in Experiment 2.

Metric	Before	After
Exact-match acc.	0.0455	0.0483
Bitwise acc.	0.7235	0.7700
Overall MCC	0.2900	0.2959

A detailed flattened-bit classification report is given in Table 5. Relative to Experiment 1, class (0) metrics stay high, whereas class (1) precision and recall improve modestly at the tuned threshold $T = 0.51$.

Table 5: Classification report for flattened bits in Experiment 2.

Class	Before tuning ($T = 0.50$)			After tuning ($T = 0.51$)		
	Prec.	Rec.	F_1	Prec.	Rec.	F_1
0	0.88	0.76	0.82	0.87	0.85	0.86
1	0.36	0.58	0.45	0.42	0.46	0.44
Macro	0.62	0.67	0.63	0.64	0.65	0.65
Weighted	0.78	0.72	0.75	0.78	0.77	0.78

2.3.3. *Experiment 3: manually specified class weights*

This experiment reproduces exactly the same preprocessing pipeline, network architecture and training settings used in the previous experiments, but changes the loss–weighting strategy. Instead of data–dependent weights we ran the code twice with two manually chosen positive weights:

- Run 1: `pos_weight` = 5;
- Run 2: `pos_weight` = 3 (negative weight implicitly = 1).

The training procedure (batch size 16, learning rate 0.001, early stopping with patience 10, etc.) was identical in both runs; only the scalar weight changed. During loss computation we passed the chosen `pos_weight`, so false negatives were penalised 5× (run 1) or 3× (run 2) more strongly than false positives.

Table 6 contrasts the key metrics of run 1 (`pos_weight` = 5) with those of run 2 (`pos_weight` = 3). For the further evaluation we proceeded with (`pos_weight` = 3).

Table 6: Impact of the manual positive weight in Experiment 3.

Metric	`pos_weight` = 5 ($T = 0.50$)	`pos_weight` = 3 ($T = 0.35$)
Neural Network		
Exact-match acc.	0.0000	0.3333
Bitwise acc.	0.5587	0.6837
Overall MCC	0.1315	0.3592

After training we evaluated the model on the test set (i) with the default threshold $T = 0.50$ and (ii) with the threshold tuned by a grid search ($T^\star = 0.35$; step 0.01) to maximise the MCC. The resulting global metrics for are summarised in Table 7.

Table 7: Performance before and after threshold tuning (run 2) ($T^\star = 0.35$) in Experiment 2.

Metric	Before	After
Exact-match acc.	0.3333	0.3333
Bitwise acc.	0.6837	0.6837
Overall MCC	0.2965	0.3592

The complete class-wise breakdown is with the final set of tuned parameters is given in Table 8. Here, we also compared CNN-LSTM model with selected baseline machine learning models. Using the identical train/test split, the same flattened-input preprocessing for traditional models, a MultiOutput wrapper to match the CNN-LSTM's multi-bit outputs, and identical evaluation criteria (simple hyperparameter tuning, exact-match accuracy, bitwise accuracy, MCC and per-class precision/recall/F1 with MCC-driven threshold search), the baseline models serve in complementary diagnostic role. Logistic Regression provides a fundamental linear benchmark that tests whether flattened features alone explain the labels, Random Forest captures non-linear interactions and threshold-like decision structure in tabular representations, and the margin-based SVM can explain, whether margin maximization outperforms probabilistic log-loss in this high-dimensional setting. In case that the CNN-LSTM model significantly outperforms all three across the same metrics, the most probable explanation is that sequence-aware convolutional and recurrent components are extracting temporally-local patterns and cross-time dependencies lost by flattening.

The experiments show, that the CNN-LSTM model dramatically improves recall and F1 for the minority class, which the baseline classifiers (Logistic Regression, Random Forest, calibrated SVM) completely ignore (precision/recall/F1 = 0.00). After optimizing thresholds by MCC, the baselines recover the recall for the minor class, but remain substantially weaker than the neural network.

Table 8: Classification report for flattened bits (run 2) in Experiment 3 compared with baseline models.

Class	($T = 0.50$)			($T = 0.35$/opt.)		
	Prec.	Rec.	F_1	Prec.	Rec.	F_1
Neural network (original)						
0	0.90	0.68	0.78	0.99	0.45	0.62
1	0.34	0.68	0.45	0.30	0.99	0.46
Macro avg.	0.62	0.68	0.62	0.65	0.72	0.54
Weighted avg.	0.79	0.68	0.72	0.86	0.55	0.59
Logistic Regression (baseline)						
0	0.80	1.00	0.89	0.81	0.95	0.88
1	0.00	0.00	0.00	0.44	0.14	0.21
Macro avg.	0.40	0.50	0.44	0.63	0.55	0.55
Weighted avg.	0.64	0.80	0.71	0.74	0.79	0.74
Random Forest (baseline)						
0	0.80	1.00	0.89	0.81	0.99	0.89
1	0.00	0.00	0.00	0.64	0.08	0.15
Macro avg.	0.40	0.50	0.44	0.73	0.54	0.52
Weighted avg.	0.64	0.80	0.71	0.78	0.80	0.74
Calibrated Linear SVM (baseline)						
0	0.80	1.00	0.89	0.81	0.98	0.89
1	0.00	0.00	0.00	0.53	0.07	0.13
Macro avg.	0.40	0.50	0.44	0.67	0.53	0.51
Weighted avg.	0.64	0.80	0.71	0.75	0.80	0.73

3. DISCUSSION AND FUTURE WORK

The viability of applying a hybrid deep learning architecture, specifically combining CNN with LSTM networks, to detect subtle, short-term VLF/LF ionospheric anomalies related to GRBs has been demonstrated in this paper. This methodological approach builds upon previous research, which initially confirmed the detectability of short-term reactions in the low ionosphere following GRB events. The established pipeline successfully processed VLF/LF recordings from six geographically distributed trans-

mitting stations into binned peak indicators, confirming the presence of statistically significant signatures in the radio signals. However, consistent with expectations for detecting low-intensity events, the extracted patterns associated with these short-term perturbations proved to be subtle and often near the background noise floor. This inherent subtlety complicates detection, as the ionosphere is continually perturbed by numerous phenomena including solar activity and lightning, making it difficult to isolate the GRB-specific perturbation.

A significant limitation encountered during this research was the reliance on a relatively small dataset, featuring only 54 short-lasting GRBs detected by the Swift satellite, which may reduce the statistical significance and generalizability of the findings. The low occurrence rate of genuine positive events, characteristic of this astrophysical phenomenon, resulted in an extreme class imbalance during training. Initial classification experiments utilizing generalized weighting schemes and default thresholding settings yielded only moderate results, indicating that classification accuracy remained below expectations when attempting to detect the rare positive class. For instance, despite achieving acceptable bitwise accuracy, the overall MCC was often close to zero before optimization, reflecting the model's initial difficulty in reliably isolating the positive signatures. Optimal detection rates, achieving an overall MCC of 0.3592, were ultimately realized only through the manual adjustment of positive-class weights and a selection of the classification threshold. This necessity for specialized hyperparameter tuning strongly suggests that while the deep learning model can perceive the weak multi-station GRB signature, it requires explicit algorithmic guidance to distinguish it reliably from ambient background noise.

In the current experimental setup, the baseline classifiers appear to fail at detecting the minority class even with the same class-weighting and threshold optimization. The achieved result may be a combination of severe label imbalance and asymmetries. Still, important differences remain even using the same weighting because the methods optimize different objectives and produce scores with different calibration properties. The neural network architecture additionally benefits from sequence-aware feature extraction that is not available to flattened-tabular estimators; consequently, weighted baselines often close much of the performance gap but do not always eliminate it, especially on metrics sensitive to temporal structure or on individual outputs with very low prevalence. To make comparisons fairer in the evaluation phase, an identical global threshold-search procedure based on maximizing the MCC was applied to every model, and probability outputs that were not naturally probabilistic were calibrated beforehand so that the threshold sweep operated on comparable probability estimates. Such post-hoc alignment reduces, but does not remove, residual differences arising from training-time weighting, calibration variance, per-output prevalence heterogeneity, and the neural model's ability to leverage temporal patterns, suggesting that the most robust fairness checks are cross-validated calibration and per-output threshold selection in addition to matched weighting during training.

To advance this proof-of-concept toward a more robust and automatically deployable detector, several critical possibilities for future research exist. Primarily, expanding the data volume and quality is crucial, requiring either an extension of the recordings duration or the incorporation of signals from additional transmitting stations to acquire a larger reference sample of true GRB events. Given the signal's subtlety and the complexity of the propagation environment, future work should investigate more sophisticated noise-filtering techniques and multi-channel signal synchronization methods to better isolate the specific GRB-associated signal from localized fluctuations. Finally, integrating independent data sources, such as satellite observations, for event cross-validation will be vital, significantly increasing confidence in the events identified by the VLF/LF anomaly detector and providing a more comprehensive understanding of the ionospheric reaction. The work presented here confirms the methodological viability of using modern machine learning to investigate indirect GRB detection and delineates a clear path forward for enhancing sensitivity and reliability.

4. CONCLUSION

During this work, a methodology was developed for collecting and pre-processing VLF/LF signals from six geographically distributed transmitting stations, followed by the application of a CNN-LSTM neural network for the classification of potential GRBs. The main idea of the research presented in this paper was to extract the peak-based representations of the raw signal data and train the proposed neural network architecture on such preprocessed representations, instead of raw signal data. The extracted patterns proved to be quite subtle and hovered near the noise floor, which is an expected outcome for such an experimental approach, given that no previous studies have simultaneously analyzed multi-station VLF/LF spikes to detect GRBs. Nevertheless, the mere ability to record these statistically significant signatures attests to the viability of the concept: the trained models are capable of detecting rare and short-lived disturbances even when they are barely discernible in any single channel.

The primary factors influencing classification performance were the low occurrence rate of genuine peak events and the heterogeneity of background noise across different locations. Standard weighting schemes and thresholding yielded only moderate results, and classification accuracy remained below expectations. However, when manual adjustment of

positive-class weights was implemented alongside a careful selection of thresholds, the model achieved its best detection rates, demonstrating that even a weak multi-station signal can be "extracted" through meticulous tuning of algorithmic hyperparameters.

To enhance the model's sensitivity, the next logical step is to expand both the volume and quality of the input data: incorporating additional transmitting stations or extending the duration of recordings to collect more reference examples of true GRB events. It is also advisable to investigate more sophisticated noise-filtering techniques and multi-channel signal synchronization to better isolate the GRB-associated signal from local fluctuations. Furthermore, combining these data with independent information sources (e.g., satellite observations) for cross-validation of events will increase confidence in the obtained results.

Thus, although the initial patterns were very faint, the conducted experiments confirmed the feasibility of applying modern machine learning methods to detect multi-station VLF/LF effects associated with gamma-ray bursts, and they have delineated promising avenues for future research.

Acknowledgements – This work was supported by the Slovak APVV agency Slovak-Serbian project SK-SRB-23-0029, and the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Serbian-Slovak project 337-00-3/2024-05/11). The authors acknowledge funding provided by the Institute of Physics Belgrade and the Astronomical Observatory (the contract 451-03-136/2025-03/ 200002) through the grants by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia.

REFERENCES

Arnaut, F., Kolarski, A. and Srećković, V. A. 2023, Univ, 9, 436

Czech, D., Mishra, A. and Inggs, M. 2018, Astronomy and Computing, 25, 52

Fishman, G. J. and Inan, U. S. 1988, Natur, 331, 418

Gehrels, N., Ramirez-Ruiz, E., and Fox, D. 2009, ARA&A, 47, 567

Hayakawa, M. 2025, Atmosphere, 16, 1312

Huang, W.-G., Gu, S.-F. and Shen, H. 2008, ChA&A, 32, 65

Hudec, R., Spurny, M., Krizek, M., et al. 2010, AdAst, 2010, 428943

Hussein, B. M. and Shareef, S. M. 2024, in ITM Web of Conferences, Vol. 64, EDP Sciences, 01003

Inan, U. S., Lehtinen, N. G., Moore, R. C., et al. 2007, Geophysical Research Letters, 34, L08103

Kerrache, F., Ammar, A., Ikhlef, R., et al. 2024, Journal of Geophysical Research: Space Physics, 129, e2023JA031721

Kolarski, A., Srećković, V. A. and Mijić, Z. 2022a, CoSka, 52, 105

Kolarski, A., Srećković, V. A. and Mijić, Z. R. 2022b, Applied Sciences, 12

Lin, Y.-C. and Wu, J.-H. P. 2021, PhRvD, 103, 063034

Maeda, K., Tomizawa, I., Shibata, T. F., et al. 2005, Geophysical Research Letters, 32, L18807

Nina, A. 2024a, Remote Sensing, 16, 397

Nina, A. 2024b, Remote Sensing, 16, 1330

Nina, A., Simić, S., Srećković, V. A. and Popović, L. Č. 2015, Geophysical Research Letters, 42, 8250

Nina, A., Biagi, P. F., Mitrović, S. T., et al. 2021, Atmosphere, 12, 444

Obi, J. C. 2023, World Journal of Advanced Engineering Technology and Sciences, 8, 308

Pal, S., Hobara, Y., Shvets, A., et al. 2023, Atmosphere, 14, 217

Price, C., Yair, Y. and Asfur, M. 2007, Geophysical Research Letters, 34, L09805

Silber, I. and Price, C. 2017, Surveys in Geophysics, 38, 407

Slosiar, R., Hudec, R., Kocka, M., Marko, R. and Zatko, M. 2011, AIP Conference Proceedings, 1358, 393

Speranza, G. 2022, AI, 6, 215, Survey of LSTM architectures, mechanisms, and modern applications

Tanaka, Y. T., Raulin, J.-P., Bertoni, F. C. P., et al. 2010, ApJ, 721, L24

Zhao, X., Wang, L., Zhang, Y., et al. 2024, Artificial Intelligence Review, 57, 1, Review of recent advances and applications of CNNs in computer vision

ДЕТЕКЦИЈА КРАТКОТРАЈНИХ ПИКОВА У VLF/LF СИГНАЛИМА ОКО ВРЕМЕНА ДЕТЕКЦИЈЕ ГАМА-ЗРАКА КОРИШЋЕЊЕМ ДУБОКОГ УЧЕЊА

M. Sarnovský[1], А. Нина[2], Л. Ч. Поповић[3], В. Срећковић[2], A. Khudevych[1], V. Krešňáková[1], P. Butka[1] and P. Bednár[1]

[1] *Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, 04200 Košice, Slovakia*

E–mail: *martin.sarnovsky@tuke.sk*

[2] *Институт за физику у Београду, Институт од националног значаја за Републику Србију, Универзитет у Београду, Прегревица 118, 11080 Београд, Србија*

[3] *Астрономска опсерваторија, Волгина 7, 11060 Београд 38, Србија*

УДК 523.98 + 551.510.535 + 004.8
Оригинални научни рад

Бљескови гама зрака (енг. *gamma ray bursts* - GRBs) су препознати као најенергичнији и најсветлији астрофизички феномени чији фотони високе енергије могу јонизовати атмосферске честице и тако утицати на електричну проводност. Сходно томе, електромагнетни таласи попут радио сигнала веома ниске и ниске фреквенције (енг. *very low and low frequency* - VLF/LF) који се користе за посматрање ниске јоносфере могу се користити за детекцију поремећаја, тј. утицаја GRB-ова на подручје у којем се ови сигнали шире. Овај рад се бави применом дубоког учења за аутоматизовано откривање краткорочних пикова у VLF/LF сигналима око времена регистрације GRB-ова сателитом. Користили смо рафинирану хибридну архитектуру која комбинује конволуционе неуронске мреже (енг. *Convolutional Neural Networks - CNN*) за аутоматизовану екстракцију карактеристика и мрежу дуге краткорочне меморије (енг. *Long Short-Term Memory network* - LSTM) за моделирање секвенцијалних зависности у вишеканалним временским серијама података. Експерименти са узорком од 54 краткотрајна GRB-а открили су ограничења због екстремне неравнотеже класа. Међутим, усавршени приступ који користи ручно подешавање тежина позитивних класа и оптимизацију прага класификације постигао је најбоље перформансе детекције. Коначни CNN-LSTM модел показао је јаке метрике перформанси, потврђујући одрживост CNN-LSTM приступа за поуздану детекцију GRB-а на основу VLF/LF јоносферских сигнала.