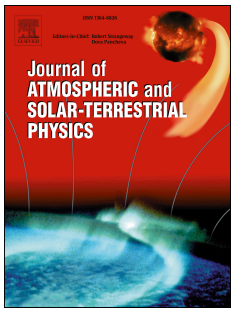
Journal Pre-proof



Automated Supervised Identification of Thunderstorm Ground Enhancements (TGEs)

Davit Aslanyan

PII: S1364-6826(26)00146-X

DOI: https://doi.org/10.1016/j.jastp.2026.106862

Reference: ATP 106862

To appear in: *Journal of Atmospheric and Solar-Terrestrial Physics*

Received Date: 9 November 2025

Revised Date: 21 May 2026

Accepted Date: 13 June 2026

Please cite this article as: Aslanyan, D., Automated Supervised Identification of Thunderstorm Ground Enhancements (TGEs), *Journal of Atmospheric and Solar-Terrestrial Physics*, https://doi.org/10.1016/j.jastp.2026.106862.

Automated Supervised Identification of Thunderstorm Ground Enhancements (TGEs)

Davit Aslanyan
Alikhanyan National Laboratory (Yerevan Physics Institute), Yerevan 0036, Armenia

Abstract

Thunderstorm Ground Enhancements (TGEs) are bursts of high-energy particle fluxes detected at Earth's surface, linked to the Relativistic Runaway Electron Avalanche (RREA) mechanism within thunderclouds. Accurate detection of TGEs is vital for advancing atmospheric physics and radiation safety, but event selection methods heavily rely on expert-defined thresholds. In this study, we use an automated supervised classification approach on a newly curated dataset of 2024 events from the Aragats Space Environment Center (ASEC). By combining a Tabular Prior-data Fitted Network (TabPFN) with SHAP-based interpretability, we attain 94.8% classification accuracy with 96% precision for TGEs. The analysis reveals data-driven thresholds for particle flux increases and environmental parameters that closely match the empirically established criteria used over the last 15 years. Our results demonstrate that modest but concurrent increases across multiple particle detectors, along with strong near-surface electric fields, are reliable indicators of TGEs. The framework we propose offers a scalable method for automated, interpretable TGE detection, with potential uses in real-time radiation hazard monitoring and multi-site atmospheric research.

1. Introduction

Charge separation in thunderclouds, driven by updrafts of warm air and interactions among hydrometeors, creates oppositely directed dipoles within the cloud. The oppositely directed atmospheric electric field (AEF) in the upper and lower dipoles accelerates free electrons toward open space and the Earth's surface. Free electrons are abundant in the atmosphere due to extensive air showers (EASs). Electric fields generated by strong thunderstorms transfer energy to these electrons, accelerating them and, under certain conditions, leading to electron-photon avalanches. These avalanches propagate through large volumes of the atmosphere, covering areas in the range of several hundred square meters to a few square kilometers (Chilingarian et al., 2022a) when they reach the Earth's surface, significantly increasing natural gamma radiation (NGR), which affects radiation safety and climate. These enhancements can last from seconds to tens of minutes (Chilingarian et al., 2020).

The ionized channels formed by relativistic electrons create pathways for lightning leaders to move toward the ground (Chilingarian et al., 2017). A key physical process behind these atmospheric particle flux increases is the Relativistic Runaway Electron Avalanche (RREA) process, introduced by Gurevich et al. (1992). RREAs are crucial for understanding a range of high-energy atmospheric phenomena, including Thunderstorm Ground Enhancements (TGEs) (Chilingarian et al., 2010, 2011), gamma-ray glows (Marisaldi et al., 2024), and both upward and downward terrestrial gamma-ray flashes (TGFs) (Fishman et al., 1994).

Recognizing the shared physical origin of brief, microsecond-scale TGFs, minute-long gamma glows, and surface-level TGEs is essential for understanding high-energy atmospheric physics and marks a step towards accepting RREA and EAS as a universal physical process that is responsible for the enhanced particle fluxes in the lower and upper atmosphere (Chilingarian et al., 2022a, 2023, 2024a).

While TGFs consist of microsecond bursts of gamma radiation originating in equatorial thunderstorms and observed from orbiting gamma-ray observatories positioned 400 to 700 kilometers above the source (Dwyer et al., 2012), TGEs manifest as intense, prolonged particle fluxes detectable at ground level that originate from accelerating electric fields located directly above particle detectors (often within 25-100 m), enabling detailed measurement of electron and gamma-ray energy spectra and characterization of thundercloud charge structures (Chilingarian et al., 2019; Aglietta et al., 1989).

Gamma glows represent gamma-ray radiation from RREA within a thunderstorm's upper dipole and are typically detected at higher altitudes by balloon or aircraft-based instruments. These emissions last from tens of seconds to several minutes, often ending with lightning discharges. Gamma-ray emissions observed at Earth's surface are sometimes also referred to as gamma glows due to their exclusive gamma-ray content, which indicates the altitude of the thunderclouds and the absorption of lower-energy particles (Tsuchiya et al., 2007; Kuriyama et al., 2022).

Given the variety of physical mechanisms that influence atmospheric particle fluxes, accurately identifying specific processes is essential. Since the initial observations at Aragats in 2009, over 1,000 TGE events have been recorded at mountain observatories worldwide, including locations in Eastern Europe, Japan, Russia, Germany, and Armenia (Chilingarian et al., 2025; Alexeenkko et al., 2002; Kudela et al., 2017; Shepetov et al., 2021; Eack at al., 2000). Recent observations extend these findings to Mt. Hermon in Israel (Mauda et al., 2025) and to sites in Finland (Leppänen et al., 2025) and Slovakia (Kísvardai et al., 2025). Along with numerical simulations, these observations offer detailed insights into the RREA mechanism and related cloud-charge distributions within the lower atmosphere.

In this work, we expand the study of TGEs by applying a supervised classification model to a newly curated dataset of 2024 events. Using the Tabular Prior-data Fitted Network (TabPFN) model and SHapley Additive exPlanations (SHAP)-based interpretability analysis, we derive data-driven thresholds for key variables and demonstrate their consistency with manual criteria developed over 15 years of observational experience.

2. Methods

The Aragats Space Environment Center (ASEC), located at the Aragats high-altitude research station of the Cosmic Ray Division, is situated at 3200 meters on Mt. Aragats in Armenia. It experiences frequent and intense thunderstorms during spring and summer, often causing thunderclouds to descend below 100 meters above the detectors. Low-altitude thunderclouds create a significant electric field gradient between the main negative (MN) charge layer and its mirror image on Earth's surface, which promotes the RREA process registered at the Earth's surface as TGE. This gives ASEC a uniquely advantageous position for detecting particle bursts.

2.1. Detectors

Throughout its 80 years of continuous operation, ASEC has integrated various instruments for detecting different species of secondary cosmic rays (primarily neutrons, muons, electrons, and gamma rays), electric field disturbances, and meteorological conditions. In this study, we analyze data from the STAND 1cm, STAND 3cm, SEVAN particle detectors, BOLTEK EFM-100 electric field mills, and automatic weather stations from DAVIS Instruments. The detailed descriptions of these detectors can be found in Chilingarian et al. (2024b). The STAND 1cm detector, illustrated in Figure 1, consists of a three-layer assembly of 1 cm-thick molded plastic scintillators with a 1-square-meter sensitive area stacked vertically, along with a 3 cm-thick scintillator positioned nearby.

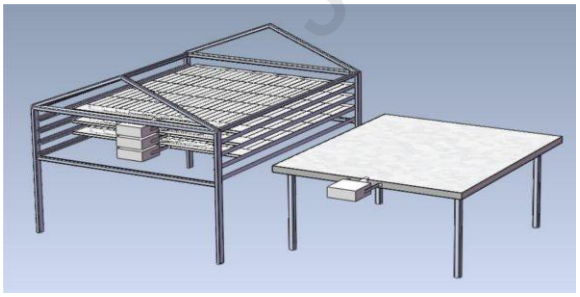


Figure 1. The stacked STAND 1cm detector consisting of three 1 cm thick, 1 m² area scintillators and a stand-alone 3cm thick scintillator of the same area.

Light from the scintillator is re-emitted in the long-wavelength range by optical spectrum-shifter fibers and then directed to the FEU-115M type photomultiplier (PMT). The peak luminescence occurs at 420 nm, with a luminescence decay time of approximately 2.3 ns. The high voltage and discrimination threshold of the photomultiplier are adjusted so that the upper scintillator of the 1 cm STAND detector has an energy threshold of about 0.8 MeV. This threshold is defined for

electrons: the detector assembly includes a 0.3 cm aluminum cover and the 1 cm thick scintillator, which effectively stops incident electrons with energies below 0.8 MeV. The plastic scintillator is optimized for detecting charged particles; however, it can also register gamma rays via Compton-scattered electrons generated inside the scintillator material. The detector is integrated into a data synchronization system that records particle count rates with a 50 ms sampling interval.

The 3 cm STAND detector, as shown in Figure 2, consists of a four-layer assembly of 3 cm-thick plastic scintillators with a 1-square-meter sensitive area stacked vertically. The detector electronics operate similarly to the 1 cm STAND detectors. GEANT4 simulations were performed to calculate the energy thresholds required for electrons to penetrate specific depths. By registering coincidences across the vertically stacked layers (e.g., signals in the top layer only, versus the top two layers), the detector differentiates electrons with energies of 10, 20, 30, and 40 MeV.

Figure 2. The Stand 3cm detector consisting of four stacked 3 cm thick scintillators with an area of 1 m².

SEVAN, illustrated in Figure 3, is a hybrid particle detector that measures gamma rays, neutrons, muons, positrons, and electrons (Chilingarian et al., 2009; Chilingarian et al., 2018). It features a three-layer system made up of plastic scintillator slabs, lead absorbers (which serve to filter low-energy background and facilitate particle identification based on penetration depth), light guides, and PMTs. Physically, SEVAN consists of two identical assemblies of 100 cm x 100 cm x 5 cm outer plastic scintillator slabs, sandwiching a thicker, spectrometric central unit measuring 50 cm x 50 cm x 20 cm. The system differentiates between particle species via coincidence logic; for example, the upper layer acts as a charged-particle veto, allowing the central unit to identify neutral particles (neutrons and gamma-rays) by their lack of a precursor signal in the top slab.

The upper scintillator has a threshold of about 7 MeV, determined via GEANT4 simulations (Karapetyan and Hovsepyan, 2024).

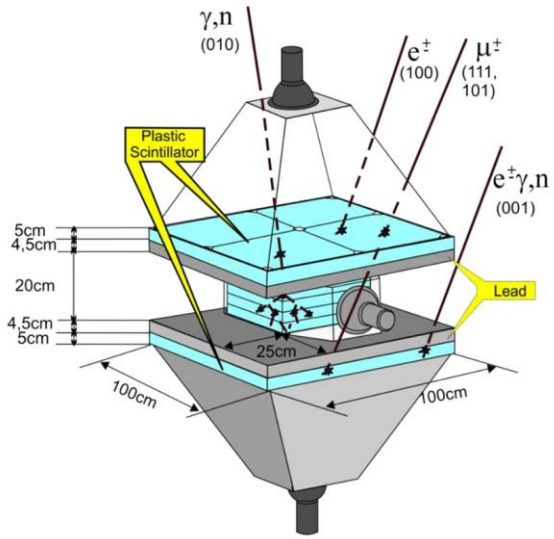


Figure 3. Assembly of SEVAN Detector

Information about the Boltek EFM-100 electric field mill and Davis Vantage Pro2 weather station (see one of the locations of these detectors on the roof of "Cuckoo's Nest" lab in Figure 4) can be found in Chilingarian et al. (2024c).

Figure 4. Electric field mill EFM 100 from BOLTEK, DAVIS weather station, and all-sky cameras installed on the roof of a small "Cuckoo's Nest" lab.

2.2. TGE Event definition

TGEs are classified using a detailed, multi-parameter method based on raw data collected over a 15-year span. This approach combines multi-detector monitoring of various cosmic ray species, near-surface electric fields (NSEF), and environmental data to confirm that each recorded event is related to thunderstorm-induced particle acceleration rather than unrelated background phenomena or equipment failures (Chilingarian et al., 2023).

A candidate TGE event is confirmed only when independent particle detectors record simultaneous and statistically significant increases in count rates. Synchronization between independent detectors is maintained via Network Time Protocol (NTP), while internal coincidences within stacked detectors are resolved within a window of 1 microsecond. Specifically, the procedure requires that at least three detectors, particularly the SEVAN, STAND1, and STAND3 scintillators, observe a flux rise exceeding three standard deviations (3σ) above pre-storm mean values. These baselines are calculated over a 30–60-minute interval of fair weather preceding a storm. Including detectors with higher energy thresholds, such as STAND3 and SEVAN, ensures that the detected increase accurately indicates the presence of RREAs and is not merely caused by lower-energy radon progeny radiation. While EAS can also trigger simultaneous responses across detectors, TGEs are distinguished by their longer duration and correlation with strong NSEF.

Along with the particle flux criteria, each TGE must also happen during a sufficiently strong atmospheric electric field. To be accepted, the absolute value of the NSEF during the event must surpass 5 kV/m. This method allows classifying TGEs based on their strength and duration into two categories: “large” TGEs (Chilingarian et al., 2024c), and “small” TGEs.

Small TGEs are defined as particle flux enhancement events where the particle count enhancement exceeds the 3σ significance threshold, but remains below a 20% increase. These events generally show only modest increases and limited high-energy electron counts. These less intense events are more common in summer, when higher temperatures cause higher cloud bases that reduce the particle flux reaching the Earth’s surface. In Figure 5, we display a summer TGE event that occurred on August 2, 2024 registered by STAND 3cm detector. The count rate increases slightly, reaching the 3σ threshold. The NSEF was quite strong, at -20 kV/m; however, the cloud base was very high, above 600 meters. Therefore, it is possible that the RREA was very intense high in the atmosphere, however, due to the high cloud base height, the avalanche was heavily attenuated. Most particles were absorbed in the thick atmosphere, and only a few gamma rays reached the detector. In any “border” situation for small TGEs, it is important to consider environmental factors and the season.

In contrast, Large TGEs are defined as particle flux enhancement events where the particle count enhancement exceeds 20% increases in particle detector counts above their normal levels across

all detectors considered. These significant events mainly happen in Spring and Fall, when lower temperatures lead to low cloud bases (often 25-50m above the surface). This proximity and the resulting short free passage distance minimize atmospheric attenuation, allowing near-surface electron accelerators to produce intense particle fluxes at ground level (Chilingarian et al., 2024c).

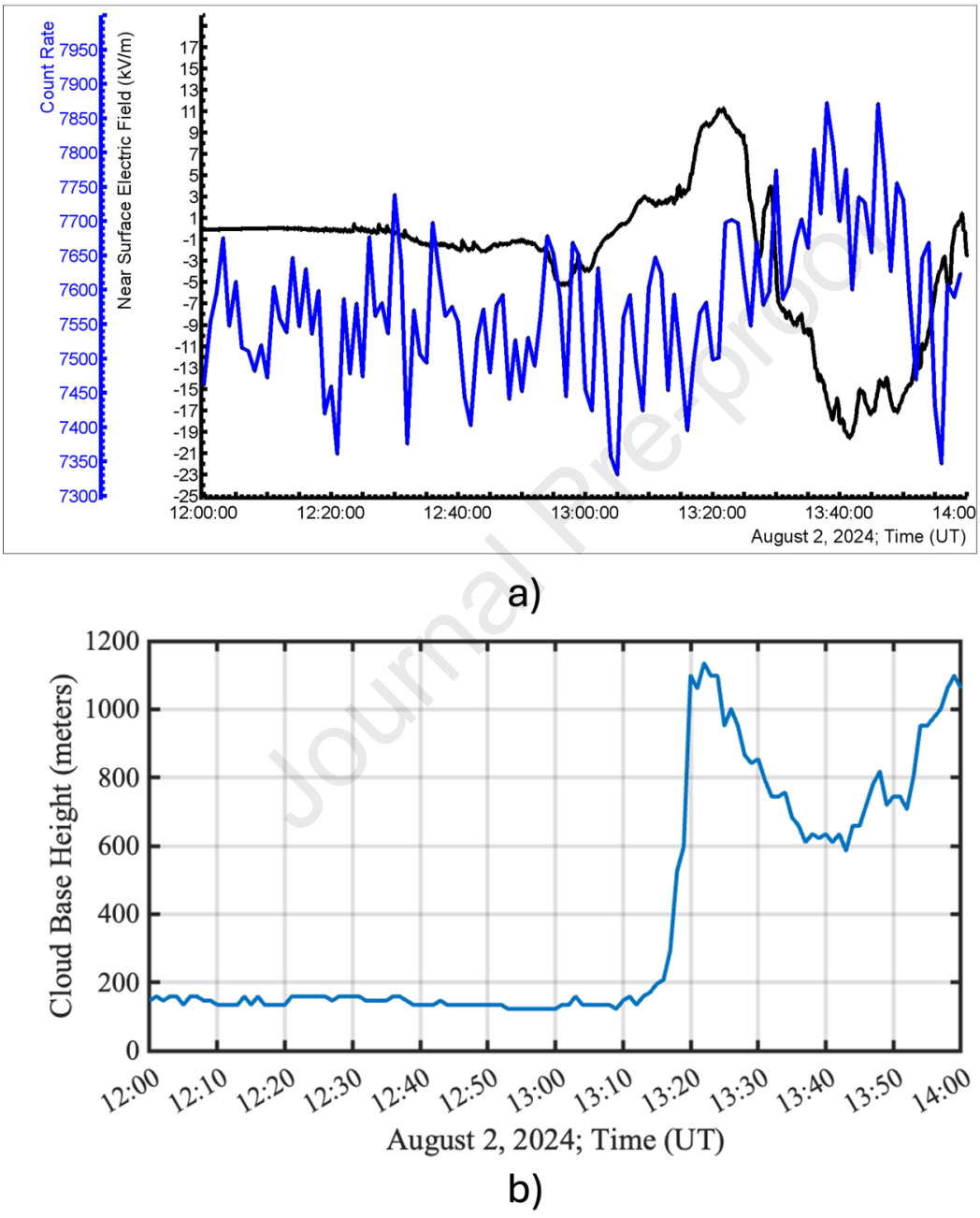


Figure 5. a) Small Summer TGE registered by STAND 3cm detector (blue curve); disturbances of electric field are shown in black. b) distance to the cloud base.

An example of a very intense TGE is shown in Figure 6. To highlight the dynamics of count rate change, we present a time series with 1-second resolution. The network of STAND 1cm detectors allowed measurement of particle fluxes with 50 ms sampling. On October 2, 2024 particle detectors on Aragats recorded a double-stage TGE with a sharp increase in gamma-ray and electron fluxes. A one-second measurement from a 1 cm thick, 1 m² outdoor scintillator shows nearly a tenfold rise (900%, 120 σ, the second peak). The black curve displays the NSEF, which is strongly negative during TGE. A broad peak began at 00:41:40, peaked at 00:42:35, then declined at 00:43:15 before immediately rising again until a cloud-to-ground (-CG) lightning flash abruptly ended the TGE at 00:43:36. The RREA electron flux in the cloud was sufficient to create an ionization channel in the lower atmosphere, providing a path for the lightning leader, a mechanism discussed in detail by Chilingarian et al., 2017. This event was the first and only TGE of 2024 with a large electron content (defined as an electron-to-gamma-ray ratio exceeding 20%). Electrons attenuate much faster than gamma-rays, so their presence implies that the acceleration region was very close to the surface. The increase in gamma-ray (primarily with energies <10 MeV) and electron flux detected by the STAND3 plastic scintillator (which is optimized to separate these species, see Chilingarian and Hovsepyan, 2023) reached 225% (125 σ). The 20-cm-thick, 0.25-m² area spectrometric scintillator of the SEVAN detector recorded a 50σ enhancement in the elementary particle flux. GEANT4 simulations of the detector response show that this scintillator (using the same spectral analysis configuration as the SEVAN light in Chilingarian et al., 2024d) detects electrons with energies above 10 MeV.

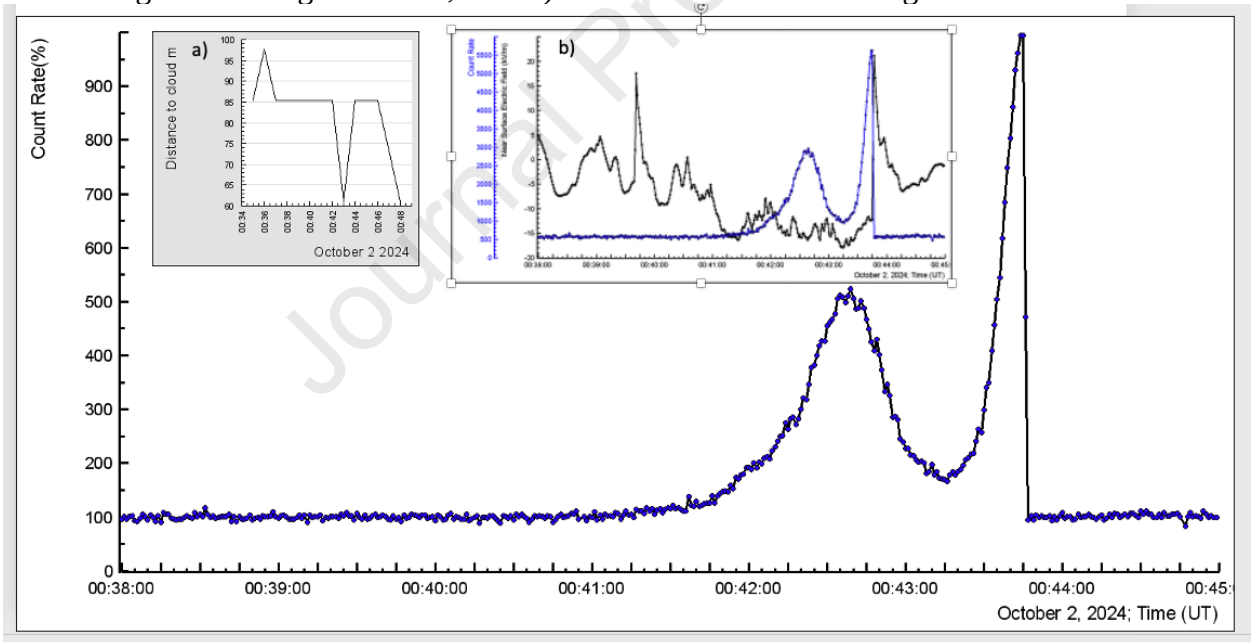


Figure 6. Time series with 1-second resolution of a large TGE event observed by the upper scintillator of the STAND 1cm detector (Note: The y-axis is normalized relative to the fair-weather background count rate, set at 100%). In the inset, a) shows the distance to the cloud base, and b) displays disturbances in the electric field.

Throughout previous studies, the application of this strict selection procedure has resulted in a consistent dataset of TGEs reported across several publications (e.g. Chilingarian et al., 2022b, 2024a, 2024b). In this study, we use the same methodology to build a dataset of TGE events

observed in 2024, which serves as the ground-truth for training and evaluating our new automated analysis methods.

2.3. TGEs observed in 2024

The 2024 core variables include data from the STAND 1cm, STAND 3cm, and SEVAN detectors, recorded as both percentage increases and statistical significance compared to pre-storm background levels. The detectors measure particle flux via their upper scintillators, which count all passing particles. To differentiate particles by energy, we analyze the coincidence of triggers across the stacked layers. When an event is recorded exclusively in the top scintillator layer and lower layers do not register a concurrent signal, we effectively select low-energy particles that are absorbed immediately upon striking the detector surface.

Measurements of NSEF are included along with a new constructed variable, ΔE, which is defined as the difference between the highest and lowest NSEF values recorded during the entire duration of the specific TGE event (from onset to termination). ΔE thus measures the dynamic variability of the electric field, offering insights into the mechanisms of particle enhancements. All meteorological parameters (temperature, humidity, cloud base height, solar radiation, atmospheric pressure, wind speed, and rain rate) were recorded to ensure correlation studies across variables.

To prevent overcounting of neighboring TGEs, only enhancements that significantly decrease after reaching the peak, before the second subsequent rise, were regarded as distinct TGEs, with at least a ten-minute gap between events. Redundancy within the ASEC network, which includes three electric field mills and two weather stations, minimized the effects of instrument outages; any gaps were filled by redundant sensors.

The dataset also includes training examples without TGE: control samples recorded under fair-weather conditions and during thunderstorms that did not meet TGE confirmation criteria. These entries (TGE = 0) provide the necessary contrast to confirmed events (TGE = 1), ensuring the model learns meaningful distinctions rather than defaulting to single-class prediction. This balanced design underpins the robustness of subsequent classification experiments.

2.4. Artificial Intelligence Model and Threshold Determination

This section describes the statistical modeling used to compare the model-derived TGE classification rule with the empirical decision rules for TGE selection from our previous work. In addition, we observed strong correlations between TGEs and atmospheric parameters during particle count bursts. We used the TabPFN, a recently developed artificial intelligence (AI) model specifically designed for small to medium-sized datasets, usually containing up to 10,000 samples (Hollmann et al., 2025). Unlike traditional machine learning methods, which require model selection, feature preprocessing, and hyperparameter tuning, TabPFN is a prior-data fitted model that is trained once and applied directly to new tasks.

2.4.1. TabPFN Architecture and Training

As detailed by Hollmann et al., 2025, TabPFN is based on a transformer architecture, originally developed for sequence modeling in natural language processing, but adapted to the tabular domain. TabPFN utilizes a 12-layer 2D transformer encoder, where each layer implements a two-way attention mechanism that alternates between 1D feature attention and 1D sample attention. This design ensures that the model is invariant to the order of features and samples. Unlike traditional supervised deep learning models that update model parameters per dataset, TabPFN uses in-context-learning (ICL) to learn how to solve prediction tasks. Its training procedure involves exposure to a collection of approximately 100 million synthetic datasets that were generated by structural causal models (SCMs). The datasets incorporate common tabular challenges, such as missing values, outliers, and varying feature scales. The extensive training on synthetically generated classification tasks, allows the model to function as an amortized Bayesian inference engine for tabular data. For a new test sample x_{test} and a labelled training set $D_{train} = (x_{train}, y_{train})$, TabPFN can approximate a Bayesian posterior predictive distribution $p(y_{test}|x_{test}, D_{train})$ over class labels conditioned on input features in a single forward pass (Hollmann et al., 2025).

In practice, this means that TabPFN can provide out-of-the-box high-quality predictions without iterative retraining or gradient-based weight updates like Adam (Kingma & Ba, 2014), making it particularly well-suited to scenarios in physics where datasets may be limited in size, costly to generate, or not amenable to extensive hyperparameter tuning. By encoding statistical relationships between features and outcomes during pretraining, TabPFN bypasses the sequential model-building process characteristic of ensemble methods such as gradient boosting.

2.4.2. Feature Selection

The input features were selected based on their signal-to-noise ratio (SNR), including percentage enhancements of peak fluxes from STAND1, STAND3, and SEVAN upper scintillators, which capture the cumulative elementary particle flux, as well as environmental parameters such as near-surface temperature and the absolute value of the NSEF.

To provide a rigorous basis for feature inclusion, we design the input as a vector $X = [\Delta_{STAND1}, \Delta_{STAND3}, \Delta_{SEVAN}, |NSEF|, T, H]^T$, where the enhancement Δ for each detector is defined as:

$$\Delta = \frac{C_{peak} - \mu_{bg}}{\mu_{bg}} \times 100$$

Here C_{peak} is the peak count rate of a detector and μ_{bg} is the pre-storm background mean. To determine which detector channels to include in the automated model, we evaluate their SNR, defined as the signal amplitude divided by the standard deviation of the background (σ_{bg}):

$$SNR = \frac{C_{peak} - \mu_{bg}}{\sigma_{bg}}$$

As described in Section 2.3, the coincidence channels are physically relevant for isolating specific particle energies. However, from a statistical modelling perspective, they inherently register lower count rates and are prone to high background noise, which can result in low SNR, so they reserved for expert review. Conversely, the upper scintillators of the STAND 1cm, STAND 3cm, and SEVAN detectors count all passing particles without coincidence restrictions (Section 2.3), which results in higher baseline count rates. This minimizes the relative background noise and results in high SNR. As such, the percentage enhancements in the upper scintillators of the detectors were selected as inputs to the TabPFN model.

While a large enhancement in the particle count rates serves as the primary, deterministic trigger for identifying a TGE, the enhancements need to be physically contextualized. As detailed in Section 2.2, manual TGE confirmation relies heavily on environmental conditions to differentiate RREAs. To allow the model to autonomously identify TGEs, we also include the absolute value of the NSEF ($|NSEF|$), as well as outside temperature (T), and cloud base height (H). Inclusion of these meteorological parameters gives the model the data necessary to perform sanity checks, such as verifying the presence of an accelerating electric field and checking that meteorological conditions are physically capable of supporting a ground-level avalanche. This ensures that the model learns to identify TGEs only when the observed particle count rate enhancements are thunderstorm related.

2.4.3. Interpretability and Threshold Determination

A limitation of TabPFN is that, like many deep learning models, it is intrinsically non-interpretable. Its decision-making relies on distributed representations within the transformer layers, which makes it non-trivial to directly attribute predictions to specific features. For applications where interpretability is essential - for instance, to uncover underlying physical mechanisms - TabPFN must therefore be complemented with post-hoc explainability techniques or contrasted with more transparent baseline models.

Unlike classical regression models, which provide coefficients reflecting the influence of each predictor, “black box” models require additional methods to understand the contributions of each feature. To address this, we utilized SHAP (Lundberg & Lee, 2017; Lundberg et al., 2018), a framework that quantifies each feature’s effect on individual predictions, allowing comparison and ranking of variables according to their relative importance. The SHAP values are grounded in game theory and are consistent: as a predictor’s effect on the model output increases, its SHAP value rises, and results are insensitive to feature scaling. For each observation, SHAP values ϕ_i across all features sum to the difference between the individual model prediction ($f(x)$) and the expected base value ($E[f(X)]$): $\sum_{i=1}^{M} \phi_i = f(x) - E[f(X)]$. Summary plots provide a global view of feature importance and directional influence. The horizontal axis indicates the SHAP value (impact on the predicted outcome), while the vertical axis ranks features by overall importance. Individual observations are represented by color-coded points reflecting feature magnitude and mean absolute SHAP values are displayed to quantify relative importance within the model.

To further explore the relationship between each variable and its SHAP value, in Figure 10, we plotted SHAP values against raw feature values. These scatterplots can be noisy, complicating threshold determination. To address this, we applied a locally weighted scatterplot smoothing (LOWESS) method (Cleveland, 1979; Cleveland and Devlin, 1988). LOWESS fits a smooth curve through the data without assuming a global functional form, capturing non-linear trends. The point where the smoothed curve crosses zero indicates the feature value at which its influence on TGE prediction changes sign, providing a practical, data-driven threshold for decision-making.

2.4.4. Validation and Performance Metrics

The model was evaluated using two complementary strategies to resolve the trade-off between the sample size and estimate stability. First, the model was trained to classify each observation in the 2024 dataset as either TGE or non-TGE using a leave-one-out cross-validation (LOOCV) strategy. LOOCV was our primary evaluation strategy and was used to produce the confusion matrix in Figure 8. In LOOCV approach, each of the N observations are held out as the test set while the model is trained on all remaining $N - 1$ observations, ensuring that every observation is used once for validation. This method reduces variance in performance estimates compared to a single 70/30 or 80/20 training/testing split and is especially valuable for datasets of modest size, where retaining as much training data as possible improves stability in learned thresholds. Next, to ensure that the results were not artifacts of the dataset composition, we performed sensitivity analysis using 10 iterations of randomized 80/20 data splits.

Given the imbalance in our dataset (69 TGEs vs. 27 non-TGEs), standard accuracy can be a misleading indicator of performance. We employed several complementary metrics to provide a robust assessment. Specifically, we utilized the Matthews Correlation Coefficient (MCC; Matthews, 1975), which is widely regarded as a balanced measure for binary classification even when classes are of very different sizes. The MCC produces a value in the range [-1:1], where “1” represents a perfect prediction, “0” corresponds to random guessing, and “-1” denotes total disagreement. We also report the F1-score (Goutte and Gaussier, 2005), defined as the harmonic mean of precision and recall, to quantify the balance between the purity of the selected sample and the completeness of the detection.

Finally, to visualize the structure of the high-dimensional data, we applied t-Distributed Stochastic Neighbor Embedding (t-SNE; Van der Maaten and Hilton, 2008). t-SNE is a non-linear dimensionality reduction technique well-suited for embedding high-dimensional data for visualization in a low-dimensional space of two or three dimensions.

3. Results

The final classifier achieved an overall accuracy of 94.8% on the test set. The MCC was calculated to be 0.87, which confirms that the model’s predictive power is robust across both positive and negative classes and is not an artifact of the dataset composition. The F1-score was 0.96, with a precision of 96% and recall of 97% for the TGE class. The model demonstrated exceptional stability during sensitivity analysis, with a mean MCC of 0.90 ± 0.06, indicating that the classification boundaries are intrinsic to the physical features rather than dependent on

specific training samples. The performance of the TabPFN model was also compared against standard baseline models using in tabular data tasks, such as XGBoost (Chen and Guestrin, 2016), CatBoost (Prokhorenkova et al., 2018), logistic regression (LaValley, 2008), and K-nearest neighbors with results summarized in Table 1.

Table 1. Comparison of TabPFN against other models

Model	Accuracy	MCC	F1-Score	Recall (TGE class)
TabPFN	94.8	0.87	0.96	0.97
XGBoost	94.8	0.87	0.96	0.99
CatBoost	94.8	0.87	0.96	0.99
Logistic Regression	93.8	0.85	0.96	0.96
KNN	80.2	0.47	0.87	0.93

The comparative analysis demonstrates that the TabPFN model performs on par with state-of-the-art models XGBoost and CatBoost. All non-linear models significantly outperform the KNN baseline (MCC = 0.47), which struggled to capture the structure of the event dataset. It is noteworthy, that TabPFN required no hyperparameter tuning or iterative training, and provides a robust baseline that is theoretically resistant to overfitting typically observed in boosted

ensembles on small datasets.

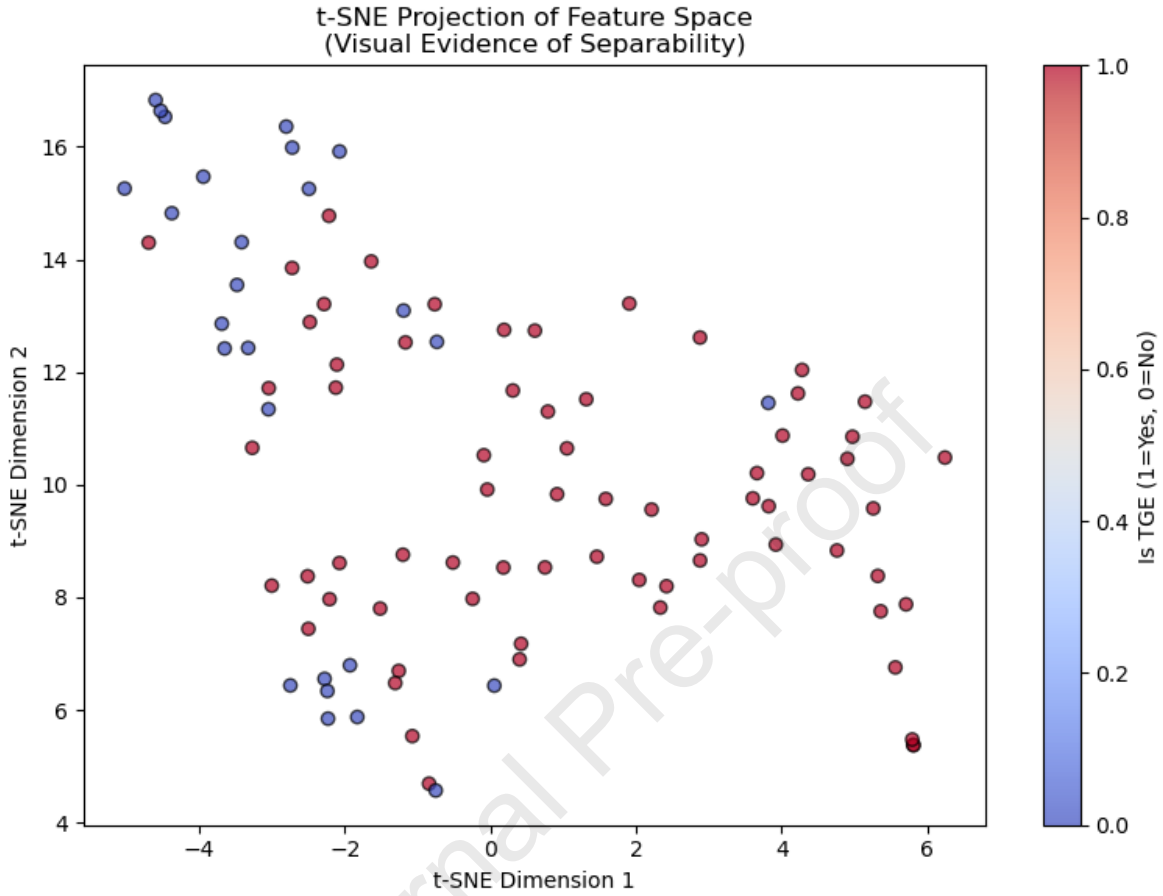


Figure 7. t-SNE visualization of the 2024 event dataset projected from the six-dimensional feature space (STAND1, STAND3, SEVAN enhancements, |NSEF|, Temperature, and Cloud Base Height). The color scale distinguishes confirmed TGEs (red, Class 1) from non-TGE background events (blue, Class 0).

The clear separation between TGE (red) and non-TGE (blue) events in Figure 7 visually confirms that TGEs occupy a distinct region in the feature space, supporting the high classification performance. The few outliers (e.g., non-TGE events appearing within the TGE region) likely correspond to 'near-miss' events—thunderstorms with strong electric fields that did not produce sufficient particle flux to meet the strict manual confirmation criteria. This overlap reflects the continuous nature of atmospheric particle acceleration rather than a failure of classification.

The final classifier achieved an overall accuracy of 94.8% on the test set, with 96% precision for the TGE class and a recall of 97%, demonstrating strong generalization performance. The confusion matrix showed two false negatives (True label 1, Predicted label 0) and three false positives (True label 0, Predicted label 1) for TGEs (Figure 8).

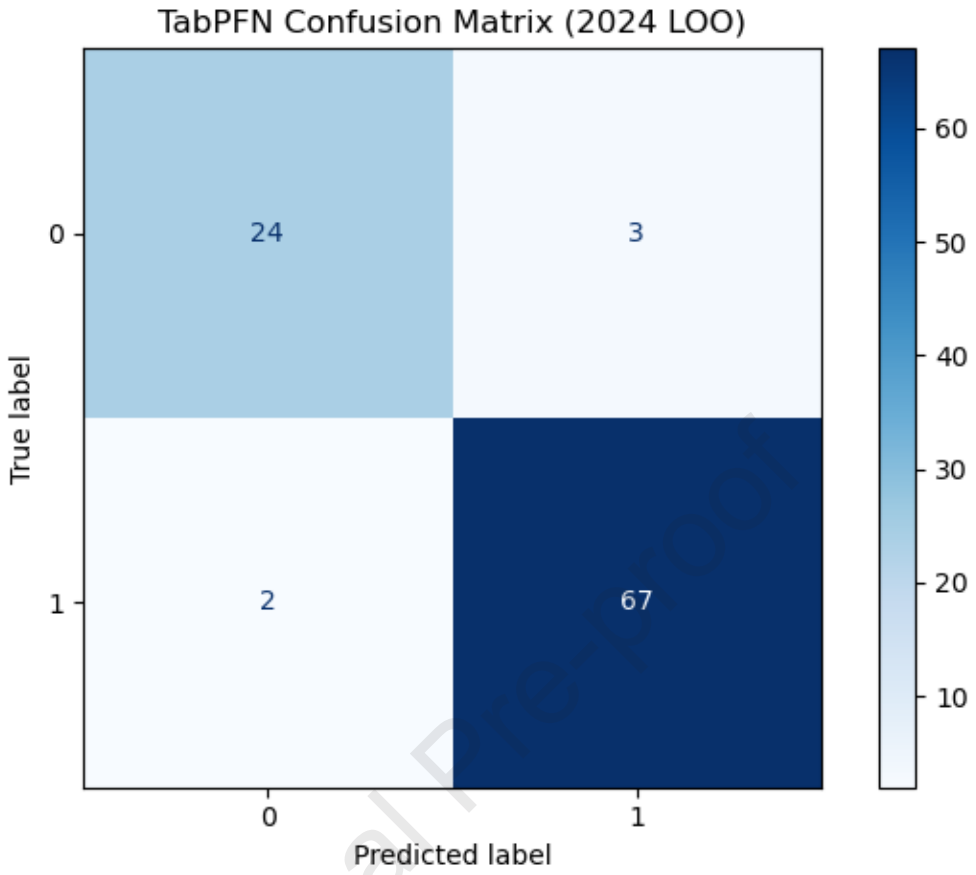


Figure 8. Confusion matrix for the TabPFN classifier on the held-out test set.

Figure 9 presents a SHAP "beeswarm" summary plot, where each point corresponds to one observation's SHAP value for a given feature. The horizontal spread reflects the magnitude of each feature's contribution to the model output, while the color scale (cool to warm) shows the raw value of the feature. Features are ordered top to bottom by their mean absolute SHAP values, indicating their overall importance in the classification task. The top three predictors are the percentage enhancements of STAND 3cm and SEVAN upper scintillators (denoted by STAND3 (%) and SEVAN (%) in Figure 9), followed by the absolute value of the near-surface electric field (denoted in Figure 9 by |NSEF|). Contributions from the cloud base height and the outside temperature are smaller but still non-negligible. Overall, increases in particle detector enhancements (represented by warm-colored points on the right) consistently result in positive SHAP values, indicating that these features tend to push the model's prediction toward the TGE class.

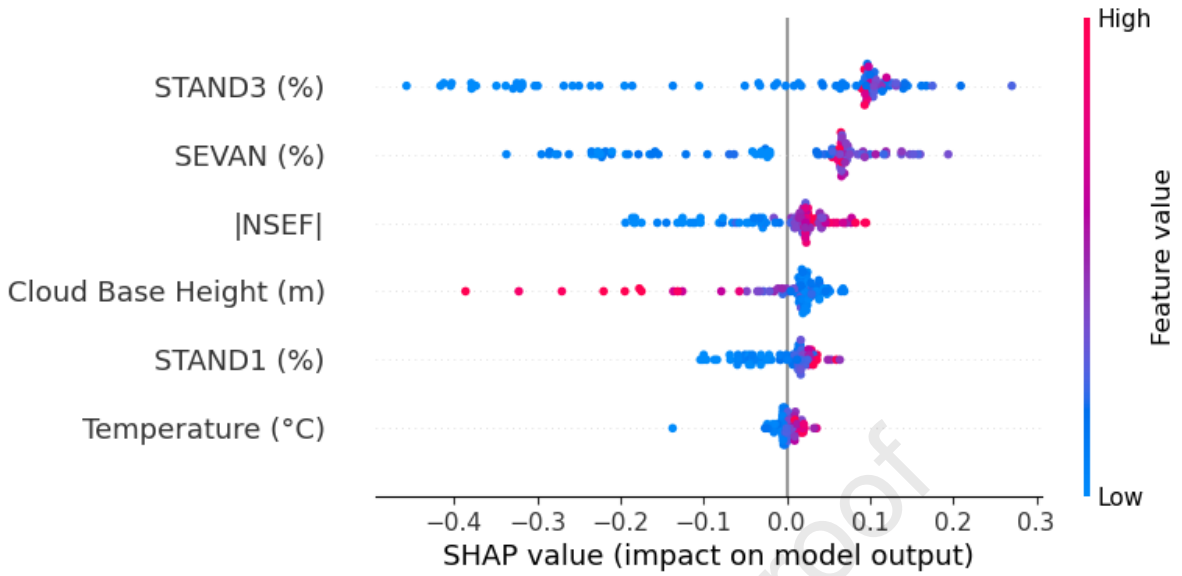


Figure 9. SHAP summary plot showing the mean absolute SHAP value for each predictor variable.

Figure 10(a) overlays a LOWESS smoothed curve on the scatter of SHAP values versus the percentage enhancement of STAND 1cm upper scintillator. Below ≈ 4.6 % the smoothed curve remains negative, indicating that small enhancements reduce the likelihood of a TGE relative to the baseline. At ≈4.6 % the curve crosses zero and rises steeply, showing that further increases in count rates of STAND 1cm upper scintillator add substantial positive weight to the TGE prediction. The steep slope beyond the threshold explains why even modest (> 4.6 %) peaks registered by the STAND 1cm scintillator are strong discriminators.

Figure 10(b) shows the SHAP dependence on the outside temperature, where the LOWESS curve crosses zero near 0.6°C. Temperatures below this value contribute negatively to TGE likelihood. Figure 10(c) presents the SHAP-LOWESS dependence for particle flux percentage enhancement as registered by the upper scintillator of the STAND 3cm detector. The curve crosses zero at approximately 2.5%, showing that relatively modest enhancements in the particle fluxes registered by STAND 3cm detector already contribute positively to TGE classification. Beyond ≈5%, the curve flattens, indicating that further increases yield little additional gain in predictive power.

Figure 10(d) shows the SHAP dependence on Cloud Base Height. The curve crosses zero at approximately 240m altitude, which is consistent with our previous observations (Chilingarian et al., 2020). Figure 10(e) shows that SEVAN percentage enhancements exceeding ≈2% result in positive SHAP contributions. The LOWESS curve increases smoothly, without abrupt changes, indicating that even relatively small enhancements in the SEVAN detector strongly influence classification. This is consistent with SEVAN's sensitivity to RREA-produced gamma rays,

which are effectively detected even at low flux levels. Figure 10(f) shows the SHAP dependence on the Near Surface Electric Field values. The SHAP curve becomes strongly positive when the absolute value of the NSEF exceeds ≈7.4 kV/m, indicating that high field variability is a key predictor of TGEs.

Taken together, these results reveal a two-stage behavior. To assess the significance of these low enhancements, we compared them to the natural background variance. The 95% confidence level (2 σ) for fair-weather fluctuations is approximately 1.3% for SEVAN, 1.4% for STAND 3cm, and 3.6% for STAND 1cm detectors. The SHAP-LOWESS analysis identifies thresholds (≈ 2% for SEVAN, 2.5% for STAND3cm, and ≈ 4.6% for STAND 1cm) that lie consistently above these noise levels.

At low to moderate flux enhancements, the SHAP-LOWESS analysis identifies clear thresholds (≈ 2–5 %) beyond which detector responses begin to contribute positively to TGE classification. Once enhancements grow larger, however, the predictive landscape changes qualitatively: for enhancements exceeding ≈10 % in any of the detectors, the classifier achieves clear distinction, with SHAP values uniformly positive and no observed misclassifications. The large percentage increases therefore correspond to an effectively strong indicator of TGE activity, highlighting that extreme flux surges are unambiguous indicators of thunderstorm-related particle events. These results align very well with the empirical minimum of 5 kV/m used in the manual selection procedure. The data-driven threshold values are summarized in Table 2.

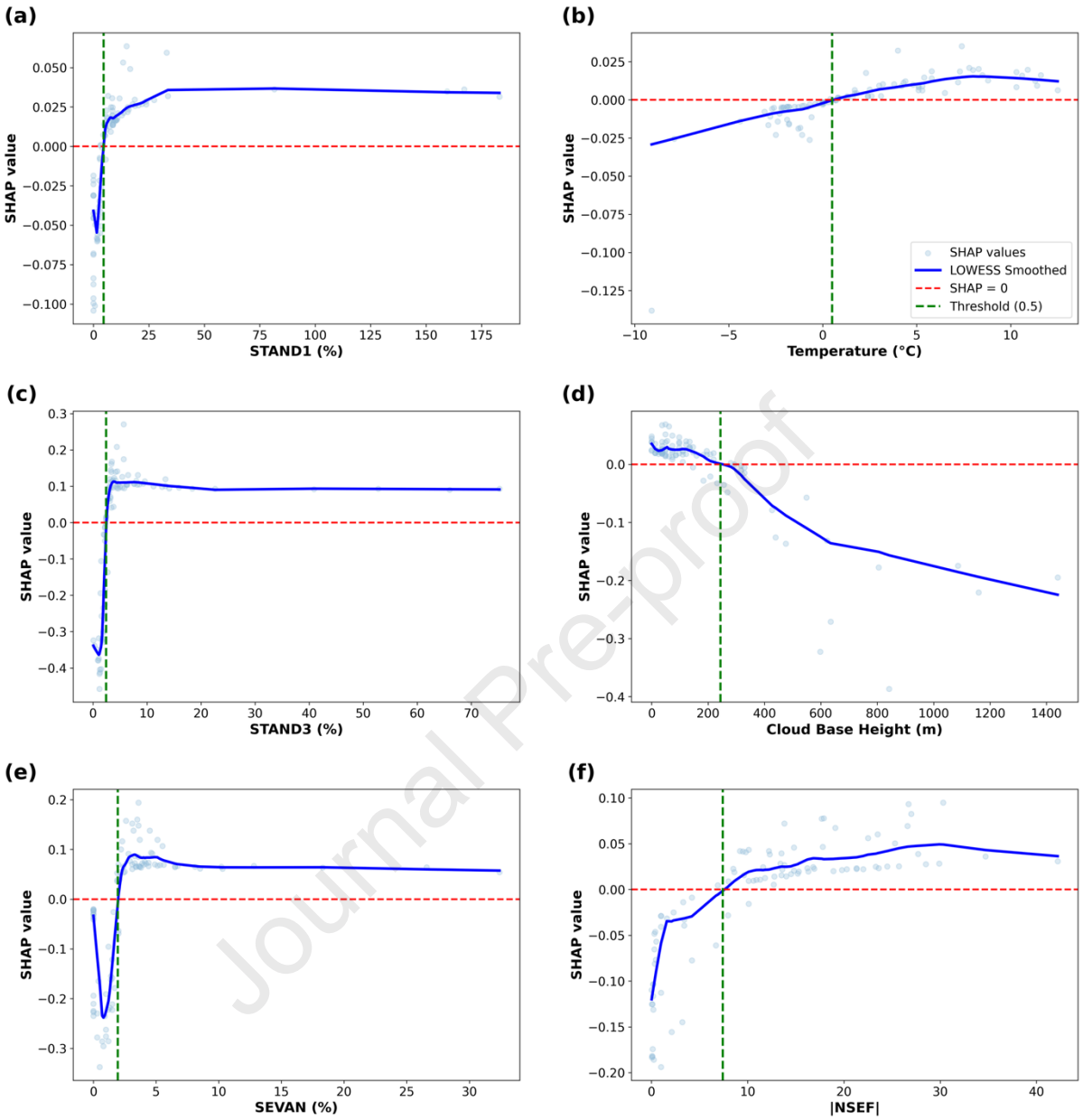


Figure 10. SHAP dependence plots with LOWESS smoothing lines for a) STAND1 (%), b) Temperature (°C), c) STAND3 (%), d) Cloud Base Height (m), e) SEVAN (%), f) Absolute value of the NSEF (kV/m).

Table 2. Summary of approximate threshold values for TGE classification based on TabPFN

Parameter	**Threshold**	**Background Noise (2σ)**
STAND1 (%)	>= 4.6	3.6

STAND3 (%)	>= 2.5	1.4
SEVAN (%)	>= 2	1.3
Temperature (°C)	>= 0.6	-
Cloud Base Height (m)	<= 240	-
|NSEF| (kV/m)	>= 7.4	-

4. Discussion and Conclusions

The SHAP summary (Figure 9) confirms that multi-detector particle enhancement observations, especially the percentage increases recorded by the STAND3 and SEVAN detectors, have the strongest explanatory influence in the TabPFN classifier. Significance-based features (defined as flux enhancements measured in units of standard deviation, σ, relative to the background mean) play a secondary role, indicating that once an enhancement is observed in percentage terms, its statistical significance improves but does not dominate the decision boundary. Environmental variables, such as the near-surface electric field and temperature, further reduce the explanatory gap, showing that even strong particle bursts are unlikely to be classified as TGEs without the proper weather and NSEF conditions. In summer, when high temperatures raise cloud heights, the electric field is also elevated, due to the increased atmospheric depth between the cloud base and ground, particle avalanches are attenuated before reaching the surface, and TGEs are rare. The SHAP-LOWESS dependence plots (Figure 10) measure these relationships and, importantly, reveal data-driven thresholds (summarized in Table 2 as the zero-crossing points) that agree well with the rules previously used in manual TGE selection. Specifically, we identify thresholds of approximately above **4.6%** for **STAND 1cm, 2.5% for STAND 3cm, and 2% for SEVAN.** These results confirming that modest enhancements in three independent channels are sufficient for automatic flagging.

Along with the confusion matrix results (Figure 8), the t-SNE analysis (Figure 7) reveals that TGE events occupy a distinct region of the feature space, separated from non-TGE background events. This observation supports the hypothesis that TGEs form a physically well-defined cluster characterized by simultaneous multi-detector enhancements and electric field excursions. Quantitatively, this separability yields a classification accuracy of 94.8% and a Matthews Correlation Coefficient (MCC) of 0.87. These metrics confirm that the selected features provide a robust basis for automated identification, independent of dataset balance. The two false negatives highlight that only a small set of borderline cases might be missed by automatic detection, where coincidence-channel review or expert judgment remains important.

The consistency between SHAP-derived thresholds and empirical selection criteria provides a quantitative corroboration of fifteen years of empirical practice at ASEC. It should be noted that because the model was trained on expert-labeled data, this alignment confirms that the model has successfully internalized the expert decision boundaries rather than discovering new physics independently. Importantly, the LOWESS-based approach produces continuous response curves rather than hard cut-offs, enabling adjustable sensitivity should operational priorities shift.

Finally, model recall/sensitivity was assessed using the independent dataset of TGEs from 2018 to 2023 (Chilingarian et al., 2022b, 2024b). The 2018–2023 catalog consists exclusively of confirmed TGE events. Therefore, this evaluation serves strictly as a test of the model's sensitivity (recall) to historical events, rather than a full classification metric. The TabPFN classifier, which was trained exclusively on 2024 data, successfully retrieved 199 out of 202 (98.5 % recall accuracy) events, demonstrating high sensitivity to known TGE signatures.

This robustness highlights the model's capacity to transfer across years and feature sets, further supporting its use in operational TGE identification.

This study combines a rigorously curated 2024 TGE dataset with an interpretable TabPFN-SHAP analytical pipeline. Our analysis yielded several key findings. First, the classification accuracy of 94.8% suggests that the selected features (particle flux enhancements and NSEF dynamics) provide strong separability between TGEs and non-TGE background events within the constructed dataset. Second, the data-driven thresholds extracted via SHAP–LOWESS closely match long-standing intuitive criteria used in 2009-2024, quantitatively corroborating the empirical rules. Variable importance analysis confirms that observing multi-detector enhancements, combined with environmental variables, are the main predictors of TGEs. Ultimately, this framework provides a scalable approach to automate screening, with SHAP thresholds serving as initial filters that identify rare and borderline cases for expert review.

Future work will expand this approach to multi-year, multi-site datasets and explore real-time deployment for operational radiation hazard monitoring and alert issuing.

References

- Gurevich, A.V., Milikh, G.M. and Roussel-Dupre, R., 1992. Runaway electron mechanism of air breakdown and preconditioning during a thunderstorm. *Physics Letters A*, *165*(5-6), pp.463-468
- Fishman, G.J., Bhat, P.N., Mallozzi, R., Horack, J.M., Koshut, T., Kouveliotou, C., Pendleton, G.N., Meegan, C.A., Wilson, R.B., Paciesas, W.S. and Goodman, S.J., 1994. Discovery of intense gamma-ray flashes of atmospheric origin. *Science*, *264*(5163), pp.1313-1316.
- Mauda, N., Yair, Y. and Reuveni, Y., 2025. Multi-parametric analysis of thunderstorm ground enhancements (TGE) and associated gamma-ray emissions on Mount Hermon, Israel. *Science of The Total Environment*, *993*, p.179988.
- Leppänen, A.P., Peräjärvi, K., Paatero, J., Joutsenvaara, J., Hannula, A., Hepoaho, A., Holm, P., Ilander, T. and Kärkkäinen, J., 2025. Thunderstorm ground enhancements in Finland: observations using spectroscopic radiation detectors. *Acta Geophysica*, *73*(3), pp.2955-2970.
- Chilingarian, A., Daryan, A., Arakelyan, K., Hovhannisyan, A., Mailyan, B., Melkumyan, L., Hovsepyan, G., Chilingaryan, S., Reymers, A. and Vanyan, L., 2010. Ground-based observations of thunderstorm-correlated fluxes of high-energy electrons, gamma rays, and neutrons. *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, *82*(4), p.043009.
- Chilingarian, A., Hovsepyan, G. and Hovhannisyan, A., 2011. Particle bursts from thunderclouds: Natural particle accelerators above our heads. *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, *83*(6), p.062001.
- Chilingarian, A., Chilingaryan, S., Karapetyan, T., Kozliner, L., Khanikyants, Y., Hovsepyan, G., Pokhsraryan, D. and Soghomonyan, S., 2017. On the initiation of lightning in thunderclouds. *Scientific Reports*, *7*(1), p.1371.

- Chilingarian, A., Hovsepyan, G., Elbekian, A., Karapetyan, T., Kozliner, L., Martoian, H. and Sargsyan, B., 2019. Origin of enhanced gamma radiation in thunderclouds. Physical review research, 1(3), p.033167.
- Chilingarian, A., Hovsepyan, G., Karapetyan, T., Karapetyan, G., Kozliner, L., Mkrtchyan, H., Aslanyan, D. and Sargsyan, B., 2020. Structure of thunderstorm ground enhancements. *Physical Review D*, *101*(12), p.122004.
- Chilingarian, A., Hovsepyan, G., Aslanyan, D., Karapetyan, T., Khanikyanc, Y., Kozliner, L., Pokhsraryan, D., Sargsyan, B., Soghomonyan, S., Chilingaryan, S. and Zazyan, M., 2022a. Thunderstorm ground enhancements: Multivariate analysis of 12 years of observations. *Physical Review D*, *106*(8), p.082004
- Chilingarian, A., Hovsepyan, G., Aslanyan, D., Sargsyan, B. and Karapetyan, T., 2022b. Catalog of Thunderstorm Ground Enhancements (TGEs) observed at Aragats in 2013–2021. *Mendeley Data*, *1*, pp.10-17632.
- Chilingarian, A., Hovsepyan, G., Karapetyan, T., Aslanyan, D., Chilingaryan, S. and Sargsyan, B., 2023. Genesis of thunderstorm ground enhancements. *Physical Review D*, *107*(10), p.102003.
- Chilingarian, A. and Hovsepyan, G., 2023. Proving "new physics" by measuring cosmic ray fluxes. *Astronomy and Computing*, *44*, p.100714.
- Chilingarian, A., Hovsepyan, G., Sargsyan, B., Karapetyan, T., Aslanyan, D. and Kozliner, L., 2024a. Enormous impulsive enhancement of particle fluxes observed on Aragats on May 23, 2023. *Advances in Space Research*, *74*(9), pp.4377-4387.
- Chilingarian, A., Karapetyan, T., Sargsyan, B., Aslanyan, D. and Chilingaryan, S., 2024b. Dataset on extreme thunderstorm ground enhancements registered on Aragats in 2023. *Data in Brief*, *54*, p.110554.
- Chilingarian, A., Sargsyan, B., Karapetyan, T., Aslanyan, D., Chilingaryan, S., Kozliner, L. and Khanikyanc, Y., 2024c. Extreme thunderstorm ground enhancements registered on Aragats in 2023. *Physical Review D*, *110*(6), p.063043.
- Chilingarian, A., Karapetyan, T., Sargsyan, B., Knapp, J., Walter, M. and Rehm, T., 2024d. Energy spectra of the first TGE observed on Zugspitze by the SEVAN light detector compared with the energetic TGE observed on Aragats. *Astroparticle Physics*, *156*, p.102924.
- Chilingarian, A., Williams, E., Hovsepyan, G. and Mkrtchyan, H., 2025. Why schonland failed in his search for runaway electrons from thunderstorms. *Journal of Geophysical Research: Atmospheres*, *130*(10), p.e2024JD042350.
- Kisvárdai, I., Štempel, F., Randuška, L., Mackovjak, Š., Langer, R., Strhárský, I. and Kubančák, J., 2025. Analysis of 42 years of cosmic ray measurements by the neutron monitor at Lomnický štít observatory. *Earth and Space Science*, *12*(1), p.e2024EA003656.
- Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.
- Lundberg, S.M., Erion, G.G. and Lee, S.I., 2018. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.
- Marisaldi, M., Østgaard, N., Mezentsev, A., Lang, T., Grove, J.E., Shy, D., Heymsfield, G.M., Krehbiel, P., Thomas, R.J., Stanley, M. and Sarria, D., 2024. Highly dynamic gamma-ray emissions are common in tropical thunderclouds. *Nature*, *634*(8032), pp.57-60.
- Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, *74*(368), pp.829-836.
- Cleveland, W.S. and Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, *83*(403), pp.596-610.
- Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S.B., Schirrmeister, R.T. and Hutter, F., 2025. Accurate predictions on small data with a tabular foundation model. *Nature*, *637*(8045), pp.319-326.

- Dwyer, J.R., Smith, D.M. and Cummer, S.A., 2012. High-energy atmospheric physics: Terrestrial gamma-ray flashes and related phenomena. Space Science Reviews, 173(1), pp.133-196.
- Aglietta, M., Badino, G., Bergamasco, L., Castagnoli, C., Castellina, A., Cini, G., Dardo, M., D'Ettorre-Piazzoli, B., Fulgione, W., Galeotti, P. and Ghia, P., 1989. The EAS-TOP array at E0= 1014− 10 16eV: Stability and resolutions. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 277(1), pp.23-28.
- Tsuchiya, H., Enoto, T., Yamada, S., Yuasa, T., Kawaharada, M., Kitaguchi, T., Kokubun, M., Kato, H., Okano, M., Nakamura, S. and Makishima, K., 2007. Detection of high-energy gamma rays from winter thunderclouds. Physical review letters, 99(16), p.165002.
- Kuriyama, E., Masubuchi, M., Koshikawa, N., Iwashita, R., Omata, A., Kanda, T., Kataoka, J., Tsurumi, M., Diniz, G., Enoto, T. and Wada, Y., 2022. Compton Camera Imaging of a Gamma-Ray Glow From a Thunderstorm. Geophysical Research Letters, 49(19), p.e2022GL100139.
- Alexeenko, V.V., Khaerdinov, N.S., Lidvansky, A.A. and Petkov, V.B., 2002. Transient variations of secondary cosmic rays due to atmospheric electric field and evidence for pre-lightning particle acceleration. Physics Letters A, 301(3-4), pp.299-306.
- Kudela, K., Chum, J., Kollárik, M., Langer, R., Strhárský, I. and Baše, J., 2017. Correlations between secondary cosmic ray rates and strong electric fields at Lomnický štít. Journal of Geophysical Research: Atmospheres, 122(20), pp.10-700.
- Shepetov, A., Antonova, V., Kalikulov, O., Kryakunova, O., Karashtin, A., Lutsenko, V., Mamina, S., Mukashev, K., Piscal, V., Ptitsyn, M. and Ryabov, V., 2021. The prolonged gamma ray enhancement and the short radiation burst events observed in thunderstorms at Tien Shan. Atmospheric Research, 248, p.105266.
- Eack, K.B., Suszcynsky, D.M., Beasley, W.H., Roussel-Dupre, R. and Symbalisty, E., 2000. Gamma-ray emissions observed in a thunderstorm anvil. Geophysical Research Letters, 27(2), pp.185-188
- Karapetyan, T. and Hovsepyan, G., 2024. SEVAN particle detector response function, purity, efficiency, and response to charged and neutral CR species.
- Chilingarian, A., Hovsepyan, G., Arakelyan, K., Chilingaryan, S., Danielyan, V., Avakyan, K., Yeghikyan, A., Reymers, A. and Tserunyan, S., 2009. Space environmental viewing and analysis network (SEVAN). Earth, Moon, and Planets, 104(1), pp.195-210.
- Chilingarian, A., Babayan, V., Karapetyan, T., Mailyan, B., Sargsyan, B. and Zazyan, M., 2018. The SEVAN Worldwide network of particle detectors: 10 years of operation. Advances in Space Research, 61(10), pp.2680-2696.
- Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure, 405(2), pp.442-451.
- Goutte, C. and Gaussier, E., 2005, March. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In European conference on information retrieval (pp. 345-359). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Maaten, L.V.D. and Hinton, G., 2008. Visualizing data using t-SNE. Journal of machine learning research, 9(Nov), pp.2579-2605.
- Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V. and Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Advances in neural information processing systems, 31.
- LaValley, M.P., 2008. Logistic regression. Circulation, 117(18), pp.2395-2399
- Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

- AI model is used to identify Thunderstorm Ground Enhancements (TGEs)
- AI model identifies TGEs with 94.79% accuracy on 2024 events registered at Aragats
- Data-driven thresholds validate the empirical TGE criteria derived over 15 years
- Multi-detector flux increases and strong electric fields are key TGE predictors
- Model trained on 2024 data achieves 98.51% accuracy on 2018-2023 dataset

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: