safety

MDPI

Article

Acoustic Fingerprint Identification of Lithium-Ion Battery Thermal Runaway and Application in Fire Investigation

Chaojie Yang [1], Jiawei Tang [2], Wenlong Chen [2] and Shengli Kong [2,*]

[1] Shanghai Fire Research Institute of the Ministry of Emergency Management, The Seventh Research Division, Shanghai 200032, China
[2] Faculty of Urban Construction and Ecological Technology, Shanghai Institute of Technology, Shanghai 201418, China
* Correspondence: kslcone@126.com

Abstract

Identifying the specific battery type that caused a fire is a critical part of forensic investigation, yet it becomes difficult when severe heat destroys physical evidence. To provide a non-contact forensic solution, this study proposes an interpretable identification framework based on eXtreme Gradient Boosting (XGBoost) and SHapley Additive exPlanations (SHAP). Thermal runaway signals from prismatic, pouch, and cylindrical cells were recorded using various devices under realistic conditions, including environmental noise and wall obstructions. By integrating time-domain statistics, frequency-domain metrics, and Mel-frequency Cepstral Coefficients (MFCCs), we characterized the unique acoustic signatures of each battery format. The results show that battery structure determines the sound. Cylindrical cells produce high-frequency impulsive sounds due to valve jetting, while pouch cells produce low-frequency tearing sounds. Using these physical features, the optimized XGBoost classifier achieved an overall accuracy of 96.39% and an F1-score of 96.21% on a mixed dataset. This model outperformed Support Vector Machine and Random Forest classifiers. It proved robust even when processing low-quality audio from surveillance equipment. Additionally, SHAP analysis interprets the decision logic, identifying Peak Frequency and specific cepstral textures as the most important features. In conclusion, acoustic fingerprinting serves as independent, quantifiable evidence, enabling accurate ignition source identification even when traditional physical analysis is impossible.

Keywords: lithium-ion battery fire; fire investigation; acoustic fingerprint; XGBoost; SHAP; machine learning

Academic Editor: Raphael Grzebieta

Received: 10 July 2026
Revised: 24 August 2026
Accepted: 15 September 2026
Published: 22 September 2026

1. Introduction

Lithium-ion batteries (LIBs) are widely used in portable electronics, electric vehicles (EVs), and grid-scale energy storage systems because of their high energy density and long cycle life [1]. However, the expanding application of LIBs introduces safety risks. Statistical data indicates a rising incidence of fire accidents associated with battery failures [2]. Under abuse or internal faults, LIBs may enter thermal runaway (TR), a process that rapidly releases combustible gases and jet flames, often leading to explosions [3,4]. After such incidents, accurate accident reconstruction is essential to determine liability and improve safety designs. However, fire investigations face a fundamental limitation: the severe thermal damage from TR often destroys the morphological evidence needed for traditional physical analysis [5]. Consequently, identifying the ignition source from carbonized debris is difficult, creating a need for non-contact forensic methods.

 https://doi.org/10.3390/safety12050121

To address these limitations, acoustic signals have emerged as a viable source of digital evidence for accident reconstruction [6–8]. The TR process generates distinct sounds resulting from safety valve opening, gas venting, and mechanical rupture [9,10]. The existing literature often uses signal processing to extract quantitative indicators from these signals [11]. Standard approaches calculate time-domain features, such as Root Mean Square (RMS) and Kurtosis, to characterize transient shocks [12]. They also use frequency-domain metrics like Spectral Centroid to analyze spectral shifts caused by gas flow. Additionally, MFCCs are increasingly used to represent audio texture for source identification [13]. Unlike Battery Management System (BMS) data, which fire can destroy, environmental audio is often captured by nearby devices and stored remotely [12]. Thus, these recordings serve as a potential data source for forensic analysis when physical sites are destroyed.

Currently, combining acoustic analysis with machine learning (ML) is standard practice in battery safety research [14]. Data-driven algorithms are effective in extracting patterns from acoustic features for anomaly detection. For instance, Zhang et al. used ultrasonic signals and Support Vector Machines (SVM) to monitor battery state of health [15]. However, distinct from active monitoring for early warning, post-accident fire investigation presents specific challenges regarding generalization and evidence transparency. First, most existing models are trained on high-fidelity data from laboratory-grade microphones. In forensic scenarios, audio evidence comes from diverse devices with varying frequency responses, often affected by noise and distortion [16,17]. Second, the straightforward stacking of high-dimensional acoustic features may introduce substantial feature redundancy, which not only increases computational burden but also hampers the physical interpretability of the resulting models—an issue that is particularly critical when linking acoustic patterns to underlying thermal runaway mechanisms [18]. Third, for formal fire investigation reports and judicial applications, classification outcomes must be accompanied by explicit and quantitative feature attribution, as black-box predictions without explainability are insufficient to support reliable accident reconstruction and evidential reasoning [19–21].

To address these forensic requirements, this paper proposes a robust and interpretable acoustic identification framework based on the XGBoost algorithm and SHAP [22,23]. We established a multi-source recording platform to capture thermal runaway sounds from three typical lithium-ion batteries. This setup integrates audio from diverse devices subject to fan noise and obstacle interference to simulate realistic conditions [24]. We extracted a feature set comprising time-domain statistics, frequency-domain characteristics, and MFCCs to characterize physical acoustic properties [25]. To ensure efficiency, we implemented a correlation-based feature optimization strategy to create a compact input vector for the classifier [26]. Furthermore, the XGBoost model is employed due to its superior performance on structured tabular data compared to traditional algorithms. Finally, we integrated the SHAP framework to interpret the black-box nature of the model, transforming probabilistic predictions into scientifically grounded evidence [21]. The main contributions of this study are as follows:

1. We established a multi-source recording platform to capture thermal runaway sounds from three typical lithium-ion batteries under realistic conditions. By integrating audio from diverse devices subject to fan noise and obstacle interference, this study validates environmental sound as a reliable physical basis for non-contact ignition source identification.
2. We extracted a comprehensive feature set comprising time-domain statistics, frequency-domain characteristics, and MFCCs to characterize physical acoustic properties. To ensure efficiency, we implemented a correlation-based optimization strategy to elimi-

nate redundancy. This constructed a compact input vector for the XGBoost classifier, enabling accurate battery-type classification.

3. By integrating the SHAP framework, this work establishes a transparent decision-making mechanism that interprets the black-box nature of the machine learning model. This approach transforms probabilistic predictions into scientifically grounded evidence. It provides a valuable non-contact tool for fire investigation, enabling the tracing of ignition sources even when physical evidence is severely compromised.

2. Methodology

To reliably identify ignition sources in complex fire environments, this study proposes a systematic acoustic fingerprint analysis framework. As shown in Figure 1, the framework consists of four stages: (1) Data Acquisition, capturing signals under interference; (2) Signal Preprocessing, using augmentation to construct the dataset; (3) Feature Extraction, covering time, frequency, and cepstral domains; and (4) Classification and Interpretation, using XGBoost and SHAP.

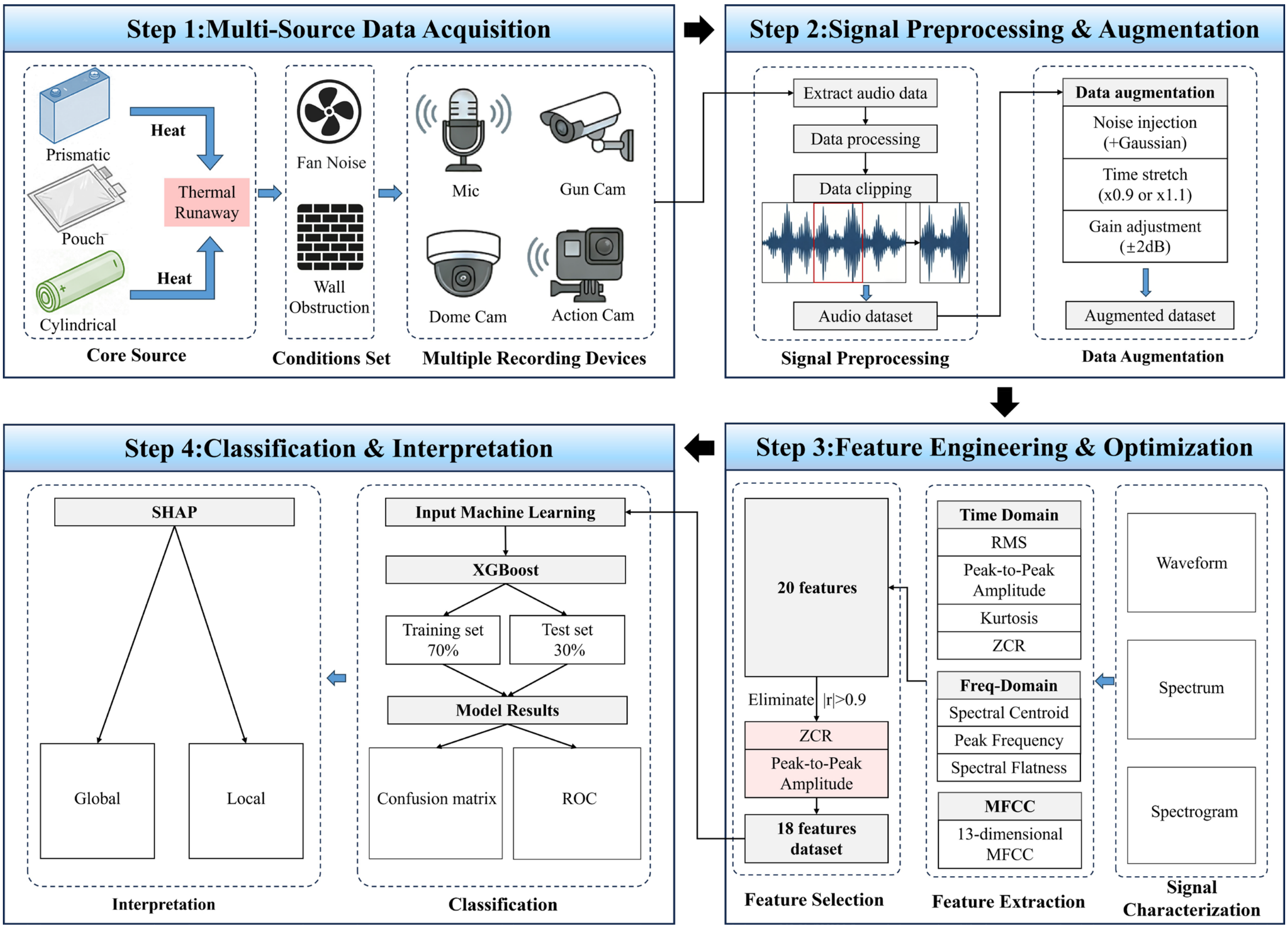

Figure 1. The overall framework of the proposed interpretable acoustic identification methodology for lithium-ion battery thermal runaway.

2.1. Experimental Setting

A multi-source recording platform was established to simulate real-world accident scenarios and construct an acoustic database for fire investigation. Figure 2 illustrates the schematic of the experimental setup.

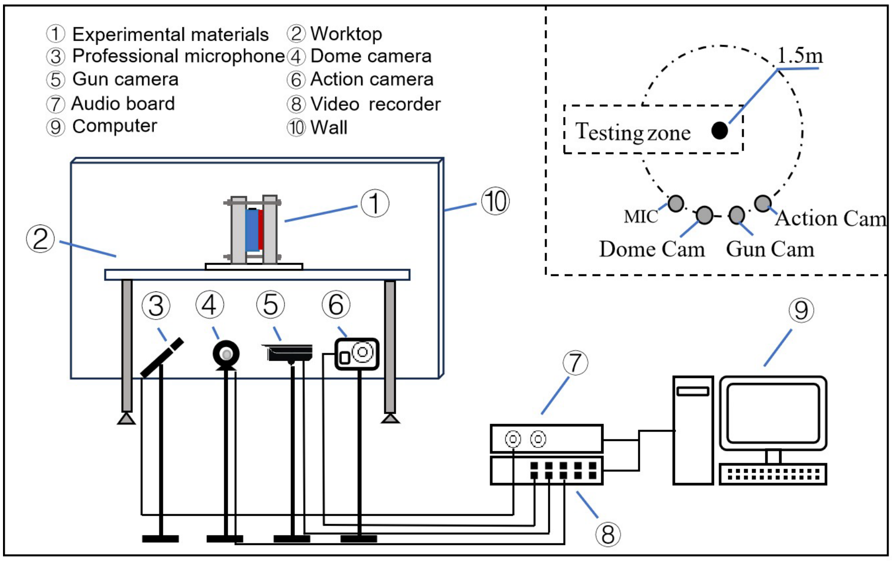

Figure 2. Schematic diagram of the multi-source acoustic signal acquisition experimental setup for lithium-ion battery thermal runaway.

2.1.1. Research Objects

Three commercial lithium-ion batteries with NCM cathode chemistry were selected as research objects: prismatic, pouch, and cylindrical cells. These batteries represent the mainstream chemical systems and structural designs currently in use while sharing the same chemical system, ensuring that acoustic differences arise primarily from structural rather than material variations. Table 1 details the battery parameters, including dimensions and casing materials, which significantly affect acoustic resonance. In all experiments, thermal runaway was triggered by an 800 W heating plate placed on the side of the cell for continuous heating. Temperature was recorded using a four-channel K-type thermocouple attached to different positions on the cell surface, with a measurement range of –200 °C to 1372 °C, capturing the full thermal runaway process in real time.

Table 1. Specifications of the lithium-ion batteries used in the experiments.

Battery Format	Cathode Material	Nominal Capacity (Ah)	Nominal Voltage (V)	Dimensions (mm) (L × W × H/D)	Casing Material
Prismatic	NCM	20	3.7	150 × 100 × 27	Aluminum alloy
Pouch		10	3.7	100.5 × 60 × 12	Aluminum plastic film
Cylindrical		0.2	3.7	Φ18 × 65	Nickel-plated steel

All NCM batteries were charged to the same state prior to testing, with open-circuit voltage maintained above 4.0 V. All thermal runaway tests were conducted under the same voltage baseline to ensure consistency of initial conditions.

To acquire thermal runaway acoustic signals of lithium batteries, experiments were conducted under three operating conditions, namely baseline, fan noise, and wall occlusion. Under each condition, three parallel experiments were carried out for prismatic, pouch, and cylindrical cells, yielding a total of 27 independent experiments. Four types of recording devices, including a professional microphone (BSWA Technology Co., Ltd., Beijing, China), a gun-type camera (Zhejiang Dahua Technology Co., Ltd., Hangzhou, China), a dome-type camera (Xiaomi Corporation, Beijing, China), and an action camera (Dongguan Shangu Intelligent Technology Co., Ltd., Dongguan, China), were used for synchronous audio acquisition, producing 108 raw audio samples. To expand the dataset and simulate signal variations in real-world scenarios, data augmentation was applied through time stretching ranging from 0.9 to 1.1 times the original duration and additive Gaussian white noise with an amplitude range of 0.001 to 0.01. A final acoustic signal sample database consisting of 540 samples was constructed and divided into a training set comprising 70% of the samples and a test set comprising the remaining 30%. The detailed distribution is presented in Table 2.

Table 2. Distribution of the acoustic signal sample database.

Battery Packaging Format	Acquisition Device Type	Raw Samples	Augmented Samples	Total Samples
Pouch cell	Professional microphone	9	45	180
	Gun-type camera	9	45	
	Dome-type camera	9	45	
	Action camera	9	45	
Prismatic cell	Professional microphone	9	45	180
	Gun-type camera	9	45	
	Dome-type camera	9	45	180
	Action camera	9	45	
Cylindrical cell	Professional microphone	9	45	
	Gun-type camera	9	45	
	Dome-type camera	9	45	
	Action camera	9	45	

2.1.2. Heterogeneous Recording Terminals

To address device heterogeneity in fire investigations, four distinct recording terminals were used for synchronous audio acquisition. Detailed technical specifications are listed in Table 3. A professional acoustic microphone captured baseline audio, serving as a standard reference for spectral analysis. A MicW M215 omnidirectional microphone (BSWA Technology Co., Ltd., Beijing, China) with a frequency response of 20 Hz to 31.5 kHz and a sampling rate of 48 kHz was used as the reference recording device, along with a Dahua gun-type camera (Zhejiang Dahua Technology Co., Ltd., Hangzhou, China), a Xiaomi dome-type camera (Xiaomi Corporation, Beijing, China), and a SARGO action camera (Dongguan Shangu Intelligent Technology Co., Ltd., Dongguan, China) for synchronous multi-source audio-video acquisition. A gun-type camera and a dome-type camera were selected to simulate outdoor and indoor surveillance environments, respectively. Additionally, an action camera was included to represent dynamic setups.

Table 3. Technical specifications of the heterogeneous recording terminals used in the experiment.

Device Type	Model (Manufacturer)	Sampling Rate (kHz)
Professional Microphone	MicW (M215)	48 kHz
Gun-type Camera	Dahua (DHP40A1AIL2)	8 kHz
Dome-type Camera	Xiaomi (MJSXJ03HL)	8 kHz
Action Camera	SARGO (A9)	48 kHz

2.1.3. Operating Condition Settings

Three operating conditions were established to evaluate the method's adaptability to complex fire environments. First, a baseline test was conducted in an ideal environment to construct a feature benchmark. Second, fan noise was introduced to simulate low-frequency background noise from ventilation systems. Finally, wall occlusion interference was implemented by positioning a solid wall structure between the sound source and the recording terminals. This scenario simulates obstructed surveillance devices to investigate acoustic transmission loss and high-frequency attenuation caused by barriers.

2.2. Signal Preprocessing

The raw audio streams captured during experiments contain long periods of silence. To isolate thermal runaway events, a segmentation strategy based on joint auditory–visual inspection was employed. Specifically, by analyzing waveforms and spectrograms, the onset and offset of each event were precisely localized. A 0.5-s buffer was retained before and after the signal to preserve transient onset and decay characteristics.

Since fire experiments are destructive, the number of original acoustic samples is limited. To prevent overfitting and enhance generalization, a data augmentation strategy was implemented while preserving original feature characteristics. The dataset was expanded using three techniques: (1) Gaussian noise injection with a variance factor of 0.005 to simulate sensor thermal noise; (2) time stretching at rates of 0.8× and 1.2× to account for slight variations in reaction speed; and (3) volume gain adjustment within a range of ±2 dB to mimic signal attenuation differences. This process expanded the dataset size while maintaining the core physical properties of the acoustic signatures [13].

Following augmentation, this study adopts a minimal intervention preprocessing paradigm. Distinct from conventional acoustic diagnosis approaches that utilize aggressive denoising filters to enhance signal clarity, this study retains the raw noise characteristics. In real investigations, audio evidence is often compromised by environmental noise and distortion. Excessive denoising risks attenuating subtle high-frequency features critical for identification. Therefore, using minimally processed signals ensures the model adapts to environmental interference, enhancing its practical applicability [27].

2.3. Feature Extraction

A discriminative feature set was constructed to characterize the acoustic signatures of different battery failure modes. This study extracts 20 quantitative indicators derived from the time, frequency, and cepstral domains.

2.3.1. Time and Frequency Statistical Features

Key statistical indicators were calculated from time and frequency domains to quantify signal energy evolution and spectral distribution. Time-domain metrics describe transient explosion dynamics, while frequency-domain metrics characterize the spectral structure determined by battery geometry. Table 4 lists the selected features and their physical interpretations.

Table 4. List of extracted statistical features and their physical implications [28–32].

Domain	Feature Name	Physical Interpretation
Time domain	Root Mean Square (RMS)	Represents the effective sound pressure level, quantifying the overall intensity of the explosion.
	Peak-to-Peak Amplitude	Captures the maximum dynamic range of the shock wave generated by casing rupture.
	Kurtosis	Measures signal impulsiveness; highly sensitive to transient shocks caused by sudden safety valve opening.
	Zero-Crossing Rate (ZCR)	Correlates with the high-frequency turbulence noise produced by high-velocity gas venting.

https://doi.org/10.3390/safety12050121

Table 4. *Cont.*

Domain	Feature Name	Physical Interpretation
Frequency domain	Spectral Centroid	Describes the brightness of the timbre; differentiates high-pitched whistling from dull rupture sounds.
	Peak Frequency	Identifies the dominant frequency component, corresponding to the fundamental resonant frequency of the battery structure.
	Spectral Flatness	Quantifies tonality; distinguishes between broadband noise and tonal components.

2.3.2. Mel Frequency Cepstral Coefficients

Additionally, Mel-frequency Cepstral Coefficients (MFCCs) were extracted to capture the fine-grained acoustic texture of the explosion. MFCCs emulate the non-linear frequency perception of the human auditory system and serve as a robust digital fingerprint for source identification. The extraction process involves pre-emphasis, framing, windowing, Fast Fourier Transform (FFT), Mel-scale filtering, logarithmic compression, and Discrete Cosine Transform (DCT) [33]. We retained the first 13 static coefficients as the cepstral feature vector. The MFCC extraction parameters were as follows: pre-emphasis coefficient 0.97; framing with 25 ms window length and 10 ms frame shift, followed by Hanning windowing; FFT to the frequency domain; filtering through 128 Mel-scale triangular filter banks with logarithmic compression; DCT to obtain the first 13 coefficients. All features were normalized using Z-score standardization. Feature selection was performed using the Pearson correlation coefficient with a threshold of $|r| \geq 0.8$ to remove highly correlated features, retaining 16 effective features.

2.4. Classification and Interpretability Framework

After feature extraction, the raw acoustic signals are transformed into a structured tabular dataset comprising 20-dimensional feature vectors. To achieve precise identification and ensure decision transparency for forensics, we constructed a framework combining XGBoost and SHAP.

2.4.1. Battery-Type Classification Based on XGBoost

Battery-type identification is modeled as a multi-class supervised learning task. XGBoost was selected as the core classifier. As a scalable Gradient Boosting implementation, XGBoost is known for its efficiency on structured data, outperforming other deep learning models in this context.

We chose XGBoost for two reasons. First, its L1 and L2 regularization controls complexity, preventing overfitting on noisy forensic data [34]. Second, its tree-based structure offers better interpretability than black-box neural networks [35]. The dataset was divided into training and test sets at a 7:3 ratio using stratified sampling to ensure consistent class distribution. Hyperparameter optimization was conducted using the Optuna framework with 5-fold cross-validation, using mean classification accuracy as the optimization objective. The training set is partitioned into five nonoverlapping subsets. One subset is sequentially selected as the validation set, while the remaining four subsets are utilized as the training set to complete model training and evaluation. The search ranges for XGBoost were: n_estimators [100, 1000], learning_rate [0.01, 0.3], max_depth [3, 10], subsample [0.6, 1.0], colsample_bytree [0.6, 1.0], and gamma [0, 5].

2.4.2. Model Interpretation via SHAP

In fire investigations, the opacity of black-box models limits their credibility as evidence [36]. To address this issue, we integrated the SHAP framework to interpret XGBoost outputs. SHAP assigns an importance value to each feature, representing its contribution to the classification. Using SHAP values, this study achieves interpretability at two levels. Globally, the model identifies dominant acoustic indicators for distinguishing battery types. Locally, it explains individual cases by attributing decisions to specific feature values [20]. Ultimately, this mechanism transforms probability scores into scientifically grounded investigation evidence.

3. Results and Analysis

3.1. Acoustic Characterization of Thermal Runaway Signals

To understand the physical differences in acoustic emissions, we analyzed signals captured by the professional microphone under baseline conditions. Figure 3 compares the time-domain waveforms, Power Spectral Densities (PSD), and spectrograms for Pouch (Figure 3a–c), Cylindrical (Figure 3d–f), and Prismatic (Figure 3g–i) cells.

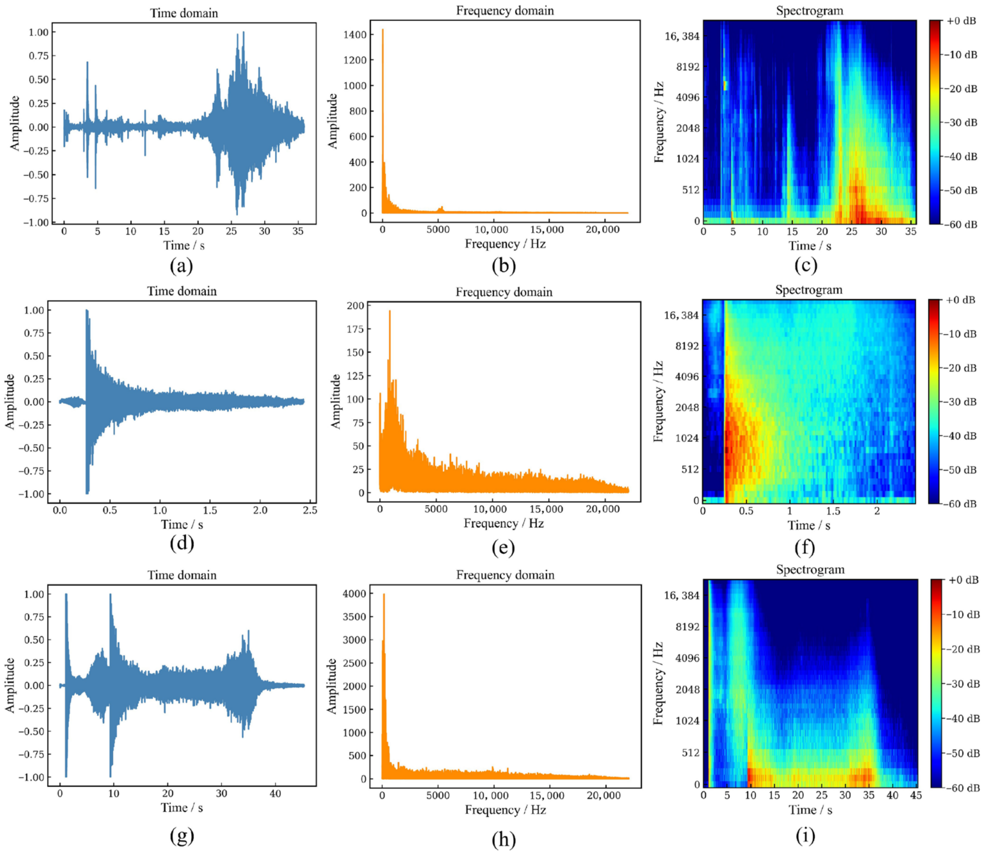

Figure 3. Multi-domain visualization of thermal runaway acoustic signals. (**a**–**c**) Pouch cell; (**d**–**f**) cylindrical cell; (**g**–**i**) prismatic cell. The columns correspond to time-domain waveforms (**left**), frequency spectra (**middle**), and spectrograms (**right**), respectively.

As the waveforms show (Figure 3a,d,g), the signal's temporal evolution depends heavily on casing geometry and pressure relief mechanisms. The cylindrical cell (Figure 3d) exhibits a classic impulsive characteristic. The event is extremely short, lasting less than 2.5 s, with a near-instantaneous rise time. This signature resembles a gunshot, caused by the sudden opening of the safety valve acting as a point source. In contrast, the pouch cell (Figure 3a) shows a chaotic, multi-burst pattern lasting approximately 35 s. Due to the lack of a rigid valve, its waveform reflects the progressive tearing of the aluminum-plastic film along sealing edges rather than a clean explosion. The prismatic cell (Figure 3g) represents a complex intermediate state with multi-stage evolution. An initial impulse from valve opening occurs at 2 s, followed by eruptions at 10 and 30 s, indicating a stepwise release of internal energy.

The spectral analysis further elucidates these structural differences. The cylindrical cell (Figure 3e,f) exhibits the broadest frequency response with a spectral tail extending significantly into the high-frequency region up to 20 kHz. This brightness is caused by the nozzle effect, where high-velocity gas ejecting through a narrow vent creates intense high-frequency turbulence and metallic resonance. Conversely, the pouch cell (Figure 3b,c) is dominated by low-frequency components below 2 kHz. The spectrogram shows vertical broadband bursts without harmonic structures, consistent with the dull sound of soft material rupture. The prismatic cell (Figure 3h,i) combines these features. It shows dominant low-frequency energy due to large gas volume, but retains broadband content up to 11 kHz during violent combustion.

In summary, thermal runaway acoustic fingerprints are determined by battery structure. Cylindrical cells produce high-frequency, impulsive jetting noise due to the valve mechanism. Pouch cells generate prolonged, low-frequency sounds caused by seam rupture and material tearing. Prismatic cells exhibit high-energy, multi-stage explosions. These distinct physical differences provide a robust basis for feature extraction and classification.

3.2. Statistical Evaluation of Feature Separability and Correlation

3.2.1. Comparative Analysis of Time-Domain and Frequency-Domain Features

Figure 4 presents a radar chart to compare time-domain and frequency-domain features based on the representative instances from Section 3.1. The feature values were computed using a two-step aggregation process: first, the mean was calculated for every group of 5 consecutive frames, and then the average of these group means was taken as the final representative value.

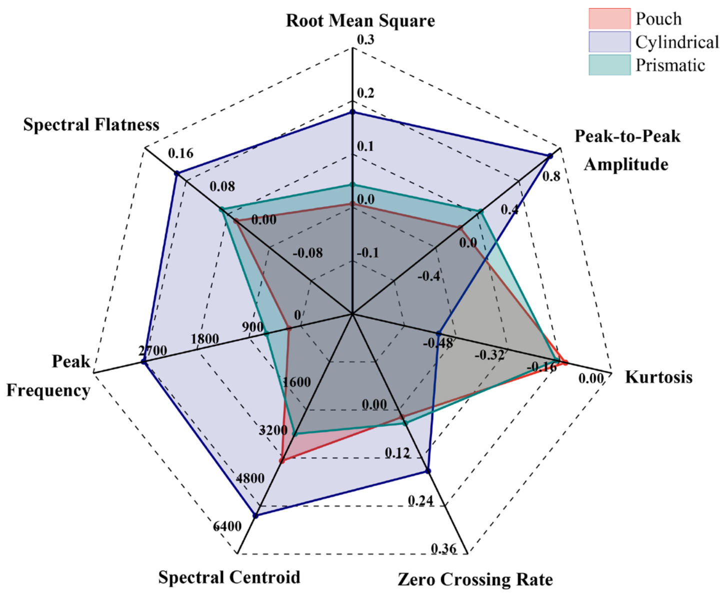

Figure 4. Radar chart comparison of acoustic statistical features for different battery types.

As shown, the three battery types exhibit a clear hierarchical distribution. The cylindrical cell dominates the feature space, showing the highest values in nearly all dimensions, particularly Spectral Centroid and RMS. This corresponds to the intense, high-frequency noise from safety valve jetting. In contrast, the pouch cell is confined to the center, exhibiting the lowest Peak-to-Peak Amplitude and Peak Frequency. This reflects the muted and low-frequency nature of soft material rupture. The prismatic cell occupies an intermediate position, bridging the gap between the rigid cylindrical and soft pouch formats. These distinct, non-overlapping boundaries confirm that the selected features effectively distinguish the battery types.

3.2.2. Distribution Characteristics of MFCC Features

We analyzed the distributions of the 13-dimensional MFCCs using boxplots (Figure 5) to investigate acoustic timbre. This visualization allows for a direct comparison of the numerical range and consistency of the acoustic features across different battery types.

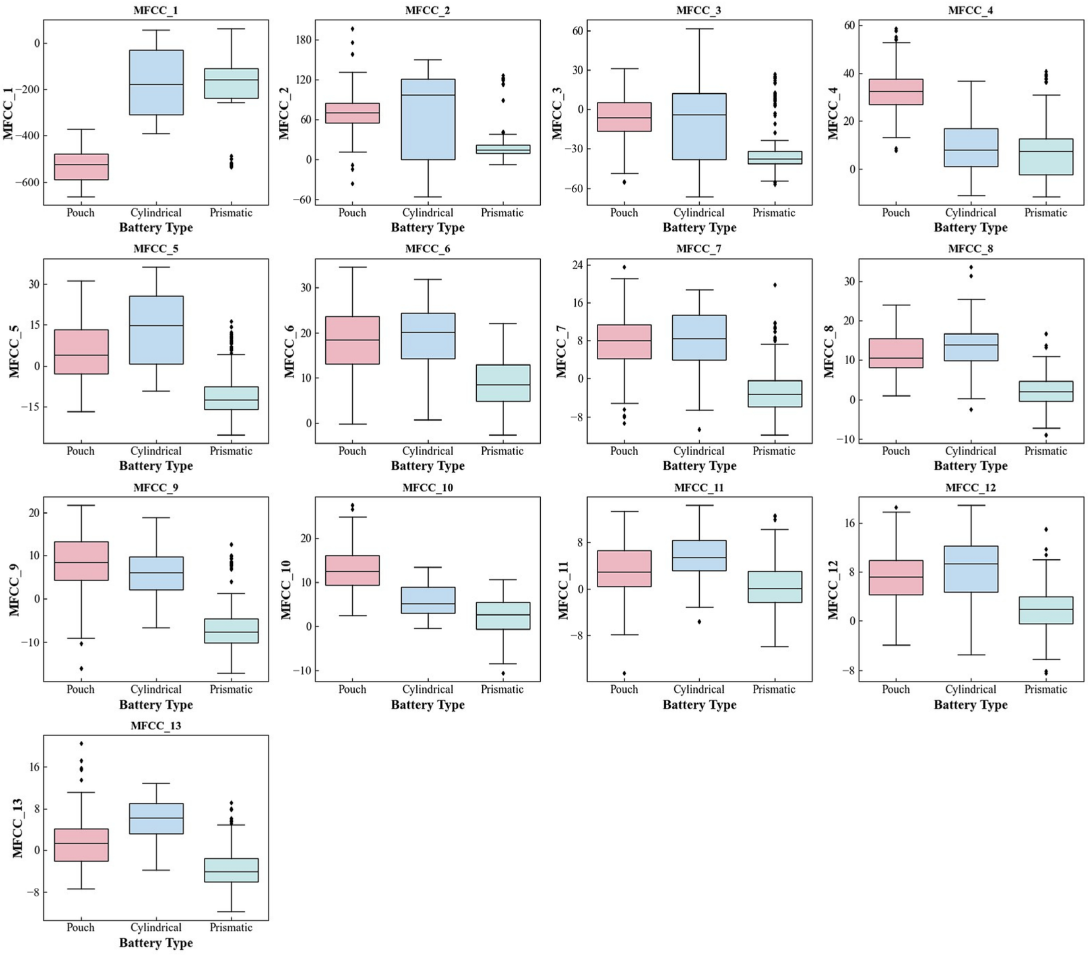

Figure 5. Boxplot distributions of MFCC features for different battery types.

Low-order coefficients show clear separation based on casing type. As shown in MFCC_1, the pouch cell is concentrated in a significantly lower value range with a median below −500, showing no overlap with the other two types. In contrast, the rigid cylindrical and prismatic cells exhibit much higher values, mostly above −300. This indicates that the spectral envelope of soft aluminum-plastic packaging differs from that of rigid metal casings. Beyond material properties, specific coefficients differentiate individual battery formats. Pouch cells show high values in MFCC_4 and MFCC_9. Prismatic cells show a unique pattern in MFCC_5, shifting largely into the negative range. This suggests that these specific MFCC dimensions capture the unique acoustic signatures associated with different battery geometries. Furthermore, the box height reflects signal consistency. By comparing the interquartile ranges, it is observed that the cylindrical cell generally has a taller box height, particularly in MFCC_1 and MFCC_2, indicating a wider fluctuation in feature values across samples. Conversely, the prismatic cell often shows a more compact distribution, suggesting that its acoustic signals are relatively more stable and consistent. However, notable outliers represented by black dots are present in features like MFCC_2, indicating occasional signal deviations. Finally, in certain dimensions such as MFCC_7 and MFCC_11, the value ranges of the three battery types overlap significantly. This overlap confirms that single features are insufficient for identification, validating the need for the multi-feature XGBoost model.

3.2.3. Feature Correlation Analysis and Optimization

To systematically evaluate the redundancy within the constructed feature space, a Pearson correlation analysis was conducted on the 20-dimensional feature set, as illustrated in Figure 6. The heatmap reveals clusters of high collinearity among similar indicators. Notably, the correlation between RMS and Peak-to-Peak Amplitude reaches 0.97, indicating they convey nearly identical information about explosion intensity. Similarly, the Zero-Crossing Rate exhibits a strong correlation of 0.93 with the Spectral Centroid, suggesting that the time-domain oscillation rate acts as a direct proxy for the spectral center of mass in this dataset. Meanwhile, Kurtosis and most MFCC dimensions maintain low correlations, capturing orthogonal information about impulsiveness and timbre.

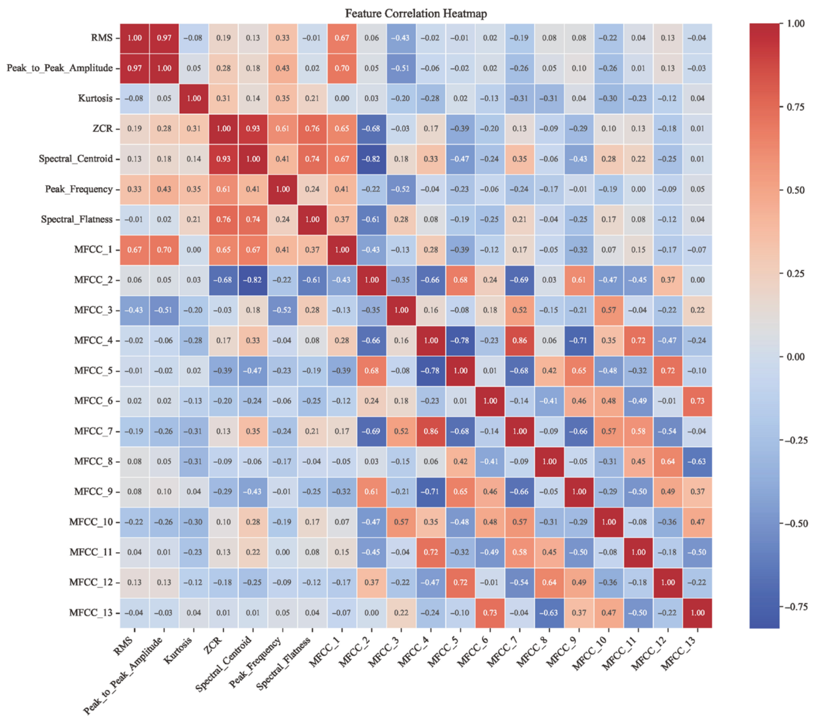

Figure 6. Correlation analysis of the extracted acoustic features.

To balance information retention with model interpretability, a screening threshold of $|r| > 0.9$ was applied. Although tree-based models are generally robust to multicollinearity, the presence of nearly identical features can dilute SHAP importance scores and obscure physical explanations. Based on this criterion, specific decisions were made to prioritize robust features over sensitive ones. We removed Peak-to-Peak Amplitude in favor of RMS. RMS provides an average of effective energy, making it less susceptible to noise spikes. Furthermore, Zero-Crossing Rate was excluded while Spectral Centroid was retained. This decision is based on the fact that Spectral Centroid incorporates information on amplitude to provide a weighted representation of the frequency distribution, whereas Zero-Crossing Rate lacks energy weighting and is prone to distortion by low-energy high-frequency background noise. This refinement produced an optimized 18-dimensional feature set, focusing on stable, information-dense indicators to maximize efficiency.

3.3. Comparative Analysis of Classification Models

In this study, the battery types were numerically encoded for analysis: Class 0 represents pouch cells; Class 1 represents cylindrical cells; and Class 2 represents prismatic cells. We benchmarked the XGBoost classifier against three mainstream algorithms: Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN). The dataset was randomly partitioned into a training set (70%) and a testing set (30%). To objectively evaluate the classification performance on the multiclass dataset, Accuracy, Precision, Recall, and F1-score were selected as the evaluation metrics. According to the quantitative results in Table 5 and the visualizations in Figure 7, the XGBoost model demonstrates superior performance across all indicators. It achieved 96.39% accuracy and a 96.21% F1-score, significantly outperforming the baseline models.

Table 5. Comparison of classification performance metrics among four machine learning models.

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	96.39%	96.52%	95.99%	96.21%
SVM	90.96%	90.79%	91.33%	90.96%
RF	93.37%	93.44%	93.14%	93.25%
KNN	90.36%	90.47%	90.06%	90.15%

The confusion matrices in Figure 7 show that XGBoost accurately distinguishes pouch cells and identifies prismatic cells with a 100% recognition rate. In contrast, SVM and KNN showed weaker performance. Specifically, the SVM model often misclassified prismatic cells as cylindrical ones. This confusion likely results from the spectral similarities between these two rigid-casing batteries. Similarly, the KNN model exhibited scattered misclassifications across all categories, resulting in the lowest accuracy of 90.36%. It is worth noting that although the Random Forest model performs better than SVM and KNN with an accuracy of 93.37%, it still lags behind XGBoost by approximately 3%. ROC analysis confirms these findings, with the AUC for XGBoost approaching 1.00 for all battery types. Therefore, XGBoost's ability to handle complex non-linear feature boundaries makes it the most effective solution for classification.

3.4. SHAP-Based Interpretability Analysis

While XGBoost performs well, fire investigations require classification results supported by transparent, reproducible reasoning. To explicitly quantify the contribution of individual acoustic features to battery-type identification, SHAP was applied to interpret the trained multi-class XGBoost model. We analyzed interpretability at four levels: global feature importance, directional influence, non-linear responses, and instance-level consistency. The results correspond directly to the visualizations.

3.4.1. Global Feature Importance Across Battery Types

Figure 8 shows the SHAP global feature importance rankings for pouch, cylindrical, and prismatic batteries, where importance is quantified using the mean absolute SHAP value.

Across all battery types, Peak Frequency consistently ranks as the most influential feature, serving as the primary discriminative cue. However, the importance of secondary features depends on the battery type.

https://doi.org/10.3390/safety12050121

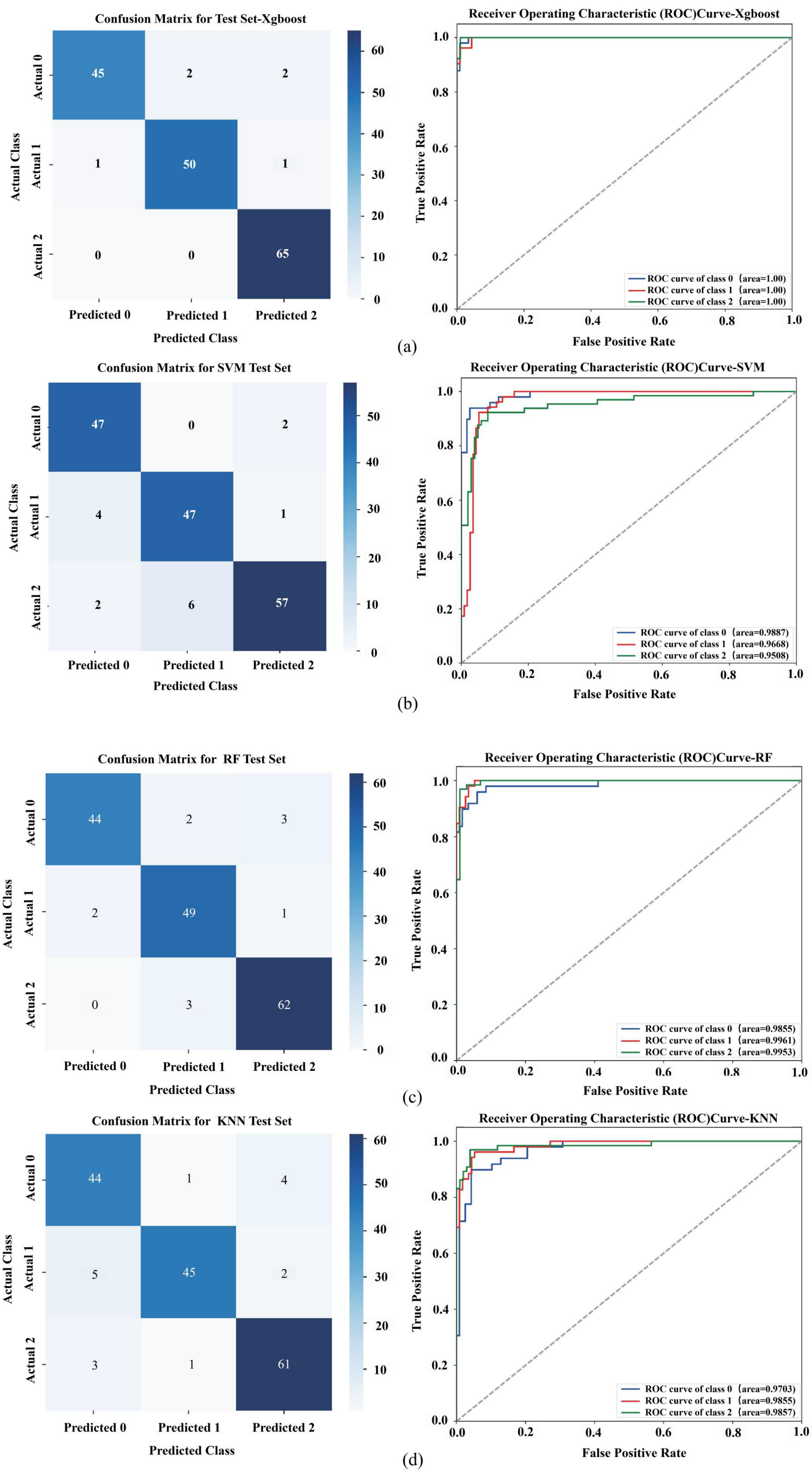

Figure 7. Confusion matrices and ROC curves of different classification models. (**a**) XGBoost; (**b**) SVM; (**c**) RF; (**d**) KNN.

For pouch batteries, Peak Frequency is followed by MFCC_3, Kurtosis, and contributions from MFCC_6, MFCC_8, and MFCC_4. This distribution indicates that pouch battery identification relies primarily on dominant frequency location and cepstral descriptors capturing the overall spectral envelope, while impulsiveness-related information plays a supplementary role. Energy-related features like RMS show comparatively low influence. In cylindrical batteries, Peak Frequency remains dominant, but the SHAP magnitude for Kurtosis increases noticeably. This highlights the model's sensitivity to impulsive acoustic behavior in cylindrical cells. Several MFCC components like MFCC_8 and MFCC_13 refine classification, while RMS and Spectral Centroid contribute moderately. In prismatic batteries, Peak Frequency is significantly more important than other features, creating a frequency-driven discrimination pattern. Secondary contributions come from MFCC_3, MFCC_6, MFCC_4, and Spectral Flatness. In contrast, RMS, Kurtosis, and Spectral Centroid exert limited influence.

These results show that the model uses battery-type-dependent feature combinations, with Peak Frequency as a universal core indicator and MFCCs refining discrimination.

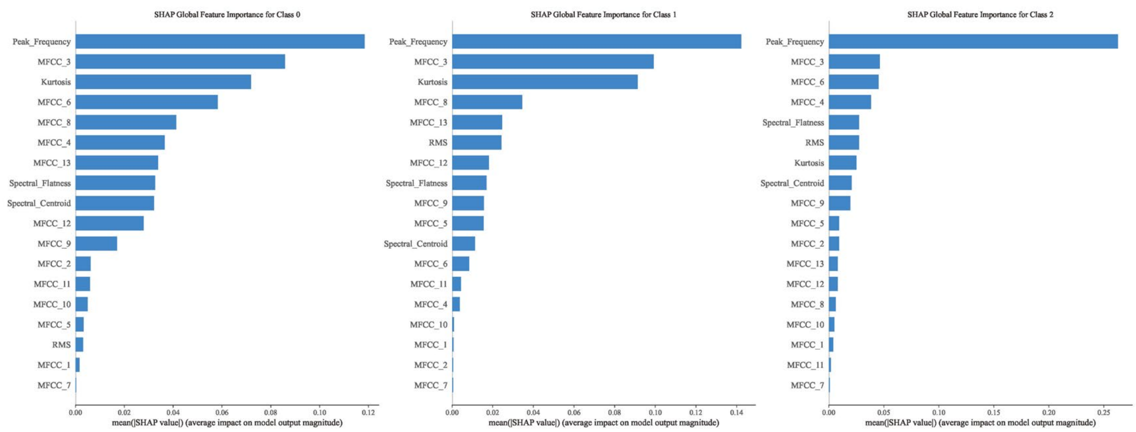

Figure 8. Class-wise SHAP feature importance for battery-type identification.

3.4.2. Directional Influence of Acoustic Features

The SHAP summary plots in Figure 9 show how feature values influence classification outcomes by visualizing SHAP distributions and value ranges.

For pouch batteries, lower Peak Frequency values produce positive SHAP contributions, increasing identification likelihood. MFCC_3 shows a clear trend where higher values contribute positively. Kurtosis displays a dispersed distribution, suggesting impulsiveness does not dominate the decision.

Cylindrical battery predictions are driven by high Peak Frequency values, which consistently yield positive SHAP contributions. A pronounced directional effect is also observed for Kurtosis, where larger values correspond to substantially increased SHAP contributions, underscoring the decisive role of strong impulsive acoustic events. MFCC_3 also contributes directionally, though with greater variability.

For prismatic batteries, intermediate-to-high Peak Frequency values contribute positively, while extreme values reduce prediction probability. MFCC_3 and MFCC_6 provide moderate directional influence, while most other features cluster near zero, indicating limited impact.

These patterns confirm that features exert distinct directional effects depending on battery type, reinforcing the need for class-specific analysis.

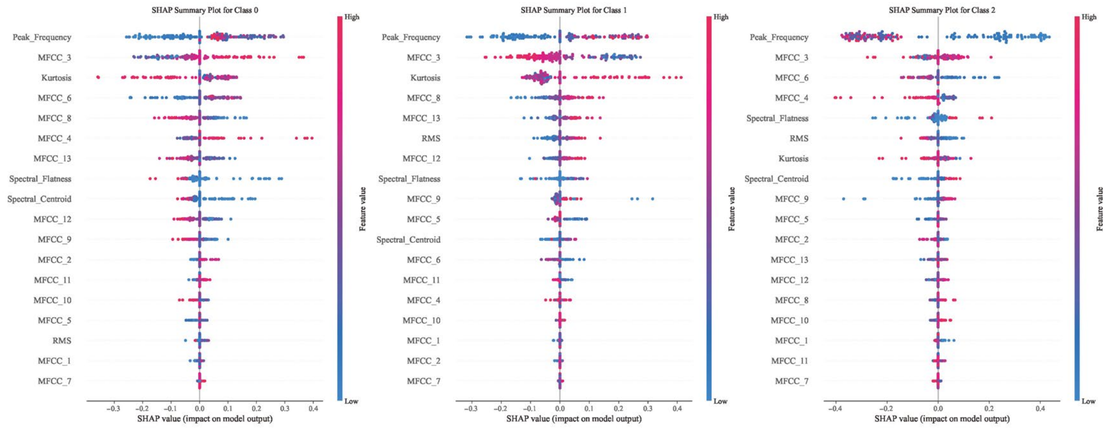

Figure 9. SHAP summary plots showing the distribution and directional impact of acoustic features.

3.4.3. Non-Linear Feature Response Relationships

The SHAP dependence plots in Figure 10 further reveal non-linear relationships between feature values and model output. Peak Frequency exhibits non-monotonic SHAP responses for all battery types, indicating a non-linear trend. Instead, each battery type corresponds to a specific effective frequency interval.

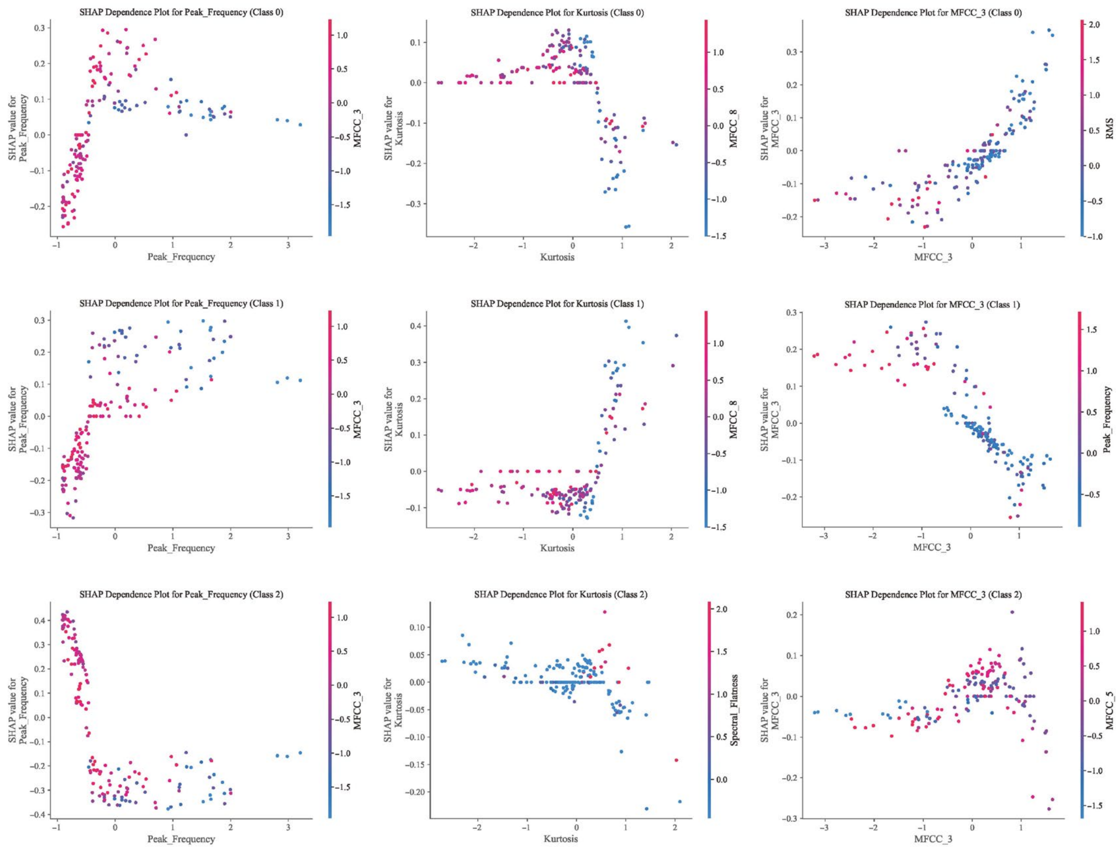

Figure 10. SHAP dependence plots illustrating non-linear feature response relationships.

In cylindrical batteries, SHAP values increase sharply at high Kurtosis levels, strongly biasing predictions toward this type. In contrast, Kurtosis remains close to zero or contributes weakly in pouch and prismatic batteries, confirming its selective relevance.

MFCC_3 shows gradual variation in SHAP values, with slopes dependent on battery type. Color-coded distributions suggest partial interaction with other features, but the lack of a dominant pattern indicates MFCC_3 contributes primarily independently.

Overall, these results indicate that the XGBoost model captures battery-type-specific, non-linear response patterns rather than simple linear thresholds.

3.4.4. Instance-Level Contribution Consistency

The SHAP heatmaps in Figure 11 visualize decision stability at the sample level. These heatmaps show SHAP contributions of major acoustic features across individual instances.

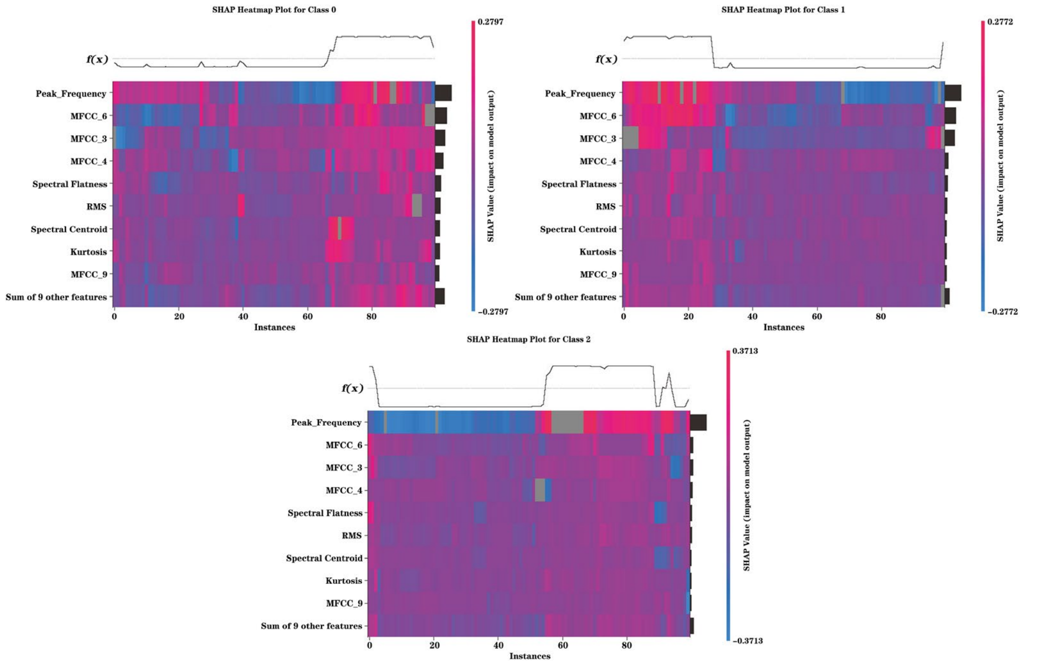

Figure 11. SHAP heatmaps of instance-level feature contributions.

Instances displayed were randomly selected from the test dataset. These examples illustrate typical contribution patterns consistent with the global rankings and summary plots discussed above. Within each category, dominant features like Peak Frequency and key MFCCs show consistent contribution directions across most samples. Although contribution magnitudes vary, no systematic sign reversal occurs, indicating stable decision logic.

This consistency is crucial for fire investigations, where acoustic evidence is collected under heterogeneous conditions. The SHAP-based analysis therefore demonstrates that the proposed XGBoost framework produces not only accurate predictions but also reproducible and interpretable decision behavior, supporting its applicability for acoustic-based battery fire source identification.

4. Conclusions

This study established an acoustic fingerprint identification system for post-accident traceability, utilizing environmental audio from heterogeneous devices. By integrating multi-domain feature extraction with XGBoost and SHAP, this work validates the effectiveness of acoustic evidence in determining ignition sources. The main conclusions are as follows:

(1) Experimental analysis confirmed that battery structure determines distinct and stable acoustic signatures during thermal runaway. Cylindrical cells exhibit high-frequency impulsive characteristics, while pouch cells produce low-frequency tearing sounds. These spectral differences remain consistent across varying conditions, providing a physical basis for identification.

(2) The XGBoost model, using hybrid statistical and cepstral features, effectively distinguished battery formats. Achieving 96.39% accuracy on a mixed dataset, the

model maintained robustness under interference like fan noise and wall obstruction, validating the reliability of acoustic features.

(3) SHAP analysis revealed that the model's decisions are driven by physically meaningful indicators, primarily Peak Frequency. The consistency between global importance and individual explanations confirms that the classification mechanism is stable, transparent, and traceable.

The single-cell design was adopted to obtain clearer data under laboratory conditions, laying the foundation for subsequent studies in complex scenarios. However, multi-cell thermal propagation is more common in real fires and involves more complex operating conditions. Therefore, we plan to conduct dedicated experiments and analyses specifically targeting this issue in our future research.

Collectively, this study establishes a non-contact acoustic forensic methodology. In scenarios where traditional morphological evidence is rendered unavailable by severe fire damage, this approach equips investigators with objective, data-driven supplementary evidence to accurately determine the format of the battery responsible for ignition.

Author Contributions: C.Y.: Conceptualization, Methodology, Data curation, Writing—original draft. J.T.: Investigation, Data curation, Software. W.C.: Investigation, Formal analysis, Validation. S.K.: Supervision, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key Research and Development Program of China (Special Project: Key Technologies for Fire Prevention and Control of Electrochemical Energy Storage Stations, Grant No.2023YFC3009900), and Shanghai Fire Research Institute of MEM project [Grant number: 25SX19].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: During manuscript preparation, generative AI tools were only used for partial Chinese-English translation and academic English polishing. No generative AI participated in experimental design, data calculation, result analysis and core academic writing of this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, M.; Lu, J.; Chen, Z.; Amine, K. 30 years of lithium-ion batteries. *Adv. Mater.* **2018**, *30*, 1800561. [CrossRef]
2. Sun, P.; Bisschop, R.; Niu, H.; Huang, X. A review of battery fires in electric vehicles. *Fire Technol.* **2020**, *56*, 1361–1410. [CrossRef]
3. Feng, X.; Ouyang, M.; Liu, X.; Lu, L.; Xia, Y.; He, X. Thermal runaway mechanism of lithium ion battery for electric vehicles: A review. *Energy Storage Mater.* **2018**, *10*, 246–267. [CrossRef]
4. Wang, Q.; Ping, P.; Zhao, X.; Chu, G.; Sun, J.; Chen, C. Thermal runaway caused fire and explosion of lithium ion battery. *J. Power Sources* **2012**, *208*, 210–224. [CrossRef]
5. Finegan, D.P.; Scheel, M.; Robinson, J.B.; Tjaden, B.; Hunt, I.; Mason, T.J.; Millichamp, J.; Di Michiel, M.; Offer, G.J.; Hinds, G.; et al. In-operando high-speed tomography of lithium-ion batteries during thermal runaway. *Nat. Commun.* **2015**, *6*, 6924. [CrossRef]
6. Hsieh, A.G.; Bhadra, S.; Hertzberg, B.J.; Gjeltema, P.J.; Goy, A.; Fleischer, J.W.; Steingart, D.A. Electrochemical-acoustic time of flight: In operando correlation of physical dynamics with battery charge and health. *Energy Environ. Sci.* **2015**, *8*, 1569–1577. [CrossRef]
7. Espinoza Ramos, I.; Coric, A.; Su, B.; Zhao, Q.; Eriksson, L.; Krysander, M.; Tidblad, A.A.; Zhang, L. Online acoustic emission sensing of rechargeable batteries: Technology, status, and prospects. *J. Mater. Chem. A* **2024**, *12*, 23280–23296. [CrossRef]
8. Su, T.; Lyu, N.; Zhao, Z.; Wang, H.; Jin, Y. Safety warning of lithium ion battery energy storage station via venting acoustic signal detection for grid application. *J. Energy Storage* **2021**, *38*, 102498. [CrossRef]

9. Noll, L.; Rueß, V.; Göhrmann, M.; Schröder, D. Listening to batteries: Systematic analysis of acoustic emission features with ML. *J. Power Sources* **2026**, *661*, 238648. [CrossRef]
10. Noll, L.; Mrowetz, J.; Kretschmer, K.; Schröder, D. Acoustic emissions testing as a complementary tool to understand chemical and electrochemical changes in battery electrodes. *J. Power Sources* **2025**, *629*, 235978. [CrossRef]
11. Zhang, K.; Yin, J.; He, Y. Acoustic emission detection and analysis method for health status of lithium ion batteries. *Sensors* **2021**, *21*, 712. [CrossRef]
12. Pan, Y.; Xu, K.; Wang, R.; Wang, H.; Chen, G.; Wang, K. Lithium-ion battery condition monitoring: A frontier in acoustic sensing technology. *Energies* **2025**, *18*, 1068. [CrossRef]
13. Salamon, J.; Bello, J.P. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Process. Lett.* **2017**, *24*, 279–283. [CrossRef]
14. Ng, M.F.; Zhao, J.; Yan, Q.; Conduit, G.J.; Seh, Z.W. Predicting the state of charge and health of batteries using data-driven machine learning. *Nat. Mach. Intell.* **2020**, *2*, 161–170. [CrossRef]
15. Zhang, G.; Chen, X.; Li, X.; Li, G.; Yang, Y. Online state of health estimation for lithium-ion batteries based on ultrasonic reflection signals and support vector machine. *J. Energy Storage* **2021**, *38*, 102517.
16. Valenzise, G.; Gerosa, L.; Tagliasacchi, M.; Antonacci, F.; Sarti, A. Scream and gunshot detection and localization for audio-surveillance systems. *IEEE Trans. Audio Speech Lang. Process.* **2014**, *22*, 705–719.
17. Mesaros, A.; Heittola, T.; Virtanen, T.; Benetos, E.; Foster, P.; Lagrange, M.; Plumbley, M.D. Detection and classification of acoustic scenes and events: Outcome of the DCASE 2016 challenge. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *26*, 379–393. [CrossRef]
18. Ding, C.; Peng, H. Minimum redundancy feature selection from microarray gene expression data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238.
19. Roscher, R.; Bohn, B.; Duarte, M.F.; Garcke, J. Explainable machine learning for scientific insights and discoveries. *IEEE Access* **2020**, *8*, 42200–42216. [CrossRef]
20. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608.
21. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30, pp. 4768–4777.
22. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
23. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurorobot.* **2013**, *7*, 21. [CrossRef]
24. Mesaros, A.; Heittola, T.; Virtanen, T. Metrics for polyphonic sound event detection. *Appl. Sci.* **2016**, *6*, 162. [CrossRef]
25. Davis, S.B.; Mermelstein, P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust. Speech Signal Process.* **1980**, *28*, 357–366. [CrossRef]
26. Yu, L.; Liu, H. Efficient feature selection via analysis of relevance and redundancy. *J. Mach. Learn. Res.* **2004**, *5*, 1205–1224.
27. Piczak, K.J. Environmental sound classification with convolutional neural networks. In *Proceedings of the IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*; IEEE: Piscataway, NJ, USA, 2015.
28. McFee, B.; Raffel, C.; Liang, D.; Ellis, D.P.; McVicar, M.; Battenberg, E.; Nieto, O. Librosa: Audio and music signal analysis in Python. In *Proceedings of the 14th Python in Science Conference*; SciPy: Austin, TX, USA, 2015; pp. 18–25.
29. Antoni, J. The spectral kurtosis: A useful tool for characterising non-stationary signals. *Mech. Syst. Signal Process.* **2006**, *20*, 282–307. [CrossRef]
30. Lu, L.; Zhang, H.-J.; Li, S.Z. Content-based audio classification and segmentation by using support vector machines. *Multimed. Syst.* **2003**, *8*, 482–492. [CrossRef]
31. Tzanetakis, G.; Cook, P. Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.* **2002**, *10*, 293–302. [CrossRef]
32. Fahy, F.; Gardonio, P. Sound and structural vibration: Radiation, transmission and response. *J. Sound Vib.* **2005**, *285*, 547–557.
33. Rabiner, L.R.; Schafer, R.W. *Digital Processing of Speech Signals*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1978.
34. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
35. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2009.
36. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef]