كلية الهندسة بشبرا
FACULTY OF ENGINEERING AT SHOUBRA
ERJ
PRINT ISSN 3009-6049
ONLINE ISSN 3009-6022

ENGINEERING RESEARCH JOURNAL (ERJ)
Volume (54) Issue (2)
April. 2025, pp:87-99
https://erjsh.journals.ekb.eg

A Supervised Machine Learning Approach: Towards The Automatic Classification of Infrasound Events

Eman Lotfy Habbak *[1] ;Mohamed Salah Abdalzaher [2] ;Hala HA Mansour [3]

[1] ENDC Department, National Research Institute of Astronomy and Geophysics, Cairo 11421, Egypt.
[2] Department of Seismology, National Research Institute of Astronomy and Geophysics, Cairo 11421, Egypt.
[3] Department of Electrical Engineering, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt.
[*] Corresponding Author.
E-mail: eman_habbak@nriag.sci.eg ; msabdalzaher@nriag.sci.eg; hala.mansour@feng.bu.edu.eg

Abstract: Infrasound, which refers to low-frequency sound waves below 20 Hz, originates from various natural and human-made sources. These signals travel through a dynamic atmosphere that might change within minutes of an incident, so their classification process is challenging and sometimes time-consuming. Moreover, accurate classification of infrasound events is crucial for monitoring nuclear test bans and detecting natural disasters. Recently, the use of machine learning (ML) for complex environments and different signals has been merged. This paper presents a supervised ML approach to classify infrasound signals, utilizing feature selection methods - “SelectKBest” and “SelectFromModel”- and feature importance to identify the eight most relevant features of these signals. Traditional machine learning methods were selected over deep learning due to their interpretability, lower computational cost, and effectiveness in handling the available dataset size and variability. The model is trained and examined by a real dataset from the infrasound reference event database (IRED) in time and frequency domains, which are processed using the progressive multi-channel correlation (PMCC) algorithm. Throughout, the system model, feature selection is adopted to use only eight features to reduce the complexity. To ensure the models’ robustness, we have examined them by several evaluation metrics. The results show the model’s effectiveness in the events classification with an accuracy of 92.87% among the benchmarks ensuring the capability of ML for automatic classification. The proposed framework demonstrates significant potential for real-world applications, particularly in nuclear monitoring and natural disaster prediction, where timely and accurate decision-making is critical.

Keywords: Event discrimination, Real-time monitoring, Earthquakes, Explosions, Artificial Intelligence.

1. INTRODUCTION

Infrasound, which is usually known as low-frequency audio energy that lies below 20 Hz [1]. Its sources can vary from natural sources (e.g., microbaroms, bolides, landslides, earthquakes, volcanos, and storms) to anthropogenic (e.g., mining blasts, power plants, explosions, and supersonic flights) [2–6]. Due to its properties, infrasound is a crucial technique for monitoring natural hazards, chemical and nuclear explosions, and other anthropogenic and natural sources of interest at the regional and global scale [7,8].

The detection of these infrasonic signals often involves using multiple sensors spread across an array to distinguish true acoustic signals from background noise caused by turbulence and similar non-acoustic disturbances [9]. Since background noise fluctuates rapidly over short distances, multiple sensors spaced tens to hundreds of meters apart are necessary to capture consistent infrasonic signals. A global infrasound network, established as part of the International Monitoring System (IMS) [10], enables worldwide detection of infrasound signals. The Comprehensive Nuclear Test-ban Treaty Organization (CTBTO) has developed IMS in order to monitor nuclear experiments around the world. It deploys infrasound technology among four different technologies to establish a dependable verification regime for the treaty [11]. Its infrasound network is under development and will reach 60 stations worldwide as planned. The International Data Centre (IDC) collects, processes, and analyzes data from the IMS network, using the Progressive Multi-Channel Correlation (PMCC) algorithm [12] to detect and analyze infrasound signals. This technique separates small amplitude coherent signals from incoherent noise even in situations when signals cannot be distinguished individually from the noise.

The classification of infrasound signals has recently become one of the important procedures deployed to discriminate between different event sources [13–16]. However, atmospheric variability due to wind and temperature changes significantly complicates infrasound signal classification. These variations can alter signal morphology [17], duration [18], frequency characteristics [19], and amplitude [20]. Although infrasound signal characteristics are affected by rapid atmospheric changes [18,21,22] each type of event originates from a distinct physical mechanism. Analysts often rely on seismic data for ground truth, but this approach can be inefficient when seismic data is unavailable or event locations are unknown.

The power and effectiveness of machine learning (ML) have recently been demonstrated in a variety of scientific and daily domains. It has been applied in infrasound research to predict transmission loss due to propagation [23], to monitor nuclear reactors [24], and to categorize signals. Numerous studies have explored the use of ML for classifying infrasound signals. Most studies focus on feature extraction from infrasound data in either the time or frequency domain, comparing different methods to estimate classification accuracy. For instance, wavelet transform has been widely used for characterizing different signals [25,26], while other studies have combined Fourier transform with fuzzy logic for classification [27]. SVM has been a popular choice, often combined with various techniques such as Hilbert-Huang Transform [28,29] and spectral entropy [30] to enhance classification performance. Additionally, it was combined with various techniques to enhance classification performance [31–33]. Random forest (RF) and SVM were also deployed with several time and frequency domain features to classify infrasound signals from storms, power plants, and quarry blasts recorded by a single station [34].

Artificial neural networks (ANN) and deep learning techniques, including convolutional neural networks (CNNs), have seen growing applications in infrasound research [35,36]. For instance, a study proposed the use of mel-frequency coefficients with ANN to classify three different event sources detected by one station [37]. Additionally, CNNs have been used to differentiate rocket signals and investigate using spectrograms for classification [38]. They have also been deployed with Fast Fourier Transform (FFT) to distinguish the infrasound signals resulting from volcanos, earthquakes, and tsunamis [39]. Moreover, they have been utilized to identify stationary and nonstationary signals in an infrasound array using the waveform data as input [40]. A recent study showed promising results of using CNNs for automatic feature extraction and classification of two types of infrasound events [41]. Besides, the augmentation technique has been employed with ResNet architectures to classify signals of chemical explosions and earthquakes and the classification accuracy was compared to SVM and extreme gradient boosting (XGB) [42]. ML techniques were also employed to enhance the detection accuracy of snow avalanches and reduce false alarms [43]. In another direction, a study used k-nearest neighbor (KNN) to identify infrasound signals associated with landslides [44]. In contrast to these studies, which primarily focus on single-station data, our work extends the application of CNNs to multi-station datasets, allowing for a more robust classification of diverse infrasound events.

The research works that have been mentioned above are examples of the most advanced approaches. They obtained a high range of accuracy, ranging from 55% to 100%. However, a major challenge remains the lack of systematically labeled, high-quality, and comprehensive infrasound datasets. This constraint not only affects the performance of supervised ML models but also restricts the exploration of more advanced techniques and a wider application of ML techniques. That is why, the development of large-scale, standardized, and systematically labeled infrasound event catalogs is a need.

Key contributions of this research include:

- Implemented a supervised machine learning workflow to classify infrasound events into four categories: earthquakes, explosions, mines and quarries, and volcanic activity, achieving a classification accuracy of 92.87%, outperforming a previous approach on the same dataset.
- Conducted feature selection and importance analysis to identify the eight most significant features, enhancing model performance while reducing complexity.
- Utilized the IRED catalog, a global infrasound dataset with signals from multiple stations, addressing challenges related to propagation variability.
- Explored and implemented data preparation techniques, including outlier management, to enhance model performance.
- Provided a practical framework for automating infrasound classification, advancing nuclear test monitoring, and natural disaster detection. Also, future work is proposed to expand datasets and integrate atmospheric models for greater accuracy.

2.MATERIALS AND METHODS

This research utilized international data from the CTBTO. This infrasound data represents the largest labeled worldwide infrasound event catalog that is currently accessible. Our objective was to collect a set of features that have been used in different previous studies on classifying infrasound signals and extract them from the IRED data [30,34]. We applied a subset of these features and performed feature selection and importance analysis. After that, we trained various supervised machine-learning models and chose the best-performing one. Here, the classification task discriminates between four classes: earthquakes, explosions, mines and quarries, and volcanic eruptions. This approach aimed to achieve acceptable classification accuracy of global infrasound signals and to identify the most significant features for this task. The proposed workflow is shown in Fig. 1. Most of the work was done using Python [45] with all required packages like sklearn,obspy, pisces, and array processing.

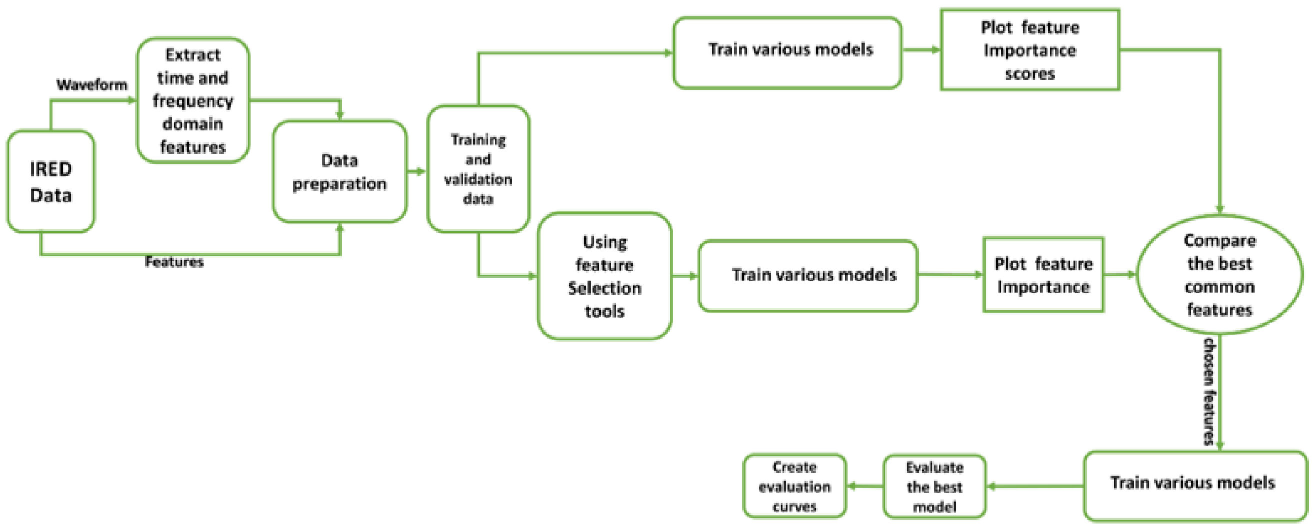

Fig 1. Workflow diagram outlining the detailed steps in our work, from data preparation and feature extraction to model training, feature importance analysis, and model evaluation.

2.1 Utilized Dataset Description

The IDC completed (IRED) by June 2010. This bulletin integrates signals detected from infrasound events identified at the International Monitoring System IMS stations across the globe (see Fig. 2). At that time, 42 infrasound stations were operational, actively detecting and collecting infrasound data. This catalog serves as a vital reference point during routine IDC analysis and is used to test and validate software and atmospheric models. What is notable about this catalog is that all its contents are meticulously analyzed and verified by analysts using reliable sources such as seismic and satellite data. The various sources of the 772 detected events in this catalog are listed in Table 1. Each source with the corresponding number of detected signals is illustrated. Notably, some types have few detections, so our study concentrated on the most abundant signals, including those from mines and quarries, chemical/accidental explosions, earthquakes, and volcanic activity, resulting in 663 signals from 569 infrasound events. As some events are detected at multiple stations, each event is represented more than once with the same label to ensure a substantial amount of data for input to machine learning models. For the most part, the signals were recorded at distances ranging from regional to global scales (15 to over 250 km), with some detected at even greater distances. That's what is considerable about this catalog, in contrast to typical catalogs employed for training predictive models that work in limited geographic areas. This actually presents additional complexity due to dynamic travel paths and makes it challenging to achieve high classification accuracy. The IRED contains a comprehensive set of tables that provide information about the events and the recording IMS stations, as well as all the corresponding waveforms for each event.

2.2 Data Preprocessing

This study implemented some of the event parameters found in IRED data and some features extracted from the event waveform. Most features that came directly from the IRED event tables were computed using PMCC by IDC. There were eleven features that we decided to use directly from the catalog. These features include backazimuth, trace velocity(V_trace), duration, consistency, correlation, root mean square amplitude(A_rms), fundamental frequency, fundamental frequency skewness, Fisher Fstatistic (FSTAT), and family size (FAMSIZE).

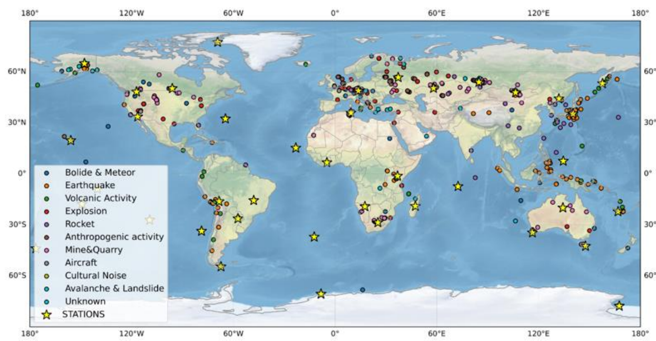

Fig 2: Geographical distribution of IRED events and the IMS infrasound stations by June 2010.

TABLE 1. IRED events grouped by source type.

Type of sources	No. of events
Mines and Quarries	265
Chemical/accidental explosions	132
Earthquakes	104
Rocket launch/re-entry &	73
Volcanic activity	68
Anthropogenic activity	56
Bolides and meteorites	27
Unknown	19
Aircraft	13
Cultural noise	8
Avalanches and landslides	7

The distance between each event and the IMS station was derived using the location information given in the IRED tables and used as a feature. These features were chosen based on their physical relevance to infrasound signals and their propagation, prior research, and classification significance. Other available IRED parameters, including date and time, were excluded since we believe that they do not provide meaningful classification information. In addition to these eleven features, an additional five features were extracted from the waveform in time and frequency domains. To prepare the waveform for this step, it passed through a preprocessing stage. At first, we cut each event's waveforms to the detection duration. Then, beamforming was applied for the available waveforms corresponding to each event. Beamforming is a signal processing technique used to enhance signals coming from a specific direction while suppressing noise and interference from other directions. This technique is particularly useful for infrasound arrays, as each IMS station contains multiple microphones arranged in an array. By aligning and summing these signals, beamforming helps to emphasize coherent infrasound arrivals while suppressing background noise. This step was performed using the metadata in the IRED catalog. The resulting beam representing each event was then filtered using a bandpass filter. The filter lower and upper limits were selected according to the sampling rate of the instruments at the station. This filtered beam representing the event was then used to extract the remaining four features. All sixteen features were combined with a type column to create the training dataset. The "type" column represents the class label, with values ranging from 0 to 3, where 0 for earthquakes, 1 for explosions, 2 for mines and quarries, and 3 for volcanic eruptions. The four classes' signals and the corresponding spectrograms are represented in Fig. 3.

2.3 Feature extraction

Three main categories of features were utilized: time domain, frequency domain, and PMCC-related features. The PMCC-related features were directly obtained from the IRED tables. Additionally, some time and frequency domain features were extracted from the available waveforms following the preprocessing stage, which included steps such as filtering and beamforming. These different features were used in order to evaluate the similarity between adjacent detections and to characterize signals' properties.

- The time domain features that were utilized were the A_rms, the signal-to-noise ratio (SNR), and the root mean squared energy (E_rms) which is calculated using this formula:

$$E_{rms} = \frac{1}{L} \sum_{J=0}^{L} A_j^2 , \qquad (1)$$

where A_j represents the j-th amplitude value in the event's signal and L is the number of samples.

- Entropy-based methods quantify the uncertainty of signal energy distribution across different domains [30,31]. That is why we extracted two of them as described. Power spectrum entropy (PSE): It is defined as the Shannon entropy of the power spectrum [46] and can be formulated as follows:

$$PSE = \sum_{f_s=0}^{\frac{f_s}{2}} P_S(f) \,.\, log_2 P_S(f), \qquad (2)$$

where fs refers to the sampling frequency and PS is the power spectrum. This feature evaluates how evenly the power spectrum is distributed. It reaches its lowest value for a monochromatic signal and rises as more frequency components are present.

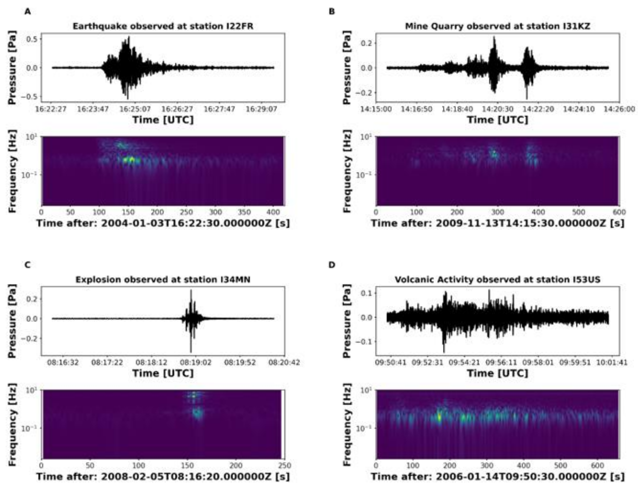

Fig 3. The four classes' examples of signals and their spectrograms.

Wavelet Singular Spectrum Entropy (WSSE) The WSSE information measure is a feature extraction technique that combines the singular value decomposition of the wavelet transform with entropy-based feature extraction. This method effectively characterizes the uncertainty of infrasound signal energy distribution in the time-frequency domain. The Wavelet Singular Spectrum Entropy is formulated as follows [30]:

$$WSSE = -\sum_{j=1}^{M} p_j . \ln p_j \, , \tag{3}$$

where the term $p_j = \frac{\lambda_j}{\sum_{i=1}^{M} \lambda_i}$ signifies the normalized value. Here, λ_j is the j-th singular value that makes up the singular spectrum and M stands for the quantity of effective singular characteristic values that are utilized in the wavelet singular entropy calculation.

- The frequency domain feature that was extracted is the spectral band energy ratio (BER). For the j-th frame it can be calculated as follows:

$$BER_j = \frac{\sum_{m=1}^{F_d-1} S_j(m)^2}{\sum_{m=F_d}^{M} S_j(m)^2} \, , \tag{4}$$

where M is the number of frames and $S_j(m)$ is the spectrogram's magnitude in the j-th frame at frequency bin m. Here, F_d is the frequency value that was chosen at which the split is performed. For this calculation, we set this value at 3 Hz, which was determined through a combination of empirical testing and prior research [34]. The empirical evaluation showed that this threshold provided the best class separability, particularly distinguishing explosions and quarry blasts from other sources. BER was derived to be the median of BER_j for each event's beam.

2.4 Data Preparation

At that moment, the preparation process was performed to handle missing values, the number of these values was small so they were imputed using scikit-learn's SimpleImputer [47]. Outlier estimation is also an essential step in this process. This step is used to identify data points that deviate significantly from the expected pattern or distribution in a dataset. Several techniques are commonly employed for detecting outliers, such as isolation forest, local outlier factor (LOF), and the interquartile range (IQR) method [48–50]. To optimize performance, we adopted a class-specific approach for outlier detection. The scikit-learn LOF algorithm was used to identify outliers within each feature data set for each class. This algorithm compares the local density of a value or a point to the local densities of its neighbors and decides if it is an outlier [51]. LOF provides better interpretability and consistency in preserving essential data structures. It is also effective for datasets with varying densities or clustered structures, making it suitable for the used dataset.

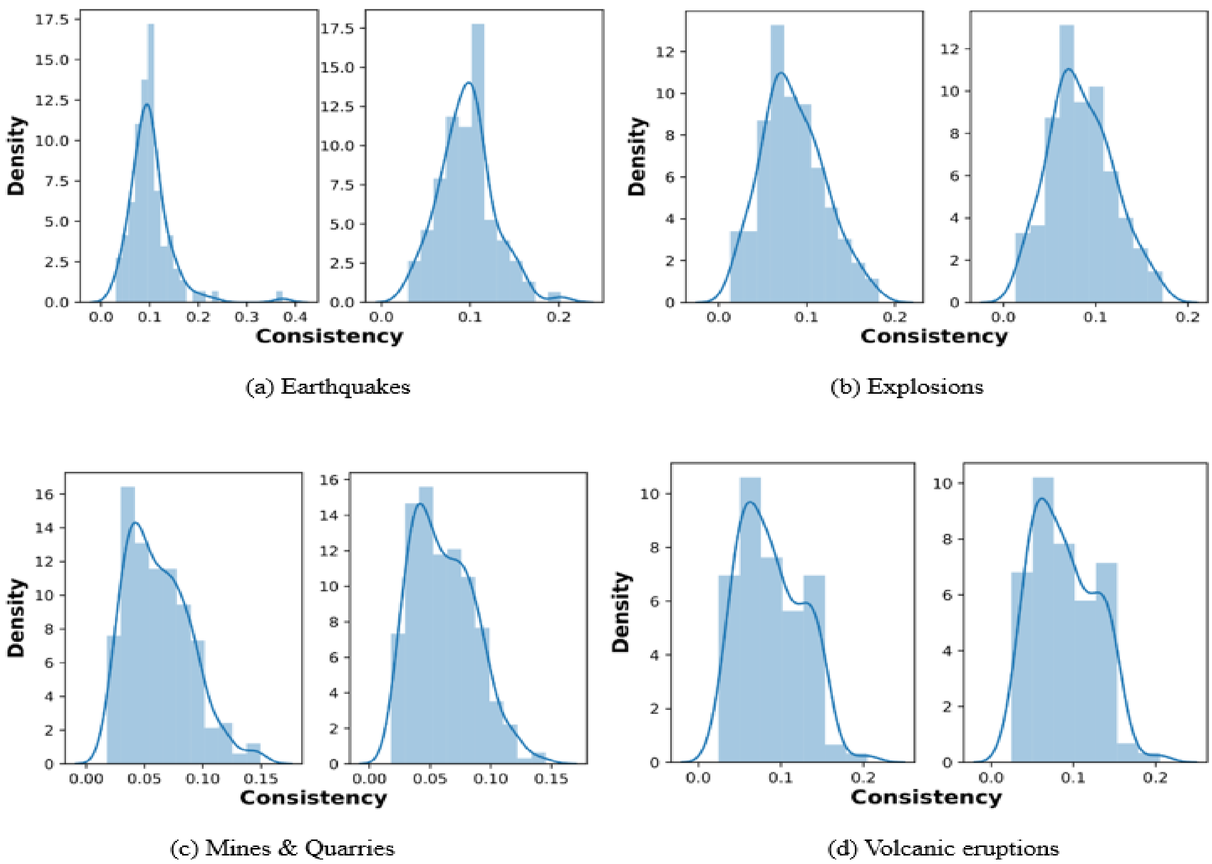

FIG 4. Distribution of one feature (Consistency) for all classes before and after outlier replacement.

Outliers were not deleted but instead replaced with the nearest quartiles Q1 or Q3 of the corresponding feature within their respective class to preserve data integrity. These quartiles Q1 and Q3 are statistical measures that explain the division of the data into four quarters. The strategy that we used proved effective in enhancing model accuracy. Fig. 4 shows one of the features' dataset distributions before and after outlier replacement for each class. This plot demonstrates that the replacement process has been successful in achieving a balance between improving feature quality and maintaining distribution consistency, thereby supporting better model performance. The last step applied to the training dataset involved normalization using the MinMaxScaler from the scikit-learn library.

2.5 Models Training and Evaluation

After the preprocessing and preparation of the data, including the extraction of additional features, the workflow proceeded to the model training step. At this stage, we followed two main paths to evaluate and refine the ML models for classifying infrasound events, focusing on identifying the most efficient features and the best-performing models. The workflow for each path is described as follows:

- Initially, using the full set of 16 features, we trained and evaluated both linear and nonlinear supervised machine learning models. These models, which are listed in Table 2 and have been explained and utilized in previous classification research studies [52,53], were selected based on their proven effectiveness in similar tasks. After evaluating the models, a feature importance plot was generated using the most effective model. Feature importance refers to the method of assigning scores to features based on their contribution to the predictions of an ML model. These scores are relative, effectively ranking the features according to their impact on the model's performance. Various techniques exist to measure feature importance, including the permutation-based method, the coefficient method, and the decision trees-based method [54,55]. Although the permutation method is available in Scikit-Learn, it requires significant computational time. In our study, we utilized the "coef_" and "feature_importance_" attributes after fitting the model. These attributes provide scores that indicate the influence of features on the model's predictions. The "coef_" attribute is used with linear models, while the "feature_importance_" attribute is applied to tree and ensemble models. This analysis provided insights into the relative significance of each feature across the models.
- In the second path, we tried to apply feature selection techniques to refine the feature set. We employed two widely used feature selection methods, "SelectFromModel" and "SelectKBest". Feature selection can be defined as identifying the most pertinent independent variables that significantly influence the prediction of the target variable. The "SelectFromModel" method is a feature selection technique that selects a subset of features based on their importance as determined by an underlying model. This method which is available in the Scikit-learn library utilizes an estimator to choose features with an importance score above a specified threshold or within a defined maximum number of features [56]. XGB was the underlying model that we selected, and we set "max_features" to nine. However, "SelectKBest" is a feature selection technique that uses univariate statistical tests to determine which k features have the highest scores. Then it selects the k-features with the highest scores based on these tests. We chose the statistical chi-squared test and set the feature count to nine in this Scikit-learn library application. To assess the robustness of our feature selection process, we compared the results of both methods to verify consistency in feature importance. The top features identified by "SelectFromModel" and "SelectKBest" were analyzed, and we observed significant overlap in the selected features. This analysis helped determine whether both methods identified a similar set of significant features or if certain features were uniquely emphasized by one technique. As before, we plotted the feature importance scores for the best model after applying both linear and nonlinear ML models on the same dataset. While feature selection methods focus on selecting the most relevant features, alternative dimensionality reduction techniques, such as Principal Component Analysis (PCA), could further simplify the feature space. However, PCA creates transformed features that may reduce interpretability in a geophysical context. In future work, we plan to explore PCA to assess whether it offers additional advantages for infrasound classification, particularly in optimizing computational efficiency.

Model	Abbreviation	Advantage
Light Gradient Boosting Machine	LGB	Efficient and scalable boosting algorithm for large datasets.
Extreme Gradient Boosting	XGB	Powerful boosting method with high predictive accuracy.
Gradient Boosting Classifier	GBC	Provides high accuracy and handles various data types well.
Extra Trees Classifier	ET	Fast, reliable, and robust to noisy data.
Random Forest Classifier	RF	Ensemble method that improves accuracy and controls overfitting.

K Neighbors Classifier	KNN	Simple and effective for classification problems with clear separation.
Decision Tree Classifier	DT	Easy to interpret and good for baseline performance.
Linear Discriminant Analysis	LDA	Effective for classification with linear separability.
Quadratic Discriminant Analysis	QDA	Suitable for non-linear decision boundaries.
Ridge Classifier	RIDGE	Useful for multicollinearity and regularization.
Logistic Regression	LR	Simple and interpretable
Support Vector Machine	SVM	Effective for non-linearly separable data.
Naive Bayes	NB	Fast and works well with high-dimensional datasets.
Ada Boost Classifier	ADA	Works with complex classification tasks and improves overall performance.
Dummy Classifier	Dummy	Baseline model for performance comparison.
CatBoost Classifier	CB	Handles categorical features efficiently

• Feature importance was evaluated using three different methods, and the top eight features that consistently appeared across all importance plots were selected for final model training. The choice of eight features was determined through a systematic evaluation of model performance. Reducing the number of features below eight led to a noticeable decline in accuracy, indicating that these eight features collectively capture the essential information required for effective classification. This threshold was maintained to ensure a balance between model simplicity and performance. This reduction in features was found to improve classification accuracy, indicating that the removal of less relevant features enhanced the model's efficiency. Using k-fold cross-validation with 13 folds, the evaluation metrics of the top-performing model were assessed. This approach ensured a robust evaluation of model performance and provided reliable insights into its generalizability. The performance metrics that were implied in this work are accuracy, F1-Score, the Cohen Kappa Score (Kappa) [57], the Matthews Correlation Coefficient (MCC) [58], and Confusion Matrix [59].

3.RESULTS

The results of this study can be presented in three main stages, corresponding to the workflow: (1) evaluation of models using the full feature set (16 features), (2) refinement of features using feature selection techniques, and (3) final model training and performance assessment using the optimized feature set. Using the complete set of 16 features, both linear and nonlinear supervised machine learning models were trained and evaluated. With an accuracy of 91.43%, GBC was the model that performed the best at this point. Following that, the feature importance score was plotted for this model, as shown in Fig. 5.
To refine the feature set, two widely used feature selection techniques were applied, “SelectFromModel” and “SelectKBest”. Using “SelectFromModel”, LGB achieved the highest accuracy of 92.08%, and the corresponding feature importance plot is presented in Fig. 6. On the other hand, the “SelectFromModel” yielded an accuracy of 90.57% with the XGB model. The feature importance plot for this model is illustrated in Fig. 7. Comparing the three feature importance plots (from the full feature set, SelectFromModel, and SelectKBest) revealed a set of eight features that appeared across analyses. Model accuracy decreased when the number of features was reduced below eight, indicating that keeping these features is important for best results. Using the refined set of eight features- backazimuth, V_trace, duration, Distance, E_rms, *WSSE*, *BER*, and *PSE* - all the ML models were retrained and evaluated to obtain the top-performing model.

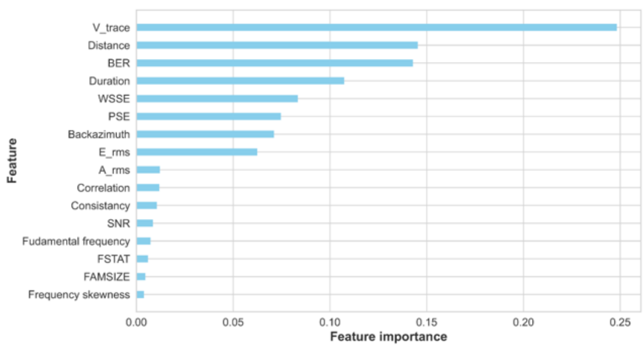

FIG 5, Feature importance plot of the full set of features.

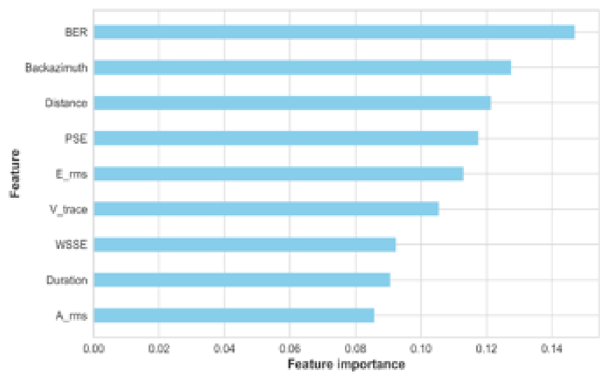

Fig 6: Feature importance plots using the feature selection ‘SelectFromModel’ method.

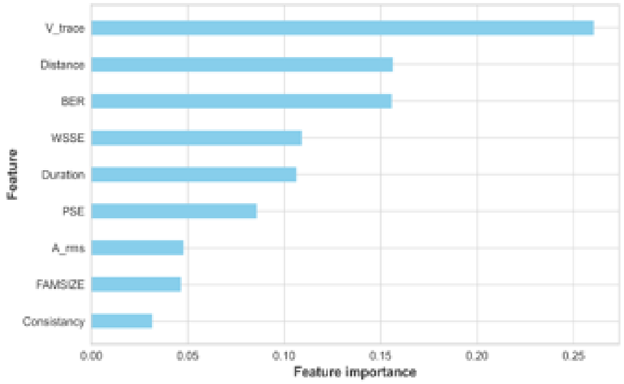

Fig 7: Feature importance plots using the feature selection 'SelectKBest' method.

A comparison of their performance is illustrated in Fig. 8. After fine-tuning the hyperparameters of the best-performing classifier, the GB model, it was evaluated using k-fold cross-validation with 13 folds. This approach ensured a robust assessment of the model's generalizability and performance. The model achieved an average accuracy of 92.87% for the four classes. The evaluation metrics, including accuracy, precision, recall, MCC, Kappa score, and F1-score, were calculated to provide a comprehensive understanding of the model's effectiveness, as summarized in Table 3. When evaluating a classification task, the confusion matrix is a crucial tool that shows how well the predicted classes perform. The confusion matrix of the final classifier is plotted in Fig. 9. The plot illustrates that the model showed acceptable classification accuracy across all categories, with classification rates exceeding 80% for earthquakes and exceeding 90% for explosions, quarries, and volcanoes. Earthquakes show the highest rate of misclassification (19.47%), primarily being confused with quarries (10.49%). This suggests a similarity in the feature space of these two event types, potentially due to overlapping characteristics in their signal patterns. Besides, the best classifier's ROC/AUC curve which shows how well it can differentiate between the four classes, is shown in Fig. 10. The validation and learning curves are also in Figs. 11 and 12, respectively.

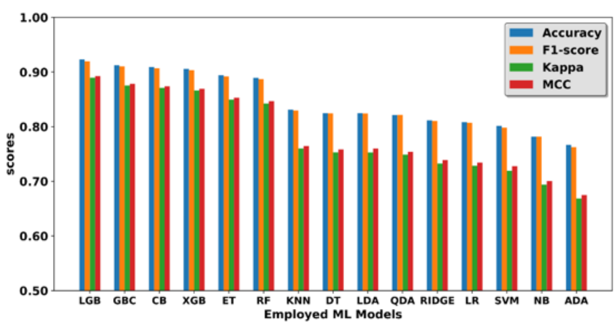

Fig 8. A comparison of Metrics for linear and nonlinear ML models relying on 8 features.

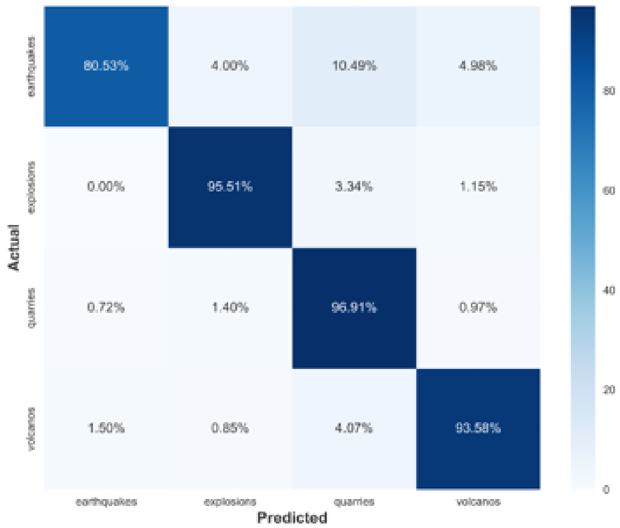

Fig 9. Confusion Matrix of LGB model.

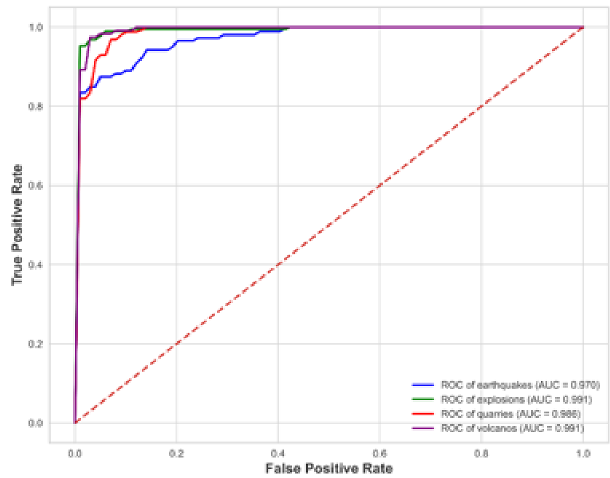

Fig 10. ROC/AUC curves of the optimal classifier model.

Table 3: Metrics of LGB classifier.

Metric	value
Accuracy	92.87%
Recall	91.21%
F1 score	91.86%
Precision	93.83%
MCC	90.05%
Kappa score	89.91%

4.DISCUSSION

In this study, we explored the application of supervised machine learning (ML) techniques for the classification of infrasound events, focusing on feature selection and model optimization. Unlike previous studies that utilized CNNs [31,38,39] or focused on only one or two models [29,30,32,34], our approach relied solely on supervised ML models, including linear and nonlinear algorithms, to classify infrasound signals. We utilized infrasound-labeled signals from the IRED catalog and extracted 16 features for analysis. Initially, we examined all features, and the corresponding feature importance plot revealed that eight features were particularly significant. Following the application of feature selection techniques, we compared their performance and corresponding feature importance plots with one. This comprehensive and comparative analysis enabled us to identify the optimal model and the most impactful features, which included distance and duration. These features highlight the influence of propagation on infrasound signals. Upon evaluating the performance of the best classifier, LGB, earthquakes were the most frequently misclassified class, with a higher rate of false positives and false negatives compared to other classes. This misclassification likely stems from the shared characteristics between earthquakes and other event types. This trend likely stems from overlapping feature distributions, particularly in the frequency content, and propagation characteristics, which can exhibit similarities among these event types.

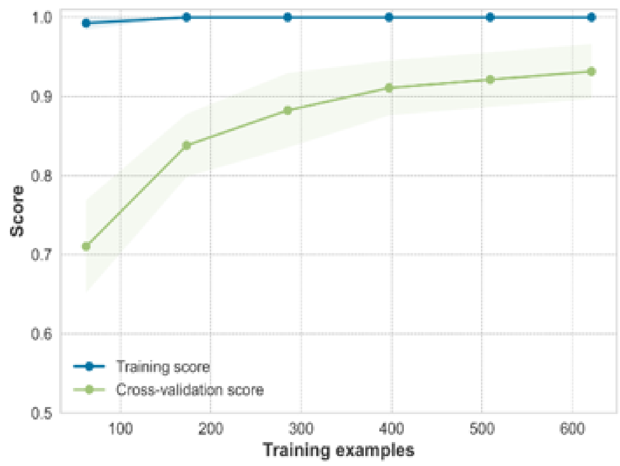

Fig 11: The learning curve of the best classifier

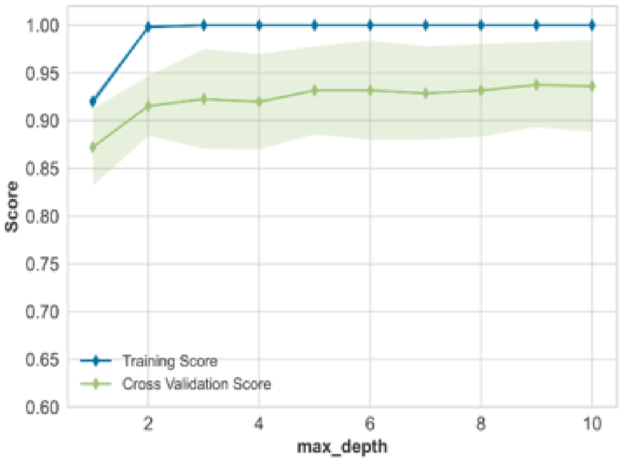

Fig 12: Validation curve of the best classifier.

These similarities may arise from atmospheric variability that can alter the morphology of earthquake signals, causing them to resemble other event types. To further investigate, we analyzed the ROC/AUC curve, which indicated lower discrimination performance for earthquakes compared to other classes. This aligns with the observed misclassifications and suggests that certain extracted features may not be sufficient to fully distinguish earthquakes from other sources. In contrast, quarry events achieved the highest accuracy, likely due to their occurrence at shorter or local distances, which simplifies their detection and classification and lowers the effect of propagation. Despite this, the model metrics demonstrate robust overall performance. Additionally, the learning and validation curves suggest that increasing the volume of data, particularly from international stations, could further enhance model performance. These findings suggest that augmenting the dataset with more earthquake samples, particularly from diverse stations and propagation conditions, could improve classification accuracy. Additionally, exploring more advanced feature extraction techniques or using deep learning may enhance the separation between earthquakes and other event types and this is targeted in future work.

A previous study that utilized the same IRED dataset focused on classifying infrasound events using both SVM and CNN [31]. The study evaluated binary classification (earthquakes vs. volcanic activity) and multi-class classification tasks. For binary classification, SVM achieved an accuracy of 75%, while CNN achieved 74%. In the multi-class classification task, SVM achieved 55% accuracy, and CNN achieved 56%. The SVM models relied on time and frequency domain features, while the CNN models used spectrograms as input. Before training, the authors selected eight important features using random forests, with the top two features being the distance between the source and receiver and waveform duration, indicating a range dependency in classification. In contrast, our study focused exclusively on supervised ML models and achieved an accuracy of 92.87% using a refined set of eight features. Unlike the previous study, which employed CNNs and spectrograms, our approach relied on feature selection techniques such as “SelectKBest” and “SelectFromModel” to identify the most impactful features. We also used a different method for calculating feature importance scores, in which distance also showed high importance. Additionally, we incorporated features that were not included in the previous study. Furthermore, our workflow included preprocessing and outlier management steps, which allowed us to achieve significantly higher accuracy without the need for complex deep-learning architectures. Although this accuracy may be lower than some research works mentioned in the introduction section, we believe that our goal is to address a more complicated issue. Unlike some prior studies, which often relied on single-station recordings or used the same station for all event classes, our approach incorporates waveforms from multiple stations that may capture the same event. This approach introduces greater variability and complexity, as it accounts for diverse source-station combinations. In contrast, studies using single-station signals or limited source station configurations simplify the classification task, which may lead to higher reported accuracies. However, such approaches limit the model's applicability and reduce its generalizability to different event catalogs. By incorporating multi-station data, our work aims to enhance the robustness and generalizability of the model.

During this work, the initial step of applying all features was to gain a comprehensive understanding of the ML models' behavior with all features and their impact on performance or prediction. This approach also provided insights into how preprocessing steps and data preparation can effectively improve the ML model's accuracy. Replacing outliers proved to be more effective in enhancing model accuracy without significantly reducing the data size than removing them. Subsequently, the implementation of the "SelectFromModel" and "SelectKBest" methods, along with Scikit-learn's feature importance attributes, enabled us to optimize the best discriminative parameters from the available data. Among the various ML models applied, the LGB model demonstrated the best performance with this dataset.

Despite the growing popularity of deep learning, our traditional ML-based approach outperformed CNN-based methods as follows:

- Traditional ML models are more data-efficient than CNNs, which require large amounts of labeled data to achieve optimal performance.
- Traditional ML models offer greater interpretability compared to CNNs. By identifying and analyzing the importance of specific features, we gained valuable insights into the factors influencing infrasound classification and this is crucial for practical applications.
- The preprocessing steps improved data quality, leading to better classification performance compared to the CNN-based method on the same dataset.

In spite of its advantages, our approach encountered a number of restrictions and difficulties while conducting. It was difficult to find and choose the most pertinent features. Feature selection techniques significantly enhanced interpretability and improved accuracy while also reducing training time by eliminating redundant features. Although feature selection introduces an initial computational overhead, the reduction in feature dimensionality ultimately optimized training efficiency. This study was conducted using a PC with an Intel Core i7-1165G7 @ 2.80 GHz processor and 16 GB RAM. Despite this modest hardware, our best-performing model (LGB) achieved 92.87% accuracy, demonstrating that our feature selection strategy effectively balanced model complexity and computational efficiency. Training time for this model was approximately 0.4300 seconds using 13-fold cross-validation. As the dataset size increases, in future studies, computational demands are expected to grow, however, strategies such as parallelized training can mitigate these challenges. The size of the dataset was one of the difficulties, which might have made it more difficult to fully validate the model's efficiency. For this reason, cross-validation was employed to a certain degree to address this problem; nevertheless, larger-scale validation using separate datasets will be employed in further studies to validate the findings' robustness. The quality of the data presents another difficulty, which was addressed during the preprocessing phase and in the management of outliers that can affect the performance of the model. Further, Infrasound signals are highly influenced by atmospheric changes, such as temperature variations, wind patterns, and pressure fluctuations, which highly affect signal propagation and detection. These changes make it essential to combine or integrate atmospheric models with ML techniques to accurately analyze and predict the behavior of infrasound signals.

To sum up, although our approach achieved higher accuracy than the CNN-based model held on the same IRED catalog, we acknowledge that deep-learning techniques still hold potential, particularly if larger datasets become available. Future work could explore hybrid approaches, integrating feature-based ML with deep learning to combine the advantages of both methodologies. Additionally, expanding the dataset by using new labeled data or augmentation techniques, and including atmospheric modeling could further improve classification performance.

5.CONCLUSION

In this paper, we conducted a comprehensive comparative analysis to evaluate the effectiveness of ML techniques for infrasound signal classification on a global scale. The proposed workflow involved data preprocessing, feature extraction, the application of both linear and nonlinear ML models, the use of feature selection methods, and the calculation of feature importance for effective infrasound classification. Although the lack of systematically labeled, and sufficient infrasound dataset quality, the promising obtained results prove the effectiveness of the proposed ML approach for infrasound classification. To ensure the model's robustness, we have evaluated it by several evaluation metrics (accuracy, precision, recall, MCC, Kappa score, and F1 score). Our model achieved significant classification with an accuracy of 92.8%, outperforming a previous study that utilized the same IRED catalog. This improvement comes from our broader feature exploration, the application of feature selection techniques, and the integration of multiple ML models. These refinements enhanced classification accuracy and model robustness, demonstrating the effectiveness of our approach for infrasound signal classification and its application in nuclear test monitoring and natural disaster detection.

In future work, we aim to expand the dataset by integrating infrasound signals from the past ten years and comparing them with seismic catalogs to construct a new large-scale labeled infrasound event catalog. This catalog will be incorporated with the existing dataset used in this study to create a more comprehensive global dataset suitable for advanced classification tasks. Additionally, we plan to investigate deep learning-based classification techniques by applying raw waveform analysis instead of feature extraction. This approach will allow deep learning models, such as transformers, to autonomously learn feature representations and enhance classification accuracy. By developing a more extensive and diverse dataset, we expect to improve model generalization and achieve higher classification performance for real-world applications. This effort will require substantial data collection and preprocessing, forming the foundation of our next research phase.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) for providing access to invaluable data that made this research possible.

REFERENCES

[1] G. Leventhall, P. Pelmear, and S. Benton, "A review of published research on low-frequency noise and its effects," *Department for Environment, Food and Rural Affairs (DEFRA)*, London, UK, 2003.

[2] A. J. Bedard Jr. and T. M. Georges, "Atmospheric infrasound," *Physics Today*, vol. 53, no. 3, pp. 32–37, 2000.

[3] J. P. Mutschlecner and R. W. Whitaker, "Infrasound from earthquakes," *Journal of Geophysical Research: Atmospheres*, vol. 110, no. D1, 2005.

[4] R. R. Zhang, "Characterizing and quantifying earthquake-induced site nonlinearity," *Soil Dynamics and Earthquake Engineering*, vol. 26, no. 8, pp. 799–812, 2006.

[5] J. Assink, L. Evers, I. Holleman, and H. Paulssen, "Characterization of infrasound from lightning," *Geophysical Research Letters*, vol. 35, no. 15, 2008.

[6] A. Cannata, P. Montalto, M. Aliotta, C. Cassisi, A. Pulvirenti, E. Privitera, and D. Patanè, "Clustering and classification of infrasonic events at Mount Etna using pattern recognition techniques," *Geophysical Journal International*, vol. 185, no. 1, pp. 253–264, 2011.

[7] N. Brachet, D. Brown, R. Le Bras, Y. Cansi, P. Mialle, and J. Coyne, "Monitoring the Earth's atmosphere with the global IMS infrasound network," in *Infrasound Monitoring for Atmospheric Studies*, Springer, Dordrecht, pp. 77–118, 2009.

[8] J. Liu, W. Tang, X. Wang, and H. Tang, "Analysis of mechanism to produce infrasound signals and its characteristics," *Environmental Engineering*, vol. 28, no. 4, pp. 92–96, 2010.

[9] V. E. Asming, A. V. Fedorov, I. S. Fedorov, and S. V. Asming, "Algorithms for the detection, location, and discrimination of seismic and infrasound events," *Izvestiya, Atmospheric and Oceanic Physics*, vol. 58, pp. 1398–1417, 2022.

[10] D. R. Christie and P. Campus, "The IMS infrasound network: Design and establishment of infrasound stations," in *Infrasound Monitoring for Atmospheric Studies*, Springer, Dordrecht, pp. 29–75, 2009.

[11] L. Zhang, X. Wang, and B. Xia, *Introduction of Nuclear Test Ban Verification Technology*, National Defense Industry Press, Beijing, China, 2005.

[12] Y. Cansi, "An automatic seismic event processing for detection and location: The PMCC method," *Geophysical Research Letters*, vol. 22, no. 9, pp. 1021–1024, 1995.

[13] S. Park, F. M. Ham, and C. G. Lowrie, "Discrimination of infrasound events using parallel neural network classification banks," *Nonlinear Analysis*, vol. 63, pp. e859–e865, 2005.

[14] I. Stancu, D. Ghica, M. Radulian, and C. Ionescu, "Identification of near quarry blasts using Plo¸stina infrasound array—Case study for Dobrogea (South-East Romania)," *Romanian Reports in Physics*, vol. 67, no. 3, pp. 1192–1205, 2015.

[15] M. Kiszely, B. Süle, C. Czanik, P. Mónus, and I. Bondár, "Discrimination of microearthquakes and quarry blasts based on the observations of the PSZI infrasound array and seismic station PSZ in Piszkés-tet˝o, Hungary," *EGU General Assembly Conference Abstracts*, pp. 7162, 2019.

[16] C. Czanik, M. Kiszely, P. Mónus, B. Süle, and I. Bondár, "Identification of quarry blasts aided by infrasound data," *Pure and Applied Geophysics*, vol. 178, pp. 2287–2300, 2021.

[17] S. J. Gibbons, V. Asming, L. Eliasson, A. Fedorov, J. Fyen, J. Kero, E. Kozlovskaya, T. Kværna, L. Liszka, S. P. N¨asholm, et al., "The European Arctic: A laboratory for seismoacoustic studies," *Seismological Research Letters*, vol. 86, no. 3, pp. 917–928, 2015.

[18] D. N. Green and A. Nippress, "Infrasound signal duration: The effects of propagation distance and waveguide structure," *Geophysical Journal International*, vol. 216, no. 3, pp. 1974–1988, 2019.

[19] L. Ceranna, A. Le Pichon, D. Green, and P. Mialle, “The Buncefield explosion: A benchmark for infrasound analysis across central Europe,” *Geophysical Journal International*, vol. 177, no. 2, pp. 491–508, 2009.

[20] P. S. Smets and L. G. Evers, “The life cycle of a sudden stratospheric warming from infrasonic ambient noise observations,” *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 21, pp. 12–084, 2014.

[21] S. Kulichkov, “Long-range propagation and scattering of low-frequency sound pulses in the middle atmosphere,” *Meteorology and Atmospheric Physics*, vol. 85, no. 1, pp. 47–60, 2004.

[22] D. C. Bowman and S. A. Albert, “Acoustic event location and background noise characterization on a free flying infrasound sensor network in the stratosphere,” *Geophysical Journal International*, vol. 213, no. 3, pp. 1524–1535, 2018.

[23] Q. Brissaud, S. P. N¨asholm, A. Turquet, and A. Le Pichon, “Predicting infrasound transmission loss using deep learning,” *Geophysical Journal International*, vol. 232, no. 1, pp. 274–286, 2023.

[24] C. Chai, C. Ramirez, M. Maceira, and O. Marcillo, “Monitoring operational states of a nuclear reactor using seismoacoustic signatures and machine learning,” *Seismological Society of America*, vol. 93, no. 3, pp. 1660–1672, 2022.

[25] J. Chilo, T. Lindblad, R. Olsson, and S.-E. Hansen, “Comparison of three feature extraction techniques to distinguish between different infrasound signals,” in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1–4, 2007.

[26] J. Chilo, R. Olsson, S.-E. Hansen, and T. Lindblad, “Classification of infrasound events with various machine learning techniques,” in *4th International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA)*, pp. 191–195, 2007.

[27] N. W. Wang, N. S. Wei, N. Q. Liao, N. Y. Xia, N. D. Li, and N. J. Li, “Fuzzy K-means clustering on infrasound sample,” in *IEEE International Conference on Fuzzy Systems*, pp. 756–760, 2008.

[28] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, 1998.

[29] X. Liu, M. Li, W. Tang, S. Wang, and X. Wu, “A new classification method of infrasound events using Hilbert-Huang Transform and support vector machine,” *Mathematical Problems in Engineering*, vol. 2014, pp. 1–6, 2014.

[30] M. Li, X. Liu, and X. Liu, “Infrasound signal classification based on spectral entropy and support vector machine,” *Applied Acoustics*, vol. 113, pp. 116–120, 2016.

[31] S. Albert and L. Linville, “Benchmarking current and emerging approaches to infrasound signal classification,” *Seismological Research Letters*, vol. 91, no. 2, pp. 921–929, 2020.

[32] H. Rezaei, A. K. Razlighi, and A. Koochari, “Infrasound signal classification based on combining spectral and sound features,” *Journal of Distributed Computing and Systems (JDCS)*, vol. 4, no. 2, pp. 28–33, 2021.

[33] Q. Lu, M. Wang, and M. Li, “Infrasound signal classification based on ICA and SVM,” *Archives of Acoustics*, vol. 48, no. 2, pp. 123–130, 2023.

[34] M. Pásztor, C. Czanik, and I. Bondár, “A single array approach for infrasound signal discrimination from quarry blasts via machine learning,” *Remote Sensing*, vol. 15, no. 6, p. 1657, 2023.

[35] Y. Wu, S. Zou, X. Pang, and X. Chen, “Experimental study on atmospheric infrasound signal recognition using SVM and ANN,” *Journal of Applied Acoustics*, vol. 39, no. 2, pp. 207–215, 2020.

[36] X. Leng, L. Feng, O. Ou, X. Du, D. Liu, and X. Tang, “Debris flow infrasound recognition method based on improved LeNet-5 network,” *Sustainability*, vol. 14, no. 23, p. 15925, 2022.

[37] F. Ham and N. S. Park, “A robust neural network classifier for infrasound events using multiple array data,” in *International Joint Conference on Neural Networks (IJCNN)*, pp. 2615–2619, 2003.

[38] K. E. Smith, M. L. Solomon, K. J. Bryan, A. O. Smith, and A. M. Peter, “Near-field infrasound classification of rocket launch signatures,” in *Chemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XIX*, vol. 10629, pp. 131–141, SPIE, 2018.

[39] L. Quanbo and L. Mei, “VMD and CNN-based classification model for infrasound signal,” *Archives of Acoustics*, vol. 48, no. 3, pp. 403–412, 2023.

[40] J. W. Bishop, P. S. Blom, J. Webster, W. Reichard-Flynn, and Y. Lin, “Deep learning categorization of infrasound array data,” *The Journal of the Acoustical Society of America*, vol. 152, no. 4, pp. 2434–2445, 2022.

[41] H. Li, X. Li, X. Tan, C. Niu, J. Liu, and T. Liu, “Infrasound event classification fusion model based on multiscale SE-CNN and BiLSTM,” *Applied Geophysics*, vol. 21, no. 3, pp. 579–592, 2024.

[42] X.-F. Tan, X.-H. Li, C. Niu, X.-N. Zeng, H.-R. Li, and T.-Y. Liu, “Classification method of infrasound events based on the MVIDA algorithm and MS-SE-ResNet,” *Applied Geophysics*, vol. 21, no. 2, pp. 123–135, 2024.

[43] T. Thüring, M. Schoch, A. Van Herwijnen, and J. Schweizer, “Robust snow avalanche detection using supervised machine learning with infrasonic sensor arrays,” *Cold Regions Science and Technology*, vol. 111, pp. 60–66, 2015.

[44] D. Liu, D. Tang, S. Zhang, X. Leng, K. Hu, and L. He, “Method for feature analysis and intelligent recognition of infrasound signals of soil landslides,” *Bulletin of Engineering Geology and the Environment*, vol. 80, no. 2, pp. 917–932, 2021.

[45] G. Van Rossum and F. L. Drake Jr., *Python Reference Manual*, Centrum voor Wiskunde en Informatica, Amsterdam, 1995.

[46] T. Inouye, K. Shinosaki, H. Sakamoto, S. Toi, S. Ukai, A. Iyama, Y. Katsuda, and M. Hirano, “Quantification of EEG irregularity by use of the entropy of the power spectrum,” *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 3, pp. 204–210, 1991.

[47] S. Prakash, S. Singh, and A. Mankar, “Bridging data gaps: A comparative study of different imputation methods for numeric datasets,” in *2024 International Conference on Data Science and Network Security (ICDSNS)*, pp. 1–7, IEEE, 2024.

[48] R. Bansal, N. Gaur, and S. N. Singh, “Outlier detection: Applications and techniques in data mining,” in *2016 6th International Conference on Cloud System and Big Data Engineering (Confluence)*, pp. 373–377, IEEE, 2016.

[49] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-based anomaly detection,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 1–39, 2012.

[50] O. Alghushairy, R. Alsini, T. Soule, and X. Ma, “A review of local outlier factor algorithms for outlier detection in big data streams,” *Big Data and Cognitive Computing*, vol. 5, no. 1, p. 1, 2020.

[51] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: Identifying density-based local outliers,” in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pp. 93–104, 2000.

[52] M. S. Abdalzaher, S. S. R. Moustafa, M. Abd-Elnaby, and M. Elwekeil, "Comparative performance assessments of machine-learning methods for artificial seismic sources discrimination," *IEEE Access*, vol. 9, pp. 65524–65535, 2021.

[53] M. S. Abdalzaher, S. S. R. Moustafa, H. E. Abdelhafiez, and W. Farid, "An optimized learning model augment analyst decisions for seismic source discrimination," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[54] J. Brownlee, *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*, Machine Learning Mastery, 2020.

[55] A. M. Musolf, E. R. Holzinger, J. D. Malley, and J. E. Bailey-Wilson, "What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics," *Human Genetics*, vol. 141, no. 9, pp. 1515–1528, 2022.

[56] A. Samantaray and S. R. Dash, "Feature selection techniques to predict the religion of a country from its flag," in *Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics*, vol. 1, pp. 191–201, Springer, 2020.

[57] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia Medica*, vol. 22, no. 3, pp. 276–282, 2012.

[58] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 405, pp. 442–451, 1975.

[59] C. Sammut and G. I. Webb, *Encyclopedia of Machine Learning and Data Mining*, Springer Publishing Company, Incorporated, 2017.