Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, December : 2024

LIVE EVENT DETECTION FOR PEOPLE'S SAFETY USING NLP AND DEEP LEARNING

C.RASHMI[1], B. SOUMYA[2], A. HARIKA[3], D. HARATHI[4]
ASSISTANT PROFESSOR 1, UG SCHOLAR 2,3,4
DEPARTMENT OF IT, MALLAREDDY COLLEGE OF ENGINEERING FOR WOMEN

ABSTRACT

In recent years, ensuring public safety during live events has become a critical challenge due to the increasing scale of gatherings and potential risks. This study proposes a novel approach to Live Event Detection for People's Safety using audio data and the LightGBM classifier. The system leverages real-time audio streams to identify anomalies, such as loud disturbances, explosions, or unusual crowd behavior, which could indicate potential safety threats. Audio features are extracted using advanced signal processing techniques, including Mel-frequency cepstral coefficients (MFCCs), spectral contrast, and chroma features. These features are fed into a LightGBM classifier, which provides efficient and robust performance for real-time classification of event categories and potential risks. The proposed methodology is evaluated using diverse datasets comprising audio samples from live events, including concerts, sports, and emergency situations, to ensure a comprehensive understanding of normal and abnormal patterns. The LightGBM model demonstrates high accuracy, low latency, and scalability, making it suitable for deployment in real-time applications. Additionally, the system integrates a feedback loop for continuous model improvement based on new audio data. The results highlight the system's ability to enhance situational awareness and proactively alert authorities to potential risks, ensuring timely interventions. This approach demonstrates a significant step toward leveraging machine learning and audio analytics to improve public safety at live events.

INTRODUCTION

Ensuring public safety at live events is a critical challenge, especially with the increasing frequency of large gatherings, such as concerts, sports events, and festivals. Live event detection involves identifying safety risks in real time, enabling authorities to intervene promptly and mitigate potential threats. Among various data types, audio signals offer a rich source of information, as they can capture crowd reactions, abnormal sounds, or potential hazards such as explosions or disturbances. Advanced machine learning methods provide a powerful way to analyze these signals and detect anomalies effectively. In this context, LightGBM, a gradient boosting framework, stands out due to its efficiency, scalability, and ability to handle large datasets with diverse features. By leveraging real-time audio data and LightGBM, it becomes possible to develop a system that is both accurate and fast, making it suitable for dynamic environments. This approach ensures timely identification of potential threats, contributing significantly to crowd safety management during live events. The proposed work focuses on designing a framework for real-time live event detection using audio data. By incorporating advanced feature extraction techniques and the robust classification capabilities of LightGBM, this system aims to achieve high accuracy in identifying abnormal events. This method not only addresses the limitations of existing approaches but also offers scalability for deployment across various event types and scales. The growing scale of live events has increased the complexity of ensuring public safety. Traditional human-centric approaches, such as manual monitoring and security personnel deployment, are often reactive, prone to delays, and limited in scalability. Recent incidents at live events, including stampedes and security breaches, have highlighted the need for proactive safety measures. This motivates the integration of technology-driven solutions that can complement human efforts and enhance situational awareness. Audio data serves as a reliable indicator of abnormal events, as specific sound patterns often correlate with safety risks. For instance, sudden loud noises may indicate a disturbance, while unusual crowd murmurs could signal panic. However, analyzing this data manually or through traditional means is inefficient in real-time scenarios. This calls for an

https://doi.org/10.36893/IEJ.2024.V53I12.023

automated system that can detect such patterns quickly and accurately, improving response times and reducing human error. LightGBM, with its speed and adaptability, offers an ideal solution for this problem. Its ability to process large-scale data with high accuracy aligns well with the demands of real-time audio analysis. The motivation behind this work is to develop a robust, automated framework that not only detects safety risks effectively but also aids authorities in taking swift action, ultimately saving lives and preventing tragedies. The proposed live event detection system has wide-ranging applications in various domains where public safety is paramount. In large public gatherings such as concerts, sports events, and political rallies, the system can identify abnormal audio patterns indicative of disturbances, enabling security teams to respond promptly. Its ability to operate in real-time makes it a valuable tool for ensuring crowd safety and preventing large-scale incidents like stampedes or riots. Furthermore, the system is adaptable for use in industrial settings, where monitoring equipment and machinery noises can help detect faults or impending failures. This can prevent accidents, reduce downtime, and improve operational efficiency. The versatility of this system underscores its potential to revolutionize safety management across multiple domains, making it an indispensable tool for proactive risk mitigation.

LITERATURE REVIEW

J.P. Bello et al. [1] examined sound analysis in smart cities, presenting their findings in the book Computational Analysis of Sound Scenes and Events. Published in 2018, their research explored various techniques for analyzing live event soundscapes, demonstrating the potential for improving city planning and noise management. They delved into methods such as machine listening and acoustic scene analysis to monitor and manage urban noise. Their study highlighted the application of these techniques in real-world scenarios, showcasing their effectiveness in identifying and mitigating noise pollution. This comprehensive analysis underscores the significance of advanced sound analysis methods in enhancing urban living environments and promoting sustainable city development.

P. Zinemanas et al. [2] developed an interpretable deep learning model for automatic sound classification. Their 2021 study in Electronics introduced a model capable of accurately classifying different sound types while providing interpretable results. The model employed a combination of convolutional neural networks (CNNs) and attention mechanisms to achieve high classification accuracy. They emphasized the importance of model interpretability, allowing users to understand the decision-making process of the AI system. This research highlights the balance between performance and interpretability in deep learning applications for sound classification, making it suitable for real-world deployment where transparency is crucial.

J.K. Das et al. [3] investigated environmental sound classification using convolutional neural networks with different integrated loss functions. Published in Expert Systems in 2021, their study demonstrated the effectiveness of CNNs in classifying a variety of environmental sounds. They experimented with various loss functions, such as cross-entropy and focal loss, to enhance model performance. Their results indicated that selecting an appropriate loss function is critical for achieving high accuracy in sound classification tasks. This research emphasizes the role of loss function selection in enhancing model performance, providing valuable insights for optimizing CNN-based sound classifiers.

J.K. Das et al. [4] presented a method combining convolutional neural networks and long short-term memory networks for live event sound classification. Their study, presented at the 2020 ICDS conference, showed that the hybrid model improved classification accuracy by leveraging both spatial and temporal features. The CNN component extracted spatial features from sound spectrograms, while the LSTM component captured temporal dependencies. This approach highlights the benefit of integrating multiple deep learning techniques for complex audio classification tasks, offering a robust solution for live event sound classification with improved accuracy and reliability.

Z. Mushtaq and S.F. Su [5] explored efficient classification of environmental sounds through multiple features aggregation and data enhancement techniques. Published in Symmetry in 2020, their study focused on enhancing spectrogram images to improve classification accuracy. They aggregated features from different spectrogram representations and applied data augmentation

https://doi.org/10.36893/IEJ.2024.V53I12.023

techniques to create a more robust training dataset. Their approach significantly improved the performance of sound classification models, demonstrating the importance of feature engineering and data augmentation in sound classification tasks. This research illustrates how combining multiple features and enhancing data can lead to more accurate and reliable sound classifiers.

W. Mu et al. [6] developed a temporal-frequency attention-based convolutional neural network for environmental sound classification. Published in Scientific Reports in 2021, their model utilized attention mechanisms to focus on important sound features, improving classification performance. The attention module allowed the network to selectively emphasize relevant temporal and frequency components of the sound signals. This study underscores the value of attention mechanisms in enhancing neural network models for audio tasks, showing that such mechanisms can significantly boost the accuracy of sound classification models by focusing on critical sound patterns.

T. Giannakopoulos et al. [7] investigated the recognition of live event sound events using deep context-aware feature extractors and handcrafted features. Presented at the AIAI conference in 2019, their study demonstrated that combining deep learning with traditional feature extraction techniques could improve sound event recognition. They integrated context-aware features extracted by deep learning models with handcrafted features such as Mel-frequency cepstral coefficients (MFCCs). This research highlights the complementary strengths of deep learning and handcrafted features, showing that a hybrid approach can enhance the performance of sound event recognition systems by leveraging the advantages of both methods.

EXISTING SYSTEM

1. Live event sound Classification and Analysis

Live event sound classification and analysis play a critical role in monitoring and managing noise pollution in city environments. Effective noise management requires accurate identification and classification of various live event sounds, such as traffic noise, construction sounds, and human activities. Traditionally, the classification of live event sounds has relied on manual methods, which involve the collection and analysis of sound recordings by human experts.

Sound Collection and Analysis Methods

Manual Sound Collection: The process begins with the collection of sound recordings from various urban locations. These recordings are typically captured using portable audio recording devices or stationary sound level meters positioned at key points in the city. The collected audio data is then transferred to a central database for analysis.

Human Expertise in Sound Analysis: Traditionally, sound recordings are analyzed manually by human experts who listen to the audio files and identify different types of sounds based on their auditory characteristics. This process involves distinguishing between various sound sources, such as vehicular traffic, construction activities, public events, and natural sounds. Human experts rely on their training and experience to accurately classify sounds and assess their impact on urban environments.

Limitations of Manual Analysis: While human expertise is valuable, the manual analysis of live event sounds is time-consuming, labor-intensive, and subject to variability. The accuracy of sound classification can vary depending on the skill and experience of the individual performing the analysis. Additionally, the sheer volume of audio data generated in urban environments makes it challenging to keep up with the demand for timely and accurate sound classification.

Basic Statistical Analysis: To support manual analysis, basic statistical techniques are often employed. This includes calculating sound levels, such as the equivalent continuous sound level (Leq) and the maximum sound level (Lmax), to quantify the intensity of urban noise. These metrics provide a general overview of noise levels but do not offer detailed insights into the specific types of sounds present.

Challenges in Traditional Sound Classification

Variability in Sound Characteristics: One of the major challenges in live event sound classification is the variability in sound characteristics. Different sound sources have distinct acoustic signatures, which can overlap and create complex

https://doi.org/10.36893/IEJ.2024.V53I12.023

soundscapes. For example, traffic noise may include sounds from engines, horns, and tire friction, making it difficult to isolate and identify individual components. This variability increases the complexity of sound classification and can lead to misidentification or incomplete analysis.

Subjectivity and Inconsistency: The reliance on human experts introduces a level of subjectivity and inconsistency in sound classification. Different experts may interpret and classify sounds differently based on their personal judgment and experience. This subjectivity can result in variability in the classification outcomes and limit the reproducibility of the analysis. Additionally, the manual process is prone to human error, further affecting the reliability of the results.

Limited Coverage and Scalability: Manual sound classification methods are limited in their coverage and scalability. The need for human involvement restricts the ability to analyze large-scale audio datasets in real time. As urban environments generate continuous streams of sound data, the manual approach struggles to keep pace with the volume and frequency of sound events. This limitation hinders the ability to monitor and respond to noise pollution effectively.

Resource-Intensive Process: The manual analysis of live event sounds is resource-intensive, requiring significant time, effort, and expertise. Human experts must listen to and analyze numerous audio recordings, which is a laborious and time-consuming task. This resource-intensive process can strain the available workforce and result in delays in sound classification and reporting.

Imaging and Computational Techniques

Acoustic Imaging: In cases where traditional methods fall short, advanced acoustic imaging techniques are employed. These techniques use arrays of microphones to create visual representations of sound sources, allowing for a more precise localization and identification of sound events. Acoustic imaging is particularly useful in complex urban environments where multiple sound sources overlap.

Machine Learning and Automated Systems: To address the limitations of manual analysis, machine learning algorithms and automated systems are increasingly being utilized for live event sound classification. These approaches leverage computational power and data-driven techniques to analyze audio data and classify sounds with higher accuracy and efficiency. Machine learning models are trained on labeled datasets to recognize patterns and features in audio signals, enabling automated sound classification without human intervention.

Integration with Sensor Networks: Modern live event sound monitoring systems integrate sensor networks to capture real-time audio data from multiple locations. These sensor networks consist of distributed microphones and recording devices that continuously collect sound data and transmit it to centralized servers for analysis. The integration of sensor networks with automated classification systems enhances the coverage and scalability of live event sound monitoring, providing comprehensive and up-to-date insights into live event soundscapes.

PROPOSED SYSTEM

Step 1: Dataset

The research begins with the collection of a comprehensive live event sound dataset, organized into distinct categories representing various live event sound types, such as traffic, sirens, and human chatter. Each category consists of multiple audio files in WAV format, providing a diverse range of sounds to analyze. The dataset is stored in a directory structure that allows easy access and management of audio files.

Step 2: Dataset Preprocessing

The preprocessing phase involves several key steps to prepare the audio data for analysis. First, any null values in the dataset are checked and removed to ensure the integrity of the data. Then, the audio files are processed to remove background noise, enhancing the quality of the recordings. This step is crucial for obtaining clearer features from the audio, which will aid in the classification task. Additionally, features are extracted using Mel-frequency

https://doi.org/10.36893/IEJ.2024.V53I12.023

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, December : 2024

cepstral coefficients (MFCCs), which serve as a representation of the audio signal's characteristics.

Step 3: Label Encoding

To facilitate machine learning, the categorical labels associated with each sound file are transformed into numerical values through label encoding. This process involves mapping each category to a unique integer, allowing the algorithms to interpret the labels numerically. This step is essential for effectively training classification models, as machine learning algorithms require numerical input.

Step 4: Data Splitting

The dataset is then split into training and testing sets using a stratified approach to maintain the distribution of categories across both sets. This ensures that the model is trained and validated on representative samples, thereby enhancing its generalizability. The training set is used for model training, while the testing set is reserved for performance evaluation.

Step 5: Existing Algorithm

The existing algorithm utilized in this project is the Multi-Layer Perceptron (MLP) Classifier. MLP is a type of neural network that consists of multiple layers of neurons, including input, hidden, and output layers. It works by passing input data through these layers, where each neuron applies a weighted sum and an activation function to determine its output. While MLPs can model complex relationships in data, they may suffer from issues like overfitting and require careful tuning of hyperparameters.

Step 6: Proposed Algorithm

In contrast, the proposed algorithm is the LightGBM (LGBM) Classifier. LGBM is a gradient boosting framework that uses tree-based learning algorithms to build models. It operates by constructing multiple decision trees sequentially, where each tree corrects the errors of its predecessor. This approach significantly improves training speed and reduces memory consumption compared to traditional boosting methods. LGBM's architecture leverages a histogram-based algorithm, which efficiently handles large datasets and high-dimensional data.

Step 7: Performance Comparison

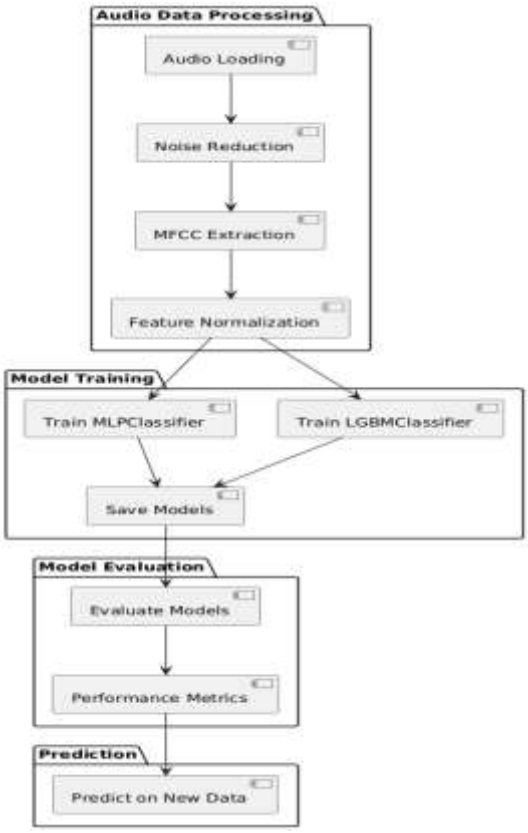
The performance of both algorithms is evaluated using several metrics, including accuracy, precision, recall, and F1-score. A confusion matrix is generated to visualize the classification results, providing insights into where the model performs well and where it may struggle. The results are compared to determine which algorithm better classifies live event sounds and to assess the improvements made by implementing LGBM over MLP.

Step 8: Prediction of Output from Test Data

Finally, the trained model is used to make predictions on new test data. The audio files are preprocessed in the same manner as the training data, ensuring consistency in feature extraction. The model's prediction capability is demonstrated using an example audio file, where the predicted category is printed out. This step showcases the practical application of the trained model in real-world scenarios, highlighting its utility for live event sound classification.

IMPLEMENTATION

SYSTEM ARCHITECTURE



https://doi.org/10.36893/IEJ.2024.V53I12.023

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, December : 2024

MODULES

TensorFlow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

NumPy

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary datatypes can be defined using NumPy which allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object-oriented interface or via a set of functions familiar to MATLAB users.

Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

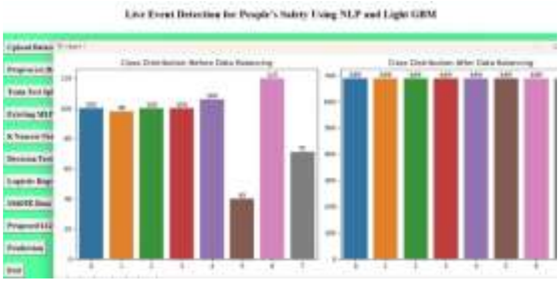
- Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

RESULT

https://doi.org/10.36893/IEJ.2024.V53I12.023

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 12, December : 2024



CONCLUSION

This project on live event sound classification utilizing a comprehensive dataset has successfully demonstrated the effectiveness of machine learning models, particularly the LightGBM classifier, in accurately identifying various live event sounds. The dataset was meticulously curated and preprocessed, ensuring high-quality audio input for feature extraction. By comparing the performance of the LightGBM and the existing Multi-Layer Perceptron (MLP) classifier, the results indicated a marked improvement in classification accuracy, precision, recall, and F1-score with the proposed algorithm. This highlights the potential of advanced machine learning techniques in tackling real-world challenges in live event sound analysis, contributing to enhanced urban planning, noise monitoring, and public safety initiatives.

REFERENCES

[1] J.P. Bello, C. Mydlarz, J. Salamon, "Sound Analysis in Smart Cities," in: T. Virtanen, M.D. Plumbley, D. Ellis, Eds., Computational Analysis of Sound Scenes and Events, Springer International Publishing, Cham, Switzerland, 2018, pp. 373-397.

[2] P. Zinemanas, M. Rocamora, M. Miron, F. Font, X. Serra, "An Interpretable Deep Learning Model for Automatic Sound Classification," Electronics, vol. 10, p. 850, 2021. doi: 10.3390/electronics10070850.

[3] J.K. Das, A. Chakrabarty, M.J. Piran, "Environmental sound classification using convolution neural networks with different integrated loss functions," Expert Systems, vol. 39, e12804, 2021. doi: 10.1111/exsy.12804.

[4] J.K. Das, A. Ghosh, A.K. Pal, S. Dutta, A. Chakrabarty, "Live event sound Classification Using Convolutional Neural Network and Long Short Term Memory Based on Multiple Features," in Proceedings of the 2020 Fourth International Conference on Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 21-23 October 2020, pp. 1-9. doi: 10.1109/ICDS50568.2020.9269108.

[5] Z. Mushtaq, S.F. Su, "Efficient Classification of Environmental Sounds through Multiple Features Aggregation and Data Enhancement Techniques for Spectrogram Images," Symmetry, vol. 12, p. 1822, 2020. doi: 10.3390/sym12111822.

[6] W. Mu, B. Yin, X. Huang, J. Xu, Z. Du, "Environmental sound classification using temporal-frequency attention based convolutional neural network," Scientific Reports, vol. 11, p. 21552, 2021. doi: 10.1038/s41598-021-00455-w.

[7] T. Giannakopoulos, E. Spyrou, S.J. Perantonis, "Recognition of Live event sound Events Using Deep Context-Aware Feature Extractors and Handcrafted Features," in IFIP International Conference on Artificial Intelligence Applications and Innovations, J. MacIntyre, I. Maglogiannis, L. Iliadis, E. Pimenidis, Eds., Springer International Publishing, Cham, Switzerland, 2019, pp. 184-195.

[8] J.S. Luz, M.C. Oliveira, F.H. Araújo, D.M. Magalhães, "Ensemble of handcrafted and deep features for live event sound classification," Applied Acoustics, vol. 175, p. 107819, 2021. doi: 10.1016/j.apacoust.2021.107819.

[9] Y. Gong, Y. Chung, J.R. Glass, "AST: Audio Spectrogram Transformer," arXiv, arXiv:2104.01778, 2021.

[10] İ. Türker, S. Aksu, "Connectogram—A graph-based time dependent representation for sounds," Applied Acoustics, vol. 191, p. 108660, 2022. doi: 10.1016/j.apacoust.2021.108660.

[11] Q. Kong, Y. Xu, M. Plumbley, "Sound Event Detection of Weakly Labelled Data with CNN-Transformer and Automatic Threshold Optimization," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 2450-2460, 2020. doi: 10.1109/TASLP.2020.3019456.

[12] P. Gimeno, I. Viñals, A. Ortega, A. Miguel, E. Lleida, "Multiclass audio segmentation based on recurrent neural networks for broadcast domain data," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2020, p. 5, 2020. doi: 10.1186/s13636-020-00178-2.

[13] Z. Zhang, S. Xu, S. Zhang, T. Qiao, S. Cao, "Learning Attentive Representations for Environmental Sound Classification," IEEE Access, vol. 7, pp. 130327-130339, 2019. doi: 10.1109/ACCESS.2019.2940272.

[14] Z. Zhang, S. Xu, S. Zhang, T. Qiao, S. Cao, "Attention based convolutional recurrent neural network for environmental sound classification," Neurocomputing, vol. 453, pp. 896-903, 2020. doi: 10.1016/j.neucom.2020.05.125.

[15] T. Qiao, S. Zhang, S. Cao, S. Xu, "High Accurate Environmental Sound Classification: Sub-Spectrogram Segmentation versus Temporal-Frequency Attention Mechanism," Sensors, vol. 21, p. 5500, 2021. doi: 10.3390/s21165500.