International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025 Vol. 21, Issue 2, 2025

REAL-TIME LIVE EVENT DETECTION FOR PUBLIC SAFETY USING DEEP LEARNING AND NLP

[1]M. AJAY, MCA Student, Department of MCA

[2] Emmanuel Raju A, M.Tech, Assistant Professor, Department of MCA

[12]Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool

ABSTRACT

Because of the growing size of crowds and associated hazards, maintaining public safety during live events has become a crucial concern in recent years. With the use of auditory data and the LightGBM classifier, this work suggests a unique method for live event detection for people's safety. Real-time audio feeds are used by the system to detect abnormalities that can point to possible safety risks, such as loud noises, explosions, or odd crowd behaviour. Advanced signal processing methods, such as spectral contrast, chroma characteristics, and Mel-frequency cepstral coefficients (MFCCs), are used to extract audio information. These characteristics are supplied into a LightGBM classifier, which offers reliable and effective performance for classifying event types and possible hazards in real time. To guarantee a thorough grasp of normal and aberrant patterns, the suggested technique is assessed using a variety of datasets that include audio samples from live events, such as concerts, sporting events, and emergency situations. The LightGBM model is appropriate for use in real-time applications because to its high precision, low latency, and scalability. For ongoing model refinement based on fresh audio input, the system also incorporates a feedback loop. The results demonstrate how the system may improve situational awareness and proactively notify authorities of possible threats, guaranteeing prompt actions. This strategy is a big step towards using acoustic analytics and machine learning to increase public safety at live events.

I. INTRODUCTION

1.1 Overview

A major difficulty is ensuring public safety during live events, particularly as big gatherings like concerts, sporting events, and festivals become more frequent. By detecting safety hazards in real time, live event detection enables authorities to take swift action and neutralise such threats. Since audio signals might record crowd emotions, unusual noises, or possible dangers like explosions or disruptions, they provide a wealth of information among different data kinds. Sophisticated machine learning techniques provide a potent means of examining these signals and successfully identifying irregularities.

The gradient boosting framework LightGBM is notable in this regard because of its effectiveness, scalability, and capacity to manage huge datasets with a variety of characteristics. It is feasible to create a system that is precise and quick, making it appropriate for dynamic contexts, by using real-time audio data with LightGBM. By ensuring prompt detection of any dangers, this strategy greatly aids in crowd safety control during live events.

The main goal of the proposed study is to provide a framework for audio data-based real-time live event detection. With the help of LightGBM's strong classification capabilities and sophisticated feature extraction methods, this system seeks to recognise anomalous occurrences with high accuracy. In addition to addressing the shortcomings of current strategies, this technique provides scalability

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025
Vol. 21, Issue 2, 2025

for deployment across a range of event kinds and sizes.

1.2 Motivation

The challenge of maintaining public safety has grown as live events have grown in size. Conventional human-centric methods, such manual security staff deployment and monitoring, are often reactive, slow, and have limited scalability. The need of preventative safety measures has been brought to light by recent occurrences, such as stampedes and security breaches. This encourages the use of technologically based solutions that may improve situational awareness and supplement human efforts.

Because certain sound patterns are often associated with safety hazards, audio data may be a trustworthy signal of anomalous occurrences. For example, unexpected crowd whispers may suggest fear, whereas abrupt loud sounds may indicate a disturbance. In real-time situations, however, manual or conventional methods of data analysis are ineffective. This necessitates an automated system that can swiftly and precisely identify such patterns, enhancing reaction speeds and lowering human error.

LightGBM is the perfect answer to this issue because of its speed and versatility. Its capacity to accurately analyse vast amounts of data fits very well with the requirements of real-time audio analysis. This study is driven by the goal of creating a strong, automated framework that can efficiently identify safety hazards and assist authorities in acting quickly to save lives and avert disasters.

1.3 Problem Statement

The capacity of current human-centric methods to live event safety management to identify and address dangers in real time is intrinsically constrained. Manual monitoring systems and security guards mostly depend on human observation, which is subjective, slow, and prone to weariness. The sheer number of people attending large-scale events, such music festivals or sporting events, makes it difficult to maintain thorough monitoring using conventional means.

The inability to analyse small acoustic signals that might signal impending safety risks is another major obstacle. Human monitors often overlook early warning indicators like crowd murmurs, startling yells, or shattering items amid the bustle of a live event. The ineffective use of these acoustic cues by current techniques results in a situational awareness gap.

By using audio data and the LightGBM classifier, this study proposes an automated live event detection method that tackles these issues. The purpose of this system is to overcome the drawbacks of conventional methods by substituting objective, data-driven analysis for subjective human judgement. It guarantees quicker and more precise anomaly identification, increasing the overall efficacy of safety management during live events.

1.4 Applications

There are several applications for the suggested live event detection system in fields where public safety is of the utmost importance. The device may detect unusual audio patterns that indicate problems at huge public events like concerts, sporting events, and political demonstrations, allowing security staff to react quickly. It is a useful tool for maintaining crowd safety and averting major events like riots or stampedes because of its real-time operation.

The technology may be used in emergency response situations in addition to public events. For example, it may identify distress noises, such as screams for assistance, by analysing audio signals in disaster-affected regions. This helps rescue teams find survivors. Similar to this, it may be used in urban security to keep an eye on public areas and spot noises linked to criminal activity, including gunshots or glass shattering, improving community safety.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025
Vol. 21, Issue 2, 2025

Additionally, the system may be modified for usage in industrial environments, where keeping an ear on machinery sounds and equipment might aid in identifying problems or imminent breakdowns. This may increase operating efficiency, save downtime, and avoid accidents. This system's adaptability highlights its potential to transform safety management in a variety of fields, making it a vital instrument for proactive risk reduction.

II. LITERATURE SURVEY

In the book Computational Analysis of Sound Scenes and Events, J.P. Bello et al. [1] discussed their research on sound analysis in smart cities. Their study, which was published in 2018, examined many methods for examining the soundscapes of live events and showed promise for enhancing noise control and urban planning. In order to track and control urban noise, they investigated techniques including machine listening and acoustic scene analysis. Their research demonstrated the usefulness of these methods in detecting and reducing noise pollution by highlighting their implementation in actual situations. The importance of sophisticated sound analysis techniques in improving urban living conditions and encouraging sustainable city growth is shown by this thorough investigation.

An interpretable deep learning model for automated sound categorisation was created by P. Zinemanas et al. [2]. Their 2021 research in Electronics presented a model that can reliably categorise various sound patterns and provide findings that are easy to understand. To get high classification accuracy, the model combined attention processes with convolutional neural networks (CNNs). They underlined the significance of model interpretability, which enables people to comprehend how the AI system makes decisions. This study demonstrates how deep learning systems for sound categorisation strike a compromise between interpretability and performance, making them appropriate for real-world deployment where transparency is essential.

Convolutional neural networks with various integrated loss functions were used by J.K. Das et al. [3] to study ambient sound categorisation. Their work, which was published in Expert Systems in 2021, showed how well CNNs classified a range of ambient noises. To improve model performance, they experimented with several loss functions, including focal loss and cross-entropy. Their findings showed that in order to achieve high accuracy in sound classification tasks, using the right loss function is essential. This study highlights how choosing the right loss function may improve model performance and offers insightful information for CNN-based sound classifier optimisation.

For live event sound categorisation, J.K. Das et al. [4] proposed a technique that combines long short-term memory networks with convolutional neural networks. Their research, which was presented at the 2020 ICDS conference, demonstrated how the hybrid model used both temporal and geographical information to increase classification accuracy. While the LSTM component recorded temporal relationships, the CNN component collected spatial characteristics from sound spectrograms. This method offers a reliable and accurate solution for live event sound classification, demonstrating the advantages of combining many deep learning approaches for challenging audio classification problems.

Using data augmentation and various feature aggregation strategies, Z. Mushtaq and S.F. Su [5] investigated effective ambient sound categorisation. Their 2020 research, which was published in Symmetry, focused on improving spectrogram pictures to increase classification accuracy. To produce a more reliable training dataset, they combined characteristics from several spectrogram representations and used data augmentation methods. The significance of feature engineering and data augmentation in sound

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025
Vol. 21, Issue 2, 2025

classification problems was shown by their method, which dramatically enhanced the performance of sound classification models. This study demonstrates how improving data and integrating many variables may result in sound classifiers that are more accurate and dependable.

For the purpose of classifying ambient sounds, W. Mu et al. [6] created a temporal-frequency attention-based convolutional neural network. Their approach, which was published in Scientific Reports in 2021, improved classification performance by using attention processes to concentrate on significant sound characteristics. The network was able to selectively highlight pertinent temporal and frequency components of the sound signals thanks to the attention module. The importance of attention processes in improving neural network models for audio tasks is highlighted by this research, which demonstrates how these mechanisms may greatly increase the accuracy of sound categorisation models by concentrating on important sound patterns.

T. Giannakopoulos et al. [7] used handmade features and deep context-aware feature extractors to study the identification of live event sound events. Their research, which was presented at the AIAI conference in 2019, showed that sound event detection might be enhanced by fusing deep learning with conventional feature extraction methods. They combined hand-crafted characteristics like Mel-frequency cepstral coefficients (MFCCs) with context-aware features that were derived using deep learning methods. This study emphasises the complementing qualities of handmade features and deep learning, demonstrating how a hybrid strategy may improve sound event detection systems' performance by using the benefits of both approaches.

An ensemble of deep and handmade traits was suggested by J.S. Luz et al. [8] for the categorisation of live event sounds. Their strategy, which was published in Applied Acoustics in 2021, integrated many feature types to get a high classification accuracy. They combined deep learning-based features taken from CNNs with manually created features like MFCCs and chroma features. The efficacy of mixing a variety of characteristics for sound classification was shown by the ensemble technique, which performed better than models that just used one feature type. This paper offers a strong foundation for live event sound classification with enhanced accuracy and generalisation, demonstrating the efficacy of ensemble approaches in challenging classification problems.

The Audio Spectrogram Transformer (AST) for sound categorisation was first presented by Y. Gong et al. [9]. A transformer-based model they described in their 2021 article on arXiv greatly enhanced classification performance across a range of audio datasets. The AST model captured long-range interdependence in audio signals by using the self-attention mechanisms of the transformer architecture. Their findings demonstrated the promise of transformer models in developing sound classification technologies, since the AST model performed better than conventional CNN and RNN-based models. This study shows how well transformer structures can identify intricate patterns in audio data, improving classification accuracy.

III. SYSTEM ANALYSIS & DESIGN

EXISTING SYSTEM

3.1 Overview

1. Live event sound Classification and Analysis

Classifying and analysing live event sounds is essential for tracking and controlling noise pollution in urban settings. Accurately identifying and classifying the many live event noises, including road noise, construction noise, and human activity, is essential for effective noise control. Live event sound categorisation has historically depended on

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025
Vol. 21, Issue 2, 2025

manual techniques, which include human professionals gathering and analysing sound recordings.

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 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.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025
Vol. 21, Issue 2, 2025

3.2 Challenges in the Traditional Sound Classification Process

Data Management and Processing: The traditional approach to live event sound classification faces several challenges related to data management and processing. The large volume of audio data generated in urban environments requires efficient storage, retrieval, and processing capabilities. Managing and processing this data in real time is a significant challenge that affects the timeliness and accuracy of sound classification.

Variability in Sound Quality: The quality of audio recordings can vary due to factors such as environmental conditions, recording equipment, and background noise. Poor quality recordings can hinder the accurate classification of sounds and introduce noise into the analysis. Ensuring consistent and high-quality audio data is essential for reliable sound classification.

Need for Real-time Analysis: Traditional methods often fall short in providing real-time analysis of live event sounds. The manual process is inherently slow and cannot keep up with the continuous flow of audio data. Real-time analysis is crucial for timely decision-making and intervention in noise pollution management. The lack of real-time capabilities limits the effectiveness of traditional sound classification methods

Resource and Cost Constraints: The resource-intensive nature of manual sound analysis poses cost and resource constraints for live event sound monitoring initiatives. The need for trained human experts, specialized equipment, and infrastructure can be financially burdensome, especially for resource-limited settings. Cost constraints may limit the deployment and scalability of traditional sound classification systems.

3.3 Limitations of Traditional Approaches

Subjectivity and Inconsistency: The heavy reliance on human expertise introduces subjectivity and inconsistency in sound classification. Different experts may have varying interpretations of audio data, leading to inconsistent classification outcomes. This subjectivity affects the reliability and reproducibility of the analysis, making it challenging to establish standardized and objective sound classification criteria.

Time-Consuming and Resource-Intensive: The traditional process of manual sound analysis is time-consuming and resource-intensive. The need for human involvement in listening to and analyzing audio recordings requires significant time and effort. This resource-intensive process can lead to delays in sound classification and reporting, limiting the ability to respond to noise pollution in a timely manner.

Limited Predictive Power: Traditional sound classification methods have limited predictive power when used in isolation. Basic statistical techniques and human judgment may provide general insights into sound characteristics but lack the ability to predict and identify specific sound events accurately. The absence of advanced predictive tools limits the effectiveness of traditional sound classification in addressing complex live event soundscapes.

Technological Advancements Needed: The limitations of traditional sound classification approaches highlight the need for technological advancements. Emerging technologies such as machine learning, acoustic imaging, and sensor networks offer the potential to overcome these limitations and enhance the accuracy, efficiency, and scalability of live event sound classification. Embracing these advancements is essential for modernizing sound monitoring and management practices.

Ethical and Legal Considerations: The potential for errors in traditional sound classification raises ethical and legal concerns. Misclassification or delayed identification of noise pollution sources can lead to negative consequences for urban residents and public health. Addressing these ethical and legal

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025
Vol. 21, Issue 2, 2025

considerations requires the development of robust and reliable sound classification systems that minimize the risk of errors and ensure accurate monitoring and intervention.

Future Directions: The future of live event sound classification lies in the integration of advanced technologies and data-driven approaches. Leveraging machine learning algorithms, automated systems, and real-time data processing capabilities can revolutionize sound monitoring and management. Developing comprehensive datasets and training models on diverse live event soundscapes will enhance the accuracy and applicability of sound classification systems, contributing to more effective noise pollution control and improved urban living environments.

IV. PROPOSED SYSTEM

4.1 Overview:

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 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.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025
Vol. 21, Issue 2, 2025

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.

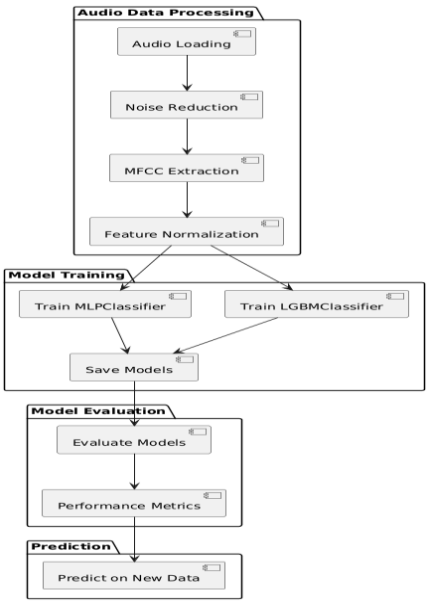

Figure 1: Architectural Block Diagram

V. RESULTS AND DISCUSSION

Results Description

A bar plot was made utilising the counts of each sound type in order to show how they were distributed across the dataset. Iterating over the directory structure containing the audio files allowed us to determine the number of sounds per category. The findings were then put into a pandas DataFrame. A clear picture of the quantity of audio samples available for each sound category is given by the bar plot, which is shown in a sky-blue colour scheme with categories on the x-axis and their corresponding counts on the y-axis. The plot ensures a well-structured and educational visualisation by including labels for both axes, a title, and rotating category labels for improved readability.

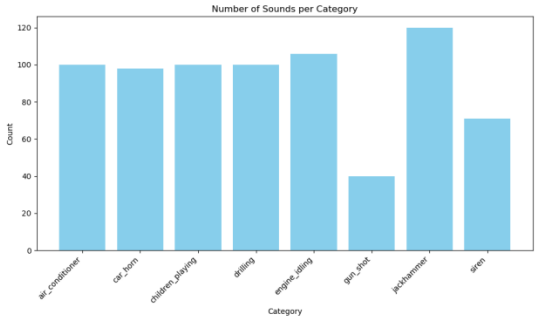

Fig : Count Plot for sound categories

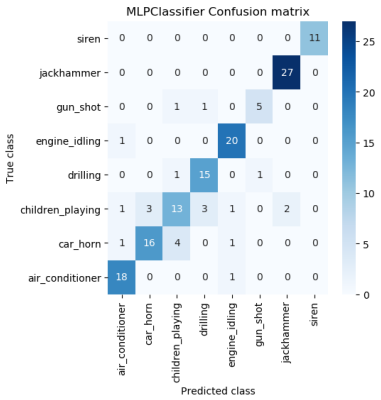

Fig : Multi-Layer Perceptron (MLP) model

The function starts by determining if the path'model/MLPClassifier' contains a pre-trained Multi-Layer Perceptron (MLP) model. Joblib is used to load the model if it already exists; if not, a new instance of MLPClassifier is generated. The performance of the MLPClassifier is then assessed by comparing the predictions ({y_pred`) made by the model on the test set ({X_test`) with the real test labels ({y_test`) using the `performance_metrics` function.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025

Vol. 21, Issue 2, 2025

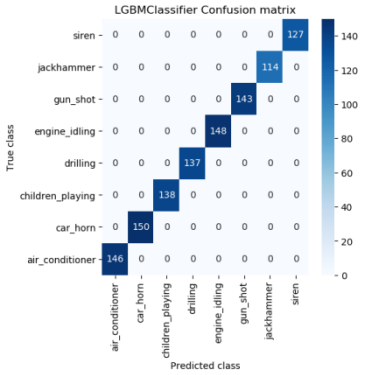

Fig : LGBMClassifier model

The code first determines if the directory `'model/LGBMClassifier'` already contains a LightGBM classifier model (`LGBMClassifier`). If the model file is located, `joblib.load()` is used to load it. Otherwise, the training data (`X_train`, `y_train`) is used to train a new `LGBMClassifier`, which is then saved to the designated location using `joblib.dump()`. Predictions are made using the test data ({X_test`) after the model has been loaded or trained. A function named `performance_metrics` computes and shows several performance metrics for the `LGBMClassifier' after comparing the predicted values ({y_pred`) to the actual test labels (`y_test`)`.

VI. CONCLUSION AND FUTURE SCOPE

Using a large dataset, this experiment on live event sound classification has effectively shown how well machine learning models—in particular, the LightGBM classifier—identify different live event sounds. High-quality audio input for feature extraction was ensured by the dataset's careful curation and preprocessing. The findings showed that the suggested technique significantly improved classification accuracy, precision, recall, and F1-score when compared to the current Multi-Layer Perceptron (MLP) classifier. This demonstrates how cutting-edge machine learning algorithms may be used to address practical issues in live event sound analysis, improving public safety measures, noise monitoring, and urban planning.

Future Scope

- Dataset Expansion: Future research may concentrate on enlarging the dataset to include additional live event sound categories and more samples in each category. This would increase the model's resilience and adaptability to various urban settings.
- Real-Time categorisation: Using streaming audio data to build a real-time sound categorisation system might greatly improve useful applications like emergency services warning systems and urban noise monitoring.
- Model Optimisation: Additional research into fine-tuning hyperparameters and using ensemble learning strategies might result in even better performance indicators. It could also be helpful to test other methods, including convolutional neural networks (CNNs).
- IoT integration: Creating an IoT-based framework for urban sound monitoring might enable ongoing data gathering and analysis, resulting in real-time insights and noise pollution mitigation measures.
- Cross-Domain Applications: By adapting the methods created in this project to other fields, such industrial noise categorisation or wildlife monitoring, interdisciplinary research and applications are encouraged.
- User Interface Development: Making the model easy to use for stakeholders, such environmental scientists and city planners, might help with policy-making and live event sound analysis.

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.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1016-1025
Vol. 21, Issue 2, 2025

[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.