IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

LIVE EVENT DETECTION FOR PUBLIC SAFETY USING SPARSE LSTM NETWORKS IN HAZARD MONITORING SYSTEMS

[1]Miss. Pooja Popat Barve, [2]Dr. Sachin Sukhadev Bere

[1]Student ME-II, [2]Associate Professor
[1]Department Of Computer Engineering,
[1]Dattakala Group Of Institutions, Faculty Of Engineering , Swami-Chincholi (Bhigwan), Tal-Daund, Dist-Pune-413130, India

Abstract : This study introduces a real-time audio classification system utilizing a Sparse LSTM (Long Short-Term Memory) network to detect potentially hazardous sounds. The system processes two types of audio data: pre-existing labeled datasets containing various audio classes and live-recorded audio. Both datasets undergo data cleaning and preprocessing to ensure high-quality input for model training and live prediction. The Sparse LSTM, optimized for reduced computational overhead, extracts temporal patterns from the audio data to enable accurate classification. During live prediction, the system analyzes the incoming audio and, if classified as "Danger," triggers an alert message. Otherwise, the process terminates without further action. This framework demonstrates efficient and timely audio-based hazard detection, making it suitable for safety monitoring and alert systems.

IndexTerms - ***Real-time audio classification; Sparse LSTM; Hazard detection; Temporal pattern extraction; Live prediction; Data preprocessing; Audio-based safety monitoring; Computational efficiency; Alert systems; Deep learning; Audio datasets; Audio signal processing; Sound classification; Safety alerts.***

I. INTRODUCTION

In dynamic and potentially hazardous environments, ensuring public safety relies on the rapid detection of threats. Traditional video surveillance systems, while valuable, are limited in their ability to operate effectively in low-visibility conditions, such as during night-time or in smoke-filled areas. In such scenarios, audio-based detection systems provide a complementary and often superior alternative, as they can identify critical sounds indicative of danger, such as alarms, screams, or explosions, irrespective of visual obstructions. This review paper introduces and analyses advancements in real-time audio classification systems, with a focus on the innovative use of Sparse Long Short-Term Memory (Sparse LSTM) networks. These systems are designed to detect hazardous sounds efficiently, offering significant advantages in resource-constrained settings, such as industrial sites, public spaces, and smart homes. Sparse LSTM models are particularly effective for temporal pattern recognition in sequential data, while their optimized architecture minimizes computational overhead, making them well-suited for real-time applications on edge devices and IOT platforms.

Figure No.1 Live Event Detection for Public Safety

INTRODUCTION

The paper highlights the importance of pre-processing techniques like noise reduction and feature extraction, including Mel-Frequency Cepstral Coefficients (MFCC), to enhance audio classification accuracy in diverse acoustic environments. Furthermore, the scalability and efficiency of Sparse LSTM-based systems make them an ideal solution for modern safety applications, where rapid detection and response to hazards are critical. This review examines the background, relevance, and evolution of audio classification systems, emphasizing their integration into intelligent safety monitoring frameworks. It also discusses key challenges, including noise interference and resource constraints, and explores potential enhancements such as hybrid CNN-LSTM architectures for improved spatial and temporal feature extraction. By evaluating current methodologies and proposing future directions, this paper aims to contribute to the development of robust, scalable, and efficient audio-based safety systems.

II LITERATURE REVIEW

Jain et al. [1] In the paper "A Symphony of Sentiments using Log-Melspectrogram Techniques for Emotional Classification," Jain et al. present a novel approach for emotion classification in audio signals by utilizing Log-Melspectrogram features. The study employs deep learning models to effectively classify emotions, demonstrating enhanced accuracy and robustness. The authors highlight the potential applications of this technique in real-time sentiment analysis and emotion-aware AI systems, particularly in interactive voice response (IVR) platforms.

Agarwal et al. [2] in their work titled "Emotional Resonance Unleashed by exploring Novel Audio Classification Techniques with Log-Melspectrogram Augmentation," explore the use of augmented Log-Melspectrogram features to improve audio emotion classification. The research focuses on enhancing model training by diversifying audio data through augmentation techniques, resulting in improved recognition of subtle emotional cues. This approach is particularly beneficial for applications in automated customer support and virtual assistants.

Ma et al. [3] In the paper "A CNN-LSTM Neural Network Model for Upper Limb Joint Angle Estimation Based on Mechanomyograph," Ma et al. introduce a hybrid model combining CNNs and LSTMs to estimate joint angles using mechanomyographic (MMG) signals. The proposed model effectively captures both spatial and temporal features, offering promising results for applications in prosthetic control and physical rehabilitation, where precise motion tracking is crucial.

Agarwal et al. [4] The study "Harmonizing Emotions: A Novel Approach to Audio Emotion Classification using Log-Melspectrogram with Augmentation" by Agarwal et al. presents an enhanced method for emotion recognition using augmented Log-Melspectrogram features. The authors demonstrate improved model performance by leveraging data augmentation, making it effective for use in personalized media experiences and adaptive learning environments.

He et al. [5] In their research titled "Text Sentiment Analysis of Douban Film Short Comments Based on BERT-CNN-BiLSTM-Att Model," He and Abisado propose a hybrid deep learning model integrating BERT, CNN, BiLSTM, and attention mechanisms for sentiment analysis. The study focuses on analyzing short film reviews, achieving high accuracy in capturing nuanced sentiments, which is particularly useful for content recommendation systems and social media analytics.

Jalili et al. [6] in the paper "Automatic Recognition System for Public Transport Robberies Based on Deep Learning," introduce a real-time surveillance system that uses deep learning models to detect and respond to criminal activities in public transportation. The system utilizes convolutional neural networks to analyze video feeds, aiming to enhance public safety through automated anomaly detection and rapid incident reporting.

Ricketts et al. [7] In the survey paper "A Scoping Literature Review of Natural Language Processing Application to Safety Occurrence Reports," Ricketts et al. review the use of NLP techniques in analyzing safety reports. The paper highlights the potential of NLP to automate the extraction of safety-critical information, thereby improving incident management in high-risk industries like aviation and healthcare.

Mredula et al. [8] in their review "A Review on the Trends in Event Detection by Analyzing Social Media Platforms' Data," discuss recent advancements in using social media data for event detection. The authors analyze the shift from traditional keyword-based methods to more sophisticated deep learning models, emphasizing the importance of real-time data analysis for applications in disaster management and social monitoring.

Singh et al. [9] in the paper "Methods to Detect an Event Using Artificial Intelligence and Machine Learning," Singh et al. provide a comprehensive overview of AI and ML techniques for event detection in sensor networks. The study covers various approaches, highlighting the advantages of deep learning in enhancing the detection capabilities of smart city infrastructure and industrial automation systems.

Gannina et al. [10] present their findings in "A New Approach to Road Incident Detection Leveraging Live Traffic Data: An Empirical Investigation," where they propose a machine learning-based system for detecting road incidents using real-time traffic data. This approach integrates sensor analytics to improve traffic flow management and reduce congestion by enabling faster incident responses.

Suma et al. [11] proposed an IoT-based system focused on women's safety, integrating screaming detection and video capturing functionalities. This system utilizes sensors and sound classification techniques to detect distress signals like screams. Once

detected, the device activates a camera to capture video evidence, which can be transmitted to emergency contacts and law enforcement for immediate action. The study emphasizes the real-time application of IoT in enhancing personal security, specifically for women, by leveraging sound analysis and IoT connectivity to improve response times in emergency situationsnterpretable Deep Learning Model for Automatic Sound Classification.

Zinemanas et al. [12] introduced a deep learning approach for sound classification, focusing on model interpretability. Their model uses convolutional neural networks (CNNs) to classify various sounds and provides insights into the decision-making process, making it transparent and understandable to users. This interpretability is crucial for applications requiring high reliability and trust, such as medical diagnostics or security systems, where understanding why a particular decision was made is as important as the decision itself. The authors demonstrated the effectiveness of their model using a diverse sound dataset, achieving high classification accuracy .

Malova et al. [13] explored the use of neural networks for recognizing emotions in verbal messages. This research highlights how speech signals can be analyzed to detect emotional states, which can be beneficial in human-computer interaction, virtual assistants, and mental health monitoring. The study employed recurrent neural networks (RNNs) to capture the sequential nature of speech, enhancing the system's ability to identify subtle emotional cues embedded in verbal communication .

Tsiktsiris et al. [14] presented a novel AI service combining both image and audio data for security applications in autonomous vehicles. The proposed system utilizes advanced deep learning models to analyze sounds and visual cues in real-time, enabling the vehicle to detect potential threats or irregularities. This multimodal approach enhances the robustness of security systems in autonomous driving, ensuring passenger safety by integrating contextual information from multiple sensory inputs .

Yang et al. [15] investigated a hybrid sound classification model combining multihead attention mechanisms and Support Vector Machines (SVM). The study leverages the attention mechanism to focus on the most relevant parts of the audio signal, improving the feature extraction process, while SVM is used for efficient classification. This approach demonstrated significant improvements in classification accuracy over traditional methods, especially in noisy environments, making it suitable for applications in surveillance, healthcare, and smart home systems .

2.1 Research Gap

The literature review highlights several research gaps in various domains, including audio emotion classification, joint angle estimation, surveillance systems, sentiment analysis, event detection, traffic incident detection, and Natural Language Processing (NLP) applications in safety reporting. Existing models struggle with real-world generalization, requiring cross-domain adaptation techniques. Joint angle estimation models using MMG signals lack real-time capabilities, requiring optimization in prosthetics and robotics. Surveillance systems face scalability issues, requiring distributed video analysis solutions. Multi modal data integration for sentiment analysis is also lacking. Lightweight AI models for edge devices are underdeveloped, requiring model compression for real-time deployment in resource-constrained environments. Ethical and privacy considerations are also lacking. Current models often focus on single-modal data, lacking robust multimodal integration for accuracy and contextual understanding. Limited research on real-time, edge-computing implementations for wearable safety devices and autonomous vehicles is also lacking. Addressing these gaps could improve the scalability, reliability, and efficiency of sound classification systems.

III PROBLEM STATEMENT

As automation and smart technologies continue to expand across various sectors, the need for real-time hazard detection systems becomes increasingly critical. In environments like industrial sites, public spaces, and smart homes, identifying potentially dangerous sounds—such as alarms, sirens, or breakages—can prevent accidents and save lives. However, current solutions often rely on manual monitoring or resource-intensive models that struggle to meet the demands of real-time processing, especially in systems with limited computational resources. The key challenge is to create a system that can accurately classify audio signals in real-time, detect hazardous sounds, and issue timely alerts without requiring heavy computational power. This project aims to tackle this challenge by developing a real-time audio classification system using Sparse Long Short-Term Memory (Sparse LSTM) networks. The objective is to design a system capable of efficiently processing audio data, identifying dangerous sounds, and triggering appropriate alerts in real-time, all while reducing computational load. The system must be optimized for deployment on embedded devices or IOT platforms, making it an ideal solution for safety monitoring in resource-constrained environments.

IV PROPOSED SYSTEM

The proposed system utilizes an LSTM-based model for real-time detection of dangerous sounds. It involves training on a data set of various audio classes, applying data pre-processing techniques like feature extraction (e.g., MFCCs), and leveraging LSTM's ability to capture temporal patterns. The system continuously monitors live audio, processes it to detect danger, and triggers alerts if a hazardous sound is classified. This approach ensures rapid response to potential threats, making it suitable for security and surveillance applications, with potential enhancements like CNN-LSTM models and edge computing for improved accuracy and performance. additionally, the system can be integrated with IoT devices for real-time audio monitoring in smart cities, enhancing public safety by detecting sounds like gunshots, explosions, or distress calls. By leveraging LSTM's sequential learning capabilities, it can distinguish between normal and abnormal sounds even in noisy environments, ensuring robust performance. Future enhancements may include the use of CNN-LSTM hybrid models to capture both spatial and temporal

features, improving classification accuracy. Implementing the model on edge computing platforms would reduce latency, enabling faster threat detection and response. This scalable solution can also be adapted for home security systems, automating emergency responses for incidents like break-ins or accidents.

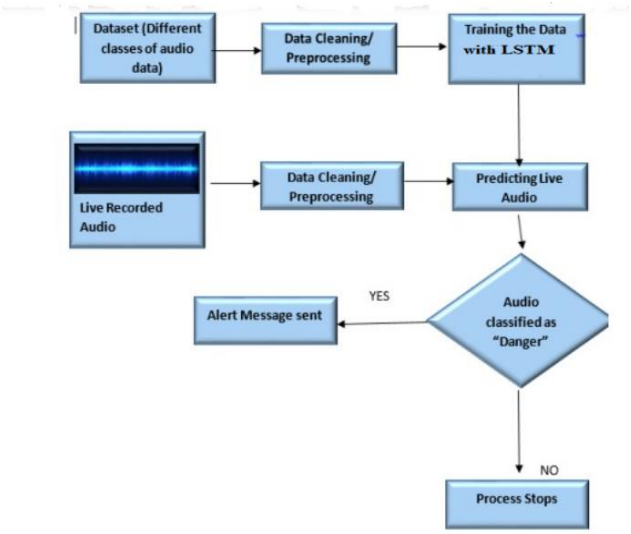

Figure No.2 System Architecture

The methodology for the audio-based danger detection system employs a robust methodology designed to identify hazardous sounds in real-time using machine learning, specifically an LSTM (Long Short-Term Memory) network. The process begins with the collection and preparation of a diverse audio data set, which includes both normal and dangerous sounds like alarms, screams, and explosions. These audio samples undergo data cleaning to remove background noise and are then transformed through feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCCs) and Spectrograms, which capture key audio characteristics essential for accurate classification. The pre-processed data is used to train the LSTM model, leveraging its capability to recognize temporal patterns in sequential data. During training, techniques like dropout regularization and hyper parameter tuning are applied to optimize the model's performance, which is evaluated using metrics such as accuracy and F1-score. After training, the system is deployed for real-time monitoring, where live audio feeds are captured, cleaned, and processed in the same way as the training data. The live audio features are fed into the trained LSTM model for classification; if a sound is identified as dangerous, the system promptly sends an automated alert via SMS, email, or other notification channels to inform security personnel or emergency responders. If no danger is detected, the system continues its monitoring loop, providing continuous surveillance.

The system is designed for flexibility, allowing it to run indefinitely or terminate based on specific conditions. Enhancements like combining CNN-LSTM architectures and deploying the system on edge devices can further improve accuracy and responsiveness, making it ideal for critical applications in public safety and security. to enhance scalability, the system can be integrated with cloud platforms to handle large volumes of audio data from multiple sources, enabling centralized monitoring for smart cities and large facilities. Incorporating adaptive noise cancellation algorithms further improves detection accuracy in dynamic environments with fluctuating background noise. Additionally, the system's modular design allows easy integration with existing surveillance infrastructure, making it a cost-effective solution for upgrading current security measures. Future iterations could explore transfer learning to adapt the model for specific environments, such as schools or hospitals, optimizing its sensitivity to context-specific sounds. The integration of real-time analytics dashboards provides stakeholders with actionable insights, enhancing decision-making in critical situations.

V ALGORITHM DEEP DENSE NETWORK

Sparse Long Short-Term Memory (Sparse LSTM) networks represent an enhancement in artificial intelligence, particularly in the field of recurrent neural networks (RNNs). Building upon the foundational LSTM architecture developed by Hochreiter and Schmidhuber in 1997, Sparse LSTMs address the limitations of traditional RNNs by not only effectively capturing long-term dependencies in sequential data but also optimizing computational efficiency. This efficiency makes Sparse LSTMs suitable for applications that involve large-scale time series data, such as speech recognition, language translation, and forecasting tasks like water demand prediction, where the ability to handle vast amounts of data without heavy computational costs is critical. The Sparse LSTM architecture shares the core components of traditional LSTMs, including the memory cell, input gate, forget gate, and output gate. However, Sparse LSTMs introduce a more efficient way of processing information by scarifying the connections within the network. This reduces the number of active parameters and computational complexity while maintaining the LSTM's ability to learn complex temporal patterns. The input gate controls the flow of information into the memory cell, deciding what new data should be stored based on the current input and previous memory state. The forget gate helps manage the retention of information, determining what should be discarded from the previous memory cell state, ensuring that only relevant data is kept. The output gate generates the output based on the updated memory, guiding the model's predictions.

In applications such as water demand forecasting, Sparse LSTMs prove particularly effective. The input layer of an Artificial Neural Network (ANN) collects data from multiple sources, including historical water usage, weather conditions, and demographic changes like population growth. As this data passes through the Sparse LSTM, intermediate layers process the

information more efficiently due to the reduced computational overhead, allowing the model to detect important patterns and make accurate predictions. The architecture of Sparse LSTMs makes them adept at learning both immediate and historical trends, which is essential for tasks like water demand forecasting, where future consumption patterns depend on both recent and long-term factors. Sparse Long Short-Term Memory networks offer a powerful and computationally efficient approach to time series analysis. By leveraging sparsity within the LSTM architecture, they can process complex datasets, capture long-term dependencies, and make accurate predictions in real-time applications such as environmental resource management and natural language processing. This makes Sparse LSTMs an attractive tool for advancing machine learning applications that require both high accuracy and low computational cost.

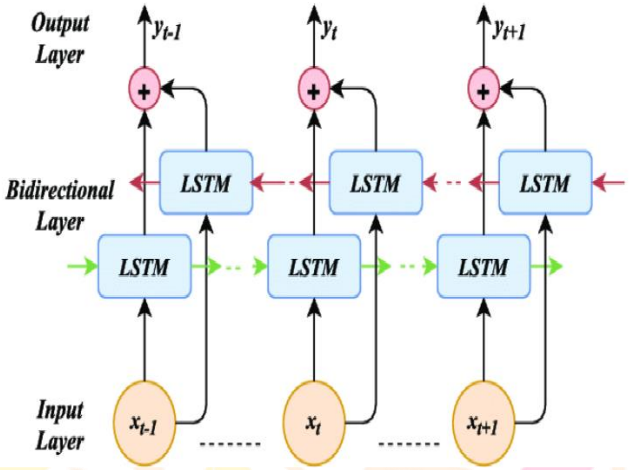

Fig 3: Architecture of LSTM

Building on their efficiency, Sparse LSTM networks are particularly advantageous for applications where both speed and accuracy are crucial. For example, in real-time monitoring systems, Sparse LSTMs can swiftly analyse sensor data to detect anomalies, such as in predictive maintenance for industrial equipment or fault detection in smart grids. The sparsity within the network also allows for energy-efficient implementations, making them suitable for edge devices like IOT sensors and wearables, where computational resources are limited. Moreover, Sparse LSTMs have shown promise in healthcare analytics, especially for analysing physiological time series data such as ECG signals, enabling early detection of irregularities without overburdening the system with computations. in financial market predictions, these networks can efficiently process vast streams of stock prices and economic indicators, identifying trends that could inform trading strategies. Furthermore, Sparse LSTMs enhance the scalability of natural language processing (NLP) tasks, such as real-time sentiment analysis, by reducing latency in processing large volumes of social media and customer feedback data. They are also being explored in environmental monitoring, like forecasting air quality and climate conditions, which involves processing complex time series data from multiple sources. This versatility across domains underscores the adaptability of Sparse LSTM models in scenarios that demand scalable, high-performance solutions. Their ability to handle diverse datasets with varying temporal dynamics positions them as a pivotal technology in the future of machine learning-driven automation and predictive analytics.

VI CONCLUSION

In conclusion, the real-time audio classification system utilizing Sparse Long Short-Term Memory (Sparse LSTM) networks showcases the transformative potential of advanced deep learning in enhancing safety monitoring across diverse environments. By balancing computational efficiency and classification accuracy, the system is optimized for resource-constrained devices like IOT platforms, ensuring seamless operation in real-time applications. Its ability to process both pre-labelled datasets and live audio streams enables adaptability to varied acoustic conditions, making it a reliable solution for industrial safety, smart home security, and public safety systems. The system's capability to effectively distinguish hazardous sounds and trigger immediate alerts positions it as a proactive and scalable tool for addressing safety challenges. Its cost-effectiveness further strengthens its feasibility for wide-scale deployment. This work highlights the advantages of Sparse LSTM networks for real-time hazard detection while laying a foundation for future improvements, such as integrating hybrid architectures or multi-modal sensory inputs. Ultimately, the project contributes significantly to creating smarter, safer, and more responsive environments.

REFERENCES

1. Jain, Eshika, et al. "A Symphony of Sentiments using Log-Melspectrogram Techniques for Emotional Classification." 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT). Vol. 1. IEEE, 2024.
2. Agarwal, Muskan, et al. "Emotional Resonance Unleashed by exploring Novel Audio Classification Techniques with Log-Melspectrogram Augmentation." 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0. IEEE, 2024.
3. Ma, Zhenjiu, et al. "A CNN-LSTM neural network model for upper limb joint angle estimation based on mechanomyograph." 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI). IEEE, 2024.

4. Agarwal, Muskan, et al. "Harmonizing Emotions: A Novel Approach to Audio Emotion Classification using Log-Melspectrogram with Augmentation." 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT). IEEE, 2024.
5. He, Aixiang, and Mideth Abisado. "Text Sentiment Analysis of Douban Film Short Comments Based on BERT-CNN-BiLSTM-Att Model." IEEE Access (2024).
6. Jalili, Laura, et al. "Automatic Recognition System for Public Transport Robberies Based on Deep Learning." Workshop on Engineering Applications. Cham: Springer Nature Switzerland, 2024.
7. Ricketts, Jon, et al. "A scoping literature review of natural language processing application to safety occurrence reports." Safety 9.2 (2023): 22.
8. Mredula, Motahara Sabah, Noyon Dey, Md Sazzadur Rahman, Imtiaz Mahmud, and You-Ze Cho. "A review on the trends in event detection by analyzing social media platforms' data." Sensors 22, no. 12 (2022): 4531.
9. Singh, Arjun, et al. "Methods to detect an event using artificial intelligence and Machine Learning." 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM). IEEE, 2022.
10. Gannina, Aswin Ram Kumar, et al. "A New Approach to Road Incident Detection Leveraging Live Traffic Data: An Empirical Investigation." Procedia Computer Science 235 (2024): 2288-2296.
11. Suma, T. P., and G. Rekha. "Study on iot based women safety devices with screaming detection and video capturing." International Journal of Engineering Applied Sciences and Technology 6.7 (2021): 257-262.
12. Zinemanas, Pablo, et al. "An interpretable deep learning model for automatic sound classification." Electronics 10.7 (2021): 850.
13. Malova, Inna S., and Daria V. Tikhomirova. "Recognition of emotions in verbal messages based on neural networks." Procedia Computer Science 190 (2021): 560-563.
14. Tsiktsiris, Dimitrios, et al. "A novel image and audio-based artificial intelligence service for security applications in autonomous vehicles." Transportation Research Procedia 62 (2022): 294-301.
15. Yang, Lei, and Hongdong Zhao. "Sound classification based on multihead attention and support vector machine." Mathematical Problems in Engineering 2021.1 (2021): 9937383.