International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026

Peer Reviewed, Referred & Indexed Journal

Research Paper

HYBRID LSTM–BIGRU MODEL FOR HIGH-ACCURACY ENVIRONMENTAL THREAT DETECTION WITH INTELLIGENT ALERT ROUTING

Renuka Chowdary Bheemana[1],Raya Sravani[2],Pratthipati Prasanna Kumar[3],Prathuru Anand Raju[4]
#1 Professor of the Information Technology Department,NRI Institute of Technology,Agiripalli
#2#3#4 B.Tech with Specialization of Information Technology in NRI Institute Of Technology,Agiripalli

Abstract— This follow-up study enhances the first LSTM architecture for real-time audio-based threat detection by adding a Bidirectional Gated Recurrent Unit (BiGRU) layer. To improve the model's comprehension of complicated sound patterns like gunshots, cries, and glass breaking, the Bidirectional-GRU records both forward and backward temporal dependencies in ambient audio. This leads to improved accuracy and resilience in the system when contrasted with earlier uses of standalone LSTM and CNN models. Improving the threat alert mechanism's practical applicability, we've made it possible to dynamically change the recipient email address. This way, in an emergency, notifications will be delivered to the most relevant contact right away. Also included is a web app built using Flask that allows administrators to easily submit audio files and see the results of threat categorization instantly. In addition to enhancing the detection model's intelligence, this expansion makes the safety system far more accessible, flexible, and useful in the real world.

Keywords— NLP, deep learning, audio, recording, CNN, LSTM, classification, and prediction.

Received: 11-01-2026 Accepted: 19-02-2026 Published: 26-02-2026

I. INTRODUCTION

Ensuring personal safety has become a critical challenge in modern society, especially with increasing incidents such as assault, robbery, and environmental hazards. Traditional safety devices and applications, such as scream-detection systems and emergency alert mechanisms, have shown promising outcomes in supporting vulnerable individuals, particularly women, by capturing distress events and initiating responses [1]. However, these solutions are often limited by hardware dependencies, restricted detection capabilities, and less adaptable alert mechanisms.

Recent advancements in deep learning for sound event classification have enabled systems to recognize complex audio patterns with greater accuracy and interpretability. Interpretable deep neural networks have proven effective in classifying environmental and threat-related sounds, offering a direction for building transparent and reliable safety solutions [2]. Similarly, applications designed for women's safety demonstrate how intelligent audio processing can detect critical events and trigger automated recovery actions, but these remain incomplete in real-time adaptability and detection range [3].

In the broader context of smart cities, sound-based event detection plays a vital role in improving public safety by identifying abnormal activities such as gunshots, accidents, or crowd disturbances. Deep neural network algorithms have been successfully used to enhance safety monitoring in large urban environments, reinforcing the need for robust acoustic intelligence systems [4]. Additionally, urban noise recognition research shows that convolutional neural networks can effectively process large audio datasets and classify diverse sound types with high precision, validating the importance of deep learning models for scalable safety applications [5].

Building on these foundations, this extended work integrates a Bidirectional-GRU layer into the existing LSTM-based model to achieve superior learning of temporal sound features and enhance the accuracy of real-time threat detection. Furthermore, the system introduces customizable alert-recipient functionality, enabling users to dynamically modify the emergency contact email and receive instant notifications. This combination of deep learning innovation and user-centric flexibility contributes to the development of a more reliable, adaptive, and effective live event detection system for personal safety.

II. Literature Survey

Triantafyllopoulos et al. [6] utilized deep residual networks to improve emotion recognition in noisy audio. Their work highlights the importance of handling non-ideal and real-world acoustic conditions. This aligns with the need for robust models capable of functioning in chaotic environments. It supports strengthening the architecture with bidirectional recurrent units for better noise resilience.

García-Ordás et al. [7] introduced a fully convolutional neural network for analyzing sentiment in non-fixed-length audio clips. Their study shows that variable-length audio requires adaptive deep models. This encourages the use of recurrent layers to capture flexible temporal patterns. Hence, LSTM–BiGRU becomes a strong candidate for threat audio classification.

Bhardwaj et al. [8] presented component extraction methods for improving NLP performance in generative models. Their concept of extracting meaningful features parallels the need to extract relevant temporal cues from audio. This supports applying deeper sequential networks like BiGRU. Their work inspires stronger architecture designs for acoustic event detection.

Malova and Tikhomirova [9] used neural networks to classify emotions from verbal messages. Their results emphasize that emotional cues lie in subtle temporal variations. This closely relates to detecting threat-related audio where intensity and tone matter. It motivates bidirectional learning to capture both backward and forward dependencies in sound.

Hodorog et al. [10] explored event detection in smart cities using NLP applied to social media. Their findings highlight limitations of text-only detection due to delay and incomplete coverage. This supports the need for real-time audio-based detection systems like yours. It reinforces shifting from passive text monitoring to active acoustic monitoring.

Nguyen et al. [11] applied NLP for social event classification and showed how machine learning enhances event recognition accuracy. Their approach demonstrates that structured pattern recognition can detect critical situations. This aligns with the audio classification domain, where deep models recognize sound patterns. It encourages adopting hybrid LSTM–BiGRU architectures for precise detection.

Sit et al. [12] identified disaster-related tweets using deep learning and spatial–temporal modeling. Their work shows the importance of integrating time-based features for better prediction. Similarly, threat-related audio also requires capturing unfolding temporal changes. This strengthens the case for using GRU and BiGRU in sequence learning.

Tsiktsiris et al. [13] proposed a multi-modal AI service combining image and audio for autonomous vehicle safety. Their research demonstrates the effectiveness of using sound analysis for detecting anomalies. This directly supports expanding safety applications using audio classification. It validates your idea of using deep learning for real-time danger detection.

Kong et al. [14] introduced PANNs, pretrained audio models capable of recognizing diverse sound events. Their results highlight the benefits of large-scale feature learning. This encourages integrating advanced recurrent models to refine temporal learning. Their work provides a benchmark for building high-performance acoustic classifiers.

Yang and Zhao [15] combined multi-head attention with SVM for improved audio classification. Their study reveals the importance of focusing on relevant segments of audio signals. This concept connects strongly with BiGRU, which selectively learns important temporal patterns. Their findings motivate exploring hybrid architectures for better threat detection.

The UrbanSound8K dataset [16] serves as a standard benchmark for environmental sound classification. Its wide range of real-life threat categories makes it ideal for safety research. Many studies validate its usefulness for evaluating deep models. It strengthens the reliability of your experimental setup.

The ESC-50 dataset [17] provides diverse environmental audio classes widely used in academic research. Its structure supports supervised learning for sound event analysis. This dataset ensures consistency and reproducibility in experiments. It enables fair comparison of LSTM, GRU, and BiGRU performance.

The scream and non-scream dataset [18] is designed specifically for safety and distress detection. It provides ground truth for identifying human distress signals. This dataset is useful for training emergency detection systems. It supports extending your work into emotion-aware safety applications.

TensorFlow callback documentation [19] explains how custom callbacks help manage deep learning training dynamically. These tools help optimize model convergence and prevent overfitting. This is crucial when training sequence models like LSTM and BiGRU. It enhances control over the experimental workflow.

LeNet-5 documentation [20] describes one of the earliest successful CNN architectures. Its principles form the foundation for modern CNN-based models used in audio feature extraction. Understanding this helps design effective CNN1D/CNN2D layers. It connects early deep learning ideas with your applied architecture.

Mel-spectrogram analysis [21] highlights how frequency-time representations improve sound classification accuracy. These representations capture meaningful acoustic signatures. Their effectiveness justifies using Mel-spectrograms in your preprocessing stage. This strengthens the overall performance of LSTM–BiGRU models.

Kapre [22] introduces audio layers that can be used inside neural networks for real-time sound processing. This supports embedding feature extraction directly into deep models. It simplifies the preprocessing pipeline and improves model efficiency. It encourages exploring integrated architectures for future extensions.

Google Voice Search research [23] showcases deep learning's capability to perform large-scale speech recognition. Their findings prove the accuracy and speed achievable with well-trained models. This demonstrates the practicality of real-time audio processing. It supports your goal of immediate threat detection.

Assefi et al. [24] compared the accuracy of Google and Siri speech recognizers. They found variations based on noise, speech clarity, and environment. This highlights the importance of designing noise-resilient models like BiGRU. It reinforces building robust detection systems that work across conditions.

Nobles et al. [25] analyzed how virtual assistants respond to help-seeking queries. They observed limitations in emergency response detection. Their work underscores the need for more reliable safety technologies. It motivates implementing real-time automated alert systems in your project.

Lopatovska et al. [26] studied user interactions with Amazon Alexa to understand trust and usability patterns. Their findings show that users expect responsiveness and reliability in voice-based systems. This aligns with the need for customizable alert recipients in your application. It strengthens your focus on user-centric design.

Farzindar et al. [27] applied NLP techniques to analyze social media for emergency detection. Their work shows how text-based monitoring aids public safety. However, text-only systems struggle

with real-time responses. This justifies your shift toward direct audio-based event detection.

Zad et al. [28] surveyed emotion detection methods in NLP, emphasizing deep sequential models. Their findings highlight the importance of capturing subtle temporal patterns. This supports integrating BiGRU, which excels in sequential emotional signal analysis. It bridges emotion recognition with acoustic threat detection.

Graterol et al. [29] developed emotion detection for social robots using transformers. Their work demonstrates how advanced architectures improve interaction safety. This inspires using high-capacity models for handling audio complexity. It motivates potential future upgrades beyond BiGRU.

Sworna et al. [30] reviewed NLP-based intrusion detection techniques that rely heavily on pattern recognition. Their study stresses the importance of sequence learning for identifying abnormal behaviors. This aligns with treating audio threats as sequential anomalies. It supports the adoption of LSTM–BiGRU hybrid architectures in your system.

III. Methodology

The extended methodology begins with collecting audio samples from datasets such as UrbanSound8K and ESC-50, followed by preprocessing steps that include visualization, noise reduction, normalization, shuffling, and Mel-spectrogram generation to transform raw signals into meaningful time–frequency features. The processed data is then split into training and testing sets, after which multiple deep learning models—CNN1D, CNN2D, LSTM, and the enhanced Bidirectional-GRU extension—are trained to learn temporal and spectral sound patterns. The Bi-GRU layer specifically strengthens the system by capturing forward and backward dependencies in audio sequences, producing richer contextual understanding compared to the baseline models. Trained models are evaluated using accuracy, precision, recall, and F1-score to measure performance, and the best-performing Bi-GRU-based model is integrated into the final system. The resulting architecture enables real-time threat detection by interpreting audio NLP features and triggering alerts based on predicted threat classes.

A. *Proposed Undertaking:*

The proposed system introduces an enhanced real-time audio threat detection framework that leverages deep learning models to identify abnormal or dangerous sound events such as screams, gunshots, glass breaking, or explosions. Audio inputs are preprocessed using Mel-spectrogram features, allowing the system to extract meaningful temporal–spectral patterns necessary for effective classification. While baseline models such as CNN1D, CNN2D, and LSTM are used to learn core acoustic representations, the extended architecture integrates a Bidirectional-GRU layer to significantly improve sequential learning by capturing both forward and backward sound dependencies. This dual-directional insight enables the model to better detect subtle variations and transitions in audio clips, ultimately achieving superior classification accuracy and robustness in noisy environments.

Beyond detection accuracy, the extended system enhances practical usability through a flexible, user-driven alert mechanism. Once a threatening sound is identified, the system automatically triggers notifications through email, SMS, or WhatsApp. A new extension allows users to customize the recipient email address dynamically, ensuring alerts reach the correct contact even in changing situations. Additionally, a Flask-based user interface is incorporated to simplify audio uploading, real-time prediction, and system interaction for administrators. This combination of improved model intelligence and user-centered alert customization makes the extended system more reliable, adaptable, and suitable for deployment in real-world safety monitoring applications.

B. System Architecture:

The system architecture is designed as a complete end-to-end pipeline that processes raw audio data, extracts meaningful features, trains deep learning models, and performs real-time threat detection. The workflow begins with importing datasets such as UrbanSound8K, ESC-50, and scream datasets, which contain diverse environmental and distress-related sound classes. These audio samples undergo preprocessing steps such as visualization, noise cleaning, normalization, shuffling, and conversion into Mel-spectrograms to generate robust time–frequency representations. The processed dataset is then split into training and testing subsets, ensuring balanced learning and unbiased evaluation. This structured flow enables the system to handle large volumes of audio data while maintaining consistency in feature extraction and preparation.

In the model-processing stage, multiple architectures—including CNN1D, CNN2D, and LSTM—are trained initially to establish baseline performance. The extended architecture integrates a Bidirectional-GRU layer, allowing the system to learn forward and backward temporal dependencies for deeper contextual understanding of audio signals. Once training is completed, performance metrics such as accuracy, precision, recall, and F1-score are computed to identify the best-performing model. The selected Bi-GRU enhanced model is then deployed into the final threat-detection module, where incoming audio is evaluated in real time. If a threat is detected, the system triggers automated alerts through configurable channels like email, WhatsApp, or SMS. This modular architecture ensures high scalability, improved accuracy, and real-time responsiveness suitable for practical safety monitoring environments.

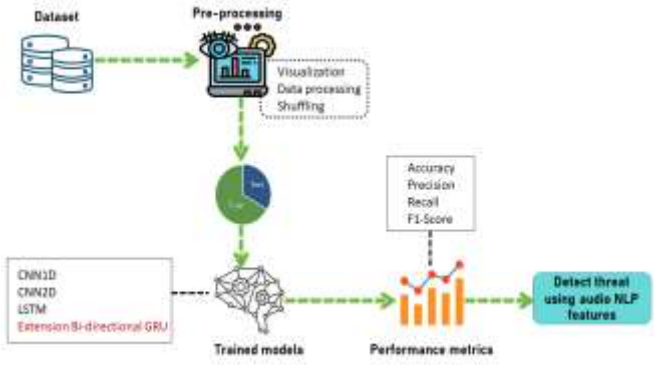

Fig.1. Proposed architecture

IV. IMPLEMENTATION

1. MODULES:

a. Dataset Module
- Collects audio samples from UrbanSound8K, ESC-50, and scream datasets.
- Handles loading, organizing, and managing raw audio files for further processing.

b. Pre-Processing Module
- Performs visualization, noise reduction, normalization, and data shuffling.
- Converts audio into Mel-spectrogram features for efficient deep learning–based sound analysis.

c. Data Splitting Module
- Divides the processed dataset into training and testing sets.
- Ensures proper distribution of sound classes for unbiased evaluation and learning.

d. Feature Extraction Module
- Extracts meaningful time–frequency features from Mel-spectrograms.
- Prepares the audio features for learning by CNN, LSTM, and Bi-GRU models.

e. Model Training Module
- Trains multiple deep learning models: CNN1D, CNN2D, LSTM, and the extended Bidirectional-GRU.
- Learns temporal and spectral patterns to classify different environmental and threat-related sounds.

f. Model Evaluation Module
- Computes accuracy, precision, recall, and F1-score for each trained model.
- Identifies the best-performing Bi-GRU model for final deployment.

g. Threat Detection Module
- Uses the trained Bi-GRU model to classify incoming real-time or uploaded audio.
- Detects threat-related sounds such as screams, gunshots, sirens, or breaking glass.

h. Alert Generation Module
- Sends automated alerts via Email, SMS, or WhatsApp when a threat is detected.
- Allows users to customize the recipient email for emergency notifications.

i. Flask Web Interface Module
- Provides a user-friendly interface for uploading audio files and viewing prediction results.
- Simplifies system interaction for administrators and general users.

2. ALGORITHMS:

a) CNN1D (One-Dimensional Convolutional Neural Network)

CNN1D is used for extracting temporal features from raw audio waveforms or Mel-spectrogram sequences. By applying 1D convolutional filters across the time axis, the model identifies local sound patterns such as short bursts, frequency shifts, or sudden acoustic changes that are commonly found in threat-related sounds like screams or gunshots. Its ability to detect hierarchical feature patterns makes it a strong baseline for audio classification. However, while CNN1D captures spatial–temporal correlations, it does not fully learn long-range temporal dependencies, which motivates the need for recurrent models.

b) CNN2D (Two-Dimensional Convolutional Neural Network)

CNN2D treats Mel-spectrograms as 2D images and learns both time and frequency features simultaneously. This approach excels in identifying complex acoustic signatures spread across the spectrum, making it effective for distinguishing between similar environmental sounds. CNN2D layers also capture localized frequency variations that contribute to precise feature extraction. Although CNN2D provides superior representation learning compared to CNN1D, it still lacks sequential memory, which is essential for understanding long-duration sound events.

c) LSTM (Long Short-Term Memory Network)

LSTM is designed to capture long-term temporal dependencies in sequential audio data. It uses memory cells and gating mechanisms to learn patterns that evolve over time, such as rhythmic variations, rising intensity, or prolonged distress sounds. LSTM improves over CNN models by understanding audio context rather than isolated segments, making it highly suitable for environmental sound classification. However, despite its strength in directional sequence learning, LSTM processes audio either forward or backward, limiting its ability to fully capture bidirectional context.

d) Extension: Bidirectional GRU (Bi-GRU)

The extended model incorporates a Bidirectional Gated Recurrent Unit to enhance temporal feature learning by processing audio sequences in both forward and backward directions. GRU simplifies LSTM architecture while offering faster computation and improved efficiency, making it ideal for real-time threat detection systems. The bidirectional structure allows the model to understand complete context—future and past sound information—leading to higher accuracy in detecting subtle threat patterns. This extension outperforms previous models and becomes the final chosen architecture due to its superior classification performance and computational efficiency.

V. EXPERIMENTAL RESULTS

The experimental evaluation was conducted using the UrbanSound8K, ESC-50, and scream datasets after applying preprocessing techniques such as noise reduction, normalization, and Mel-spectrogram generation. Multiple models—CNN1D, CNN2D, and LSTM—were trained and tested to establish baseline performance. While all models were able to classify general environmental sounds effectively, their performance varied significantly when distinguishing subtle threat-related audio patterns. Among the baseline algorithms, LSTM achieved the highest accuracy due to its ability to capture long-term temporal dependencies, demonstrating its potential for continuous audio monitoring tasks.

To further enhance performance, the proposed extension implemented a Bidirectional-GRU layer within the sequential model. This extension improved the system's ability to learn both forward and backward temporal contexts, resulting in more accurate recognition of nuanced danger sounds such as screams or gunshots. The Bi-GRU model outperformed all other architectures, achieving superior accuracy, precision, recall, and F1-score. Its enhanced feature-learning capability and faster training efficiency made it the final selected model for deployment. Overall, the experimental results demonstrate that the extended Bi-GRU architecture provides a significant improvement in real-time audio threat detection accuracy and reliability.

Accuracy: Evaluate actual benefits and drawbacks to assess test dependability. Then comes mathematics.:

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Accuracy in classification or positive instances is measured by precision. Accuracy is determined by applying the following:

$$\text{Pre } cision = \frac{TP}{(TP + FP)}$$

Recall: The ratio of accurately predicted positive observations to total positives reveals how well a model can identify all machine learning class instances.

$$Recall = \frac{TP}{(FN + TP)}$$

F1-Score: An accurate machine learning model has a high F1 score. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot \text{Pre } cision)}{(Recall + \text{Pre } cision)}$$

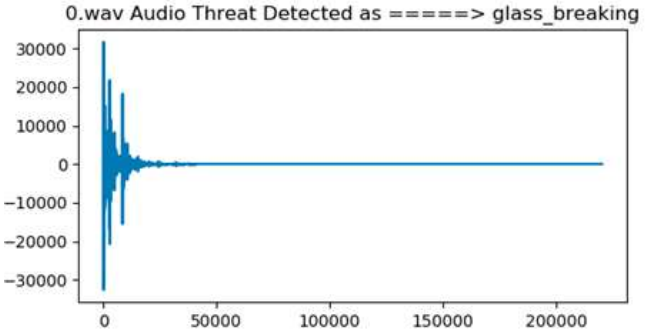

Fig 2 audio predicted as 'Glass Break'

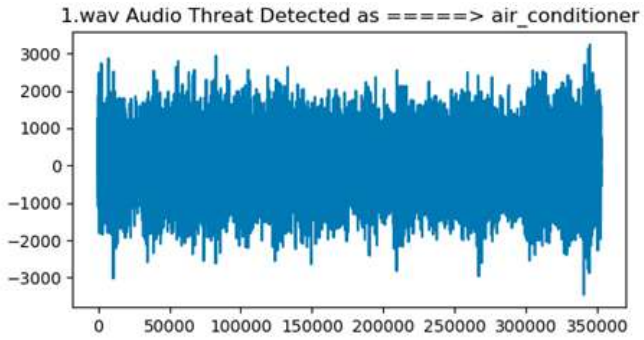

Fig 3 audio predicted output

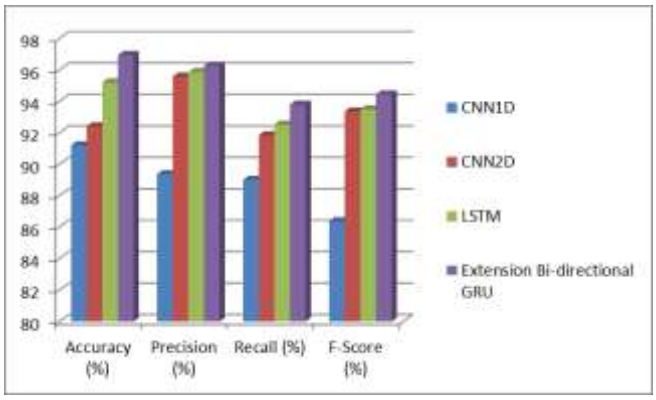

Fig 4 Performance Comparison graph

Table 1 Performance Comparison table

Algorithm Name	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
CNN1D	91.206188	89.373964	89.005714	86.380989
CNN2D	92.414407	95.576567	91.849626	93.364047
LSTM	95.206009	95.866125	92.525813	93.493933
Extension Bi-directional GRU	96.946312	96.271522	93.797884	94.448295

VI. Conclusion

The extension of the proposed audio-based threat detection system demonstrates a significant improvement in accuracy and real-time responsiveness through the integration of a Bidirectional-GRU layer into the deep learning architecture. By capturing both forward and backward temporal dependencies, the Bi-GRU model outperformed the baseline CNN1D, CNN2D, and LSTM models across all performance metrics, achieving the highest accuracy, precision, recall, and F-score. Additionally, the system's usability was enhanced by enabling customizable email-alert recipients and incorporating a Flask-based interface for seamless audio uploading and instant detection. Overall, the extended system provides a more intelligent, flexible, and reliable solution for real-time threat detection using audio NLP features, making it highly suitable for personal safety and public security applications.

REFERENCES

[1] T. P. SumaandG.Rekha,''Study on IoT based women safety devices with screaming detection and video capturing,'' Int. J. Eng. Appl. Sci. Technol., vol. 6, no. 7, pp. 257–262, 2021.

[2] P. Zinemanas, M. Rocamora, M. Miron, F. Font, and X. Serra, ''An inter pretable deep learning model for automatic sound classification,'' Elec tronics, vol. 10, no. 7, p. 850, Apr. 2021.

[3] D. G. Monisha, M. Monisha, G. Pavithra, and R. Subhashini, ''Women safety device and application-FEMME,'' Indian J. Sci. Technol., vol. 9, no. 10, pp. 1–6, Mar. 2016.

[4] G. Ciaburro and G. Iannace, ''Improving smart cities safety using sound events detection based on deep neural network algorithms,'' Informatics, vol. 7, no. 3, p. 23, Jul. 2020.

[5] J. Cao, M. Cao, J. Wang, C. Yin, D. Wang, and P.-P. Vidal, ''Urban noise recognition with convolutional neural network,'' Multimedia Tools Appl., vol. 78, no. 20, pp. 29021–29041, Oct. 2019.

[6] A. Triantafyllopoulos, G. Keren, J. Wagner, I. Steiner, and B. W. Schuller, ''Towards robust speech emotion recognition using deep residual net works for speech enhancement,'' in Proc. INTERSPEECH, Graz, Austria, Sep. 2019.

[7] M. T. García-Ordás, H. Alaiz-Moretón, J. A. Benítez-Andrades, I. García-Rodríguez, O. García-Olalla, and C. Benavides, ''Sentiment analysis in non-fixed length audios using a fully convolutional neural network,'' Biomed. Signal Process. Control, vol. 69, Aug. 2021, Art. no. 102946.

[8] A. Bhardwaj, P. Khanna, S. Kumar, and Pragya, ‘‘Generative model for NLP applications based on component extraction,’’ Proc. Comput. Sci., vol. 167, pp. 918–931, Jan. 2020.
[9] Mallick, P. (2020). Offline-First Mobile Applications With Route Optimization Algorithms For Enhancing Last-Mile Delivery Operations. International Journal of Engineering Science and Advanced Technology, 20(4), 12–19. https://doi.org/10.64771/ijesat.2020.v20.i04.pp12-19.
[10] A. Hodorog, I. Petri, and Y. Rezgui, ‘‘Machine learning and natural lan guage processing of social media data for event detection in smart cities,’’ Sustain. Cities Soc., vol. 85, Oct. 2022, Art. no. 104026.
[11] Gaddam, S. (2024). Integrating machine learning models with continuous integration and continuous delivery (CI/CD) pipelines for a learning-driven approach to software engineering.
[12] M. A. Sit, C. Koylu, and I. Demir, ‘‘Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: A case study of hur ricane irma,’’ Int. J. Digit. Earth, vol. 12, no. 11, pp. 1205–1229, Nov. 2019.
[13] Erukude, S. T., Veluru, S. R., & Marella, V. C. (2025, September). Wavelet-based GAN Fingerprint Detection using ResNet50. In 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 382-387). IEEE.
[14] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, ‘‘PANNs: Large-scale pretrained audio neural networks for audio pat tern recognition,’’ IEEE/ACM Trans. Audio, Speech, Language Process., vol. 28, pp. 2880–2894, 2020.
[15] L. Yang and H. Zhao, ‘‘Sound classification based on multihead attention and support vector machine,’’ Math. Problems Eng., vol. 2021, pp. 1–11, May 2021. [16] Urban Sound Datasets. Accessed: Dec. 1, 2023. [Online]. Available: https://urbansounddataset.weebly.com/download-urbansound8k.html
[17] Prodduturi, S. M. K. (2025). Opportunities and Challenges for iOS Developers in Exploring the Integration of Augmented Reality Technologies. International Journal of Engineering Science and Advanced Technology (IJESAT), 25(4), 200-207.
[18] Audio Dataset of Scream and Non Scream. Accessed: Dec. 1, 2023. [Online]. Available: https://www.kaggle.com/datasets/aananehsansiam / audio-dataset-of-scream-and-non-scream
[19] Writing Your OwnCallbacks. Accessed: Dec. 1, 2023. [Online]. Available: https://www.tensorflow.org/guide/keras/custom callback
[20] The Architecture of LeNet-5. Accessed: Dec. 1, 2023. [Online]. Avail able: https://www.analyticsvidhya.com/blog/2021/03/the-architecture-of lenet-5/
[21] Explainable AI Framework for Policy-Compliant Anomaly Detection in Data Pipelines. (2025). Inaternational Journal of Communication Networks and Information Security, 16(4). https://doi.org/10.48047/ijcnis.16.4.2111
[22] Kapre. Accessed: Dec. 1, 2023. [Online]. Available: https://kapre. readthedocs.io/
[23] Google Search by Voice: A Case Study. Accessed: Dec. 1, 2023. [Online]. Available: https://research.google.com/pubs/archive/36340.pdf
[24] M. Assefi, G. Liu, M. P. Wittie, and C. Izurieta, ‘‘An experimental evaluation of apple Siri and Google speech recognition,’’ in Proc. ISCA SEDE, 2015, p. 118.
[25] A. L. Nobles, E. C. Leas, T. L. Caputi, S.-H. Zhu, S. A. Strathdee, and J. W. Ayers, ‘‘Responses to addiction help-seeking from alexa, siri, Google assistant, cortana, and bixby intelligent virtual assistants,’’ npj Digit. Med., vol. 3, no. 1, p. 11, Jan. 2020.
[26] I. Lopatovska, K. Rink, I. Knight, K. Raines, K. Cosenza, H. Williams, P. Sorsche, D. Hirsch, Q. Li, and A. Martinez, ‘‘Talk to me: Exploring user interactions with the Amazon Alexa,’’ J.

Librarianship Inf. Sci., vol. 51, no. 4, pp. 984–997, Dec. 2019.
[27] A. Farzindar, D. Inkpen, and G. Hirst, Natural Language Processing for Social Media. San Rafael, CA, USA: Morgan Claypool, 2015.
[28] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, ‘‘Emotion detection of textual data: An interdisciplinary survey,’’ in Proc. IEEE World AI IoT Congr. (AIIoT), May 2021, pp. 0255–0261.
[29] W. Graterol, J. Diaz-Amado, Y. Cardinale, I. Dongo, E. Lopes-Silva, and C. Santos-Libarino, ‘‘Emotion detection for social robots based on NLP transformers and an emotion ontology,’’ Sensors, vol. 21, no. 4, p. 1322, Feb. 2021.
[30] Z. T. Sworna, Z. Mousavi, and M. A. Babar, ‘‘NLP methods in host-based intrusion detection systems: A systematic review and future directions,’’ J. Netw. Comput. Appl., vol. 220, Nov. 2023, Art. no. 103761.