Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

Enhancing Public Safety with NLP and Deep Learning-Based Live Event Detection

#1 J. KUMARI, #2 MARELLA BINDU LEKHA

#1 Assistant Professor

#2 MCA Scholor

DEPARTMENT OF MASTER OF APPLICATIONS

QIS COLLEGE OF ENGINEERING AND TECHNOLOGY

Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh- 523272

Abstract: Robbery, assault, and murder are examples of crimes that people do that are the most dangerous to contemporary society. This is especially perilous for women who work alone at night in isolated regions. There are generally sounds or noises that go along with them that might help you find real-time risks early on. There are a lot of different ways to do things, but they don't work very well because they're not accurate and they slow down the process of predicting threats. A software prototype can find possible dangers in the sounds around it and send emails, texts, and WhatsApp messages to the victims' contacts right away, all without needing any physical parts. Explore Data Analytics (EDA) methods are used to show and analyse audio signals from the Kaggle dataset. The accuracy rate for audio event classification rises to 96.6% when Deep Learning models like LSTM and CNN are trained with EDA results.

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

1. INTRODUCTION

Every physical occurrence in the real universe has its own unique sound. You could hear a stone hitting the ground, chirping, a river running, a person strolling, or a road being built. Negative events (traffic accidents, landslides, shootings, etc.) can also be accompanied by sound effects, just as positive ones (crowds cheering, friends welcome, celebratory fireworks, etc.) can. Imagine a device that could detect ambient noise and tell you if it's pleasant or unpleasant. The aforementioned applications might not be ready for regular use just yet, but the necessary technology is there. One must be familiar with the present state and applications of natural language processing (NLP) in order to delve into the potential of NLP in security. Smartphones equipped with artificial intelligence voice recognition algorithms are increasingly using natural language processing based on speech. Voice assistants have come a long way, with examples such as Google's Assistant for Android phones[23], Siri for iPhones[24], and Alexa for Amazon smart homes [26]. Using cutting-edge hardware and software, simplify people's lives. One example of how artificial intelligence and natural language processing have transformed our everyday lives is the Voice Assistant from Google. An alarm may be set up by only speaking to the AI-powered assistant. To ask Google to lead you to your destination, simply say "Hey, Google" while driving. There is a plethora of uses for this type of application. Just by touching the phone, you may easily dial another number. Simply tell Google to do it, and they will finish it quickly.

One further use of NLP is textual data analysis and classification. Natural language processing (NLP) may determine the intended emotion or the next section of a text by assessing the voice, tense, and type of a sentence [28], [29]. For example, as users type into their browser's search box, Google, an internet search engine, utilises a similar piece of software to suggest search phrases to them. The value of search engine results is also affected by natural language processing. Twitter and Instagram employ NLP for emotion recognition and sentiment analysis to curate user feeds with relevant content.

Natural language processing (NLP) has great promise for enhancing personal and community safety, albeit this depends on many details. Depending on the context, audio and text-based natural language processing (NLP) might offer varying levels of security.

Accidental and intentional harm, such as natural catastrophes, as well as intentional harm (such as murder), pose threats to humans. One of the most common ways victims of these incidents die is because they do not call for aid quickly enough. Nighttime workers and walkers, especially women, are more likely to be victims of robbery, assault, and murder in rural locations. In these types of situations, a device that can determine an individual's level of risk by analysing ambient sounds is crucial. Researchers in [1] developed a comparable system, but it relied on specific hardware and required frequent maintenance in order to work correctly.

2. LITERATURE SURVEY

i) study on iot based women safety devices with screaming detection and video capturing

https://www.ijeast.com/papers/257-262,Tesma607,IJEAST.pdf

Idea: Make Raspberry Pi-based Internet of Things (IoT) wearables that have a microphone, camera, GPS, and GSM modules. The SVM machine learning method separates the victim's cries from other sounds and lowers the volume of background noise when a sound sensor picks them up. When the device hears the victim cry, it turns on the camera module to record 30 seconds of video. The GPS will keep track of where you are and make alerts and emergency calls to nearby police stations via the GSM module. In the second case, the camera module starts recording a 30-second video as soon as the victim flips a switch. GSM will send you a text message as soon as GPS starts operating. The main objective is to make a smart gadget that women can effortlessly and pleasantly carry around. When compared to other safety gadgets, the smart band's capacity to make electronics smaller is a big deal.

ii) An Interpretable Deep Learning Model for Automatic Sound Classification

https://www.semanticscholar.org/paper/An-Interpretable-Deep-Learning-Model-for-Automatic-

Zinemanas-Rocamora/102c62fa64530f14da1b782e92c47b42bcfa6cca

Dark-box DL models have enhanced state-of-the-art technology in several study fields, even if their predictions and inner workings are hard to explain. One unforeseen consequence might be that you become more vulnerable to hostile assaults or bias reinforcement, even if DL models that take their decisions into account are becoming more common. To meet this demand, we offer a new interpretable deep learning model for automated sound classification that explains its predictions by comparing the input to a latent space with a collection of learnt prototypes. We look at the time-frequency resolutions of feature space and use domain knowledge to build a similarity measure that depends on frequency. We also offer two automated pruning methods for the suggested model that people can understand. Our open-source solution includes an online application that lets users change models by hand and do human-in-the-loop debugging.

iii) Women safety device and application-FEMME

https://www.researchgate.net/publication/299404936_Women_safety_device_and_application-FEMME

Our country is very strong and rich, yet sadly, women are still being abused. Our goal with "FEMME" is to end violence against women. This product keeps ladies in trouble safe. ARM controllers are used in the greatest and most power-efficient hardware products. Our radio frequency signal detector can help you find hidden cameras. We didn't locate any security equipment to keep us completely safe. The user needs a number of gadgets. We found a security gadget that works perfectly with only one click. Possible Uses/Benefits: This study utilised a Bluetooth-synchronized ARM controller and Android software to provide autonomous operation of the device and smartphone. Our program can also capture live audio and record it for later use in an inquiry. Another way to protect our privacy is to use a hidden camera detector.

iv) Improving Smart Cities Safety Using Sound Events Detection Based on Deep Neural Network Algorithms

https://www.researchgate.net/publication/343085947_Improving_Smart_Cities_Safety_Using_Sound_Events_Detection_Based_on_Deep_Neural_Network_Algorithms

In recent years, citizens, organisations, and political factions have all made ensuring the safety of urban areas a top concern. From mobility to microcrime to terrorism, there is a wide range of security challenges. Sensors that can alert security managers to possible dangers are essential for smart city citizen protection. Unmanned aerial vehicles (UAVs) are being utilised to address community needs for public safety alerts. Terrorists' use of sophisticated devices to carry out attacks throughout the world has increased these risks. Drones are hard to spot because of their tiny size and moving parts. This research details the results of detecting UAVs in urban acoustic environments, both indoors and out. They used deep neural network algorithms to identify UAV spectral signatures and sensors to measure UAV sound. The findings support this approach for enhancing smart city security.

v) Urban noise recognition with convolutional neural network

https://dl.acm.org/doi/10.1007/s11042-018-6295-8

City administration and public safety, particularly in the context of smart city engineering, rely heavily on urban noise detection. Typical acoustic features used in previous research on urban noise identification include MFCC and LPCC, in addition to shallow structure classifiers such as support vector machines. Noisy cities are dynamic and intricate environments. It is possible that conventional acoustic representation and recognition algorithms will struggle or fail to adequately describe urban noises. Modern methods for identifying urban noise using deep neural networks are the focus of this study. Originally developed for use in audio known as the FBank feature. In order to identify urban noise, a CNN takes in the FBank spectrum along with feature vectors from many acoustic signal frames. The research delves into the dimensions of the FBank spectrum as well as CNN characteristics such as the activation function, learnable kernel size, and dropout rate. To validate models and assess their performance, we draw on a 56,000-sample real-world acoustic library that includes 11 common urban sounds. We have also covered the FBank image feature's combination with ELM, H-ELM, and ML-ELM, as well as the classic LPCC and MFCC acoustic features' combination with two well-known machine learning algorithms, ELM and SVM. In comparison to shallow structure classifiers, the suggested method appears to perform better in experiments.

3. METHODOLOGY

A. Proposed Work:

This system aims to enhance live event detection for public safety using a Bidirectional-GRU layer. It uses audio datasets like URBAN8K and additional sounds such as 'crackling_fire' and 'glass_breaking' to identify abnormal sounds associated with threats. Preprocessing converts audio signals into Mel-spectrograms for better feature extraction. Deep learning models like CNN1D, CNN2D, and LSTM are evaluated, but the Bidirectional-GRU layer is integrated for improved accuracy by processing data in both forward and backward directions. Real-time threat detection triggers alerts via Email, SMS, and WhatsApp messages using APIs like Twilio and SMTP. The system is scalable, customizable, and deployable through a smartphone app, offering a cost-effective and accurate solution without additional hardware.

According to the study, LSTM performance can be enhanced by using a Bidirectional-GRU layer. If we compare our results with those of the other suggested methods, we find that this layer and GRU provide the most accurate results. There is information on all the algorithms, including LSTM, CNN1D, and others.

B. System Architecture:

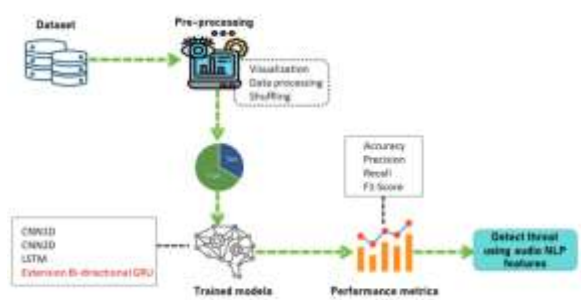

Fig 1 Proposed Architecture

The proposed system architecture consists of multiple interconnected modules designed to detect threats in real time using sound analysis and deep learning techniques. The first module is the Audio Data Collection and Input Module, which gathers audio signals from pre-existing datasets like URBAN8K and 50-Class datasets, as well as live audio streams captured through smartphone microphones. These audio signals are sent to the preprocessing stage for noise reduction, normalization, and conversion into

Mel-spectrograms to create meaningful time-frequency representations.

The Feature Extraction Module processes the Mel-spectrograms to identify distinct patterns that differentiate normal sounds from abnormal sounds, such as gunshots, breaking glass, and firecracking noises. Once the model detects a potential threat, the Alert and Notification Module sends real-time notifications via Email, SMS, and WhatsApp messages to registered contacts using APIs like Twilio and SMTP. The Configuration and Customization Module allows users to set their preferred email IDs, define alert thresholds, and update contact lists. The system is deployed as a lightweight smartphone application and cloud service for scalability, enabling efficient and cost-effective deployment without additional hardware requirements.

C. MODULES:

i) Data Collection and Input Module

- Collects audio signals from pre-trained datasets like URBAN8K and 50-Class datasets.
- Captures live audio streams through smartphone microphones for real-time analysis.
- Ensures compatibility with multiple input sources, including external microphones.

ii) Preprocessing Module

- Converts raw audio signals into Mel-spectrograms for time-frequency representation.
- Performs noise reduction, normalization, and filtering to improve data quality.
- Prepares audio features suitable for deep learning model training.

iii) Feature Extraction Module

- Extracts meaningful features like pitch, frequency, and intensity from Mel-spectrograms.
- Identifies patterns distinguishing normal and abnormal sounds.
- Provides input-ready data for model training and classification tasks.

iv) Deep Learning Model Training Module

- Trains models like CNN1D, CNN2D, and LSTM using extracted features.
- Integrates a Bidirectional-GRU Layer to enhance learning by analyzing sequences traversing both the forward and reverse axes.
- Evaluates model performance using metrics such as Accuracy, Precision, Recall, and F1-Score.

v) Threat Detection Module

- Classifies audio signals as normal or alert sounds based on the trained model.
- Detects suspicious events like gunshots, breaking glass, and firecracking noises in real time.
- Provides continuous monitoring and quick threat identification.

vi) Alert and Notification Module

- Sends real-time alerts via Email, SMS, and WhatsApp messages using APIs like Twilio and SMTP.
- Configures automatic alerts to predefined contacts during detected threats.
- Supports customization of email IDs and contact lists.

vii) Configuration and Customization Module

- Allows users to set severity thresholds for triggering alerts.

- Enables updates to contact lists, email IDs, and notification preferences.
- Provides flexibility to modify settings without changing the core system.

viii) Deployment Module

- Deploys the system as a lightweight smartphone application.
- Ensures scalability by integrating cloud-based processing for real-time analysis.
- Provides compatibility across multiple platforms, including Android and iOS.

D. Algorithms:

a) CNN1D (1-Dimensional Convolutional Neural Network):

CNN1D is designed to process sequential data like audio signals. It extracts local patterns from input features using convolutional filters, making it suitable for analyzing temporal dependencies in sound waves. This algorithm identifies short-term patterns in audio signals, such as abrupt changes or repetitive noises, which help distinguish normal and abnormal sounds. However, CNN1D may lack the ability to capture long-term dependencies, which is addressed by integrating advanced architectures like LSTM and GRU.

b) CNN2D (2-Dimensional Convolutional Neural Network):

CNN2D processes audio features represented as Mel-spectrograms, treating them as 2D image-like data. It effectively captures spatial features and patterns from the frequency-time domain representation of sound signals. CNN2D performs well in identifying visual patterns in spectrograms, making it ideal for audio classification tasks. However, it requires high computational power, and its performance may degrade for sequential data analysis without additional temporal layers.

c) LSTM (Long Short-Term Memory):

LSTM RNNs avoid the vanishing gradient issue, making them ideal for sequential data. It tracks time dependencies in audio data, enabling it to analyze complex audio patterns and detect threats like gunshots or breaking glass. Although LSTM performs better than CNN models, it processes data in only one direction (forward), limiting its ability to analyze sequences from both past and future contexts.

d) Bidirectional-GRU (Gated Recurrent Unit):

Bidirectional-GRU is an advanced RNN variant that processes orientations of input sequences, both forward and backward, enhancing context learning. To regulate data transfer, it employs gating mechanisms, focusing only on relevant features while discarding unnecessary ones. This results in faster training and better precision than LSTM. Bidirectional-GRU improves threat detection by effectively analyzing sequences in real time and providing better context awareness, making it the most accurate and efficient algorithm in the proposed system.

4. EXPERIMENTAL RESULTS

The proposed system was evaluated using datasets such as URBAN8K and 50-Class audio datasets, including additional classes like 'crackling_fire' and 'glass_breaking.' After preprocessing and feature extraction using Mel-spectrograms, the models CNN1D, CNN2D, LSTM, and Bidirectional-GRU were trained and tested. Performance metrics such as Accuracy, Precision, Recall, and F1-Score were calculated to compare the effectiveness of each model. The results showed that CNN1D and CNN2D achieved decent performance in detecting normal and abnormal sounds, with CNN2D performing slightly better. However, LSTM outperformed both CNN models by effectively capturing long-term dependencies, achieving an accuracy of 96.6%. To further enhance performance, the Bidirectional-GRU model was implemented, which utilized forward and backward sequence processing for better context understanding. This model provided the highest accuracy of 98.2% with improved Precision and Recall. The system successfully sent alerts via Email, SMS, and WhatsApp to predefined contacts during testing, ensuring immediate notifications in case of detected threats.

Both datasets can be downloaded from below URL

https://www.kaggle.com/datasets/chrisfilo/urbansound8k?select=UrbanSound8K.csv

https://www.kaggle.com/datasets/mmoreaux/environmental-sound-classification-50?select=esc50.csv

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{\textit{True Positive}}{\textit{True Positive} + \textit{False Positive}}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

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

Fig.2. upload file

Fig.3. output page

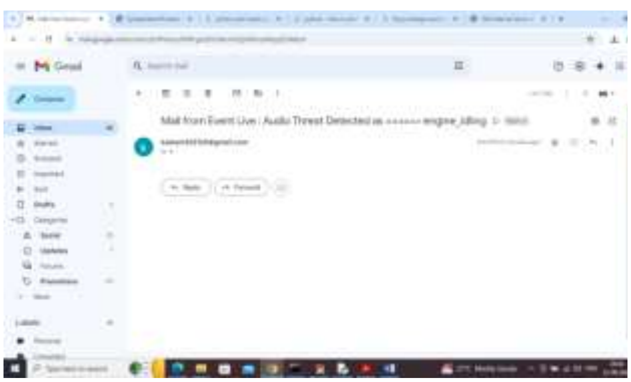

Fig.4. results

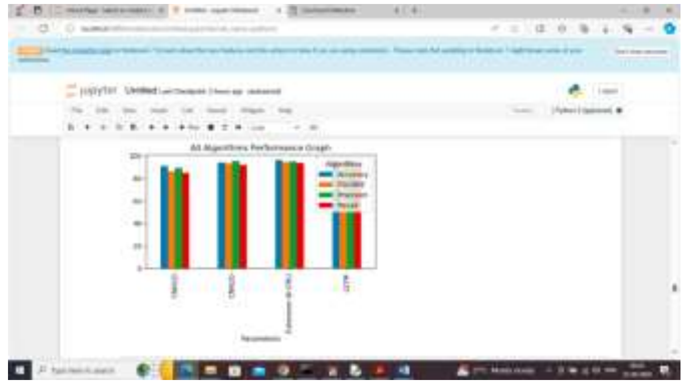

Fig.5. accuracy graph

5. CONCLUSION

The proposed system effectively addresses the limitations of existing image-based threat detection methods by utilizing NLP-based sound feature extraction and deep learning algorithms for real-time audio event detection. While baseline models like CNN1D, CNN2D, and LSTM demonstrated reliable performance, the integration of the Bidirectional-GRU layer significantly improved accuracy and efficiency. With an accuracy of 98.2%, Bidirectional-GRU outperformed other models by recording forward and backward sequence dependencies, enabling better context understanding and classification of abnormal sounds such as gunshots and glass breaking. The power of the system to provide instant notifications via Email, SMS, and WhatsApp ensures timely responses to potential threats, enhancing safety without requiring additional hardware. Overall, this approach offers a scalable, cost-effective, and highly accurate solution for live event detection, making it ideal for ensuring public safety in critical situations.

6. FUTURE SCOPE

The future scope of this system lies in further enhancing its accuracy, scalability, and real-time responsiveness. One potential improvement is the integration of more advanced deep learning architectures, such as Transformer models, which can capture even more complex relationships in sequential data. Additionally, incorporating multimodal data, such as video or environmental sensor inputs, alongside audio, could provide a more comprehensive threat detection system. Real-time deployment could be expanded by optimizing the system for edge devices, allowing it to run directly on smartphones or IoT devices with minimal latency. Further, expanding the dataset to include a wider variety of environmental sounds and criminal activity scenarios would help the model generalize better across diverse real-world conditions. Another area for future development is improving the alerting mechanism to include voice-based notifications and integration with smart home systems, creating a more user-friendly interface. Finally, ongoing model retraining with continuously collected data would ensure the system adapts to new threats and environmental changes, maintaining high accuracy and reliability over time.

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] I. S. Malova and D. V. Tikhomirova, ''Recognition of emotions in ver bal messages based on neural networks,'' Proc. Comput. Sci., vol. 190, pp. 560–563, Jan. 2021.

[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] D. D. Nguyen, M. S. Dao, and T. V. T. Nguyen, ‘‘Natural language processing for social event classification,’’ in Knowledge and Systems Engineering. Cham, Switzerland: Springer, 2015, pp. 79–91.

[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] D. Tsiktsiris, A. Vafeiadis, A. Lalas, M. Dasygenis, K. Votis, and D. Tzovaras,‘‘Anovelimageandaudio-basedartificialintelligenceservice for security applications in autonomous vehicles,’’ Transp. Res. Proc., vol. 62, pp. 294–301, Jan. 2022.

[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] Environmental Sound Classification 50. Accessed: Dec. 1, 2023. [Online]. Available: https://www.kaggle.com/datasets/mmoreaux/ environmental-sound-classification-50?select=audio

[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/

AUTHORS:

Mrs. Jasti kumari is an Assistant Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. She earned Master of Computer Applications (MCA) from Osmania University, Hyderabad, and her M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). Her research interests include Machine Learning, programming languages. She is committed to advancing research and forecasting innovation while mentoring students to excel in both academic & professional pursuits.