

Sniper Gunshot Detection through Mel-Spectrogram using Deep Learning

Ahmed Shafique

Artificial Intelligence Intern at AL Khwarizmi Institute of Computer Science (KICS)
NWN Lab

ABSTRACT

The approach utilizes the Mel spectrogram to extract key audio features, which are then analyzed by a neural network to identify gunshots. The system is trained on a dataset of gunshot and non-gunshot audio recordings, and its performance is assessed using standard metrics such as accuracy, precision, and recall. The results of this study demonstrate the effectiveness of the proposed method in accurately detecting sniper gunshots, making it a viable solution for enhancing surveillance and security measures.

INTRODUCTION

The detection of sniper gunshots is a critical task in various fields, including law enforcement, military operations, and surveillance systems. The ability to accurately identify and locate gunshot sources can help prevent crimes, reduce casualties, and enhance public safety. However, the detection of sniper gunshots is a challenging problem due to the complexity of acoustic signals and the presence of background noise.

Traditional methods for gunshot detection rely on acoustic sensors and signal processing techniques, which can be limited by their sensitivity to environmental factors and their inability to distinguish between different types of sounds. Recent advances in

deep learning and audio signal processing have opened new opportunities for developing more accurate and robust gunshot detection systems.

This project aims to develop a novel approach to sniper gunshot detection using mel spectrogram analysis and deep learning techniques. The proposed system leverages the Mel spectrogram's ability to extract relevant acoustic features from audio signals, which are then fed into a deep neural network for classification. The system is designed to be robust, accurate, and efficient, making it suitable for real-time surveillance and security applications.

OBJECTIVE

Design and implement a system that utilizes Mel spectrogram analysis and deep learning techniques to detect sniper gunshots with high accuracy.

METHODOLOGY

The project's initial stage was hindered by the scarcity of datasets specifically related to the light sniper rifle, a crucial component for developing an accurate gunshot detection system. Despite efforts to obtain relevant datasets, the acquired .wav files, when converted to frequency plots, yielded unrealistic representations of the rifle's acoustic signature. To overcome this limitation, a local dataset of 7 .wav files was

compiled, but its size was insufficient for robust model training. To augment this dataset, additional audio samples were sourced from diverse environments, including first-person shooter games and YouTube videos, resulting in a total of 20 .wav files. This eclectic dataset, although not ideal, provides a foundation for preliminary model development and testing, with the understanding that further data collection and refinement will be necessary to achieve optimal performance.

Here are one of the results of the frequency plots of the .wav files.

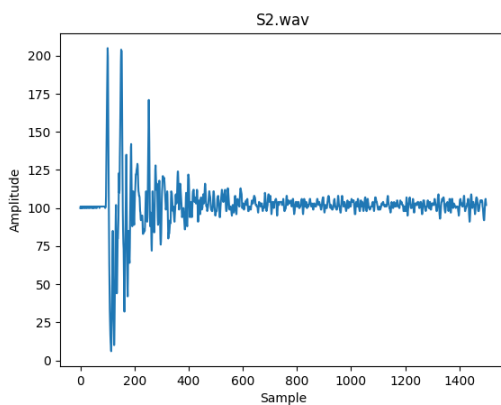


Figure 1: Frequency plot

After this I again converted the frequency plots to the .wav files to check that from frequency to .wav audio conversion can be possible or not. Then all the frequency plots are converted to the .wav files with the sampling rate of 12000.

After the conversion from frequency plots to the .wav files, I converted the .wav audio files into the Mel Spectrograms using python library librosa, NumPy, Pandas and SciPy. Here are the one of the Mel Spectrograms plots.

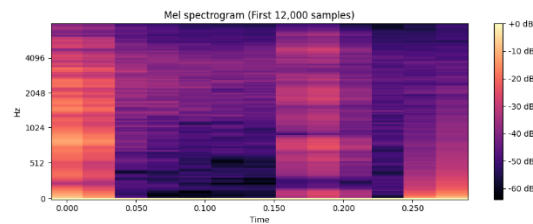


Figure 2: Mel Spectrogram of local dataset

Now, here is one of the Mel Spectrogram image plots from diverse sources such as games, YouTube videos.

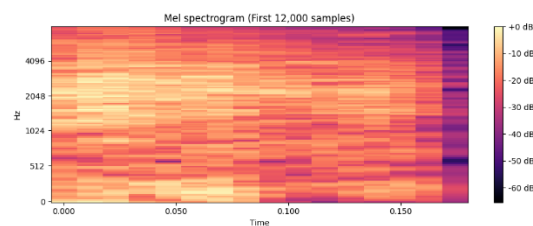


Figure 3: Mel Spectrogram of Internet sample

Following the conversion of the audio files to Mel Spectrograms, the dataset was augmented to increase its size and diversity. The resulting dataset was deemed sufficient for training a deep learning model. A Convolutional Neural Network (CNN) was applied to the dataset, which was split into 70% training and 30% testing sets. However, the training process revealed an overfitting issue, as CNN achieved an accuracy of 100% on the training set. This suggested that the model was memorizing the training data rather than learning generalizable features. Furthermore, when presented with Mel Spectrogram plots of non-sniper rifle sounds, the CNN incorrectly classified them as LSR sniper gunshot Mel Spectrograms, indicating a lack of robustness and generalizability. This outcome highlights the need for a more extensive and diverse dataset to train a reliable and accurate gunshot detection model.

After CNN I applied the Recurrent Neural Network to the same dataset and I added some feature engineering in it and for training I used 3 features, one is for the label = 1 which is for the LSR gunshot Mel Spectrograms and other 2 labels are set to the label = 0 for the non-sniper gunshot as well as the clap sound. Because without the clap sound feature engineering and labeling the RNN predicting to the LSR gunshot. So, from these datasets along with the feature engineering I trained the RNN and tested it and it is giving accuracy of about 92%.

RESULTS

Here are the results of the RNN with feature engineering and labeling as label =1 for sniper gunshot and label = 0 for the non-sniper and clap sound. I arranged the test dataset in order to clarify both labels.

Epoch 1/50

6/6 —————
 ————— 8s 273ms/step - accuracy: 0.6452 -
 loss: 0.6709

.....

Epoch 50/50

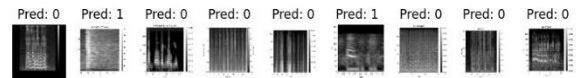
6/6 —————
 ————— 2s 279ms/step - accuracy: 0.4883 -
 loss: 81234608.0000

Here are the predicted outputs from the RNN.

1/1 —————
 ————— 1s 634ms/step

The prediction of the test image is [0, 1, 0, 0,
 0, 1, 0, 0, 0]

Here is the output predicted images of the Mel Spectrograms.



CONCLUSION

This project aimed to develop a system to detect gunshots from a light sniper rifle using deep learning models. Initially, we had a limited dataset of 7 audio files, but we augmented it with more samples from different sources, resulting in 20 files. We then converted these files into Mel Spectrograms and increased their diversity.

We tried using a Convolutional Neural Network (CNN) but it didn't work well, achieving 100% accuracy on the training set but failing to distinguish between sniper rifle and other sounds. We then used a Recurrent Neural Network (RNN) with feature engineering and labeling, which resulted in an accuracy of 92%. This project shows the potential of deep learning for gunshot detection, but also highlights the need for more diverse and robust datasets to improve the model's accuracy and generalizability.