Classification of audio data using deep learning approaches

*Abstract*—This paper explores the potential of using machine learning approaches, specifically using audio data, to improve crime detection and prevention. This research proposed a machine learning model for identifying crime and anti-social behavior using audio recordings in a local area. The paper includes a discussion of data pre-processing, model architecture selection, and evaluation metrics. The dataset used in this study is an urban sound dataset with 8732 labelled sound recordings. The deep learning model architectures has been proposed in this study are Long Short-Term Memory (LSTM), Convolutional Recurrent Neural Networks (CRNNs), and 1D Convolutional Neural Networks (1DCNN). During the implementation of these architectures results will be compared to determine which architecture yields the best results. The evaluation metrics for the determining the model performance will include accuracy, precision, recall, specificity, F1 score, and confusion matrix.

Keywords—ML (Machine Learning), Deep learning, crime, anti-social behavior, WHO, EDA, UK, LSTM, CRNN, 1DCNN, evaluation metrics, audio data, classification.

# Introduction

A crime and antisocial behaviour are intentional acts that cause physical or mental injury, as well as property loss or damage. Depending on the severity of the offence, a state or other authority may give punishment [1]. According to the [2], World Health Organization (WHO), crime and antisocial behaviour are global problems that impact both individuals and communities. It refers to any act that is punishable by law, such as theft, harassment, assault, or murder. Victims of crime and antisocial behaviour may suffer huge consequences, including physical harm, psychological trauma, and financial loss. It can also have broader social and economic consequences, such as higher healthcare expenses, decreased property prices, and slower economic growth. In addition, crime can instil fear and insecurity in communities, resulting in lower trust in governmental institutions and social cohesion [3]. As reported by [4], Crime Survey for England and Wales (CSEW) projects that in the year ending in June 2022, adults aged 16 and older will have committed 9.4 million crimes. According to the [5], United Nations Office on Drugs and Crime (UNODC), over 1.5 million killings occured worldwide in 2019. As written by [6], it is crucial to know the causes of crime and antisocial behaviour and to develop effective prevention and intervention strategies to address these problems. Collaboration between law enforcement, community organisations, and government organisations is required as a commitment to prevention-focused, evidence-based initiatives.

There are several ways, both traditional and using machine learning, to report crimes and other anti-social behaviour. The traditional way to report a crime or anti-social behaviour is to call the police or emergency services or go to a police station in person. As mentioned, [7] United Kingdom (UK) police, there are several traditional, easy ways to tell the police about a crime. In an emergency, like when someone's life is in danger or when a crime is happening, the victim must call 999. For things that aren't emergencies, you can call 101 or go to the nearest police station. You could also go to a police station in person or use the police department's website to report a crime or anti-social behaviour. Additionally, In the UK [8], In the last 10 years, a lot of money has been put into CCTV equipment so that crimes and other antisocial behaviour can be reported. There are about 4 million active CCTV cameras in use right now. However, the detection of crimes does not take place instantly because it requires human involvement and ongoing CCTV screen monitoring [9]. In contrast [10], using machine learning to report a crime means using automated systems to look at crime data and find patterns. This can help law enforcement agencies better decide where to put their resources and what areas need more attention. Machine learning can also be used to find possible suspects by analysing data and making predictions.

Traditional ways of reporting crimes have a several limitations that can affect how accurate crime statistics are and how well law enforcement works. Many crimes are not reported because people are afraid of getting in trouble or do not trust the police. This means that crime is under-reported. These methods can also be slow, ineffective, and only cover a small area, which can make it take law enforcement longer to respond and make it harder to catch the criminal. Moreover, these methods rely on human judgement, which can be subjective and prone to mistakes. This makes crime statistics, and the way law enforcement resources are used less accurate and more biased [11]. These problems show the need for different ways to report and control crime that deal with these problems and make crime reporting and control more accurate and efficient. However, in accordance with [12], using machine learning or deep learning to analyse crime data can save time and help find patterns and trends that might not be obvious to human analysts. This lets law enforcement agencies better allocate resources and focus on high-crime areas.

This research is primarily focused on the task assigned by UK Home Office for the creation of a deep learning architecture or model. The model will be capable of identifying crime and anti-social behaviour using audio recordings in a local area.

The format of this research paper as follows: The reporting of criminal activity, antisocial behaviour, etc. is covered in section II. Section III explains how and why using machine learning and audio data together will result in an improved solution. The audio dataset's data description and the necessary Exploratory Data Analysis (EDA) are covered in Section IV. A design strategy for data pre-processing, model architecture selection, discussion, and justification, model training, and evaluation metrics are included in the methodological section's final section.

# Background

Crime is a real threat to everyone. Criminal activity is any action or inaction that goes against the law and gets you in trouble. Even though some crimes cause little damage, others could kill. Crimes can happen anywhere, in small towns as well as in big cities, so there is no one place to stay away from them. To protect our society from any threats, we need to find a faster way to solve this problem [13]. Crime detection is important, and machine learning is becoming more and more popular to find and stop crime. Numerous organizations across the globe have been trying out these methods.

The researchers of prior studies experimented a variety of techniques for crime or anti-social behaviour detection. Research papers from these studies includes both machine learning and cut-edging deep learning approaches. In these approaches various types of data has been used to identify crime or anti-social behaviour, such as video surveillance data, audio data, social media data, crime report data and biometric data.

In [12], the suggested system is a web-based application that analyses real-time crime data presented as online news stories and provides a report of crime-related news. The system's input comes from the content of newspaper websites. Using a Python-based crawling application, the website is scanned, and the results are stored in a temporary database. The Support Vector Machine (SVM), Multinomial NB, and Random Forest classifiers were the three used by the authors of this research. The information is divided into data linked to crime and data unrelated to crime. SVM's average accuracy is 79.16%, Multinomial NB's average accuracy is 74.16%, and Random Forest's average accuracy is 85.83%. However, traditional machine learning-based algorithms are unable to produce key prime attributes from the crime dataset, they frequently fail to accurately forecast crime patterns [9].

The authors of this study [14], proposed a system, Intention Detection system, that detects the human supervisor to take the appropriate action when crime is detected in real-time recordings and photos. To notify the managers or the closest police station of a crime, they added an SMS sending mechanism. The suggested solution is constructed using the deep learning model VGGNet-19 that has been pre-trained to recognize knives and guns in the hands of people pointing towards other people. For training, they implemented two different pre-trained models, such as GoogleNet and InceptionV3. They discovered that the VGG19 method produced results that were more accurate in terms of training accuracy and used less processing time than GoogleNet. They employed the Fast RCNN and Faster RCNN algorithms for the bounding box over objects in photos including people, guns, and knives. However, due to the amount and complexity of the data, processing video data classification demands a lot of computational resources [15].

Recent year ambient audio/sound classification and detection are among the popular topics. Numerous studies have been conducted in a variety of fields, including environmental monitoring, undersea monitoring, health, building and design, animal tracking, and crimes [16]. For the crime detection, [17] presented a model for gunshot classification. The authors of this study used the hierarchical Gaussian mixture model (GMM) classification method for gunshot detection. They achieved 90% category classification in a total of 100 shots data sets in 10 different gun types. However, the limitations encountered in the detection of gun audios and the reliability of the techniques were shown. Consequently, researchers of this study [16], proposed an improved automated gunshot audios classification method. For the implementation of automatic gunshot classification approach, a novel gun audios dataset was acquired from YouTube and novel Machine Learning (ML) method has been used. The suggested finger-pat based feature generating network and IRF feature elector have been utilized to demonstrate the success of the generated and selected features using the classifier k-nearest neighbors (KNN). They used KNN to attain a classification accuracy of 94.48%.

In this paper [18], the algorithm that the authors developed can identify potential dangers in phone calls. For the machine learning system, they developed the first dataset of Bengali voice calls. Their system receives a voice call, analyzes the call using a Deep 1D Convolutional Neural Network (1DCNN), and employs a Multi-Layer Perceptron to determine whether there are any threats. The suggested system detected the crime calls with 91% precision, recall, and F1-score. They anticipated that these technologies will be useful for future investigations, voice conversation analysis, and threat predictions and assessments.

In [19], the authors of this study presented a method that uses audio classification to provide security for women and children. The victim's screams, which is audible from quite a mile away, alerted them to the danger. Screaming in response to various audio and visual signals may indicate danger. The dataset is used to train six different classification models (MobileNetV2, InceptionV3, Xception, DenseNet121, ResNet50, and ResNet101), which classify the data into three categories. (Normal, Woman in Danger, Child in Danger). From their experiments they conclude that DenseNet121 performs better than other models when it comes to detecting normal audio with 99.47% accuracy.

According to earlier research, experts have been looking into several techniques for years to accomplish the identification of crime using machine learning.

# Considerations

When it comes to detecting anti-social behaviour or crimes, audio classification data and machine learning methods has a number of benefits over other type of data and traditional methods. In some cases, audio classification is a superior option because it can reveal more details about the environment or situation being studied. According to the [20], visual analysis is impossible or insufficient, audio categorization may be used. For instance, when analysing sounds from underwater or in dimly lit areas, audio classification can offer important information that would be difficult or impossible to acquire through visual analysis alone. Additionally, for a more complete image of the surroundings, audio signals can be combined with additional sensors [21] because they are less impacted by dim lighting or blocked views [22]. In addition, other data, like video footage may not be able to catch all the information that audio data can, including hidden or obscured behaviour. Audio analysis may be less computationally intensive and easier to conceal than other machine learning techniques, such as text or visual recognition [21], making it helpful in covert surveillance uses. Overall, audio data can provide a more comprehensive and detailed understanding of the acoustic environment, making it a better solution for the detection of anti-social behaviour or crimes.

In addition to the benefits of audio data, machine learning can improve the precision and effectiveness of systems that identify criminal activity or antisocial behaviour. Machine learning is the effective approach for classifying audio data because it can automatically discover complex patterns in huge amounts of data, which would be challenging for humans to do directly. Additionally, traditional approaches to audio classification frequently depend on manually created features and rules, which can be time-consuming and may not fully capture the information in the data. As written in [23], in contrast to earlier systems that depended solely on manual classification, for accurately identification within audio recordings, machine learning are preferable options due to their ability to automate the process and make it more efficient. Furthermore, a common challenge in acoustic monitoring is foreground and background noise, which can also be handled by feature engineering using machine learning approaches.

Overall, machine learning is an effective approach for the analysis of audio data because it provides a scalable and affordable solution for automatic acoustic classification.

# Data description

The dataset has been provided by UK Home office is an audio dataset with 8732 labelled sound recordings that are all no longer than 4 seconds each. The audio files contain urban sounds from 10 various classes includes street music, air conditioner, car horn, children playing, dogs barking, drilling, idle engines, gunshots, jack hammers. The dataset has been pre-sorted into ten folders, and there is also a CSV file with metadata about each excerpt. The metadata contains details about the audio file, such as its name, the recording's sound ID, the start and finish times of each slice, and the class ID of the sound.

A picture containing chart

Description automatically generated

**Figure – 1** Ten folds with class distribution in each fold

Figure 1 shows all the ten folds, and each fold contains all the classes. In the dataset there are 10 classes which means that it’s a multi-class problem.

It has been suggested that for the evaluation use 10-fold cross-validation because classes in the dataset are not balanced.

## Exploratory Data Analysis

#### Exploratory Data Analysis (EDA) is a process that involves analyzing datasets to identify and summarize their primary characteristics, frequently with the help of statistical and visualization techniques. EDA seeks to find patterns, trends, anomalies, relationships, and insights in the data as well as possible issues and areas suitable for additional investigation. Python is a popular language for doing EDA because it is easy to use, flexible, and have many libraries for data processing and visualization. Python has been used in this work to check the insights of the urban sounds dataset. Let’s check the insights by importing the dataset and metadata of the dataset in python.

**Dataset metadata Information**

A screenshot of a computer

Description automatically generated with low confidence

**Figure – 2** Metadata information of the dataset

Figure 2 shows the metadata information of the urban sound’s dataset. It can be seen from the figure that there are 8 columns in the metadata includes slice\_file\_name, fsID, start, end, salience, fold, classID and class name. Figure 2 also shows the number of occurrences and the datatypes. In the metadata, slice\_file\_name and class have object datatype, and start and end have float datatype while rest of the column have integer datatype.

Table

Description automatically generated with medium confidence

**Figure – 3** First five rows of metadata

Figure 3 represents the first five rows of the metadata. First column has filename of audio files with file extension wav, second column has sound Id, third and fourth column has the start and end time of the sound of specific class, for example, dog\_bark if we see the first observation. The rest of the column has salience rating of the sound, fold number of the sound file, class Id and the class name of the sound clip.

**Missing values**

**Chart

Description automatically generated**

**Figure – 4** Missing values in the metadata

Figure 4 demonstrates that there are no missing values or missing cells in the datasets, indicating that no missing values handling methods are required.

**Audio class distribution**

Chart, pie chart

Description automatically generated

**Figure – 5** Class distribution in the audio dataset

Figure 5 represents the classes of target variable “class” in the dataset. The above pie chart shows the class imbalance in the dataset. The sounds of gun\_shot is just 4.3%, car\_horn is 4.9%, siren is 10.6%, and rest of the classes has 11.5% in the urban sound’s dataset. These classes are not balanced therefore class balance is significant because it facilitates model training by preventing the model from becoming biased towards one class.

**Audio amplitude analysis**

**Timeline

Description automatically generated**

**Figure – 6** Amplitude analysis of class gunshot

Figure 6 demonstrates the three different samples of waveform of audio class gunshot, which have amplitude peaks at regular intervals. However, we can see that various samples of the same sound class can have different amplitude levels. As a result, sound amplitude is not a helpful attribute for classifying different sounds. It lacks sufficient specificity to distinguish between the distinct sound classes.

**Log magnitude spectrogram analysis**

The frequency content of an audio file can be represented over time using the log-magnitude spectrogram. Each point's intensity on the spectrogram indicates how strong the signal was at that specific frequency and moment.

**Chart

Description automatically generated with medium confidence**

**Figure – 7** Log magnitude spectrogram of class gunshot

Figure 6 demonstrates the log magnitude spectrogram of the class gunshot. The frequency content of the audio stream can be seen visually as a log-magnitude the spectrogram over time. With a logarithmic scale, the x-axis shows time in seconds and the y-axis frequency in Hz. Each point's color on the plot corresponds to the magnitude or intensity of a particular frequency component at that precise instant, with stronger components denoted by brighter colors.

Overall, the amplitude, frequency content, and general structure of the audio file can all be understood by looking at the waveform and spectrogram.

# Methodology

This section of the paper is divided into four parts. In first part there will be design plan for data pre-processing, second part will be about the chosen model architecture selection and justification, in the third part there will be a discussion on model training and hyperparameter selection, in the last part evaluation metrics will be discussed.

## Data pre-processing

An important stage in the machine learning pipeline is data pre-processing, which involves transforming data into a format that machine learning algorithms can use to learn from.

Data pre-processing's primary goal is to clean, normalize, and transform the data to eliminate any errors, inconsistencies, or missing values and to make it more usable for machine learning algorithms.

**Resampling**

In urban’s dataset some of the audio files sampled at 44.1k Hz. In the next work I will resample these audio files from 44.1 kHz to 22.05k Hz by using ‘librosa.resample’ method of ‘librosa’ library of python. The purpose of resampling to reduce noise form the high frequency data and increase processing time.

**Data augmentation**

After the re-sampling the next step is to augment the data so I can balance my target classes. From figure 5, the class distribution of audio files across different classes are not balanced. Three classes gun\_shot, car\_horn and siren need to be balanced. I will be applying data augmentation technique to balance these audio samples in the dataset.

Data augmentation will be done by shifting the pitch of sound samples and by adding the background noise in the samples.

**Feature Extraction**

Waveforms of raw audio signals are high in dimensions and redundant as well which can make it difficult to work with the signals directly. Due to this I will be extracting useful features from the audio signal to reduce the dimensionality and get relevant information. Following are the features which will be extracted from the audio signal, 'mfcc' (for speech and audio analysis), 'chroma\_stft' (for analysing harmonic data), 'rmse' (for noise reduction), 'spectral\_centroid' (for brightnes of sound), 'spectral\_bandwidth' (for making difference between two sounds), and 'spectral\_rolloff' (for upper frequency limit). These features will be extracted by using ‘librosa.feature’ method of ‘librosa’ python library.

**Normalization**

After the feature extraction the next step is normalization of the features. Normalisation can help to improve the performance of the machine learning models by reducing the impact of variations in the amplitude and scaling of audio signals.

Normalisation will be applied to MFCCs (Mel Frequency Cepstral Coefficients) of the audio signal in the dataset. The mean and standard deviation will be calculated of the MFCCs frames to compute the normalised MFCC representation of audio files.

**Splitting the dataset**

After normalisation the final step of data pre-processing will be the splitting of dataset into training, validation, and test sets. The purpose of data splitting is to evaluate the performance of the audio-based machine learning model on a new data which have not seen before.

For the urban sound’s dataset, I will be using 10-fold cross validation because it has many advantages like it can deal with limited data size, class imbalance and can reduce the impact of random initialization by training and evaluating the model multiple times.

The dataset will be split into 10 folds, where each fold will be act as validation set and rest 9 will be as training set. The model will be trained on 9 folds and will be evaluated on the remaining fold. As our dataset in in 10 folds so this process will be repeated 10 times. After this performance of the model will be calculated by taking the average of these evaluations.

## Model Architecture Selection and Justification

There are several machine learning and deep learning model architectures that have been developed so far for the audio classification. Each model has a particular application that is dependent on the type of problem and dataset.

The dataset of urban sound lies on the category of supervised machine learning. For the classification of urban sound dataset, I will be using following deep learning model architectures because deep learning models performed well in the audio classification as compared to traditional approaches.

* Long Short-Term Memory (LSTM)
* Convolutional Recurrent Neural Networks (CRNNs)
* 1D Convolutional Neural Networks (1DCNN)

**Long Short-Term Memory (LSTM)**

LSTM is an improvised form of Recurrent Neural Network (RNN) which can be used for sequential data processing problems. As audio signals have series of sequences, LSTM can be used for the audio classification datasets. LSTM can capture long-term dependencies from the audio signals which can be effective for audio classification. However, LSTM networks can be expensive in terms of computational resources if there will be large datasets.

LSTM requires audio features such as MFCCs or spectrograms as an input to the network. After this the network will process the features considering the information from the previous steps. Overall, LSTM can be very effective for the audio classification as it can get the relationship between the audio features over time.

**Convolutional Recurrent Neural Networks (CRNNs)**

CRNN is a hybrid architecture of CNN and RNN which can be used for the audio classification. CRNN use the strengths of both CNN and RNN. It uses CNN to extract the features from the audio signals and use RNN to capture the dependencies in the audio features.

CRNN will use first CNN to extract features from the audio signals and after capturing the important information it will insert this output to RNN. Then RNN will process the data into sequence to remember the information from the previous inputs. Finally, the output from RNN will be used for the classification decisions. Overall, CRNN can be used for the classification of audio signals more accurately by combining the strengths of two different architectures.

**1D Convolutional Neural Networks (1DCNN)**

1DCNN is a type of neural architecture which can also be used for the classification of audio data. 1DCNN also requires features such as MFCCs as an input to the architecture. The advantage of using 1DCNN is computationally efficient. It requires less memory as compared to the other architectures such as LSTM and CRNNs. Overall, 1DCNN can be good option for the audio classification problem especially in terms of computational resources.

## Model Training and Hyperparameter Selection

For the evaluation I will use k-fold cross validation technique having 10 folds. The dataset will be divided into 10 equal parts or folds. The model will be trained on the 9 folds and the remaining 1 will be used for the validation. The process will be repeated till each fold used once for the validation. The performance of the model will be evaluated during each iteration and average will be taken after running all the 10 iterations. During each training iteration the model’s performance will be monitored and will use techniques like early stopping if there is no enhancement. Monitoring can also be used to check the learning rate and to prevent overfitting.

I have selected LSTM, CRNN and 1DCNN deep learning architectures for the modelling. Let’s discuss the model hyperparameter section of each model one by one.

**Long Short-Term Memory (LSTM)**

The training of LSTM model for the classification of urban sounds dataset requires various hyperparameters like number of layers, number of neurons in each layer, activation function such as sigmoid and tanh, batch size, training epochs, learning rate and dropout rate.

**Convolutional Recurrent Neural Networks (CRNNs)**

The training of CRNN model for the classification of urban sounds dataset requires various hyperparameters like number and size of convolutional filters, number and size of recurrent layers, optimizer (such as Adam and SGD), activation function (such as ReLU, sigmoid and tanh), batch size, training epochs, learning rate and dropout rate.

**1D Convolutional Neural Networks (1DCNN)**

The training of 1DCNN model for the classification of urban sounds dataset requires various hyperparameters like number of convolutional layers, number of filters, optimizer (such as Adam and SGD), activation function (such as ReLU, sigmoid and tanh), batch size, pooling layer, training epochs, learning rate and dropout rate.

The hyperparameters need tuning otherwise it can impact the performance of the model if the best values are not identified. For the hyperparameter tunning there are many techniques like GridSearch and RandomSearch. GridSearch search the hyperparameters by applying all the possible combinations of hyperparameters values whereas RandomSearch randomly sample the hyperparameter values for a specified number of iterations. Consequently, I will be using RandomSearch for the hyperparameters tuning because GridSearch can be computationally expensive for the large datasets.

## Model Evaluation

When the model has been trained on the dataset, next step is to evaluate the model performance using evaluation metrics. The performance of LSTM, CRNN and 1DCNN on the urban sound dataset can be evaluated using the following metrics.

* Accuracy
* Precision
* Recall
* Specificity
* F1 score
* Confusion matrix

The above metrics, confusion\_matrix, accuracy\_score, precision\_score, recall\_score, f1\_score can be imported from ‘sklearn.metrics’, library of python, and can be used for the evaluation of the model by just passing test data and predicted results. In order to find ‘Specificity’ I will use ‘TN / (TN + FP)’ formula of it. True Positive, True Negative, False Positive and False Negative can be extracted from the confusion matrix.

The classification report will also be used to determine the evaluation of each class of urban sounds dataset.

Once the evaluation of each model will be done, I will compare the results of each deep learning architecture and determine which architecture yields the best results.

# Conclusion

In conclusion, crime and anti-social behaviour are significant issues that affect individuals and communities worldwide. The detection and prevention of crime are essential for maintaining public safety and social cohesion. Machine learning approaches, particularly those using audio data, have shown promise in improving crime detection and prevention. This paper proposed deep learning architectures (LSTM, CRNN and 1DCNN) for identifying crime and anti-social behavior using audio data in a local area. In this paper, data pre-processing, model architecture selection, and evaluation metrics has been discussed. The study also discusses evaluation metrics of proposed deep learning architectures which includes accuracy, precision, recall, specificity, F1 score, and confusion matrix. Overall, the proposed architectures have the potential to be an effective tool for crime detection and prevention.

##### References

[1] N. Shah, N. Bhagat, and M. Shah, “Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention,” *Visual Computing for Industry, Biomedicine, and Art*, vol. 4, no. 1. Springer, Dec. 01, 2021. doi: 10.1186/s42492-021-00075-z.

[2] World Health Organization, “Understanding and addressing violence against women,” 2012.

[3] J. Q. Wilson and G. L. Kelling, “Broken Windows The police and neighborhood safety,” 1982.

[4] Crime in England and Wales, “Crime in England and Wales year ending June 2022,” 2022.

[5] United Nations Office on Drugs and Crime, “GLOBAL STUDY ON HOMICIDE,” 2019.

[6] L. W. Sherman and D. Weisburd, “General deterrent effects of police patrol in crime ‘HOT SPOTS’: A randomized, controlled trial,” *Justice Quarterly*, vol. 12, no. 4, pp. 625–648, 1995, doi: 10.1080/07418829500096221.

[7] POLICE.UK, “Contact the police \_ Police.uk”, Accessed: Feb. 14, 2023. [Online]. Available: https://www.police.uk/pu/contact-the-police/

[8] P. Miller *et al.*, “Intelligent sensor information system for public transport - To safely go...,” in *Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, 2010, pp. 533–538. doi: 10.1109/AVSS.2010.36.

[9] A. Palanivinayagam, S. S. Gopal, S. Bhattacharya, N. Anumbe, E. Ibeke, and C. Biamba, “An Optimized Machine Learning and Big Data Approach to Crime Detection,” *Wirel Commun Mob Comput*, vol. 2021, 2021, doi: 10.1155/2021/5291528.

[10] E. Ingram, F. Gursoy, and I. A. Kakadiaris, “Accuracy, Fairness, and Interpretability of Machine Learning Criminal Recidivism Models,” Sep. 2022, [Online]. Available: http://arxiv.org/abs/2209.14237

[11] M. Favaretto, E. de Clercq, and B. S. Elger, “Big Data and discrimination: perils, promises and solutions. A systematic review,” *J Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0177-4.

[12] S. Ghankutkar, N. Sarkar, P. Gajbhiye, S. Yadav, D. Kalbande, and N. Bakereywala, “Modelling Machine Learning for Analysing Crime News,” in *2019 6th IEEE International Conference on Advances in Computing, Communication and Control, ICAC3 2019*, Dec. 2019. doi: 10.1109/ICAC347590.2019.9036769.

[13] R. B, “Crime Activity Detection Using Machine Learning,” *Int J Res Appl Sci Eng Technol*, vol. 10, no. 7, pp. 3791–3795, Jul. 2022, doi: 10.22214/ijraset.2022.45860.

[14] IEEE Circuits and Systems Society. India Chapter and Institute of Electrical and Electronics Engineers, *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET).*

[15] M. S. Islam, M. S. Sultana, U. K. Roy, and J. al Mahmud, “A review on Video Classification with Methods, Findings, Performance, Challenges, Limitations and Future Work,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 6, no. 2, p. 47, Jan. 2021, doi: 10.26555/jiteki.v6i2.18978.

[16] T. Tuncer *et al.*, “AN AUTOMATED GUNSHOT AUDIO CLASSIFICATION METHOD BASED ON FINGER PATTERN FEATURE GENERATOR AND ITERATIVE RELIEFF FEATURE SELECTOR A Watermarking Application for Authentication of Holy Quran View project EEG-based Emotion Recognition Using Deep Learning Techniques View project AN AUTOMATED GUNSHOT AUDIO CLASSIFICATION METHOD BASED ON FINGER PATTERN FEATURE GENERATOR AND ITERATIVE RELIEFF FEATURE SELECTOR,” 2021. [Online]. Available: https://orcid.org/0000-0002-8380-7891

[17] S. Khan, A. Divakaran, and H. S. Sawhney, “Weapon identification using hierarchical classification of acoustic signatures,” in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VIII*, May 2009, vol. 7305, p. 730510. doi: 10.1117/12.818375.

[18] K. Saifullah, M. M. Alam, M. Joy, I. Hasan, S. Islam, and N. Hossain, “Deep Learning based Crime Detection and Resource Creation Approach From Bengali Voice Calls.” [Online]. Available: https://tinyurl.com/detecThreats

[19] M. Ashikuzzaman, A. A. Fime, A. Aziz, and T. Tasnima, “Danger Detection for Women and Child Using Audio Classification and Deep Learning,” in *2021 5th International Conference on Electrical Information and Communication Technology, EICT 2021*, 2021. doi: 10.1109/EICT54103.2021.9733601.

[20] J. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research,” in *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, Nov. 2014, pp. 1041–1044. doi: 10.1145/2647868.2655045.

[21] M. A. Tayebi, U. Glausser, and P. L. Brantingham, “Learning where to inspect: Location learning for crime prediction,” in *2015 IEEE International Conference on Intelligence and Security Informatics: Securing the World through an Alignment of Technology, Intelligence, Humans and Organizations, ISI 2015*, Jul. 2015, pp. 25–30. doi: 10.1109/ISI.2015.7165934.

[22] J. Rex, “Audio Event Detection and Localisation for Directing Video Surveillance-A Survey of Potential Techniques.”

[23] C. Chalmers, P. Fergus, S. Wich, and S. N. Longmore, “Modelling Animal Biodiversity Using Acoustic Monitoring and Deep Learning.” [Online]. Available: https://www.xeno-canto.org/