Detecting Anti Social Behaviour(ASB) by Analysing Audio Recordings Using Deep Learning Algorithms

*Abstract*— This paper explores the application of Deep Learning (DL) techniques in the detection of crime and antisocial behavior using environmental audio recordings. The study explores the problems caused by anti-social behavior. Moving on it discusses how the problems have been tackled in the past to prevent ASB. The paper presents a critical evaluation of the applicability of advanced DL concepts and techniques in the development of a reliable and accurate model for detecting crime and antisocial behavior. The methodology includes pre-processing tasks, feature extraction, DL model architecture, and evaluation methods. The study aims to showcase the merits of advanced DL techniques and their appropriate utilization in developing a model for detecting crime and antisocial behavior using environmental audio recordings.

Keywords—component, formatting, style, styling, insert (key words)

# Introduction (*Heading 1*)

Deep Learning (DL), which has developed as a promising subject in machine learning research in recent years, has the potential to tackle complicated problems and extract valuable insights from massive datasets. Deep Learning has emerged as a promising topic in machine learning research. DL has been successfully implemented in a variety of domains, including natural language processing, computer vision, speech recognition, and autonomous systems, to name a few. The advent of ever-more-powerful computational capabilities and the availability of enormous volumes of data has made DL an indispensable instrument for the solution of a broad variety of issues that are encountered in the real world. The application of deep learning has already had a significant influence on a variety of industries, including medicine, aerospace, social media, the military, agriculture, and many more[1]. These applications make use of machine learning as well as deep learning models to either make predictions or make discoveries, depending on the type of answer that is required for their particular issues. The identification of criminal activity and other antisocial behaviour through the use of ambient audio recordings is one of the domains in which DL has the potential to have a big effect. The United Kingdom's Home Office has requested the creation of a model for machine learning that can analyse audio recordings in order to identify criminal activity and other antisocial conduct occurring in the surrounding region. The objective is to record the ambient sound using either ad-hoc equipment, like a mobile phone, or fixed installation equipment, such waterproof microphones.

The negative impact of anti-social behavior on individuals and communities is well-documented. The research by [2] effects of long-term antisocial behavior (ASB) victimization, which is often overlooked as a sub-criminal domain. The research highlights that ASB victimization can have similar negative impacts as violent crimes and domestic burglary. The study also refutes the idea that ASB is a "low-level" crime, emphasizing the need for a qualitative approach to prioritize victims' narratives. However, the limited literature on the experiences of ASB victims highlights the need for further research to fully comprehend the effects of victimization on a person's life and identity. It is also crucial to consider the perspectives of practitioners to identify any barriers that prevent effective responses to ASB victimization. Without this understanding, meaningful solutions cannot be created.It can lead to social isolation, fear, and decreased quality of life for those who are victimized by it. It can also contribute to a breakdown in social cohesion and trust, and may ultimately lead to the deterioration of neighborhoods and communities.

In the real world, reporting illegal activity and antisocial behaviour is typically done in the form of a whistle-blower fashion. For instance, if someone hears a gunshot, they will call the police to inform them of the incident. This method is inefficient and time-consuming, which can lead to delays in police response times and potentially dangerous situations. As a result, the creation of a model that is dependable and accurate in identifying these behaviours through the use of audio recordings has the potential to have a substantial influence on society.

This paper presents an in-depth exploration of the theoretical principles and objectives of advanced DL techniques, including a critical evaluation of their applicability in the development of a crime and antisocial behaviour detection model. Additionally, the paper presents an in-depth investigation of the theoretical principles and objectives of advanced DL techniques. In order to determine which DL strategies are most suited for our model, we will do research on the current software and hardware solutions that have been published in the past and analyse the machine learning algorithms that were employed in those articles. In addition, we will present a comprehensive analysis of the data that was utilised in this investigation, including a discussion of its structure, features, and class distributions.

In the section on methodology, the preprocessing tasks, such as the methods of data cleaning and feature extraction, will be described. After that, we will proceed to give an overview of the architecture of the DL model, which will include a discussion of the hyperparameters that were selected as well as the evaluation methods that were used to analyse the performance of the model. In addition, we will go over the analysis of time-series data by employing the suitable DL architectures, and we will defend each and every pick that was made for this model. In general, the purpose of this article is to demonstrate a profound and holistic grasp of advanced DL principles, as well as to critically analyse and establish the application of a variety of advanced DL concepts and methodologies. With the help of this study, we will demonstrate the benefits of more sophisticated DL approaches and how they can be employed effectively to create a model that is dependable and accurate for identifying criminal activity and antisocial behaviour via ambient audio recordings.

# Background

In the domains of criminology, psychology, sociology, and public policy, among others, substantial research and discussion have been conducted on the subject of crime because of its complexity and the many facets it possesses. Any action that harms or distresses other people or disrupts the social fabric of a community is considered anti-social behaviour, which is a subcategory of the crime category known as criminal activity. There is now research being conducted in the realm of crime in India by [3] that focuses on managing the many sorts of crimes to assist the police split their workload in an appropriate manner. In this work, machine learning models are utilised to classify the data that was obtained from police reports.

The investigation of criminal activity and antisocial conduct is a significant challenge that is faced by many cultures all around the world. It has enormous repercussions for the quality of life of people, as well as for the safety and security of the general public. In addition, antisocial behaviour may be characterised as activities that show a disrespect for the health and safety of other people. It encompasses a wide variety of disruptive behaviours, the degree of which ranges from annoyances like noisy or disorderly behaviour to illegal behaviours like vandalism or physical assault [4]. During this time, the anti-social behaviour was gaining a great foothold throughout social media platforms. According to the study done by [5], The appearance of antisocial behaviour online is a developing worry that discourages participation and fosters a variety of societal problems. The expression of antisocial behaviour online can take many different forms. Some of the most common types of antisocial behaviour that may be observed on the internet include a disrespect for lawful behaviour, impatience and aggression, a disregard for safety, and a lack of remorse. Cyberbullying is one form of the antisocial behaviour that occurs on the internet, which is a phenomena of everyday malice that occurs when people seem to enjoy themselves at the price of the agony and anguish that they cause to others. In addition, the research demonstrates that this pattern of behaviour may be recognised through the use of text categorization with deep learning and machine learning. Despite the fact that this is a great tool for identifying antisocial behaviour, it does have its limitations, such as not being able to detect behaviour on social media platforms; as a result, crimes in the real world that are caused by antisocial behaviour continue to occur.

Another piece of research has been conducted by [6] that focuses on differentiating between different categorical classes in various contexts, including antisocial behaviour disorders, through the application of machine learning classifiers. These classifiers are used to differentiate between different types of behaviour. This method has the benefit of being able to identify complicated combinations of traits that are highly predictive of an illness or clinical result. On the other hand, it is unable to differentiate between variables of interest and those that humans would regard to be confounding factors. The article also cites studies that have successfully used different data sources to differentiate between adults with psychopathy or antisocial personality disorder and healthy controls. These data sources include content from social media, Near Infrared Spectroscopy, speech patterns, videotaped head motion, electroencephalogram frequency data, and histories of childhood abuse and caregiving. Although though there is a lot of intricacy and controversy around psychopathy diagnoses, the information on the criteria that are used to define psychopathy is frequently left out of research that concentrate on machine learning methods rather than psychopathy, as the article points out. In conclusion, the paper proposes that classifiers based on MRI might have the potential to improve both the diagnosis of mental health illnesses and the treatment of those diseases.

As a result of investigating other relevant work and doing research by [7] The frequency of crimes committed online, known as cybercrime, is on the rise, and the internet is frequently utilised as a weapon in the commission of these offences. The authors have created a framework that can be used to identify, detect, and categorise cybercrime offences such as organised and unstructured attacks. This framework may be used to identify, detect, and classify cybercrime offences. The purpose of the framework is to give a comprehensive knowledge of cybercrime offences in society and to examine security vulnerabilities using techniques derived from machine learning. This will assist in reducing the likelihood that such offences will occur. According to the findings, the framework cuts down on the amount of time and effort needed to report occurrences. Also, it has the ability to identify hotspots in which cybercrime is prominent, which enables preventative steps to be done. This data may be utilised to make predictions regarding instances and to devise preventative actions to thwart cybercrime in particular regions.

It is generally accepted understanding that antisocial conduct is a complicated problem that is caused by a number of different variables. According to the findings of many pieces of study, this behaviour is thought to be the result of a confluence of individual, societal, and environmental elements[8]. Individual variables such as difficulty with self-regulation, impulsiveness, a lack of empathy, and a history of traumatic experiences or abuse can all contribute to the development of anti-social conduct [9]. The development of antisocial conduct can also be influenced by societal variables such as poverty, social marginalisation, and a deficiency of positive role models [10]. However, environmental variables such as poor living conditions, high crime rates, and restricted access to education and work opportunities can play a key part in the manifestation of anti-social behaviour. This can be a contributing element in the development of anti-social conduct [10].

Because antisocial conduct may have severe effects on both individuals and society, there has been a significant amount of interest in the development of effective solutions to both avoid and deal with the issue. These strategies range from early intervention and prevention programmes designed to address the underlying causes of anti-social behaviour to more traditional approaches to law enforcement, such as increased police patrols and stricter sentencing guidelines. The goal of these programmes is to reduce the likelihood that anti-social behaviour will occur in the first place [11]. In recent years, there has been a rising realisation of the need of having a comprehensive and collaborative approach to treating anti-social behaviour(Redgrave, 2022), this recognition has been particularly prevalent in the United Kingdom. This method involves bringing together a variety of stakeholders to establish comprehensive and integrated plans to prevent and treat anti-social behaviour. These stakeholders include community groups, law enforcement agencies, local governments, and providers of health and social services. In spite of the substantial amount of attention that has been paid to anti-social conduct in both study and policy, it continues to be a pervasive and difficult problem in many communities. This underscores the necessity of continuing study, assessment, and the refining of measures to avoid and address this disease[12].

## Data Description

This investigation makes use of a dataset provided by urbansound8k, which is made up of the [13]. This is a universal dataset that contains a variety of common noises that may be heard in urban environments. This research made use of a dataset that included information on the sounds of an air conditioner, automobile horn, children playing, dog barking, drilling, engine idling, pistol fire, jackhammer, siren, and street music. The age and gender of the victim, the location of the occurrence, the sort of anti-social conduct that occurred, and whether or not the incident was reported to the police are some of the factors that are contained within the dataset. In addition, the dataset contains information on the victim's perception of the impact that the event had on them, such as if the occurrence caused them bodily or mental harm and whether it disrupted their normal day-to-day activities.

### Exploratory Data Analysis

The total number of observations for the dataset is gathered amounts to 8732. The data is already organised into 10 different folders, and the metadata, which is saved in csv format, does not include any blanks or missing values. This particular data has already been processed, at least partially. The data is provided as audio files in the wav format, and there is also a csv file that contains all of the metadata associated with the dataset. Waveform Audio File Format, also known as WAVE (or WAV in relation to its file extension), is a container format that was developed by Microsoft and IBM for the purpose of storing digital audio. According to research by [14], The canonical structure of wave files Headers of the files A WAV file's header is comprised of many blocks of information. To begin, there is the declaration block, which contains information about the wav file, such as the format it uses and the size of the file itself. This section was taken from the RIFF and inherited. After that is the section that explains the format of the audio file. The second block is more intriguing since it gives information such as the sample rate, which is measured in hertz, the number of channels, which ranges from one mono channel to two stereo channels, and other relevant information. And last but not least, the data block, which is where these blocks are also located and which stores the sample bytes for each channel.

The metadata file contains the attributes such as filename labelled as 'slice file name' with its subsequent class labelled as 'class' which records the class of the audio file with it's 'classID' which is the numerical value appointed to each class. Other attributes include class labelled as 'class' which records the class of the audio file with it's 'classID' which records the numerical value appointed to each class. The description also contains the number of the folder in which the audio file is located as well as the channels of the audio file that are labelled as 'salience.' The number "1" denotes mono sound, while the number "2" denotes stereo sound. Along with the start time of the slice in the original freesound recording, which is logged as 'start' the freesound id is labelled as 'fsID,' and the end time of the recording is likewise documented as 'end.' Table 1 provides an illustration of the table.

A screen shot of a computer

Description automatically generated with low confidence

Table

Upon further exploration the shape of data was checked using the data shape which clarified that there are total eight columns in the table and each of them consists of 8732 values this is illustrated in Table 2.

Graphical user interface, text

Description automatically generated

Table

After reading the description of the classes for further processing we have to us the librosa library of python to check the class balance of the audio files. The class balance of the model is disturbed hence we need to use a strategy for creating a class balance. The class balance is displayed in the figure 2.

Chart

Description automatically generated

Figure

Progressing further, we need to analyse the time amplitude signal of our dataset to check if we can classify on its basis. To analyse it first we have to select random samples from the dataset and visually identify if the amplitudes and wave patterns are similar of at-least three random samples of each class. As illustrated in figure 3 the time amplitude graph of gun shot audio files the magnitude of the amplitude has variance in data as well as wave patterns are not similar.

Graphical user interface, table, Excel

Description automatically generated

Figure

Hence, the last resort would be to use mel spectrogram and Mel-frequency cepstral coefficients (MFCCs).the study by [15] describes mel spectrogram and MFCCs as Spectrograms, on the other hand, are a visualization of the frequency content of an audio signal over time. They are created by dividing the audio signal into short overlapping frames, computing the power spectrum of each frame using the Fourier transform, and then stacking the spectra together to form a 2D image. Spectrograms can provide valuable information about the time-frequency structure of an audio signal and can be used as input to deep learning models for audio classification. The Mel-spectrogram is a type of visual representation of audio data that uses the human auditory system as a reference. It is based on the theory that the frequencies audible to the human ear are not perceived uniformly but rather are grouped into distinct melodies. This frequency scale used is called the Mel scale, it is based on the human perception of frequency and not on a linear scale in Hertz. The audio signal is analyzed into frequencies and each frequency band is converted into an intensity in dB, which is then represented in color to create an image (HeatMap). It uses a frequency scale called the Mel scale, which is based on human perception of frequency and not a linear scale in Hertz. The audio signal is analyzed into frequencies and each frequency band is converted into an intensity in dB, which is then represented in color to create an image. Mel-frequency cepstral coefficients (MFCCs) In order to understand what Mel-frequency cepstral coefficients are, we need to start by knowing what are cepstral frequencies as well as the Fourier Transform. Cepstral frequencies are representations of the audio signal that capture the spectral characteristics of the audio signal using a mathematical transformation, such as the Fourier Transform..The spectogram of a dog barking is illustrated in figure 3.

A picture containing text, night sky

Description automatically generated

Figure

The final step is to review the sampling rates of dataset, figure 4. Upon analysis of this the majority class is

A picture containing shape

Description automatically generated

Figure

## Considerations

Instead of using the tried and true methods previously stated, we can find a better way to detect antisocial behaviour by using audio data and machine learning. There is a plethora of information that can be gleaned from audio data that may be applied to the detection of antisocial conduct. Consider the speaker's inflection, the ambient noise level, and any other contextual circumstances that may affect the dialogue. Capturing this kind of information through textual exchanges or other data sources might be challenging. Although text-based categorization is widely employed in social media for monitoring antisocial behaviour and cyberbullying, it is unable to aid in the identification of a real-world event. Nuances in language and conversation can be recorded in audio data that would otherwise be lost. A caustic tone of voice or hesitancy in speech, for instance, might provide light on the speaker's attitude or motivation. Massive volumes of audio data may be used to train machine learning algorithms to recognise the traits and patterns of antisocial conduct.

These algorithms may be applied in real-time to analyse and categorise newly recorded audio in search of signs of antisocial conduct. In times of crisis, the ability to analyse talks in real time using audio data and machine learning is invaluable. For the time being, while we are still in development, we will not be segmenting the audio files after a certain amount of time has passed. This model's development would necessitate decomposing the gathered data into dataframes. Comparatively to more time-consuming and expensive methods, such as human review of recorded conversations, hiring additional staff to monitor conversations, or police car patrols in areas with high rates of antisocial behaviour, using machine learning algorithms to analyse audio data can be a cost-effective solution. Machine learning applied to audio data can be a more thorough and time-saving method of detecting criminal activity. It's a low-cost option for real-time conversation monitoring and it enables for the acquisition of crucial contextual information.

# Methodology

The first step in methodology is to import all of the concerned libraries which include the basic libraries, Model training and classification libraries and the libraries required to analyse the audio files. Basic libraries required for experimentation would be pandas, numpy, sklearn, matplotlib and seaborn. The libraries required for model training such as tensorflow keras for ML model to analyse the audio files the pyhton library available is librosa which allows to load the audio files in jupyter notebook. Moving forward, then we perform Data ETL to ensure its accuracy and consistency by performing EDA and This involves cleaning the data to remove any irrelevant information or data that may skew our results. Additionally, for data pre-processsing we will use techniques such as data imputation to fill in any missing data values. We will also perform feature engineering to extract useful information from our data and convert it into a format that is suitable for our chosen model architecture.

The pre-processing stage is a critical step in developing a machine learning model as it involves cleaning, transforming, and preparing the data to be used in the model. The quality of the pre-processing stage significantly affects the performance of the model, and therefore, it is essential to ensure that the data is well-prepared before feeding it to the model. In this study, the pre-processing stage involves several steps, including resampling, feature extraction, class balancing and data splitting.

## Resampling

Audio files may have different sample rates. It is very necessary to resample audio data in order to guarantee that all of the files have the same sample rate. The sampling rates 16 kHz, 22.05 kHz, 44.1 kHz, and 48 kHz are the most common ones. Either I can downsample the sample that represents the majority, or I can upsample the samples that represent the minority such that they have a sampling rate of 44.1 kHz. I noticed that the vast bulk of our data had a sample rate of 44.1 kHz. In accordance with [16] If I wish to resample our data to 22.05 kHz, while 44.1 kHz has superior audio quality, 22.05 kHz is good enough to differentiate between sounds for the purpose of sound categorization. In addition, the speed of our model will improve if the size of each data set is reduced to one-half of its initial size. The librosa library will automatically adjust the sample rate to 22.05kHz if it is not specified otherwise, therefore sampling at this frequency offers a distinct benefit. Also, the librosa library will, by default, change the number of channels to mono. Because of this, it is not necessary for us to monitor the channels of our system unless we intend to utilise it for categorization.

## Feature extraction

## The extraction of features is the first step in the pre-processing procedure. At this stage, the relevant characteristics from the dataset that are going to be used in the model will be identified and selected. The audio recordings and the labels that correlate to them, which indicate if the tape involves criminal activity or other anti-social conduct, are the aspects that are important to this study. The process of feature extraction is carried out by utilising the relevant techniques, such as spectrograms, Fourier transforms, and Mel Frequency Cepstral Coefficients (MFCCs). It is possible to implement MFCCs by first transforming our existing audio files into arrays and then using the array values to pull out features.[17, p. 5]

## Data split

The data collection for this investigation consists of 10 directories of audio files. The data has to be reshuffled and partitioned for the best possible model performance. To that end, the K-fold technique from the sklearn package is employed here. This technique separates the data at random into K equal-sized subsets, with one subset serving as a test set and the remaining subsets being utilised for model training. K iterations are performed so that each partition is used as the test set exactly once. The K-fold approach makes it possible to use a cross-validation score to evaluate the trained model's efficacy. Using the test partition you just made in each folder, this score represents how effectively your model generalises to new, unknown data. As the number of folders in the dataset is 10, [18] recommends setting K to that value. As a whole, the K-fold method is a helpful tool for shuffling and splitting data in a way that guarantees the best possible performance from a model. Using a cross-validation score, we can accurately evaluate the model's generalizability to fresh data and verify its performance.

## Class balancing

## The efficiency and reliability of the classification model are both impacted by how evenly the classes are distributed in the audio data. It's crucial to the model's accuracy because of how sensitive and specific it is. Inadequate representation of minority groups might result from an imbalance in the input data [19]. Many methods exist for equitably dividing up audio data into its component groups. Upsampling is a method of increasing the quantity of samples from a minority group by replicating them at random. Second, downsampling reduces the total number of samples by arbitrarily eliminating some of the samples from the majority class. Finally, data augmentation involves the generation of new samples by altering preexisting ones in some way, for as by lengthening or shortening the duration of the sample, altering the pitch, or adding noise. In a final step, we combine models that were each trained on a unique portion of the data in order to achieve class parity. In order to fairly assess the model's efficacy across all classes, it is essential that the validation and test sets be well-balanced. Given the small size of the underrepresented groups, we have decided to use upsampling rather than downsampling, which would reduce the amount of data available for model training and produce less accurate findings. Lastly, I can't enhance the data since it would modify the features and cause drift in the features.

## Model architecture

Accurate detection of suspicious behaviour (ASB) is essential in the setting of Automatic Speaker Recognition (ASR). Both false positives and false negatives in ASB detection have the potential to have devastating results, such as the wasteful use of home office resources or, in the worst case scenario, the loss of a human life. Consequently, selecting the right model architecture is crucial to reducing these mistakes. Three main model designs are proposed for ASB identification, each with its own set of benefits and drawbacks. They range in their efficiency and accuracy on tasks like voice recognition and categorization. In order to reduce the high cost of misclassification, it is crucial to make well-informed choices on the model architecture to use.

### RNN

When it comes to processing sequential data, such as time series, audio signals, and natural language processing, recurrent neural networks (RNNs) excel. By processing incoming data sequentially and keeping a hidden state that might carry information from previous stages, an RNN is able to capture temporal relationships thanks to its design. Several examples of applications where the sequence of the input data is crucial for creating correct predictions are discussed by Phan et al. (2017), including speech recognition, music production, language modelling, sentiment analysis, and machine translation. For instance, in speech recognition, the prediction of the following phoneme can be affected by the prediction of the phoneme before it, and in machine translation, the meaning of a phrase can be affected by the order in which its words appear. Furthermore, RNNs may be integrated with other neural network designs like Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs) to extract features from sequential data and improve the model's performance. The output of a convolutional neural network (CNN) trained to extract spatial characteristics from a picture may be input into a recurrent neural network (RNN) trained to capture temporal relationships and provide a series of predictions, for instance. To sum up, RNNs are an effective method for handling sequential data and have demonstrated positive outcomes in a number of different settings. They are often used for jobs involving sequential data due to their ability to capture temporal relationships and their compatibility with other neural network topologies.

### CNN

These are commonly used for image classification but can also be adapted for audio classification. The research by [20] CNNs can learn to extract relevant features from audio signals through convolutional layers and can be used with other layers such as pooling and dropout to improve performance. 1DCNNs are mainly used for processing time-series data such as audio, speech, or signals. They apply the convolution operation only along the time axis of the input data. The advantage of 1DCNNs is that they can capture temporal dependencies in the data effectively and require fewer parameters to train compared to 2DCNNs. However, they may not perform well on spatial data such as images, as they lack the ability to capture spatial information. 2DCNNs, on the other hand, are designed to process 2D spatial data such as images. They apply the convolution operation along both the height and width dimensions of the input data and are able to capture spatial patterns in the data. This makes them highly effective for image recognition and object detection tasks. However, 2DCNNs require significantly more parameters to train compared to 1DCNNs.

### Convolutional Recurrent Neural Network (CRNN)

This model architecture is the most accurate architecture in the terms of performance. The study by [21] explores CRNN is a type of neural network that combines the power of both Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). It is particularly useful for audio classification because it can capture both the temporal and spectral information present in audio signals. In audio classification, temporal information is crucial because audio signals are time-varying and the sequence of audio frames carries important information about the audio content. Spectral information is also important because audio signals are composed of multiple frequency components, and different types of audio have distinct frequency characteristics. CNNs are good at extracting spectral features from audio signals and can identify patterns in audio frames, while RNNs are good at capturing temporal dependencies between audio frames. By combining the two, CRNNs can effectively capture both spectral and temporal features in audio signals. Therefore, CRNNs can be a powerful tool to achieve high accuracy results for our problem.

## Hyperparameter selection

Experimentation and assessment will be used to determine the optimal values for the hyperparameters of the CRNN architecture suggested for use in audio categorization. Hyperparameters like dropout rate, learning rate, batch size, and the number of LSTM units will be optimised using grid search approaches. After a 1D convolutional layer, a CRNN employs two bidirectional LSTM layers and a fully linked layer for its design. By configuring 64 filters in a 1D convolutional layer, a simple feature extractor may be created. The 1D convolutional layer's pool size is fixed at 2, and its maximum kernel size is limited at 3. To do so, we use two 64-unit, bidirectional LSTM layers to record the time-based features of the audio stream. A dropout rate of 0.2 is employed to the LSTM layers to mitigate overfitting. A fully linked layer with 64 units and the ReLU activation function is applied to the output of the LSTM layers. With a softmax activation function, the probabilities of all classes are output at the output layer. This network is optimised using the categorical cross-entropy loss function and a learning rate of 0.001. All of these settings are defaults, and the model's performance will inform future adjustments. As a recap, the proposed CRNN architecture will have a 1D convolutional layer, two bidirectional LSTM layers, and a fully connected layer, and its hyperparameters will be chosen by means of grid search algorithms. For this example, we'll use a learning rate of 0.001 and a loss function based on the categorical cross-entropy to fine-tune the network. Whether or not the model's hyperparameters need adjusting depends on how well it performs..

## Model Evaluation

## According to [22]Many measures, including cross-entropy, sensitivity, specificity, accuracy, and cross-validation score, will be used to assess the quality of the suggested model. As an added step, we'll evaluate the model's efficacy using a confusion matrix and ROC curve. To quantify the discord between the labels and predictions, we shall employ cross-entropy. To measure how well the model can classify positive and negative data, we'll utilise sensitivity and specificity. The overall success of the model will be evaluated based on its level of accuracy. To verify the efficacy of our trained model, we will additionally use cross-validation score. The number of correct, incorrect, and unsuitable classifications will be counted in the confusion matrix to provide a comprehensive assessment of the model's efficacy. The model's efficacy in terms of the compromise between sensitivity and specificity will also be evaluated using the ROC curve. Overall, we plan to measure the quality of the suggested model with several different indicators, such as cross-entropy, sensitivity, specificity, accuracy, and cross-validation score. For a more in-depth evaluation of the model's efficacy, we'll use the confusion matrix and receiver operating characteristic curve.

## Libraries and APIs

TensorFlow, Keras, scikit-learn, and librosa are just some of the libraries and APIs that we plan to use to put into practise our suggested technique. These libraries provide a wealth of useful features and utilities that will speed up the process of putting our chosen model architecture into practise. Making use of these frameworks and APIs allows us to shorten the implementation process and guarantee that our model's architecture is well-suited for rapid and accurate training. In addition, these resources include a variety of features that will make it simple to preprocess our data, train our model, and assess its efficacy.

# Future work

In this part, we detail the methodology recommended for identifying ASB in recorded audio. To help law enforcement agencies cut crime rates, we want to use this method to create a reliable ASB detection model in our future research. We will train our ASB detection model using three distinct model architectures to accomplish this goal. Because of its improved feature extraction capabilities compared to 1DCNN, 2D Convolutional Neural Networks (2DCNNs) are widely employed for audio categorization. Second, a Long Short-Term Memory (LSTM) model, which excels in voice recognition and prediction, will be trained as a Recurrent Neural Network (RNN). Ultimately, a Convolutional Recurrent Neural Network (CRNN) will be trained, as it is a combination of RNN and CNN and may be used for both audio categorization and voice recognition. We'll use assessment measures like cross-entropy, sensitivity, specificity, and cross-validation scores to compare the performance of different models. Next, we will analyse the model's efficacy with tools like the confusion matrix and the Receiver Operating Characteristic (ROC) curve. TensorFlow, Keras, scikit-learn, and librosa are just a few of the libraries and APIs we'll be using to put our strategy into action. These libraries offer a rich set of features that will facilitate the smooth deployment of our selected model architecture.

##### References

[1] S. R. Sree, S. B. Vyshnavi, and N. Jayapandian, ‘Real-World Application of Machine Learning and Deep Learning’, in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Nov. 2019, pp. 1069–1073. doi: 10.1109/ICSSIT46314.2019.8987844.

[2] V. Heap, ‘Exploring the effects of long-term anti-social behaviour victimisation’, *International Review of Victimology*, vol. 27, no. 2, pp. 227–242, May 2021, doi: 10.1177/0269758020961979.

[3] V. Gupta, S. Shukla, and R. Rawat, ‘Crime Tracking System and People’s Safety in India using Machine Learning Approaches’, Feb. 2022.

[4] G. H. Schoenmacker *et al.*, ‘Identification and validation of risk factors for antisocial behaviour involving police’, *Psychiatry Research*, vol. 291, p. 113208, Sep. 2020, doi: 10.1016/j.psychres.2020.113208.

[5] R. Singh *et al.*, ‘Deep Learning for Multi-Class Antisocial Behavior Identification From Twitter’, *IEEE Access*, vol. 8, pp. 194027–194044, 2020, doi: 10.1109/ACCESS.2020.3030621.

[6] R. Pauli and P. L. Lockwood, ‘The computational psychiatry of antisocial behaviour and psychopathy’, *Neuroscience & Biobehavioral Reviews*, vol. 145, p. 104995, Feb. 2023, doi: 10.1016/j.neubiorev.2022.104995.

[7] R. Ch, T. R. Gadekallu, M. H. Abidi, and A. Al-Ahmari, ‘Computational System to Classify Cyber Crime Offenses using Machine Learning’, *Sustainability*, vol. 12, no. 10, Art. no. 10, Jan. 2020, doi: 10.3390/su12104087.

[8] NHS, ‘Antisocial personality disorder’, *nhs.uk*, Feb. 12, 2021. https://www.nhs.uk/mental-health/conditions/antisocial-personality-disorder/ (accessed Mar. 14, 2023).

[9] Cleveland clinic, ‘Antisocial Personality Disorder: Causes, Symptoms & Treatment’, *Cleveland Clinic*, 2023. https://my.clevelandclinic.org/health/diseases/9657-antisocial-personality-disorder (accessed Mar. 14, 2023).

[10] J. Murray *et al.*, ‘Risk Factors for Antisocial Behavior in Low- and Middle-Income Countries: A Systematic Review of Longitudinal Studies’, *Crime Justice*, vol. 47, no. 1, pp. 255–364, Mar. 2018, doi: 10.1086/696590.

[11] College of Policing, ‘Targeted approaches to crime and disorder reduction’, *College of Policing*, May 01, 2021. https://www.college.police.uk/research/what-works-policing-reduce-crime/targeted-approaches (accessed Mar. 14, 2023).

[12] S. Harradine, J. Kodz, F. Lemetti, and B. Jones, ‘Defining and measuring anti-social behaviour’.

[13] ‘UrbanSound8K’, *Urban Sound Datasets*. https://urbansounddataset.weebly.com/urbansound8k.html (accessed Mar. 14, 2023).

[14] ‘Microsoft WAVE soundfile format’. http://soundfile.sapp.org/doc/WaveFormat/ (accessed Mar. 14, 2023).

[15] J. Jogy, ‘How I Understood: What features to consider while training audio files?’, *Medium*, Jun. 17, 2021. https://towardsdatascience.com/how-i-understood-what-features-to-consider-while-training-audio-files-eedfb6e9002b (accessed Mar. 14, 2023).

[16] R. Kim, ‘Urban Sound Classification — Part 2: sample rate conversion, Librosa’, *Medium*, Jun. 23, 2021. https://towardsdatascience.com/urban-sound-classification-part-2-sample-rate-conversion-librosa-ba7bc88f209a (accessed Mar. 14, 2023).

[17] ‘FIGURE 5. The output images by using Mel-Spectrogram and MFCC.’, *ResearchGate*. https://www.researchgate.net/figure/The-output-images-by-using-Mel-Spectrogram-and-MFCC\_fig5\_342794354 (accessed Mar. 14, 2023).

[18] M. Kelechava, ‘Urban Sound Classification with Librosa — tricky cross-validation’, *Medium*, Oct. 15, 2020. https://towardsdatascience.com/urban-sound-classification-with-librosa-nuanced-cross-validation-5b5eb3d9ee30 (accessed Mar. 14, 2023).

[19] H. Tripathi, ‘What Is Balance And Imbalance Dataset?’, *Analytics Vidhya*, Jan. 18, 2021. https://medium.com/analytics-vidhya/what-is-balance-and-imbalance-dataset-89e8d7f46bc5 (accessed Mar. 14, 2023).

[20] P. Nandi, ‘CNNs for Audio Classification’, *Medium*, Dec. 10, 2021. https://towardsdatascience.com/cnns-for-audio-classification-6244954665ab (accessed Mar. 14, 2023).

[21] E. Çakır, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen, ‘Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection’, *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 25, no. 6, pp. 1291–1303, Jun. 2017, doi: 10.1109/TASLP.2017.2690575.