International Journal of Knowledge-Based Development

ISSN online: 2040-4476 - ISSN print: 2040-4468
https://www.inderscience.com/ijkbd

Crime detection and crime hot spot prediction using the BI-LSTM deep learning model

A. Kalai Selvan, N. Sivakumaran

DOI: 10.1504/IJKBD.2024.10062799

Article History:

Received:	13 August 2023
Last revised:	14 January 2024
Accepted:	10 February 2024
Published online:	27 March 2024

Crime detection and crime hot spot prediction using the BI-LSTM deep learning model

A. Kalai Selvan* and N. Sivakumaran

Department of Instrumentation and Control Engineering & Head,
Centre of Excellence Emergency Response Support System (CoEERSS),
National Institute of Technology,
Tiruchirappalli, 620015, Tamil Nadu, India
Email: kalaisel@gmail.com
Email: nsk@nitt.edu
*Corresponding author

Abstract: Crime is defined as any act that is illegal and causes unpredictable discomfort to the common public by affecting quality of life and causing financial loss. The objective of this research work is to develop algorithms to predict crime using machine learning (ML) techniques in emotion data and predict future crime spots using crime incident data using deep learning (DL), then cross-check whether the future crime incidents match with the results of crime incidents detected. Voice-based emotion data is analysed using ML algorithms to detect crimes and crime incident data, includes audio and/or video captured from the scene of a crime with geographic coordinates, place names and timestamps are analysed using DL methods such as convolutional stacked bidirectional long short-term memory (LSTM). Crime detection using ML models provided an accuracy of 97.2% for ensemble classifiers and DL methods achieved an accuracy of 95.64% in crime hot spot forecasting.

Keywords: crime forecast; deep learning; machine learning; LSTM; long short-term memory; convolutional neural network; multiplicative attention.

Reference to this paper should be made as follows: Kalai Selvan, A. and Sivakumaran, N. (2024) 'Crime detection and crime hot spot prediction using the BI-LSTM deep learning model', *Int. J. Knowledge-Based Development*, Vol. 14, No. 1, pp.57–86.

Biographical notes: A. Kalai Selvan received his MS in Software Systems from Birla Institute of Technology and Science, Pilani (BITS Pilani) and currently pursuing PhD at Instrumentation and Control Engineering, National Institute of Technology, Tiruchirappalli, India. He is working as a Senior Director in Emerging Technologies Group of Centre for Development of Advance Computing (C-DAC), Thiruvananthapuram for the last 27 years. As a Project Director of Emergency Response Support System, a flagship project of the Ministry of Home Affairs, Govt. of India, he has been developing emergency support solutions for Disaster Management, Women, Child and Elderly Helplines of India. Contributed to several R&D projects including wireless communications, free & open source systems, promoting make-in-India products and technologies. His area of interests are big data analytics, mobile app development, video analytics, IoT, drones, data science & AI, crime detection, prediction and forecasting.

N. Sivakumaran received his PhD from National Institute of Technology (NIT), Tiruchirappalli, India in 2007 under AICTE-NDF scheme. He is currently a Professor with the Department of Instrumentation and Control Engineering, National Institute of Technology, Tiruchirappalli India. He was awarded YFRF-MeitY fellowship. He has received research grants from AICTE DST-SEED, DST-iPHEE and DST. He had received consultancy projects from CDAC, Trivandrum and HAPP Trichy. His research interests include process control, biomedical instrumentation, intelligent system design, data analytics and image processing.

1 Introduction

The emotional state of a person has a strong impact on the voice traits in general. An increase in the rate of respiration strongly impacts the encoding of physical stress in the speech. Stress influences the frequency of sound. In the case of anger, the speaking rate is slightly faster, the average pitch is very much higher, and the intensity is also increased. Similarly, if a person is frightened, the speech rate is considerably faster, the average pitch is also little higher, and intensity is a typical one. If a person is shocked, the rate of speech is much slower, the average pitch is very low, and intensity is also low. When a person is depressed, the speech rate is marginally slower, the average pitch is marginally lower, and the intensity is low. The voice can disclose behaviour like intelligence and personality, as well as emotional conditions like stress, honesty, and feelings. Speech based emotion detection endeavours at mining the emotion from the input speech. The emotion identification from voice depends on several factors such as: Linguistic origin (the language spoken), Paralinguistic knowledge (emotional state of the speakers), and Speaker's basic features (age, size of the body, etc.). Speech characteristics are divided into different types. In one such classification different aspects of speech are divided into three types such as prosodic, vocal tract, and excitation. Similarly, alternative characteristics of speech features are spectral and temporal features. The temporal features are the features of the time domain (like zero crossing rate) spectral features are the frequency domain features (like Mel frequency cepstral coefficients (MFCC), linear predictive cepstral coefficients (LPCC)). Similarly various obstacles found for speech emotion detection are speech is subjective; collection of data for speech emotion detection is complicated in common places; labelling of data involves high human time and cost.

Crime is described as any act that is illegal and makes common public concern affecting quality of life and financial development. Conventional crime detection and prediction is a time-consuming process and sometimes leads to wrong decision making. Now a days criminals use advanced technologies to commit crime without traceability and any evidence. Data is not just a record of wrongdoings that have occurred in the past; it also provides insightful information about certain sources and aids in investigations by revealing various trends (Lu and Luo, 2021). In fact, the only force who could halt these demon-related crimes in the past, when technical development was slower, were humans. Nevertheless, as data volume increases and people are unable to process billions of data points including emotional speech data, this conventional strategy is becoming more difficult to use and hence technology is crucial in this situation (D’Aloia, 2020). Numerous scientific groups are paying growing attention to machine learning (ML) and deep learning based predictive policing and crime analytics with a spatiotemporal emphasis, which are currently being used for information extracting tools. Authors followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) (Kounadi et al., 2020) guidelines to precipitate the most suitable classification method of spatial crime forecasting.

Snaphaan and Hardyns (2021) analysed urban emotions over the last five years, examining geospatial and temporal dimensions through social media to get an valuable insights that can contribute to the enhancement of urban spaces and the creation of emotionally fulfilling environments for the citizens. The principles of environmental criminology (Benabbou and Lee, 2019; Vogel, 2020) depend on three fundamental ideas. First, that offending activities are heavily swayed by the area in which it happens, i.e., location concerns (Sollund, 2021), because each place has unique qualities that either makes criminal behaviour worse or less likely. Second, crime patterns are not spread randomly because they are caused by things that change in space and time. Third, a significant drop in crime can be made by changing the features and sending resources (police, urban model, or societal and ethnic involvement) to those hotspot areas. Because of the growth of computer modelling, geographic information systems (GISs), and geospatial technologies (Jefferson, 2018; Saraiva et al., 2022), crime geo referencing, mapping, and finding crime hotspots have come a long way. Utilisation of spatial data to analyse and crime confinement has been called hotspot policing (Braga et al., 2019), place-based policing (Caplan et al., 2021), or even forensics GIS (Shankar et al., 2022). It is the portion of what Kuo and Lord (2019) have labelled the new pattern of ‘smart policing’, which also encourages excessive incorporation and knowledge spread among police institutes and research organisations, such as academia.

Applying emotion analysis techniques to understand the emotional content of the data can be done using natural language processing (NLP) techniques for text data, sentiment analysis algorithms, or pre-trained models etc. Emotional AI is a newer approach employed to make probabilistic predictions about the emotional states of people using data sources, such as facial (micro)-movements, body language, vocal tone, or the choice of words. For multimedia data, deep learning models like convolution neural networks (CNNs) or recurrent neural networks (RNNs) can be employed to extract emotion-related features. On the one hand, more advanced technologies like Space Syntax (He and Zheng, 2021), big data analytics, machine and deep learning processes are applied to grasp spatial models and even forecast outcomes using Linear methods or Bayesian classifiers (Aldossari et al., 2020), regression analysis (RA), Random Forest Tree (RF), decision tree, K-nearest neighbour (KNN), support vector machine (SVM), and artificial neural networks (ANN) (Chitra et al., 2022; Feng et al., 2023; Safat et al., 2021; Kim et al., 2018; Khairuddin et al., 2019).

Deep learning (DL) is a machine learning technique in which an algorithm can pull out characteristics from raw data. It goes beyond what was possible with other machine learning techniques. This advantage is clear, but it comes with a high cost in terms of computational complexity and raw data demand. Recently, Stalidis et al. (2021) chose this worldwide study trend to predict how crime rates will change every hour. So, these changes must be put in the proper context and understood in their local settings. First, evidence-based policing needs to be examined in terms of how new technology and the ability to manage data and analyse locations affect it. Second, how they could go beyond computation to help ultimately make decisions, which is in line with how new police models encourage sharing and switching of duties. Third, Jeyaboopathiraja et al. (2021) suggest these ideas, techniques, and models work on the locations where most of them were outside of major cities among various nations.

Utilisation of digital tools and smart environments for innovation, emergence of cyber-physical infrastructures, best business practices and the role of innovation and digital strategies are driven factors of smart growth of the cities (Komninos, 2016). An intelligent algorithm for crime prediction seamlessly integrated with an emergency response system serves as a crucial element contributing to the realisation of smart growth in urban areas. Also, inter-organisational networks present varied opportunities for facilitating the exchange of knowledge, information, and technology among key factors such as the police department, control rooms, and health organisations (Schwartz and Hornych, 2011). This collaborative framework enhances the coordination and delivery of emergency support services.

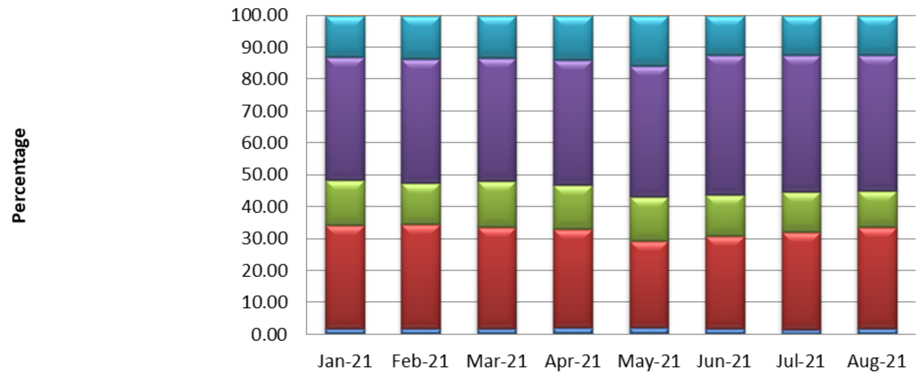
In this research work, we have employed various methods to forecast the possible feature criminal activities in geocoded location using the past crime data and correlated it with the prediction of emotional state of crime incidents happened in a similar manner. The paper is partitioned into sections. Section 2 discusses the related work. Section 3 deals with the data used for this research; Section 4 discusses the proposed model, Section 5 deals with different classification strategies that are used and Section 6 deals with results and analysis. Finally, Section 7 deals with the conclusion.

2 Relevant work

The Kerala state of Indian republic is the first place to use information technology to investigate crimes. Since the 1980s, crime data studies have been coded using GIS techniques, which are as simple as putting historical data on a map and grouping it into high-risk areas over a long period. Figure 1 shows the month wise percentage of crime that happened in various zones of Kerala state government. The percentage of crime is higher in the Thiruvananthapuram Police range when compared with other police ranges. Recent technologies in criminology in the 21st era have been directed to the time-and place-specific crime hot spot forecast (Lamari et al., 2020), using ‘Big Data’ analytics, machine learning, and deep learning techniques (Zhang et al., 2020).

In the past few years, ML techniques have grown even more critical. Aerospace, medical research, big data, time series, transportation, archaeology, finance, and even the arts have all used machine learning techniques a lot (Dargan et al., 2020). They have been used to track illegal activity, model, and predict crime, and writers often compare different approaches (Lim et al., 2021; Wang et al., 2022). For example, Gayathri et al. (2021) suggested a data-driven strategy based on the broken windows theory to find new crime hotspots. Collecting data at different time scales could improve the model’s performance. Yu et al. (2021) found that the long short-term memory (LSTM) deep learning model did better than other conventional crime forecast because it used the environmental data, places of interest and urban road network intensity as variables to improve the crime prediction output. In another research (Ye et al., 2021) the space-time patterns of theft in Manhattan were analysed using linear models and provided better results compared to the prototype application for finding crimes.

Figure 1 Month wise percentage of crimes recorded in various zones (see online version for colours)



	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21
Railway Kerala	0.10	0.19	0.19	0.09	0.06	0.06	0.09	0.07
Police Thrissur Range	12.96	13.39	13.32	13.90	15.72	12.53	12.39	12.35
Police Thiruvanathapuram Range	38.68	38.81	38.52	39.20	41.17	43.76	42.84	42.54
Police Kannur Range	14.01	12.99	14.38	13.73	13.81	12.89	12.65	11.50
Police Ernakulam Range	32.41	32.72	31.86	31.03	27.20	29.01	30.49	31.60
Medical Kerala	1.84	1.89	1.73	2.05	2.05	1.75	1.54	1.95

Instead of using universal computing to decide how to use crime prediction algorithms, local conditions should be considered (a one-size-fits-all approach). In fact, academics have used machine learning techniques to get information and predict crime data patterns based on a place's social, urban, and economic characteristics. When machine learning and deep learning techniques are used in Indian setting to predict the relationship among various crime and its occurrences related to the unemployment was found to be the most important factor for crime (Mittal et al., 2019). Experiments of Ma et al. (2020) come to the same conclusions. They demonstrated that combining various machine learning techniques such as non-linear models, gradient boost decision trees (GBDT), with GIS models have shown accurate results of how over a thousand factors affect things like population, housing, education, economy, social life, and city planning. In this situation, GBDT did better than logistic regression, SVMs, ANNs, and random forest (RF). Wang et al. (2021) explain that these area-specific crime prediction models must consider that crime patterns vary by location. This is in line with Weisburd's Law of Crime Concentration.

Wiesebud's Law (de Jesus Prado et al., 2020) is followed when spatial-temporal prediction is used for encoding crime incidents that happened in a certain area. As per the work (de Jesus Prado et al., 2020) two approaches taken to code crime incidents. The first one used statistical method based on histograms, linear discriminant analysis (LDA) and KNN classifiers, to compare the relationships with neighbourhood characteristics and the distance in time to important holidays. Since the time data got more accurate, the model's performance improved. The other approach used hierarchical density-based spatial clustering of applications with noise (HDBSCAN) to get hot points from crime hotspots for distinctive classes of crimes. This second approach used the distance connecting the cluster centroids (i.e., the crime hot points) as the attribute for classifiers. Linear

Regression and SVM have shown higher accuracy than random forest (RF) in this case. Based on this study it is possible to do spatial analysis, area-specific and space-based machine learning approach, can be done, results could be illuminated, and distributed in web GIS environments to aid local authorities and citizens make decisions (Srinivasulu Raju et al., 2022).

Topic modelling is rarely used to look at crime data, which is different from sentiment analysis. Statistical machine learning techniques are used in this method to look for patterns in a corpus or a large amount of unstructured text. For example, Carter et al. (2019) looked at Los Angeles crime reports to test the Law of Crime Concentration. They did this by comparing the coherence of the topics to the concentration of the crimes themselves. They found that crime-related topics with more coherence and spatial concentration came from latent Dirichlet allocation (LDA), while non-negative matrix factorisation (NMF) improved coherence but not spatial concentration. In a study (Wang et al., 2019) on deep learning for real-time crime forecasting and its ternarisation (DLRTCF), the authors offer a real-time spatial temporal predictor for end-to-end crime intensity prediction using data collected from many sources. The most important aspect of their forecasting technique may be summed up as follows:

- Selected adequate regional and temporal dimensions at which crime history time series hold sufficient reliable signals to conduct research, mapped the number of events that occurred at a certain time step into a picture, where each pixel value reflects the number of criminal acts that occurred inside a grid during that time.
- In order to improve the accuracy of crime predictions, they devised efficient algorithms for enhancing geographical and temporal signals. These methods also rectify the problem that CNNs have when dealing with sparse data, which is caused by weight sharing. To be more exact, they calculated the diurnal cumulative crime rate for each grid spatial area in the time dimension. Super resolution is achieved by the use of bilinear interpolation in the spatial dimension.

In order to make predictions the linear regression (LR) (McClendon and Meghanathan, 2015) based study employed linear regression, and the Akaike criteria to pick models; moreover, the system is able to operate with weighted examples. This technique of regression is straightforward, and it offers a sufficient and understandable explanation of how the factors that go into the calculation impact the results. Additive regression (AR) presented in the study (Hu et al., 2019) is a meta classifier that has the potential to improve the performance of a regression-based classifier. During each iteration of the procedure, a model is fitted to the residuals from the classification process that was completed in the previous iteration. The sum of the predictions produced by all the classifiers is the final result of the prediction process. Reducing the shrinkage parameter, also known as the learning rate parameter, helps to minimise over-fitting and has a smoothing effect, but it also extends the amount of time needed for learning. In another study a decision stump (DS) (Ruder et al., 2016) is used in conjunction with a boosting algorithm, which is a class for building regression (which is based on root-mean-squared error) or classification (based on entropy). The values that are missing are distinct values. The resilient structure of decision trees enables them to perform effectively with enormous datasets and assists algorithms in making better conclusions about the variables.

3 The proposed method

It is evident that the present models available for crime detections and predictions need advanced techniques to predict possible crime activities in advance. These shortcomings motivated us to develop a generalised model that would outperform the state-of-the-arts at the time. To achieve this, we investigated the flaws in the already available models and offered several modifications. Our goal in this research is to detect crime using emotion-based voice data with aspects and similarly, forecast the sort of crime depending on the zone. Specific definition of an aspect is either an aspect category or an aspect term, which are words or words and phrases inside a sentence (Xue and Li, 2018). Therefore, we suggest a convolution stacked bidirectional LSTM network with a multiplicative attention strategy to achieve more accuracy. We have also examined the prediction of possible crime events using the emotion embedded in voice data which is recorded during the Dial-112 calls.

The voice data of both the caller and the voice around that place are analysed to identify the emotions in the voice which may sometimes belong to the criminal. The emotion-based crime incidence prediction algorithm using geocoded voice data is explained in Figure 2 with various steps involved in the process. Figure 2 also explains the deep learning approach-based crime hot spot forecasting and comparing the results of both approaches to confirm the crime incidents in a particular hot spot. The speech data is made in to frames of 20 ms Hamming window frame, removed the artefacts and normalised the data as part of data preprocessing steps. Next meaningful time domain and frequency domain features are extracted. Machine learning techniques such as SVM, KNNs, ANN, Random Forest Trees, and Ensemble classifiers are employed to generate the models using the training set and classify the new feature matrix into different emotions.

3.1 Data collection

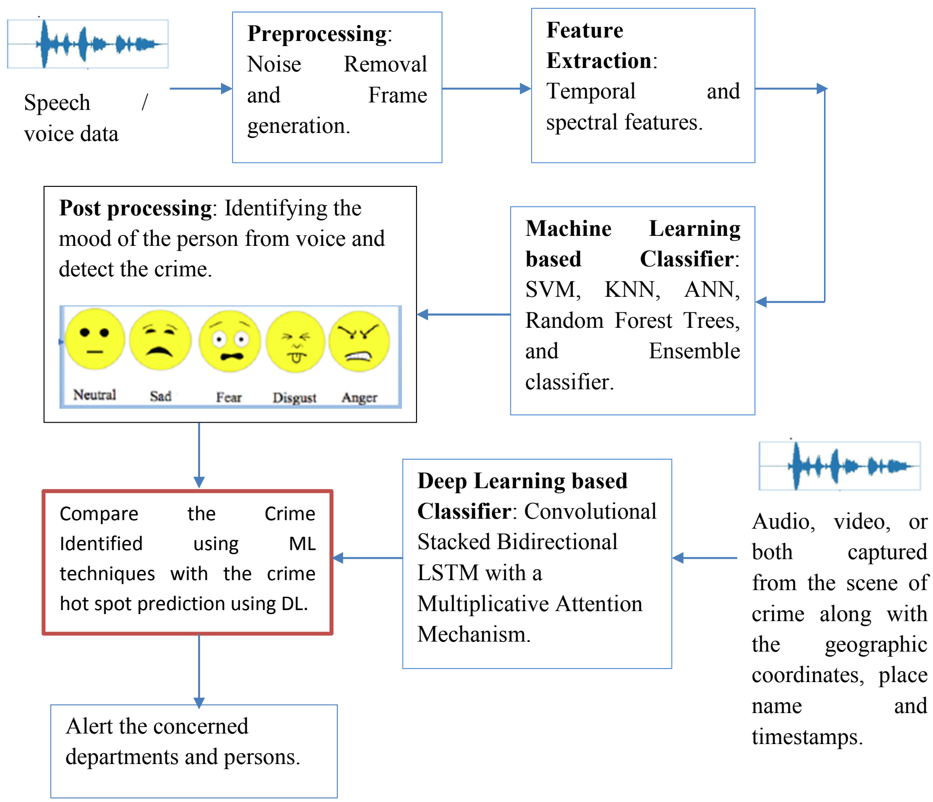
The real-time dataset is collected from ‘Emergency Response Centre’ of Kerala state, owned by Kerala Police, as part of the implementation ‘Dial-112’ facilities. When a person dials 112 to inform a criminal activity, the call lands at the Voice Gateway of the ‘State Emergency Response Centre’. The call is assigned with a unique identifier and incoming voice packets are logged in the central storage. The voice packets are analysed for emotion recognition of caller’s speech to decide if the person is in panic/fear/disgust/anger etc. The system captures the caller’s live location and stores along with incident details as geocoded data. The type of emotion, location, crime activity with timestamps is stored in the central database. Whenever a similar call is made by the same caller, it will be tagged to the earlier one. Thus, identified types of emotion help to relate and confirm the nature of crime.

Also, the ambience of the Victim/Informer of the event is captured with Ambience Listening or Discrete Listening features triggered by the Victim or Emergency Response Centre respectively. This includes audio and/or video captured from the scene of crime along with the geographic coordinates, place name and timestamps. Eventually, this helps to predict/forecast the possibility of repeating crimes and movement of crime.

This database has crime data of 6 different zones of Kerala state and around 30+ different types of crimes. The different zones are Ernakulam Police Range, Kannur Police Range, Thiruvananthapuram Police Range, Thrissur Police Range and Railway Kerala.

More than 30 types of crimes are grouped like bomb blast, crime against women, drunken atrocities, kidnapping, robbery, theft and so on. It comprises a total of 1,15,359 records, covering the period from January 1, 2021, to August 24, 2023. The Emergency Response Center equipped with IP phones with headsets including voice recording functionality. This feature enables the recording of caller voices during emergency calls, with the recorded data securely stored in a centralised database along with relevant caller details. The dataset includes attributes such as event ID, signal landing time, end time, event main type, event subtype, priority, latitude, longitude, near landmark, zone name, and district name along with the voice signals.

Figure 2 Emotion-based crime incidence prediction algorithm and implementation steps using geocoded voice data (see online version for colours)



3.2 *Methodology*

3.2.1 *Emotion based crime detection using ML methods*

Speech emotion detection can be considered as a kind of classification problem. It includes several steps including preprocessing, feature extraction, classification and post processing. Preprocessing plays a crucial role in enhancing the accuracy and efficiency of the process. The audio signal is sampled at 16 KHz to create a digital representation.

Then apply noise reduction techniques such as filtering and spectral subtraction to improve signal quality. Then normalise the audio signals to maintain the uniform input for subsequent stages of processing. After that, the continuous audio signal is divided into short frames or windows as part of preprocessing.

Once preprocessing is done then meaningful features such as time domain and frequency domain features are extracted. The scatter plot feature selection algorithm is applied to find the relevant features and the selected features are inputted to the classifier which will classify the emotions based on the separation of features. In this experiment, the emotions included are fear, anger, sadness, disgust, and neutrality because these emotions prevalence is high in crises situations. For feature extraction, the open Smile tool is used. Some of the extracted features are low-level descriptors (LLD): Loudness, Intensity, Pitch (F0), MFCC, Probability of voicing, Line Spectral Frequencies (LSF), F0 intensity, posamean (position of the algorithmic mean), flatness, range, Standard deviation, maxPos, minPos, skewness – a measure of the asymmetry of the spectral distribution around its centroid, kurtosis – a measure for the peakedness of the spectrum.

Mel-frequency cepstral coefficients (MFCCs) are extracted by converting the voice data in the time domain to the frequency domain using the Fast Fourier Transform at frame level. Signals are converted into 20–40 ms frames for computing MFCCs feature. The estimate of the power spectrum per periodogram is calculated for each frame. The filter bank of MEL to the power spectra is applied and adds the energy to each filter. All filter bank energies are converted to the logarithm scale and discrete Cosine transform (DCT) is applied to log filter bank energies as per equation (1). DCT coefficients from equations (2)–(13) are maintained, and rest are deleted.

$$c_i = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} \cos\left(\frac{\pi i(2m-1)}{2M}\right) . E_m , \tag{1}$$

where M is number of filterbanks.

Fast Fourier Transform (FFT) calculated as per equation (2) for each windowed frame to obtain the magnitude spectrum. This step transforms the signal from the time domain to the frequency domain.

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} \tag{2}$$

$X(k)$ is the complex frequency domain representation at frequency index k.

$x(n)$ is the time-domain signal at time index n.

j is the imaginary unit.

We also used linear prediction cepstral coefficients (LPCC) are one of the features for extracting spectral and cepstral features of the speech signals. LPCC is derived from linear predictive coding (LPC) as per equation (3).

$$c_m = \sum_{k=1}^{p} \frac{1}{k} \sum_{i=1}^{k} (-1)^{i-1} \binom{k-1}{i-1} a_{m+i} \tag{3}$$

where

C_m is mth LPCC coefficients

p is the order of the LPC analysis

a_i is the ith LPC coefficients.

The LPCC coefficients C_0, C_1, …, C_{M-1} represent the LPCC feature vector for a given frame of speech.

The zero-crossing rate (ZRC) is a feature which calculates the number of times in each time interval the sound signal value of zero passes through the amplitude of speech signals. The emotions in the voice data can be neutral or any of the other emotions, so the emotions selected will help to analyse and record the emotions of victims and criminals during emergencies. For voice activity detection, the ZRC feature is utilised. ZCR is nothing but the change of signal sign from positive to negative and vice versa in each time frame as per equation (4). The number of ZCR is low for voiced speech and high for unvoiced speech.

$$ZCR = \frac{1}{2(N-1)} \sum_{n=1}^{N-1} \left| sign(x(n) - sign(x(n-1)) \right| \tag{4}$$

where N is the length of the signal frame, $x(n)$ is the signal sample at time index n, sign(.) is signum function, which returns –1 for negative values, 0 for zero and 1 for positive values.

Delta coefficients are used to represent the rate of change or the gradient of a certain feature over time as per equation (5). Delta coefficients provides information about the dynamic aspects of the signal, such as the speed or acceleration of changes in the feature values.

$$\Delta_c(t) = \frac{\sum_{i=1}^{N} i\left(c(t+i) - c(t-1)\right)}{2\sum_{i=1}^{N} i^2} \tag{5}$$

where $c(t)$ is the feature of interest at time t, N is the window size for the delta computation, $\Delta_c(t)$ is the delta coefficient for the feature c at time t.

Chroma is a feature that is associated with the pitch of a sound signal. If two sound signals are almost the same, then they are seceded by one or more octaves by the same chroma. Normalisation is applied to reduce speaker and record variation without conceding the discriminative power of the features. As part of this study Z normalisation is done. Principal component attribute transformer feature selection method is employed which uses the search method based on attribute ranking. In this method first correlation matrix is created, then eigenvalue is computed, and to finish, eigenvectors are formed. Similarly, correlation based feature selection (CFS) technique is used in which the greedy forward and backward search is performed to find the relevant features. Out of 1912 features extracted most relevant 376 features are selected after applying various feature selection algorithms and the duplicate features are abolished. Table 1 listed some of the selected features.

Table 1 List of the selected features used for machine learning based approach in emotion-based crime prediction

S. no	*Feature*
1	MFCC
2	Fast Fourier transform
3	LPCC
4	Zero crossing rate
5	Fundamental frequency
6	Logarithm of energy
7	Kurtosis
8	Jitter and Shimmer
9	Delta
10	RMS signal frame energy
11	Line spectral frequencies

MFCC feature capture relevant information about spectral characteristics of audio, FFT helps for the analysing speech signals in both the time and frequency domains. This time-frequency analysis is helps for capturing dynamic changes in the speech signal, especially for phonetic transitions. LPCC features are relatively robust to additive noise, making them suitable for speech identification in noisy environments. The linear prediction model helps in separating the signal into components that are more resistant to noise. Changes in emotional expression may be reflected in the ZRC. Emotional speech often involves variations in pitch and intensity, leading to differences in zero crossing patterns. Therefore, ZCR is used as a feature for emotion recognition in speech. Fundamental Frequency(F0) is often used in emotion recognition systems to extract prosodic features contributing to emotional content in speech. Log energy is valuable for capturing overall loudness variations that are indicative of emotional states. Kurtosis is used to capture non-normality in the distribution of speech signal amplitudes, providing information about the uniqueness of emotional expressions. Both Jitter and Shimmer are associated with vocal fold irregularities and can capture the instability and variability in speech production. Delta features help capture temporal variations in acoustic parameters. Emotional expressions may involve rapid changes in pitch, intensity, or other features that can be captured by Delta. RMS energy reflects the overall energy content of a speech signal within a frame. High RMS energy may be associated with intense emotions, anger, or excitement. Low RMS energy may be linked to more subdued or calm emotional states. Changes in LSFs can reflect variations in vocal tract shape associated with different emotional expressions. LSFs are useful for capturing the unique spectral patterns that accompany emotional speech.

The extracted features set is inputted to train the classifiers and generate machine learning models. As part of this work, we have employed SVM, KNN, ANN, Random Forest Tree, and Ensemble classifiers. SVM is a discriminative classifier mostly used for pattern recognition. A SVM classifier is constructing a hyperplane which is used for regression or classification. K-nearest neighbour classifier is a laziest algorithm/instance-based learning algorithm in which k is the number of nearest neighbours and is computed initial step. It uses Euclidean distance or Manhattan distance, and it is working based on similarity measure. KNN is employed when there is no prior knowledge of the dataset available. Finely distinct classes are created using the maximum number of nearest neighbours in Fine KNN. In our research, fine KNN, Medium KNN, Coarse KNN and Cosine KNN are employed. Fine KNN provided better results. Similarly Random Forest algorithm works by constructing many tiny decision-trees and later compounding them to form a forest. Random Forest tree algorithm also uses bootstrap aggregating which also named as bagging, to reduce overfitting problem and improves the generalisation accuracy.

Ensemble modelling is another classifier technique where multiple base models are used to foretell an outcome. Stacking is an ensemble learning technique that creates a new model by employing predictions from multiple models. In our work, the stacked model is created by merging SVM, KNN, and Random Forest classifiers. Bagging joins results from multiple models and yields a generalised result. Similarly boosting is a sequential process, in which each subsequent model tries to correct the previous model's errors. Similarly, we have employed Multilayer perceptron ANN which uses backpropagation for learning. The learning rate is set as 0.03, and activation function is selected as Relu. For this ANN the number of hidden layers is set as 1283, Alpha is 0.3849, batch size is 163, maximum iteration is 100.

For carrying out our Machine Learning algorithm-based emotions detection from voice data of the crime environments which includes the voice signal from victims and others including the criminals present during the incidents. We have used an Intel(R) Core (TM) i7-8565UHz Processor with RAM size of 16 GB. We have also used software packages like IDE Anaconda Navigator, Python language with supporting libraries like Librosa, Numpy, Matplotlib and scikit-learn. Similarly, we have used MATLAB computational language for classifications.

3.2.2 Future crime spot prediction using DL methods

Research work is extended as experiments using deep learning methods and the data collected from dial 112 system for crime prediction to forecast the future crime hot spots based on the available data related to previously occurred crime. As part of this work, we have implemented CNN and LSTM with multiplicative attention. The development of CNN represents a significant step forward in picture classification, pattern detection, and the categorisation of feelings. It does this by taking the input sequences and extracting higher-level characteristics (Hu et al., 2019; Ruder et al., 2016). To be more specific, we build feature maps by using a 1D convolution layer in conjunction with several filters applied to set the window size, as shown in equation (6). Each filter is an instance of a text feature detector that belongs to the n-gram feature pattern.

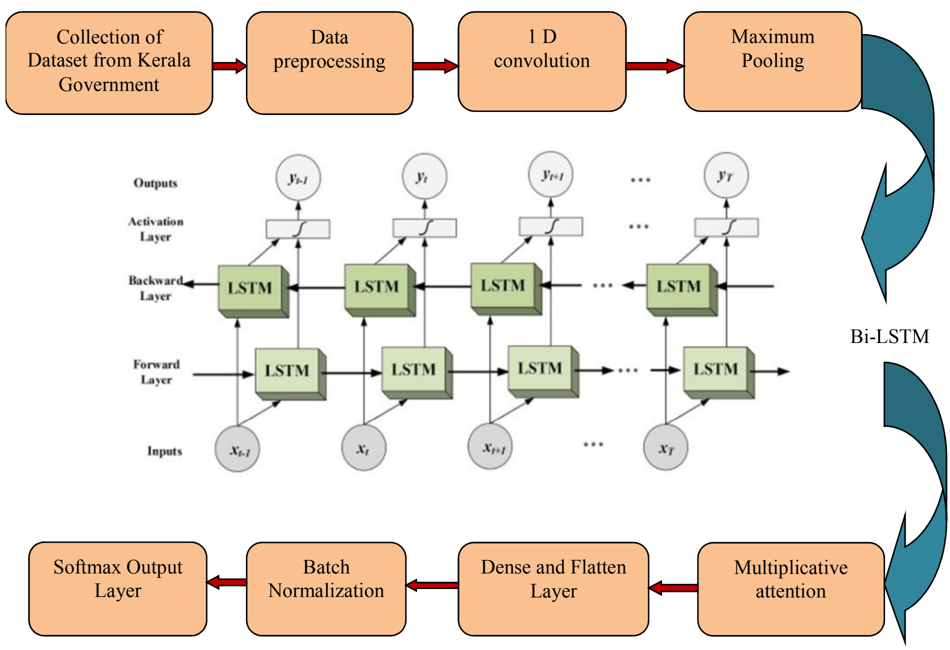
In terms of mathematics, consider the input sequence to be $V = \left(x_1 x_2 \ldots, x_l\right)$ where '*l*' is the length of the sequence that includes padding. A filter using the equation $w_c \in$ RD**f* reduces the receptive field's representation of *f* words to a single feature. The filter is moving through the sequence that is being fed in order to produce a new sequence with different characteristics. $C = \left(c_1 c_2 \ldots c_l\right)$

$$C_i = (1) f(X_i : i + K * W_c + B_C) \tag{6}$$

where *f* is a nonlinear activation function and signifies the bias and *f* also symbolises bias. In addition to this, the max-pooling layer is placed on top of the new sequence features in order to generate a new feature map that has the maximum value in a defined pool size.

These findings were sent on to the next layer, which was a stacked Bi-LSTM. The RNN base is expanded upon to create the LSTM network. Figure 3 explains about the proposed architecture of LSTM network. Hochreiter and Schmidhuber (1997) came up with the idea of using it to solve the problem of the sequence's long-range dependence. Remember gate (ft), input gate (it), and output gate (ot) are the names of the three gating mechanisms that are included inside the LSTM unit (Oueslati et al., 2020; Ma et al., 2018). These gating systems manage the flows of information at each time step, allowing for reading, writing, and updating as appropriate.

Figure 3 The proposed architecture to predict the crime hot spots using convolutional stacked bidirectional LSTM with a multiplicative attention mechanism (see online version for colours)



The LSTM unit may be mathematically expressed using equations (7)–(12).

$$F_t = (w_f\,[h_{t-1}; x_t] + b_f) \tag{7}$$

$$I_t = (w_i\,[h_{t-1}; x_t] + b_i) \tag{8}$$

$$O_t = (w_o\,[h_{t-1}; x_t] + b_o) \tag{9}$$

$$C_t = tanh\,(w_c\,[h_{t-1}; x_t] + b_c) \tag{10}$$

$$c_t = i_t * c_t + f_t * c_t - 1 \tag{11}$$

$$h_t = o_t * \tanh\,(c_t) \tag{12}$$

where x_t, c_t, and h_t, at every time step 't' represent the input vector, the cell state, and the hidden vector respectively. The weight matrices for each gate are indicated as w_f, w_i and w_o, and bias vectors are represented as b_f, b_i and b_o, and b_c respectively. When using an LSTM unit, the sequences are normally encoded in only a single fashion (past information). In light of this, an LSTM that is capable of encoding sequences in both the forward and backward orientations is referred to as a bidirectional LSTM (BiLSTM), and it is advised to be used for the purpose of employing equation (8) to preserve information from both the past and the future. To accomplish the tasks, we layered two instances of the Bi-LSTM network on top of the CNN network.

The multiplicative attention mechanism is responsible for the effective completion of tasks including natural language processing, machine translation, and picture processing. It compiles the relevant context information about a word based on the input phrase that was provided. In specifically, the attention mechanism is concentrated on both the global and the local levels of operation. On the local level, just a select number of the words from the input sequence are taken into consideration, but on the global level, selective attention is paid to each and every word in the input sequence (Dua et al., 2019; Liu and Guo, 2019). In this study, we concentrate on a narrow portion of the input sequence by using a process known as local attention, which is also known as the multiplicative attention mechanism. Let $S = \left(x_1 x_2 \ldots x_n\right)$ stand for the supplied sentence and let $h_i = \left(h_1 h_2 \ldots h_s\right)$ represent the hidden word vectors that are created by the given phrase. We produce multiplicative attention for a tiny portion of the provided input sequence, and then we use a softmax activation function to normalise the attention that we have generated. After that, the calculation for the context vector looks like this: (13)–(15).

$$h_t t' = \tanh\left(X(9)\ T_t * w_t + x_t t' * w_t + b_t\ \right) \tag{13}$$

$$e^t t' = '\left(x(10) T_t * w_a * X_t + b_a\right)` \tag{14}$$

$$a_t = \text{softmax}\ \left(e\right) \tag{15}$$

4 Results and discussions

The machine learning approach-based results are described in Table 2 for various classifiers such as SVM, KNN, ANN – Multilayer perceptron, Random Forest Tree and Ensemble classifiers using various features extracted using conventional feature extraction techniques. The total features mined for the machine learning based classifier model creation process are 1921 features and 860 instances. We have employed feature selection algorithms to shortlist only 376 relevant features. Another method for selecting the relevant feature and avoiding overfitting is cross-validation. In this experiment, we used ten-fold cross-validation. In 10-fold cross-validation, 10 equal size subsamples are obtained by random partitioning. For validation of data, one subsample is retained, the remaining 10-1 subsamples are used as training data. 10 times the cross-validation process is done, until each of the 10 subsamples is used precisely once. Finally, results of the 10 samples are averaged to obtain one estimation. The resultant accuracy after applying 10 cross-validations is displayed in Table 2 and it can be found that the multilayer perceptron shows higher accuracy than the other model. The other validation parameters such as precision, positive predictive value (recall) is also calculated. The recall value obtained for anger is 86.8%, disgust is 90.2%, fear is 87.7%, neutral is 85.3% and sad is 95.0%. Similarly, the True positive rate (precision) obtained for anger is 91.4%, disgust is 91.4%, fearful is 89.4%, neutral is 84.2% and sad is 87.4% and the overall accuracy of the model is 89.3% for Multilayer perceptron.

Table 2 Emotion classification accuracy obtained for various ML classifier

S. no	*Classifier model*	*Accuracy*	*Standard deviation of 10-fold cross validation*
1	Support vector machine	86.80%	1.2080
2	K-nearest neighbour	70.40%	0.7687
3	Multilayer perceptron	89.30%	1.2483
4	Random Forest	51.5%	0.8255
5	Artificial neural network	84.04	1.4158
6	Stacking ensemble classifier	87.60%	1.2689
7	Bagging ensemble classifier	62.01%	1.0628

The machine learning approach also employed various tuning in the classifier as per the matlab implementation of various classifiers. The results obtained in terms of accuracy for various SVM models are quadratic SVM is 90.1%, Cubic SVM is 94.7%. Similarly Fine KNN provided an accuracy of 93.8%, the bagged ensemble model showed an accuracy of 97.2%, the multilayer perceptron has provided an accuracy of 90.7%. The precision (true positive rate) obtained using Bagged ensemble model for anger is 89.4%, disgust is 85.6%, fear is 93.6%, neutral is 92.6% and sad is 91.5%. Similarly, the Positive predicted value (recall) obtained for Bagged Ensemble classifier are angry 85.4%, disgust 85.4%, fearful 84.8%, neutral is 78.9% and sad is 94.0% and the overall accuracy of the model is 97.2%.

Hyperparameter optimisation method performed to get this best result from training samples. Parameter tuning is a crucial step in enhancing the performance of machine learning models. Grid search, a systematic optimisation method, is applied to classifiers with specific hyperparameter configurations. For SVMs, the regularisation parameter (C) is tested with values 0.1, 1, 10, and 100, and the kernel function is explored using 'linear' and 'rbf'. The gamma values 'auto' and 'scale' are also considered. Optimal performance is achieved with C=10, kernel='rbf', and gamma='auto'. In KNNs, parameters such as the number of neighbours (n_neighbors), weights ('uniform' and 'distance'), and the power parameter (p) for Minkowski distance are tuned. Values of 3, 5, and 7 are tested for n_neighbors, 'uniform' and 'distance' are considered for weights, and p takes values of 1 and 2. Artificial neural network (ANN) experimentation involves testing different architectures, including a single layer with 60 neurons, two layers with (40, 60) neurons, and three layers with (30, 20, 10) neurons. Activation functions 'relu' and 'tanh' are employed, and regularisation parameters alpha take values 0.001 and 0.01.Random Forest hyperparameters include the number of trees (n_estimators), maximum depth of trees (max_depth), minimum samples required to split a node (min_samples_split), and minimum samples required in a leaf node (min_samples_leaf). Grid search explores n_estimators={50, 100, 200}, max_depth={None, 10, 20}, min_samples_split={2, 5, 10}, and min_samples_leaf={1, 2, 4}.For the Stacking Ensemble classifier, SVM, Random Forest, and ANN are combined.

Deep Learning model uses one convolution layers with 20 filters with parameters kernel size 3, padding 1 and stride 1. Then applied max pooling layer of stride 2. The resultant feature vector fed to 2-layer bi-directional LSTM. First layer contains 10 LSTM units. Finally, it passes to dense layer for prediction. It is implemented using TensorFlow library.

The results obtained from our analysis using various Deep Learning algorithms with various crime related data and its accuracy are shown in Table 3. The parameters Accuracy, Precision, Sensitivity, Specificity, FScore and Confusion Matrix are used to measure the performance of our proposed system. We have measured all parameters which are mentioned above for all crimes. It has nearly 42 tables, so for discussion here presented bomb blast, crime against women and theft. Datasets are divided into 10 chunks. Results taken for each chunk to monitor the performance of the proposed system. The percentage difference between the proposed method and existing methods of accuracy is improved 4.19%, 1.59%, 3.14% and 4.04% than the existing methods like DLRTCF, LR, AR and DS to predict the Bomb Blast. The percentage difference between the proposed method and existing methods of accuracy is improved 5.36%, 1.63%, 3.21% and 7.39% than the existing methods like DLRTCF, LR, AR and DS to predict the crime against women. The percentage difference between the proposed method and existing methods of accuracy is improved 5.44%, 1.66%, 4.35% and 3.11% than the existing methods like DLRTCF, LR, AR and DS to predict theft.

The primary performance metric is accuracy of prediction by the proposed ERSS model. Table 3 enlists the possible outcomes of ERSS in comparison to other conventional models viz. DLRTCF, LR, AR, and DS respectively for categories such as bomb blast, crime against women, and theft. The values infer that the ERSS model predicts with highly significant accuracy when compared with the conventional models. Table 1 provides the values obtained using various percentage of data from 10% to 100% and its predicted output values for Bomb blast, Crime against woman and Theft. The percentage difference between the proposed method and existing methods of accuracy is improved 4.19%, 1.59%, 3.14% and 4.04% than the existing methods like DLRTCF, LR, AR and DS to predict the Bomb Blast. The percentage difference between the proposed method and existing methods of accuracy is improved 5.36%, 1.63%, 3.21% and 7.39% than the existing methods like DLRTCF, LR, AR and DS to predict the crime against women. The accuracy difference between the proposed method and existing methods is improved 5.44%, 1.66%, 4.35% and 3.11% than the existing methods like DLRTCF, LR, AR and DS to predict the theft.

Computation of the percentage of precision is presented in Table 4. In this case too, the proposed model takes over all the four comparison models significantly. The percentage difference between the proposed method and existing methods of precision is improved 4.36%, 1.09%, 3.01% and 3.61% than the existing methods like DLRTCF, LR, AR and DS to predict the Bomb Blast. The percentage difference between the proposed method and existing methods of precision is improved 6.78%, 0.07%, 0.39% and 4.87% than the existing methods like DLRTCF, LR, AR and DS to predict the crime against women. The percentage difference between the proposed method and existing methods of precision is improved 1.31%, 1.16%, 2.10% and 0.54% than the existing methods like DLRTCF, LR, AR and DS to predict the theft.

Table 3 Accuracy (%) of deep learning based crime hot spot detection using ERSS dataset

Data (%)	*Bomb blast*					*Crime against women*					*Theft*				
	DLRTCF	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*
10	14.86	27.00	17.93	22.29	19.57	12.86	23.00	15.93	19.29	18.57	10.86	24.00	16.43	18.29	13.57
20	39.87	49.50	41.64	44.09	42.72	35.87	46.00	38.14	40.59	39.72	39.37	48.00	38.64	39.59	40.72
30	52.81	60.33	55.19	57.23	56.36	48.81	55.33	52.19	56.73	54.86	51.31	58.33	52.69	51.23	50.36
40	63.24	69.72	65.49	66.39	68.51	59.74	67.72	61.99	64.89	65.51	61.74	65.72	63.49	62.89	66.01
50	71.99	77.07	72.08	72.56	74.00	71.49	74.07	69.08	69.56	71.50	68.49	71.57	71.08	70.06	73.00
60	78.68	81.26	79.26	78.89	82.51	74.68	77.76	76.26	75.89	80.01	75.18	79.76	76.26	77.39	80.01
70	82.68	85.80	83.12	82.79	85.97	78.68	82.30	81.12	81.29	85.97	78.68	83.80	79.62	77.29	81.97
80	87.25	90.93	89.13	88.62	91.22	84.25	88.93	88.63	86.12	86.72	86.25	88.43	88.13	88.62	89.22
90	92.34	93.15	92.96	92.25	94.72	86.34	91.15	87.46	90.75	89.22	88.34	88.15	87.46	89.75	91.22
100	95.07	97.57	96.07	95.21	99.14	92.07	95.57	94.07	90.21	97.14	90.57	94.07	91.57	92.71	95.64

Table 4 Precision of different DL classifiers for different crimes

	Bomb blast					*Crime against women*					*Theft*				
Data (%)	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*
10	15.14	26.57	17.29	22.14	18.57	14.14	21.57	15.29	17.14	16.57	13.14	21.57	16.29	18.14	12.57
20	40.23	49.79	41.07	44.66	43.07	35.23	46.79	39.07	43.66	37.07	39.23	47.79	37.07	38.66	43.07
30	52.31	60.83	53.98	57.30	56.22	47.31	55.83	47.98	57.30	56.22	52.31	60.83	53.98	51.30	50.22
40	62.59	70.72	65.28	67.32	68.58	60.59	66.72	60.28	67.32	67.58	60.59	67.72	61.28	63.32	63.58
50	71.49	76.50	72.79	73.20	74.14	70.49	74.50	69.79	68.20	69.14	68.49	71.50	70.79	69.20	74.14
60	78.39	80.76	79.76	77.96	82.58	74.39	76.76	75.76	72.96	77.58	76.39	78.76	78.76	76.96	82.58
70	83.32	85.87	83.19	83.14	85.33	80.32	82.87	81.19	82.14	85.33	80.32	82.87	79.19	78.14	79.33
80	86.68	90.93	88.77	88.55	89.94	81.68	90.93	87.77	86.55	86.94	84.68	85.93	86.77	88.55	85.94
90	93.05	92.94	92.96	92.47	94.29	87.05	90.94	86.96	90.47	88.29	89.05	87.94	87.96	87.47	88.29
100	94.43	97.57	95.71	95.14	98.64	88.43	94.57	94.71	90.14	94.64	91.43	91.57	90.71	92.14	92.64

Table 5 explains the sensitivity of ERSS model while training the algorithm starting at an incremental dataset volume of 10% upto 100%. Here also the proposed model leads in performance comparatively. The percentage difference between the proposed method and existing methods of sensitivity is improved 4.07%, 2.09%, 3.30% and 4.47% than the existing methods like DLRTCF, LR, AR and DS to predict the Bomb Blast. The percentage difference between the proposed method and existing methods of sensitivity is improved 4.35%, 3.18%, 6.33% and 9.85% than the existing methods like DLRTCF, LR, AR and DS to predict the crime against women. The percentage difference between the proposed method and existing methods of sensitivity is improved 9.20%, 6.91%, 6.56% and 5.57% than the existing methods like DLRTCF, LR, AR and DS to predict theft.

The obtained specificity of the prediction model is up to 98.66% which is higher compared to the yields of other four models are shown in Table 6. The percentage difference between the proposed method and existing methods of specificity is improved 4.30%, 1.10%, 2.99% and 3.61% than the existing methods like DLRTCF, LR, AR and DS to predict the Bomb Blast. The percentage difference between the proposed method and existing methods of specificity is improved 6.17%, 0.23%, 0.26% and 5.12% than the existing methods like DLRTCF, LR, AR and DS to predict the Crime Against Women. The percentage difference between the proposed method and existing methods of specificity is improved 1.93%, 1.81%, 2.38% and 0.89% than the existing methods like DLRTCF, LR, AR and DS to predict the theft.

Another vital performance indicator of the proposed classifier is the FScore and its values are shown in Table 7. The percentage difference between the proposed method and existing methods of specificity is improved 4.22%, 1.59%, 3.15% and 4.04% than the existing methods like DLRTCF, LR, AR and DS to predict the Bomb Blast. The percentage difference between the proposed method and existing methods of specificity is improved 5.61% 1.60%, 3.09% and 7.32% than the existing methods like DLRTCF, LR, AR and DS to predict the Crime Against Women. The percentage difference between the proposed method and existing methods of specificity is improved 5.21%, 1.36%, 4.28% and 3.01% than the existing methods like DLRTCF, LR, AR and DS to predict the theft.

The percentage difference between the proposed method and existing methods of specificity is improved 4.22%, 1.59%, 3.15% and 4.04% than the existing methods like DLRTCF, LR, AR and DS to predict the Bomb Blast. The percentage difference between the proposed method and existing methods of specificity is improved 5.61% 1.60%, 3.09% and 7.32% than the existing methods like DLRTCF, LR, AR and DS to predict the crime against women. The percentage difference between the proposed method and existing methods of specificity is improved 5.21%, 1.36%, 4.28% and 3.01% than the existing methods like DLRTCF, LR, AR and DS to predict the theft. Figure 4 shows the predicted future crime spot which is successfully predicted by our algorithm using previous data.

Table 5 Sensitivity of different DL classifiers for different crimes

	Bomb blast					*Crime against women*					*Theft*				
Data (%)	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*
10	15.06	26.80	17.51	22.21	18.95	13.79	22.21	15.48	17.91	17.26	12.57	22.67	16.33	18.19	12.83
20	39.94	49.50	41.54	44.15	42.77	35.68	46.06	38.35	41.13	39.14	39.34	47.99	38.27	39.39	41.14
30	52.84	60.22	55.32	57.22	56.38	48.78	55.27	52.39	56.66	54.73	51.29	57.93	52.63	51.23	50.36
40	63.41	69.33	65.56	66.09	68.48	59.57	68.08	62.42	64.20	64.89	62.01	65.11	64.12	62.78	66.82
50	72.21	77.38	71.76	72.27	73.93	71.93	73.87	68.81	70.10	72.56	68.49	71.60	71.20	70.41	72.48
60	78.85	81.57	78.97	79.44	82.46	74.82	78.32	76.52	77.50	81.54	74.58	80.36	75.01	77.63	78.54
70	82.26	85.75	83.07	82.56	86.44	77.77	81.93	81.08	80.76	86.44	77.77	84.44	79.88	76.83	83.76
80	87.68	90.93	89.41	88.67	92.31	86.10	87.43	89.30	85.81	86.57	87.42	90.45	89.19	88.67	91.98
90	91.74	93.34	92.96	92.07	95.11	85.83	91.33	87.83	90.99	89.96	87.80	88.32	87.09	91.66	93.78
100	95.66	97.57	96.40	95.28	99.64	95.38	96.50	93.51	90.27	99.62	89.89	91.97	92.30	93.21	98.56

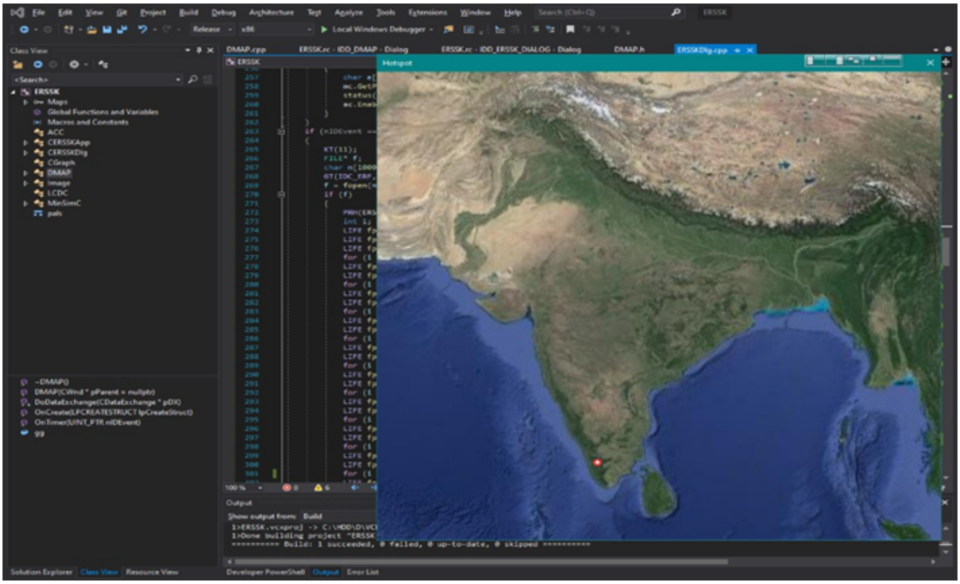
Table 6 Specificity of different DL classifiers for typical crimes

Data (%)	*Bomb blast*					*Crime against women*					*Theft*				
	DLRTCF	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*
10	14.66	27.20	18.34	22.36	20.17	11.88	23.75	16.36	20.55	19.78	8.98	25.20	16.52	18.38	14.29
20	39.80	49.50	41.73	44.02	42.67	36.05	45.94	37.91	39.97	40.23	39.40	48.01	38.98	39.78	40.26
30	52.78	60.43	55.07	57.24	56.34	48.85	55.38	52.02	56.81	55.00	51.34	58.77	52.76	51.23	50.36
40	63.07	70.12	65.43	66.70	68.53	59.91	67.37	61.59	65.65	66.18	61.47	66.37	62.92	63.00	65.27
50	71.77	76.76	72.40	72.85	74.07	71.07	74.28	69.35	69.04	70.53	68.49	71.54	70.96	69.72	73.53
60	78.52	80.95	79.56	78.36	82.55	74.54	77.21	76.00	74.46	78.62	75.81	79.17	77.64	77.16	81.63
70	83.10	85.85	83.17	83.02	85.51	79.65	82.67	81.16	81.83	85.51	79.65	83.18	79.37	77.76	80.37
80	86.83	90.93	88.85	88.56	90.19	82.57	90.55	87.98	86.43	86.88	85.14	86.60	87.12	88.56	86.81
90	92.95	92.97	92.96	92.43	94.34	86.87	90.98	87.09	90.52	88.50	88.90	87.99	87.83	88.01	88.94
100	94.50	97.57	95.74	95.15	98.66	89.21	94.68	94.65	90.16	94.90	91.28	96.39	90.87	92.23	93.06

Table 7 FScore of different DL classifiers for different crimes

Data (%)	*Bomb blast*					*Crime against women*					*Theft*				
	DLRTCF	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*
10	15.10	26.69	17.40	22.17	18.76	13.96	21.88	15.38	17.52	16.91	12.85	22.11	16.31	18.17	12.70
20	40.08	49.64	41.30	44.41	42.92	35.45	46.42	38.71	42.36	38.08	39.28	47.89	37.66	39.02	42.08
30	52.58	60.52	54.64	57.26	56.30	48.03	55.55	50.09	56.98	55.46	51.79	59.34	53.29	51.27	50.29
40	63.00	70.02	65.42	66.70	68.53	60.08	67.39	61.33	65.72	66.21	61.29	66.39	62.66	63.05	65.16
50	71.85	76.94	72.27	72.73	74.03	71.20	74.18	69.30	69.14	70.81	68.49	71.55	70.99	69.80	73.30
60	78.62	81.16	79.36	78.69	82.52	74.61	77.53	76.14	75.16	79.51	75.48	79.55	76.84	77.29	80.51
70	82.79	85.81	83.13	82.85	85.88	79.02	82.40	81.13	81.45	85.88	79.02	83.65	79.53	77.48	81.48
80	87.17	90.93	89.09	88.61	91.11	83.83	89.15	88.53	86.18	86.75	86.03	88.13	87.97	88.61	88.86
90	92.39	93.14	92.96	92.27	94.70	86.44	91.14	87.39	90.72	89.12	88.42	88.13	87.52	89.51	90.95
100	95.04	97.57	96.06	95.21	99.14	91.77	95.53	94.11	90.21	97.07	90.65	94.22	91.50	92.67	95.51

Figure 4 Shows the possible hot spot of crime which may happen in future (see online version for colours)



The classification of the proposed deep learning model is again validated using a confusion matrix. Confusion matrices are a crucial tool in evaluating the performance of a classification model, breaking down the results into True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). Figure 5(a)–(j) illustrate confusion matrices of prediction of bomb blasts crimes from training data of 10% upto 100%. In 10% data FN is 84.85% which is reduced to 5.57% on 100% data. Overall error rate reduced while more data to training. Similarly, Figure 6(a)–(j) represent the confusion matrices of crimes against women prediction from 10% of data to 100% of data with its prediction output. TP and TN increased from 14.14 to 88.42, 11.57 to 95.71 respectively from 10% data to 100% data. This implies that increasing the amount of training data has a positive impact on the accuracy of the model. Figure 7(a)–7(j) represent the confusion matrices of theft crime prediction for the data of 10–100%.

FN indicates instances where the model failed to detect a positive event that actually occurred. In a bomb blast prediction scenario, a false negative would mean that the model failed to identify an actual bomb blast. In emergency response systems, FNs may lead to delayed or inadequate responses to real emergency situations. The system's failure to recognise an actual event can have serious consequences for public safety. FP occurs when the model predicts the positive class when the true class is negative. FP implies that the model triggered an alarm or prediction for an event that did not actually occur. Responding to false alarms generated by FPs may result in the inefficient use of resources, such as emergency personnel and equipment, diverting them from where they are genuinely needed. In this crime prediction FN and FP rates are reduced by added more adding more data. It can improve a lot by feeding more data to our system.

Figure 5 Confusion matrix of bomb blast prediction at Kerala state: (a) 10% of data; (b)20% of data; (c) 30% of data; (d) 40% of data; (e) 50% of data; (f) 60% of data; (g) 70% of data; (h) 80% of data; (i) 90% of data and (j) 100% of data (see online version for colours)

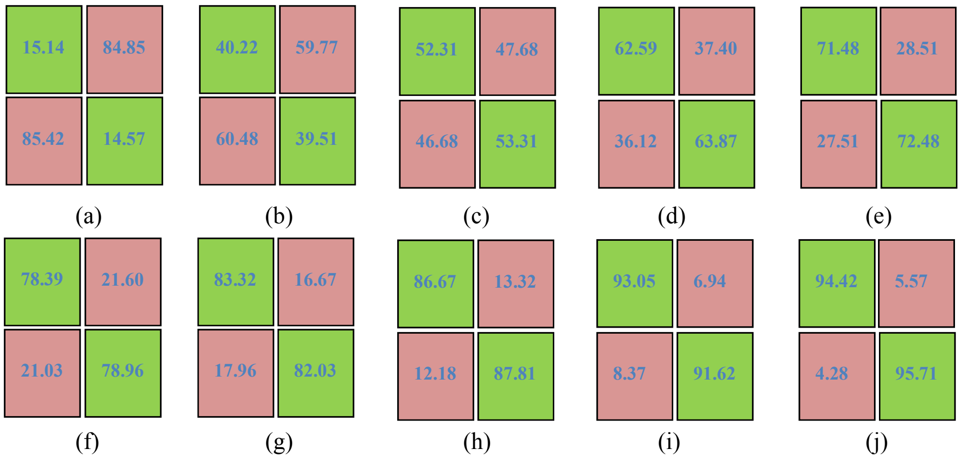
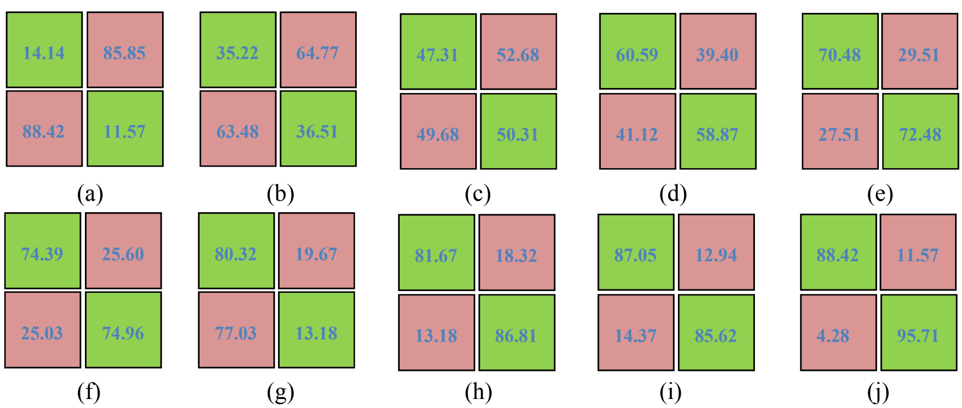


Figure 6 Confusion matrix of crime against women prediction: (a) 10% of data; (b)20% of data; (c) 30% of data; (d) 40% of data; (e) 50% of data; (f) 60% of data; (g) 70% of data; (h) 80% of data; (i) 90% of data and (j) 100% of data (see online version for colours)



Average processing time required for DLRTCF, LR, AR ,DS and ERSS is shown in Table 8. Our method ERSS take less time compared to other four methods. When we compared results of our two-research work, we found that the Deep Learning based crime spot prediction result matches the actual crime happened which is confirmed by emotion data-based crime detection using Machine Learning algorithm. So, the past data of criminals or victims with environmental sound recordings and other crime data could be effectively used for forecasting the future crime hot spots. The comparison among the crime prediction accuracy of various studies is compared with our DL based approach is shown in Table 9. From the comparative analysis our approach provided the highest crime spot prediction accuracy of 97.14%.

Figure 7 Confusion matrix of prediction of the crime theft: (a) 10% of data; (b)20% of data; (c) 30% of data; (d) 40% of data; (e) 50% of data; (f) 60% of data; (g) 70% of data; (h) 80% of data; (i) 90% of data and (j) 100% of data (see online version for colours)

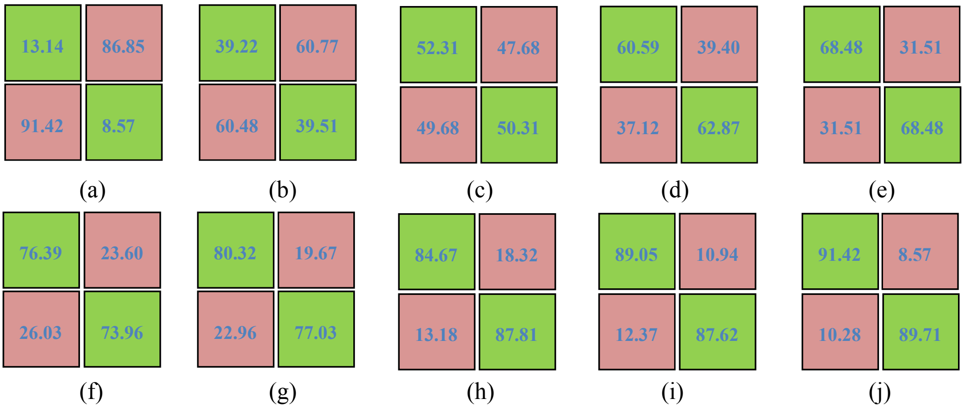


Table 8 Average processing time for algorithms

	Average processing time (ms)				
Data (%)	*DLRTCF*	*LR*	*AR*	*DS*	*ERSS*
10	1814	1532	1647	1824	1529
20	1897	1531	1654	1724	1514
30	1868	1535	1720	1841	1514
40	1831	1675	1690	1738	1565
50	1948	1565	1704	1763	1590
60	1845	1559	1606	1713	1533
70	1923	1573	1653	1829	1546
80	1878	1624	1710	1733	1455
90	1807	1578	1746	1745	1593
100	1805	1548	1659	1806	1572

Table 9 Comparison of crime hotspot forecasting of various research studies

S. no.	*Author*	*Method*	*Crime prediction accuracy in%*
1	Saraiva et al. (2020)	Logistic regression, Decision Tree, Random Forest, SVM	65, 61, 83, 80 respectively
2	Safat et al. (2021)	Logistic regression, Decision Tree, Random Forest, multi-layer perceptron (MLP), Naïve Bayes, SVM, KGBoost and KNN	90,66,77, 87, 73, 66, 94 and 98 respectively
3	Kim et al. (2018)	K-nearest-neighbour and boosted decision tree	39 and 44 respectively

Table 9 Comparison of crime hotspot forecasting of various research studies (continued)

S. no.	*Author*	*Method*	*Crime prediction accuracy in%*
4	Lamari et al. (2020)	Gradient boosting model	77
5	Lim et al. (2021)	Deep reinforcement learning (DRL), gradient boosting machine (GBM), Random Forest, and SVM	73, 65, 68, and 66 respectively
6	Srinivasulu Raju et al (2022)	Artificial neural networks	92
7	This study	Bi-LSTM with multiplicative attention method	97.14

5 Conclusion

The system that we are proposing contains a combination of technologies that will perform a variety of tasks, such as monitoring crime hotspots, detecting crimes and forecasting crime and crime hot spots. Emotional data and other data such as geolocations, type of crime etc. related to past crimes could be effectively used to forecast the future crime hotspots using deep learning algorithm is a promising field that combines the analysis of emotional cues with advanced machine learning techniques. By leveraging deep learning models and analysing emotions from diverse data sources, such as text, images, audio, or videos, it is possible to develop systems that predict the likelihood of criminal activity and its hotspots. This research work provided an overview of a framework that envisions how machine learning and neural networks might work together to assist in the development of a system that is much more beneficial to the police. The very first obstacle that will need to be overcome is creating this system, which will be followed by additional challenges such as putting it into action and using it.

The obtained promising results from our research indicate a strong potential for integrating this work into real-time applications dedicated to crime prevention and public safety. Specifically, our proposed system can seamlessly integrate to 112 emergency response system, system leverages voice-based emotion analysis to discern the authenticity of incoming calls to emergency services. Through sophisticated machine learning algorithms, the system can dynamically analyse the emotional content embedded in the caller's voice. This analysis serves as a critical determinant in identifying whether the emergency call is genuine or may require further scrutiny. The significance of this capability lies in its potential to filter and prioritise emergency calls, enabling law enforcement agencies to allocate resources more effectively. By distinguishing genuine emergency situations from non-emergency or fraudulent calls, our system contributes to streamlining the emergency response process, minimising response times, and optimising the deployment of personnel.

Conflict of interest

The authors declare that no competing interests exist.

Informed consent

Informed consent is not applicable.

Ethical approval

The conducted research did not directly involve any human or animal subjects. So ethical approval is not applicable.

Funding

The research work "CRIME DETECTION AND CRIME HOT SPOT PREDICTION USING BI-LSTM DEEP LEARNING MODEL" was funded by the Ministry of Home Affairs, Government of India. The funding certificate number is IN-DL56444902758354P.

References

Aldossari, B.S., Alqahtani, F.M., Alshahrani, N.S., Alhammam, M.M., Alzamanan, R.M. and Aslam, N. (2020) 'A comparative study of decision tree and Naive Bayes machine learning model for crime category prediction in Chicago', *Proceedings of 2020 the 6th International Conference on Computing and Data Engineering*, January, Sanya, China, pp.34–38, https://doi.org/10.1145/3379247.3379279

Benabbou, R. and Lee, H. (2019) 'Exploring the evolution of urban emotions in the city of Seoul using social media information', *International Journal of Knowledge-Based Development*, Vol. 10, No. 3, pp.232–248, doi: 10.1504/IJKBD.2019.103208.

Braga, A.A., Turchan, B., Papachristos, A.V. and Hureau, D.M. (2019) 'Hot spots policing of small geographic areas effects on crime', *Campbell Systematic Reviews*, Vol. 15, No. 3, p.e1046, https://doi.org/10.1002/cl2.1046

Caplan, J.M., Kennedy, L.W., Drawve, G. and Baughman, J.H. (2021) 'Data-informed and place-based violent crime prevention: the Kansas City, Missouri risk-based policing initiative', *Police Quarterly*, Vol. 24, No. 4, pp.438–464, https://doi.org/10.1177/10986111211003205

Carter, J.G., Mohler, G. and Ray, B. (2019) 'Spatial concentration of opioid overdose deaths in Indianapolis: an application of the law of crime concentration at place to a public health epidemic', *Journal of Contemporary Criminal Justice*, Vol. 35, No. 2, pp.161–185, https://doi.org/10.1177/1043986218803527

Chitra, D., Mahalakshmi, V., Lakshmi, B. and Isravel, Y.A.D. (2022) 'Innovation comorbidity of compulsive buying and brand addiction among the younger generation', *International Journal of Knowledge-Based Development*, Vol. 12, Nos. 3–4, pp.475–493, doi: 10.1504/IJKBD.2022.128910.

D'Aloia, A. (2020) 'Law challenged. Reasoning about neuroscience and law', *Complicated Crossings and New Perspectives*, pp.1–36, https://doi.org/10.1007/978-3-030-38840-9_1

Dargan, S., Kumar, M., Ayyagari, M.R. and Kumar, G. (2020) 'A survey of deep learning and its applications: a new paradigm to machine learning', *Archives of Computational Methods in Engineering*, Vol. 27, No. 4, pp.1071–1092, https://doi.org/10.1007/s11831-019-09344-w

De Jesus Prado, K.H., Santos Souza, L., Dias de Jesus Junior, I. and Colaço Júnior, M. (2020) 'Applied intelligent data analysis to government data related to criminal incident: a systematic review', *Journal of Applied Security Research*, Vol. 15, No. 3, pp.297–331, https://doi.org/10.1080/19361610.2020.1716511

Dua, R.D., Madaan, D.M., Mukherjee, P.M. and Lall, B.L. (2019) 'Real time attention based bidirectional long short-term memory networks for air pollution forecasting', *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, Newark, CA, USA, pp.151–158, doi: 10.1109/BigDataService.2019.00027.

Feng, Y., Xu, X. and Wang, R. (2023) 'A random forest algorithm-based customer demand forecasting model for sports enterprises in the real economy', *International Journal of Knowledge-Based Development*, Vol. 13, Nos. 2–4, pp.363–378, doi: 10.1504/IJKBD.2023.133337.

Gayathri, Y., Sri Lalitha, Y., Nag, A. and Basha, A.H. (2021) 'Data-driven prediction model for crime patterns', *Smart Computing Techniques and Applications*, Springer, Singapore, pp.47–58, https://doi.org/10.1007/978-981-16-0878-0_6

He, J. and Zheng, H. (2021) 'Prediction of crime rate in urban neighborhoods based on machine learning', *Engineering Applications of Artificial Intelligence*, Vol. 106, p.104460, https://doi.org/10.1016/j.engappai.2021.104460

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', *Neural Computation*, Vol. 9, No. 8, pp.1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735

Hu, M., Zhao, S., Guo, H., Cheng, R. and Su, Z. (2019) *Learning to Detect Opinion Snippet for Aspect-Based Sentiment Analysis*, arXiv preprint arXiv: 1909.11297, https://doi.org/10.48550/arXiv.1909.11297

Jefferson, B.J. (2018) 'Predictable policing: predictive crime mapping and geographies of policing and race', *Annals of the American Association of Geographers*, Vol. 108, No. 1, pp.1–16, https://doi.org/10.1080/24694452.2017.1293500

Jeyaboopathiraja, J. and Maria Priscilla, G. (2021) 'A thorough analysis of machine learning and deep learning methods for crime data analysis', *Inventive Computation and Information Technologies*, Vol. 173, pp.795–812, https://doi.org/10.1007/978-981-33-4305-4_58

Khairuddin, A.R., Alwee, R. and Harun, H. (2019) 'Comparative study on artificial intelligence techniques in crime forecasting', *Applied Mechanics and Materials*, Trans Tech Publications Ltd., Vol. 892, pp.94–100, https://doi.org/10.4028/www.scientific.net/AMM.892.94

Kim, S., Joshi, P., Kalsi, P.S. and Taheri, P. (2018) 'Crime analysis through machine learning', *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference* (*IEMCON)*, IEEE, November, Vancouver, British Columbia, Canada, pp.415–420, https://doi.org/10.1109/Iemcon.2018.8614828

Komninos, N. (2016) 'Smart environments and smart growth: connecting innovation strategies and digital growth strategies', *International Journal of Knowledge-Based Development*, Vol. 7, No. 3, pp.240–263, doi: 10.1504/IJKBD.2016.078536.

Kounadi, O., Ristea, A., Araujo, A. and Leitner, M. (2020) 'A systematic review on spatial crime forecasting', *Crime Science*, Vol. 9, pp.1–22, https://doi.org/10.1186/s40163-020-00116-7

Kuo, P.F. and Lord, D. (2019) 'A promising example of smart policing: a cross-national study of the effectiveness of a data-driven approach to crime and traffic safety', *Case Studies on Transport Policy*, Vol. 7, No. 4, pp.761–771, https://doi.org/10.1016/j.cstp.2019.08.005

Lamari, Y., Freskura, B., Abdessamad, A., Eichberg, S. and de Bonviller, S. (2020) 'Predicting spatial crime occurrences through an efficient ensemble-learning model', *ISPRS International Journal of Geo-Information*, Vol. 9, No. 11, p.645, https://doi.org/10.3390/ijgi9110645

Lim, M., Abdullah, A. and Jhanjhi, N.Z. (2021) 'Performance optimization of criminal network hidden link prediction model with deep reinforcement learning', *Journal of King Saud University-Computer and Information Sciences*, Vol. 33, No. 10, pp.1202–1210, https://doi.org/10.1016/j.jksuci.2019.07.010s

Liu, G. and Guo, J. (2019) 'Bidirectional LSTM with attention mechanism and convolutional layer for text classification', *Neurocomputing*, Vol. 337, pp.325–338, https://doi.org/10.1016/j.neucom.2019.01.078

Lu, Y. and Luo, L. (2021) 'Cohort variation in US violent crime patterns from 1960 to 2014: an age–period–cohort-interaction approach', *Journal of Quantitative Criminology*, Vol. 37, pp.1047–1081, https://doi.org/10.1007/s10940-020-09477-3

Ma, J., Cheng, J.C., Jiang, F., Chen, W. and Zhang, J. (2020) 'Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques', *Land Use Policy*, Vol. 94, p.104537, https://doi.org/10.1016/j.landusepol.2020.104537

Ma, Y., Peng, H., Khan, T., Cambria, E. and Hussain, A. (2018) 'Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis', *Cognitive Computation*, Vol. 10, pp.639–650, https://doi.org/10.1007/s12559-018-9549-x

McClendon, L. and Meghanathan, N. (2015) 'Using machine learning algorithms to analyze crime data', *Machine Learning and Applications: An International Journal (MLAIJ)*, Vol. 2, No. 1, pp.1–12, https://doi.org/10.5121/mlaij.2015.2101

Mittal, M., Goyal, L.M., Sethi, J.K. and Hemanth, D.J. (2019) 'Monitoring the impact of economic crisis on crime in India using machine learning', *Computational Economics*, Vol. 53, No. 4, pp.1467–1485, https://doi.org/10.1007/s10614-018-9821-x

Oueslati, O., Cambria, E., HajHmida, M.B. and Ounelli, H. (2020) 'A review of sentiment analysis research in Arabic language', *Future Generation Computer Systems*, Vol. 112, pp.408–430, https://doi.org/10.1016/j.future.2020.05.034

Ruder, S., Ghaffari, P. and Breslin, J.G. (2016) *Insight*-1 *at Semeval-2016 Task 5: Deep Learning for Multilingual Aspect-Based Sentiment Analysis*, arXiv preprint arXiv: 1609.02748, https://doi.org/10.48550/arXiv.1609.02748

Safat, W., Asghar, S. and Gillani, S.A. (2021) 'Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques', *IEEE Access*, Vol. 9, pp.70080–70094, https://doi.org/10.1109/ACCESS.2021.3078117

Saraiva, M., Matijošaitienė, I., Mishra, S. and Amante, A. (2022) 'Crime prediction and monitoring in Porto, Portugal, using machine learning, spatial and text analytics', *ISPRS International Journal of Geo-Information*, Vol. 11, No. 7, p.400, https://doi.org/10.F3390/ijgi11070400

Schwartz, M. and Hornych, C. (2011) 'Knowledge sharing through informal networking: an overview and agenda', *International Journal of Knowledge-Based Development*, Vol. 2, No. 3, pp.282–294, doi: 10.1504/IJKBD.2011.042520.

Shankar, M., Shankar, B. and Mishra, A. (2022) *Forensic GIS in Crime Mapping of North-West Region of Bihar*, India, doi: 10.5958/0974-083X.2022.00021.8.

Snaphaan, T. and Hardyns, W. (2021) 'Environmental criminology in the big data era', *European Journal of Criminology*, Vol. 18, No. 5, pp.713–734, https://doi.org/10.1177/14773708 19877753

Sollund, R. (2021) 'Green criminology: its foundation in critical criminology and the way forward', *The Howard Journal of Crime and Justice*, Vol. 60, No. 3, pp.304–322, https://doi.org/10.1111/hojo.12421

Srinivasulu Raju, S., Narasimha Swamy, G., Rejoice Angelina, M., Sai Snehitha, M., Sai Chandana, M. and Priya Mythili, M. (2022) 'Analysis and prediction of crime using machine learning techniques', *Smart and Intelligent Systems*, Springer, Singapore, pp.343–355, https://doi.org/10.1007/978-981-16-2109-3_33

Stalidis, P., Semertzidis, T. and Daras, P. (2021) 'Examining deep learning architectures for crime classification and prediction', *Forecasting*, Vol. 3, No. 4, pp.741–762, https://doi.org/10.3390/forecast3040046

Vogel, M. (Ed.) (2020) *Crime, Inequality and the State*, Routledge, London, https://doi.org/10.4324/9781003060581

Wang, B., Yin, P., Bertozzi, A.L., Brantingham, P.J., Osher, S.J. and Xin, J. (2019) 'Deep learning for real-time crime forecasting and its ternarization', *Chinese Annals of Mathematics, Series B*, Vol. 40, No. 6, pp.949–966, https://doi.org/10.1007/s11401-019-0168-y

Wang, C., Han, B., Patel, B. and Rudin, C. (2022) 'In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction', *Journal of Quantitative Criminology*, Vol. 39, pp.1–63, https://doi.org/10.1007/s10940-022-09545-w

Wang, N., Zhao, S., Cui, S. and Fan, W. (2021) 'A hybrid ensemble learning method for the identification of gang-related arson cases', *Knowledge-Based Systems*, Vol. 218, p.106875, https://doi.org/10.1016/j.knosys.2021.106875

Xue, W. and Li, T. (2018) *Aspect Based Sentiment Analysis with Gated Convolutional Networks*, arXiv preprint arXiv: 1805.07043, https://doi.org/10.48550/arXiv.1805.07043

Ye, X., Duan, L. and Peng, Q. (2021) 'Spatiotemporal prediction of theft risk with deep inception-residual networks', *Smart Cities*, Vol. 4, No. 1, pp.204–216, https://doi.org/10.3390/smartcities4010013

Yu, L., Du, B., Hu, X., Sun, L., Han, L. and Lv, W. (2021) 'Deep spatio-temporal graph convolutional network for traffic accident prediction', *Neurocomputing*, Vol. 423, pp.135–147, https://doi.org/10.1016/j.neucom.2020.09.043

Zhang, J., Wang, W., Xia, F., Lin, Y.R. and Tong, H. (2020) 'Data-driven computational social science: a survey', *Big Data Research*, Vol. 21, p.100145, https://doi.org/10.1016/j.bdr.2020.100145