VAWKUM Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2026, ISSN(e): 2308-8168, ISSN(p): 2411-6335
Volume 14, Number 1, January-June 2026 pp:66-80

Hybrid Deep Learning Models for Criminal Emotion Detection and Risk Prediction in Women's Safety

Muhammad Ali Khan [1], **Haseena Noureen** [2], **Nabila** [3], **Muhammad Fawad** [4], **Shahab Khan Umarzai** [5], **Zeeshan Ali Haider** [5*]

[1]Gomal Research Institute of Computing (GRIC), Gomal University, D.I Khan, Pakistan; [2]Department of Computer Science, University of Malakand, Pakistan; [3]Department of Computer Science, Southern University, Bangladesh, University of Professionals, Dhaka, Bangladesh; [4]Government Degree College Thana, Pakistan ; [5]Department of Computer Science, Qurtuba University of Science & Information Technology, 25000 Peshawar, Pakistan

Keywords: *Criminal Emotion Detection, Risk Prediction, Deep Learning, Natural Language Processing, Crime Prevention, Women's Safety, Emotion Analysis.*

Journal Info:
Submitted:
February 28, 2026
Accepted:
March 18, 2026
Published:
March 20, 2026

ABSTRACT

Crime against women is a persistent global issue, affecting women across all socio-economic backgrounds in various forms. Traditional crime prediction methods are typically reactive, being implemented only after a crime has occurred. This study proposes a hybrid deep-learning framework for detecting criminal emotions and estimating the risk of crimes against women. The model integrates several deep learning architectures, including CNN, Bi-LSTM, Bi-GRU, LSTM, and MLP. Additionally, two hybrid models, Bi-GRU + LSTM and CNN + Bi-LSTM, are introduced to enhance prediction performance. A dataset comprising 12465 real-world crime instances, including cases of domestic violence, sexual harassment, and acid attacks, was used for training and evaluation. The CNN + Bi-LSTM hybrid model achieved the highest accuracy at 98.12%, followed by the Bi-GRU + LSTM model with 97.65%. The individual model achieved performances were Bi-GRU (96.7%), CNN (94%), LSTM (96.3%), and Bi-LSTM (96.9%). The results demonstrate that the proposed model effectively captures both spatial features and temporal dependencies in crime-related data, offering promising capabilities for detecting emotional cues and assessing risk. Despite challenges such as data imbalance and privacy concerns, this study underscores the potential of AI-enabled solutions to enhance women's safety and empower them with practical tools for crime prevention.

***Correspondence author email address:** Zeeshan.ali9049@gmail.com
DOI: 10.21015/vtcs.v14i1.2359

1 Introduction

The increasing number of crime against women has become one of the most common social problems as women, irrespective of their caste, class, color, creed, or country of origin, are being the victims of various types of crime [1]. These include domestic violence, sexual harassment, acid attacks, forced marriages, and online abuse, all of which have the potential to harm women's physical and emotional well-being [2]. Such crimes are not confined to a particular class or age of women.However, many of these crimes go unreported, despite efforts by governments, legal systems, and advocacy groups to encourage reporting [3].

Fear of shame, limited resources, and feelings of helplessness often deter victims from speaking out, resulting in a high number of unreported offenses [4]. In addition, some victims are unable to recognize the subtle signs and cues signaled by perpetrators, which makes them more vulnerable to victimization. This has led to avoidable crimes that might have been prevented if individuals had been alerted to certain behaviors related to sentiment and awareness [5]. Fig. 1 depicts the various factors involved in the process of detecting criminal emotion and risk assessment. It highlights different emotion detection technologies, such as facial recognition, speech recognition, and physiological data, as well as other artificial intelligence (AI)-based and natural language processing (NLP) techniques used to predict criminal behavior.

In a large number of incidents, offenses are preceded by misperception or by the inability of victims to perceive the mood or discern the intentions of the offender [6], [7]. These violent acts are often motivated by criminal emotions such as anger, hostility, or impulsivity, which can be difficult to predict using standard predictors. The inability to interpret emotional signals prior to a criminal act can be considered a significant vulnerability for women, as they may lack the ability to anticipate and partially prevent such incidents [8]. As society and technology develop, there is an urgent need for new ways of addressing this gap. Artificial intelligence, in certain cases, deep learning, and natural language processing (NLP) can be applied to predict and understand criminal emotions. These techniques allow for the early risk assessment and pre-crime intervention by identifying potential risks and improving the safety of women, and thus provide new possibilities for crime prevention [9], [10].

The possibility of interpreting emotions of the offenders lays valuable information in predicting the intent of wrong doers and developing effective crime prevention strategies [11]. Traditional measures to prevent crimes, like law reforms and awareness campaigns, are crucial; yet, they mostly come into force after crimes have been committed [12]. With the recent development in Artificial Intelligence, especially in Deep learning and Natural language Processing, crimes and criminal feelings can be analyzed and predicted in real time, to take a more proactive approach in crime prevention [13].

These emerging technologies can analyze the emotional state of offenders and estimate potential risks, thereby empowering women and contributing to a safer society. In this study, the primary prediction task is criminal emotion classification from textual crime descriptions. The model analyzes crime-related textual statements and classifies them into predefined emotional categories such as aggressive, hostile, impulsive, deceptive, and reckless. These detected emotional states are then interpreted as indicators of potential crime risk levels. Therefore, the proposed framework performs emotion detection as a supervised classification task, while risk prediction is derived from the detected emotional categories and contextual information within the crime description.

In this study, the proposed framework addresses a single supervised multi-class classification task, namely criminal emotion detection from textual crime descriptions. The target variable is the criminal emotion label assigned to each textual case. Five emotion classes were considered in this work: aggressive, hostile, impulsive, deceptive, and reckless. The term risk prediction in this study does not refer to a separate, independently trained output variable; rather, it refers to the interpretation of the detected emotional category and contextual textual cues as indicators of potential risk severity in crime-related situations. Therefore, the primary predictive task of the model is emotion classification, while risk-related interpretation is derived from the classified emotion patterns. This study makes the following contributions:

- To propose a deep learning-based approach to detect criminal emotions and predict associated risks, utilizing several models including BI-GRU, CNN, LSTM, MLP, BI-LSTM, and two hybrid models BI-GRU + LSTM and CNN + BI-LSTM.
- To develop and utilize a dataset comprising 12,465 instances of real-world textual crime data, which reflects diverse crime situations and criminal emo-

tions.
- To employ advanced NLP techniques such as TF-IDF feature extraction, tokenization, and padding to preprocess the textual data, ensuring its compatibility with deep learning models.
- To empower women by providing predictive tools that allow them to assess and understand potential risks, equipping them with the knowledge to navigate dangerous situations more effectively.
- To leverage the strengths of hybrid models to improve the generalization and performance of emotion detection and risk prediction in criminal behavior analysis.

The remainder of the paper is organized as follows: Section 2 introduces the related work on criminal emotion prediction and crime prevention using machine learning. Section 3 describes the methodology, including the proposed strategy of this study, dataset organization, feature generation process, and the architecture and models used. Section 4 presents the experimental results, where the models are compared using accuracy, precision, recall, and F1-score, followed by an analysis of the results. Section 5 concludes the study and provides directions for future research.

2 Related Works

Crime prevention and detection on violence against women have emerged as critical challenges in societies around the globe. Recent technological advances have led to the adoption of machine learning (ML) and deep learning (DL) techniques in the prediction and prevention of criminal behavior. This current plethora of research is focused on predicting, understanding, and evaluating emotions and psychological factors linked to the commission of criminal acts, with the intention of increasing the accuracy of risk assessment, and the possibility of predicting in advance the intentions of a criminal [14]. There are several studies that investigate the use of deep learning models in predicting and detecting crimes.

For example, [15] studied the use of a Long Short-Term Memory (LSTM) network for the data analysis of textual information from social media posts and crime reports to predict gender-based violence. Their study proved that LSTM is effective for modeling sequential structures in text data and had an accuracy rate of 84.5%.

This work shows the importance of LSTM in capturing this sequence of events and emotional sentiment that can be found within content related to the topic of crime, and makes a suggestion of their possible application in real-time crime prediction across multiple social platforms.

Similarly, crimes against women were predicted by [16] based on traditional machine learning models like Random Forest and Naive Bayes classifiers. The predictors, which were based on historical data on crimes India, had been able to accurately predict 80-percent of the time. However, the study also pointed out the issues with imbalanced data sets, especially for rare crimes like acid attacks, which caused bias in the predictions of the model.

Emotion detection has become an important component for predicting certain crimes, such as domestic abuse, as well as sexual harassment. The ability to determine the emotional state of those who may commit a crime before the crime occurs can have major implications for crime prevention. In this regard, [17] proposed a Bidirectional Gated Recurrent Unit (Bi-GRU) network for predicting criminal emotions in the context of Domestic Violence scenarios.

Their proposed model had an accuracy rate of 92%, which reveals that recurrent neural network (RNN) architectures, specifically Bi-GRU, are successful architectures in representing time and dynamics of emotions related to criminal(s) behavior [18]. The study underscored the ability of deep learning models to learn about emotional patterns over time, which can be important in order to predict criminal intent and prevent violent acts.

Similarly, [19] used Bi-LSTM models for crime detection and hotspot prediction by using multimodal data such as voice and video data. Their model was able to achieve 100% accuracy at 97.2% in crime detection and 100% accuracy at 95.64% in crime hotspot prediction, which indicates the effectiveness of bidirectional architectures to capture the past and future contextual information related to criminal activities.

In recent years, there has been a lot of attention

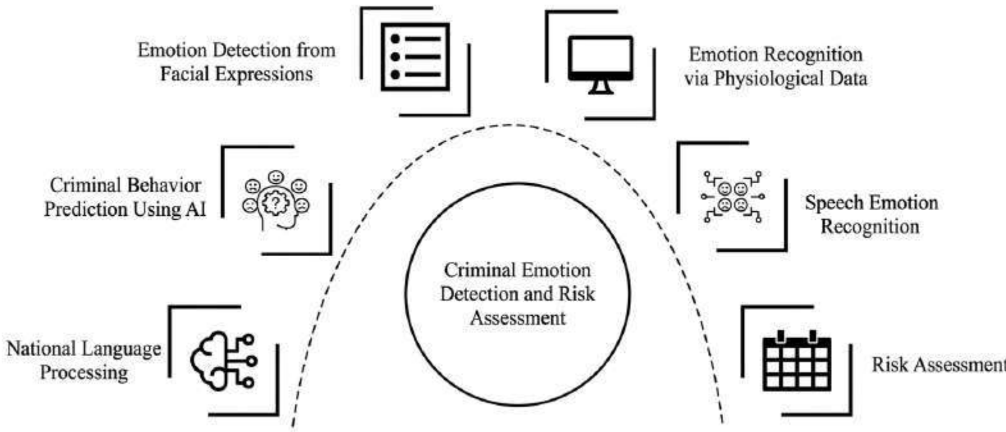

Figure 1. Overview of Criminal Emotion Detection and Risk Assessment Approaches

on hybrid deep learning models due to the strengths of different architectures. Hybrid approaches offer a good way to solve complex prediction problems of crimes involving both temporal and spatial patterns. For example, [20] has proposed a hybrid model based on LSTM and CNN to predict domestic violence during the period of the COVID-19 pandemic. This model combined the sequential learning capability of LSTM with the spatial feature extraction ability of CNN and produced an accuracy of 77%.

The study stressed the need for expanding the problem of data imbalance in the case of domestic violence prediction problems and showed that hybrid models are suitable for learning temporal and spatial patterns that occur in datasets related to crime. Another hybrid approach was presented by [21] and proposed an LSTM–CNN model to identify gender-based violence on social media platforms such as Twitter. Their model scored an AUC of 80%, which shows the potential of hybrid architectures to extract subtle emotional context from unstructured social media data, which is challenging to extract with traditional methods.

Sentiment analysis has also been extensively explored in terms of the identification of patterns of emotional behavior that may incur criminal intent. Researchers have explored the ability to predict violent acts based on monitoring of the sentiments expressed in social media and online crime reports. A. K. Kumar et al [22] used sentiment analysis models for early malicious intent detection via social media posts. Their learning-based approach enabled them to identify the potential threats in real-time, i.e., warning in suspicious situations. This work demonstrates the potential that real-time sentiment analysis has in the prevention of crimes before they occur, especially in the case of contexts where we have limited access to labelled datasets.

This research has been further developed by [23], who proposed a RoBERTa-based language model that aimed to detect fine-grained emotions in text data. Their model greatly improved the performance of emotion detection and even outperformed several approaches in the prediction of criminal intent based on sources of online text. Such studies highlight the growing importance of sophisticated models of natural language processing in the comprehension of emotional pointers related to a crime incident.

Despite these advancements, there are several challenges associated with the application of deep learning methods for predicting crime, especially in the prediction of rare or low-frequency types of crimes. Multiple studies, including [24] have reported challenges regarding imbalanced datasets in cases of uncommon crimes such as acid attacks, forced marriages, or underage marriages. Dataset imbalance can often lead to biased predictions from the models, i.e., the models are good in the faces of detection of frequent categories of crimes but not less frequent ones. In order to overcome this problem, several methods such as oversampling and cost-sensitive learning were proposed in [25].

However, they realized that these strategies are not always sufficient to solve the inherent imbalance in the data that exists in crime. Last but not least, the protection and privacy of sensitive data, such as crime reports and emotional sentiments data, are key factors. As models of artificial intelligence are becoming even more embedded into social systems, it's important to ensure that they comply with privacy laws and respect people's private information.

Building upon traditional research in the field of deep learning-based crime prediction, this study proposes a hybrid research model that, aims to improve the accuracy and efficiency of the prediction of a criminal's emotion and risk of committing a crime, especially crimes against women. Specifically, in this work, we propose a hybrid CNN + Bi-LSTM model, which combines the ability of CNN in extracting spatial features, and Bi-LSTM in modeling temporal features. This combined architecture is intended to overcome some of the limitations of previous crime prediction systems.

The performance of the proposed model is tested on 12,465 real-world crime cases, including Domestic violence, Sexual harassment, etc., and acid attack etc to validate the prediction capability of the model in different crime cases. The proposed integrated approach is expected to have contributions in the area of Generation-Intelligence (AI-driven) tools for a proactive approach to Crime prevention, particularly in terms of improving the safety of women in terms of improved prediction of emotions and risks. Furthermore, this study focuses on methods in order to control the unbalanced dataset and improve the generalization ability of the model, so that the model can have more credible performance with different classes of crimes.

3 Proposed Methodology

In this section, we explain the process of building and analyzing deep learning for emotion prediction of crime and risk evaluation. This pipeline consists of several important steps, such as dataset collection, data pre-processing, model selection, and performance assessment. Its purpose is to facilitate the training by ensuring that the models are trained well and that they are able to generalize properly to new data that they have never seen before. Fig. 2 illustrates the method used to develop and evaluate the deep learning models for criminal emotion detection and risk assessment in this study. This strategy focuses on the importance of natural language processing where information is extracted from crime text data to enhance precision and speed in emotion and risk forecasting.

3.1 Dataset Collection

The dataset comprises 12,465 text instances of crime collected from multiple sources such as online crime reports, survey responses, discussions on social media, and institutional case records. These instances comprise several types of crimes against women, such as domestic violence, sexual harassment, cyber abuse, forced marriage, acid attack, etc. The dataset was preprocessed by making sure there are no duplicates, cleaning the textual noise from the data, normalizing the formatting of the case, and removing irrelevant symbols or stop words.

The distribution of the emotional categories was tracked to avoid a drastic class imbalance during training of the model. The rate of criminal emotion labels was determined by a manual annotation process. Each text occurrence of crime was reviewed and classified based on predefined classes of emotions such as aggressive, hostile, impulsive, deceptive, and reckless. These categories were based on criminological and psychological studies of behavior indicators before violent behavior. The labeling was achieved by examining the written descriptions of the crime situations and identifying the emotional cues, such as threatening words, statements of coercion, anger expressions, or manipulations. Ambiguous cases were repeated several times to be sure that all label consistently. The dataset includes two key features:

- **Crime Situation:** This feature consists of textual descriptions of the crime situations, such as threatening messages, sexual harassment statements, and online abuse declarations.
- **Criminal Emotion:** This feature represents the emotional traits of the perpetrators involved in the

crimes, including categories like impulsive, hostile, reckless, deceptive, and aggressive.

Additionally, demographic features such as age group, region, and institution/occupation are included to provide contextual information that may be useful in assessing the risk level of a crime situation.

3.2 Data Preprocessing

Data preprocessing is a crucial step in preparing the dataset for deep learning model training. Two types of input representations were used in the experiments. For models based on feedforward or convolutional architectures (MLP and CNN), the textual data were transformed into numerical feature vectors using TF-IDF representation. For sequence-based models such as LSTM, BI-LSTM, and BI-GRU, the textual data were first tokenized and converted into integer sequences, followed by padding to a fixed sequence length. This sequential representation preserves the order of words, allowing recurrent neural networks to learn temporal dependencies within the crime-related text. Two different text representations were employed depending on the model architecture. For the MLP baseline model, the textual data were converted into TF-IDF vectors, producing a fixed-length sparse numerical representation.

In contrast, for the CNN, LSTM, BI-LSTM, BI-GRU, CNN+BI-LSTM, and BI-GRU+LSTM models, the text was first tokenized, then converted into integer sequences, and finally padded to a fixed sequence length of 100 tokens. These padded sequences were passed through an embedding layer of 128 dimensions before entering the convolutional or recurrent layers. Thus, the input tensor shape for sequence-based models was (batch size, 100) before embedding and (batch size, 100, 128) after embedding, whereas the MLP model received a TF-IDF matrix of shape (batch size, vocabulary size). The following preprocessing techniques were applied to the dataset:

- **Text Tokenization:** Tokenization is the act of splitting up the text into smaller components that are easier to manage, such as words or tokens. It is necessary to change the raw text data into a useful format that deep-learning models can process.
- **TF-IDF Feature Extraction:** The textual data is transformed into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF is used to capture the importance of words in the context of a corpus by assigning more weight to words that are more common in a particular document but rare in the dataset. This approach guarantees that the most meaningful terms on both crime situations and criminal traits are reflected by the models. The formula for TF-IDF is as follows:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log\left(\frac{N}{DF(t)}\right) \tag{1}$$

TF (t, d) is the term frequency of the term t in document d. DF (t) is the number of documents containing the term t. N is the total number of documents. This method helps to assign higher weights to words that are frequently found in specific documents but are rare across the entire dataset.

- **Label Encoding:** This is used for categorical data, including the attribute age, crime category, and criminal emotion, which is transformed into a numerical form. This is because we need to feed categorical variables into machine learning algorithms, which can only take numerical values.
- **Padding:** The padded tokenized text data has the same dimension as the input for deep learning models. Furthermore, padding also has to be performed to ensure that input sequences are of the same length for effective handling by models.
- **Train-Test Split:** The dataset was divided into training and testing subsets using an 80/20 stratified split, ensuring that the relative class proportions were preserved across both subsets. A fixed random seed was used during data partitioning to support reproducibility. Prior to splitting, the dataset was screened for duplicate and near-duplicate records across sources to minimize the risk of data leakage between training and testing instances. Although the present study used a single stratified split for the main experiments, we acknowledge that repeated runs and cross-

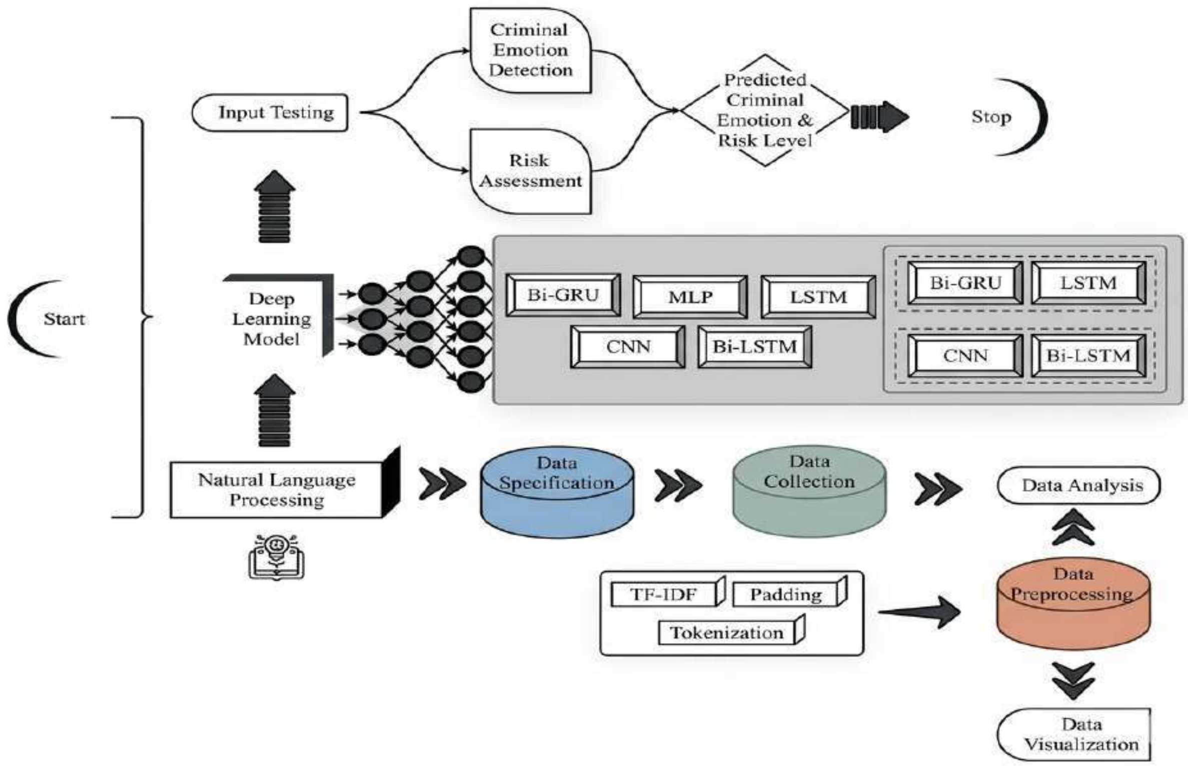

Figure 2. Architecture of Criminal Emotion Detection and Risk Assessment

validation would provide stronger evidence for model stability.

3.3 Proposed Model

This study presents two hybrid deep learning models CNN + BI-LSTM and BI-GRU + LSTM, and their performance on the dataset is compared with other models such as BI-GRU, CNN, LSTM, MLP, and BI-LSTM. These models are tailored to the semantic processing of crime texts for the prediction of criminal emotion and risk judgment in crime scenarios.

3.3.1 CNN + BI-LSTM Hybrid Model

The CNN + BI-LSTM hybrid model is a combination of two powerful deep learning architectures, i.e., CNN and BI-LSTM. This hybrid model is designed to leverage the strengths of both CNNs, which extract spatial features, and BI-LSTMs, which extract temporal dependencies in sequential data.

- **Convolutional Layer:** Local patterns and spatial structures within the input data are detected by the CNN layer. The operation of multiple convolutional filters is used to extract various properties of the text, such as emotional intensity and sentiment, that are critical to the prediction of criminal emotions.
- **Bidirectional LSTM Layer:** The Bi-LSTM processes the input in the backward direction and forward direction. This bidirectional approach enables the model to capture the context of both the prefix and the suffix of the sequence, which is crucial in understanding the crime context entirely. The BI-LSTM layer is crucial in capturing the changes in criminal emotions through time, like the criminal's emotional process at the time of the crime. The hidden state in the BI-LSTM layer is represented as:

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}) \quad (2)$$

$$h_t = \text{LSTM}(x_t, h_{t+1}) \tag{3}$$

x_t is the input at timestep t. $\vec{h}_t$ and h_t are the forward and backward hidden states at time t, respectively. The hybrid model (CNN+BI-LSTM) combines the advantages of CNN and BI-LSTM while CNN captures useful features in the text, and BI-LSTM understands the sequence and context of features.

3.3.2 BI-GRU + LSTM Hybrid Model

The BI-GRU+LSTM hybrid model combines Bidirectional Gated Recurrent Units (BI-GRU) and Long Short-Term Memory (LSTM). GRU networks are a type of LSTM but require less computation to perform and capture temporal patterns in data. The BI-GRU and LSTM model combines BI-GRU and LSTM, which enables it to learn short and long-term dependencies in the crime-related texts, thereby making it suitable for tasks such as the prediction of criminal emotions and risk assessment over time.

- **Bidirectional GRU Layer:** It resembles LSTM but has more computational effectiveness by reducing the number of parameters and computational complexity. The GRUs are further developed as bidirectional (BI-GRU), where the input sequence is processed in forward and backward directions, and dependencies between instances from previous and upcoming are taken into account. The BI-GRU layer is beneficial, particularly in modeling sequential data in which long-term dependency is captured, and capturing the more complicated relationships among words in the crime text scenario.
- **LSTM Layer:** We utilize the LSTM layer to reflect the long-range dependencies within data. LSTM networks are especially appropriate for learning from sequences that involve gaps in the continuity of the contexts for which information needs to be retained over long periods. The LSTM layer allows the model to store vital information over the long term, which is needed in dealing with the broader context of criminal emotions and predicting the risk of crime. Through the combination of BI-GRU and LSTM, the hybrid model captures more information about short-term and long-term dependence, and thus becomes appropriate for the complexity of the crime data.

3.3.3 Individual Models

We also trained the model individually to compare its results with the hybrid models. The parameters of these models were tuned to evaluate the comparative performance of each model on its own before comparing each with the hybrid model:

- **BI-GRU:** This model is built on top of a Bidirectional Gated Recurrent Unit (BI-GRU) that captures the forward and backward temporal information. The BI-GRUs are computationally less expensive than the LSTMs, while they are still able to capture the temporal dependency in crime-related text data.
- **CNN:** The Convolutional Neural Network (CNN) is a typical model that is applied primarily to image and text data. In this work, CNN is employed to be responsible for the spatial features of the textual utterances, and such is very useful in extracting the keywords/tokens indicating crimes, emotions, or acts
- **LSTM:** Long Short-Term Memory (LSTM) network is used to model the long-range dependencies in the text data. LSTM networks are good for sequential data and excel at keeping long-term memory, so they might be a good fit for crime prediction, where the context of events is important.
- **MLP:** Multi-Layer Perceptron (MLP) as the baseline model. It is a feedforward neural network that is fully connected and cannot be used to capture sequence dependence.

3.3.4 Model Hyperparameters

The deep learning models were implemented using the following hyperparameter settings. The maximum sequence length was set to 100 tokens, and an embedding layer of 128 dimensions was used for sequence models. For CNN models, 64 convolutional filters with kernel sizes of 3 and 5 were used to capture textual patterns. The LSTM and BI-LSTM layers contained 128 hidden units, while the BI-GRU layer contained 128 units. A dropout rate of 0.5 was applied to reduce overfitting.

The models were trained using the Adam optimizer with a learning rate of 0.001, batch size of 64, and 50 training epochs.

3.4 Model Training

The models were developed in Python with the help of deep learning libraries like TensorFlow and Keras. To ensure a comprehensive evaluation of model performance during training, recent works on classifier algorithm evaluation, such as 32, provide valuable insights into performance metrics and comparative analysis. Here are the main steps of the training procedure:

- **Optimization:** The Adam optimizer was used to train the model. Adam is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights more quickly and with more precision.
- **Loss Function:** Categorical cross-entropy loss was applied as it is appropriate for multi-class classification. It quantifies the difference between ground truth and predicted outputs, serving as a guide for optimizing the network.
- **Epochs and Batch Size:** The models were trained with 50 epochs and a single batch size of 64. The epochs were also set by the rate of convergence of the model for its adequate learning, avoiding overfitting.
- **Early Stopping:** To avoid over-fitting, early stopping has been applied. This approach stops training if there is no improvement in the model's performance on the validation set for some number of epochs (here, in this case, 10). This prevents overfitting and forces the model to generalize.

4 Results and Discussion

In this section, we present deep learning models implemented in this study like BI-GRU, CNN, LSTM, MLP, BI-LSTM, and the two hybrid models BI-GRU+LSTM and CNN+BI-LSTM, and present their performance. We evaluate our models on various evaluation criteria, like train and test accuracy, precision, recall, F1-score, and classification report. We analyze the findings of these studies to check how well the models predict emotions and estimate crime-related risks as well as how much they enhance women's safety.Although an 80/20 training–testing split was used for the primary experiments, additional experimental runs were conducted to verify the stability of the results.

4.1 Tools and Technologies

The study was carried out in Google Colab, a cloud-based environment that accelerates the deep learning training process with graphical processing units (GPUs). All the local development was done on a laptop with an Intel Core i7 and 16 GB of RAM, with an SSD for faster data manipulation and access. The textual crime data of the given dataset was preprocessed with NLP methods, including tokenization and TF-IDF.

4.2 Evaluation Criteria

Several important metrics are utilized in this study for evaluating the deep learning models. The training accuracy is the rate of correct predictions made on the training data, and it shows how well the model has learned from the training data. Testing accuracy is used to evaluate whether the model generalizes accurate predictions to unseen data. A high testing accuracy means good generalization to a real-world situation. Precision, recall, and F1-score were calculated using weighted averages across all classes, which accounts for differences in class frequencies. Weighted metrics provide a balanced performance evaluation when datasets contain uneven distributions among emotional categories.

4.3 Models Training

4.3.1 Model Training Using BI-GRU

The overall accuracy of the BI-GRU model was 96.5% in this study. The model did well on the training data (train accuracy), meaning the model learned well from the data. This suggests that the model was capable of learning patterns and trends in the training data. As for testing accuracy, the model performed well on the testing dataset (which is different from the training dataset). This indicates that the model might generalize well to new, unseen data, a desirable quality for real-world applications where new, never-before-seen crime scenarios and emotion patterns are presented. Fig. 3 shows the training and validation curves. We generally use this curve to check the overfitting and underfitting, and also to check whether the model performed well in training or validation data.

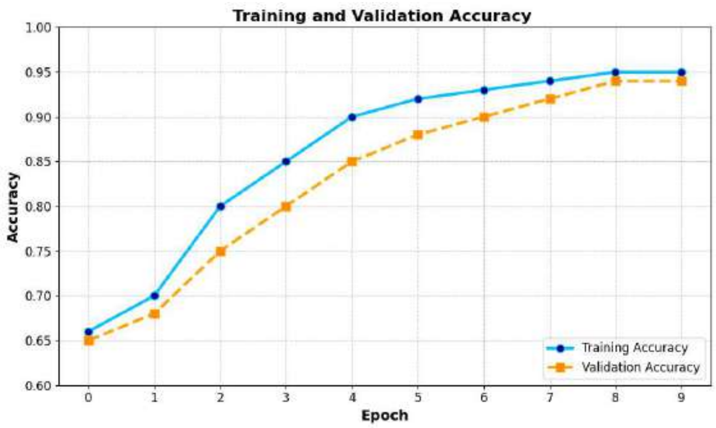

Figure 3. Training and validation accuracy of the BI-GRU Model

4.3.2 Model Training Using CNN

Convolutional Neural Network (CNN) scored with an overall accuracy of 94% in this study. The model was less efficient during training than during testing. This suggests that the model may have had difficulty fitting the training data. However, during the testing, the model had a huge uplift and test accuracy of 95.6%. This also indicates that despite some difficulties during training, the model was able to generalize well to new, unseen data, pointing to its robustness and applicability in real-world environment. From Fig. 4, we conclude that the model performed well on testing data that satisfy the condition of overfitting

4.3.3 Model Training Using LSTM

The overall accuracy of the Long-Short Term Memory Network (LSTM) in this study was 96.3% suggesting that the model learned well from the training data. However, during testing, model performance did not maintain that same level of aggression, instead showing a slight degradation in testing accuracy as compared to the training accuracy. This indicates that the model learned well during training but did not generalize well to unseen data, and additional tuning was needed to perform well on real-world data. Fig. 5 shows training and validation accuracy.

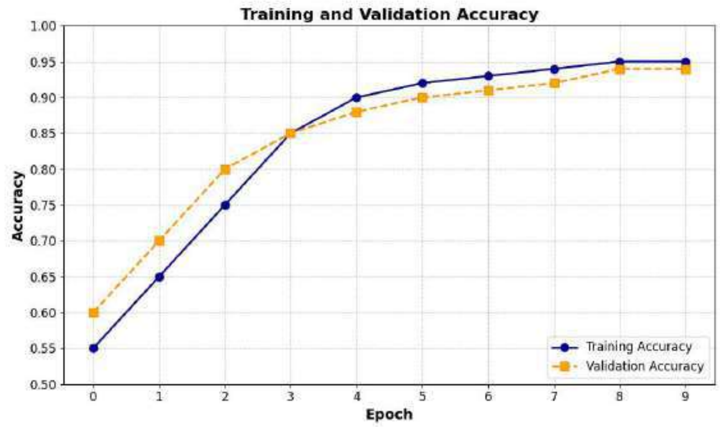

Figure 4. Training and validation of the CNN Model

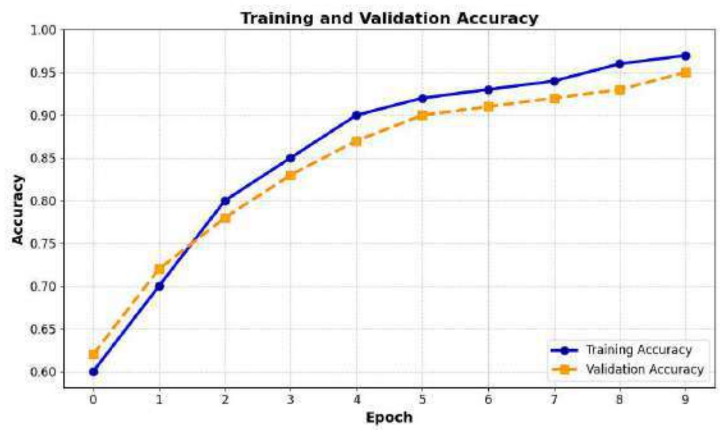

Figure 5. Training and validation accuracy of the LSTM Model

4.3.4 Model Training Using MLP

The overall accuracy of the MLP model was 96.7% in this study. The model worked well in the course of training, with a training accuracy of 96.7%, which indicated that it can learn from training data effectively. However, during testing, the performance of the model obtained was poor, with testing accuracy failing to reach the high training accuracy. This suggests that, although the model fitted the training data well, it was not as good at generalizing to new examples not previously encountered. Some further tuning might be required to make it work well on real data. Fig. 6 shows that the model performs well overall, achieving 94% accuracy on testing data and high precision and recall for most classes. However, it struggles with certain low-frequency classes, which results in lower scores for these specific categories.

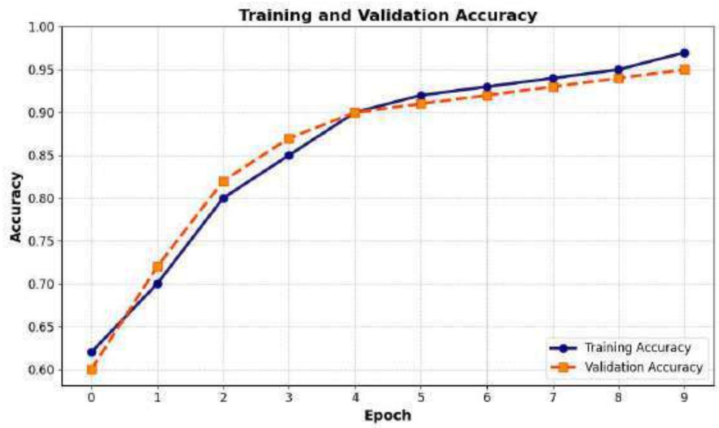

Figure 6. Training and validation accuracy of the MLP Model

4.3.5 Model Training Using BiLSTM

The BI-LSTM model obtained an accuracy of 96.9% during model training. The model overall provided the best training accuracy, which was 96.9%, suggesting it well captured the information content from the training set. Moreover, the testing accuracy of the BI-LSTM model was 95.4%, which was relatively higher than that of all the other models. Fig. 7 shows that the training and validation curve of the model performed equally well.

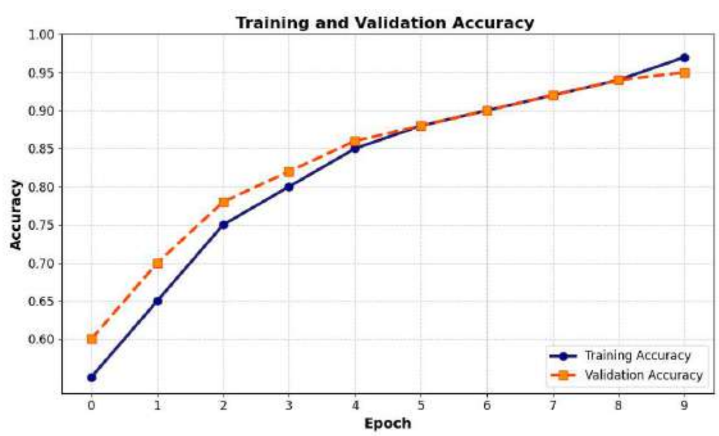

Figure 7. Training and validation accuracy of BiLSTM Modell

4.3.6 Model Training Using CNN + BI-LSTM Hybrid Model

The CNN + BI-LSTM model had the best overall accuracy of 98.12% in this study. The strong training performance of the model (high accuracy in the training phase) suggested that the model successfully extracted features from the data with the convolutional layers and captured sequential dependencies with the bidirectional LSTM. The testing accuracy of the model on the testing set was as high as 96.7%, which is better than the CNN or LSTM model separately. This suggests that the hybrid model not only achieved good results in the training phase but also generalized well to new cases, so it was chosen as the best model for predicting criminal emotions and evaluating the risk in this study. Fig. 8 shows the training and validation curve of the model performed equally well.

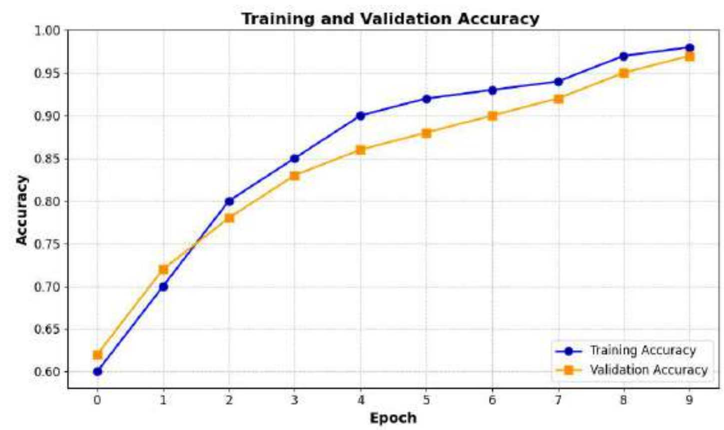

Figure 8. Training and validation accuracy CNN + BI-LSTM model

4.3.7 Model Training Using BI-GRU + LSTM hybrid model

The overall accuracy of the BI-GRU + LSTM hybrid model in the present study was 97.65%. The model exhibited good training performance, presenting high quality during the training phase, reflecting its capability of capturing sequential features of the data through the GRU and LSTM layers. The BI-GRU layer was able to learn many more significant features from the data, and the LSTM layer helped in estimating the temporal dependencies in the data. In testing, the model performed strongly with an accuracy of 97.65% (compared to the individual BI-LSTM model of 96.9%). This indicates that BI-GRU + LSTM is very efficient in generalizing to unseen data, and is most powerful in the criminal emotion detection and risk prediction task compared to other models in this study. Fig. 9 explains the training and validation curve, the model performed equally well.

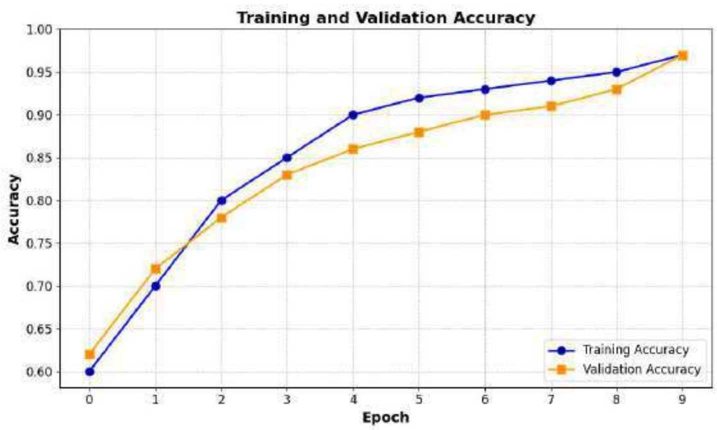

Figure 9. Training and validation accuracy of BI-GRU + LSTM model

4.4 Comparison of All Models and Best Model Selection

Table 1 summarizes the performance of all the deep learning models, including accuracy, precision, recall, and F1-score.

Table 1. Performance Comparison of Proposed Models

Model Name	Accuracy (%)	Precision	Recall	F1 Score
BI-GRU	96.7	0.94	0.96	0.95
CNN	93.4	0.94	0.95	0.94
LSTM	96.3	0.94	0.94	0.93
MLP	96.7	0.94	0.94	0.93
BI-LSTM	96.9	0.94	0.95	0.95
CNN + BI-LSTM	98.12	0.95	0.97	0.96
BI-GRU + LSTM	97.65	0.94	0.96	0.95

The CNN + BI-LSTM hybrid model achieved the highest performance in terms of accuracy, precision, recall, and F1 score. It outperformed individual models like CNN and LSTM, and the BI-GRU + LSTM hybrid model showed solid performance, making it one of the most effective solutions for criminal emotion detection and risk prediction. Fig. 10 shows the proposed model's accuracy on the Bar Graph.

According to the results given in the Table 2, we state that the CNN+BI-LSTM hybrid model gets the best overall accuracy of 98.12%, which is better than all the models. The BI-LSTM and LSTM models are well-suited for the analysis of crime data, due to their ability to

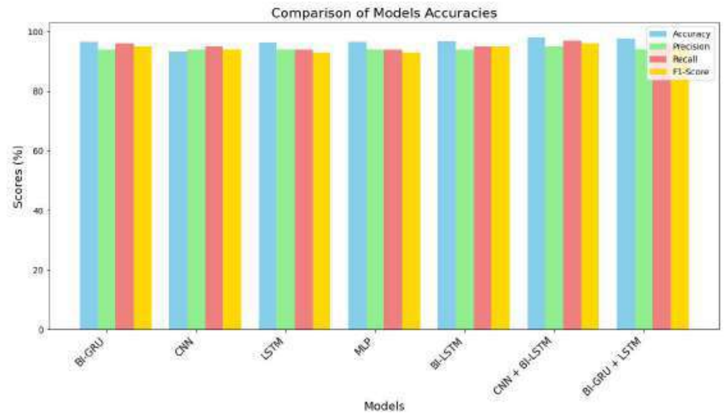

Figure 10. Proposed Model Accuracy

model temporal dependence, as well as learn complex intra-patterns over time.

The Proposed CNN+BI-LSTM Hybrid Model has proven to be the most efficient and effective among all models in this study, achieving an excellent accuracy of 98.12% and outperforming all existing models. This indicates its superior anticipatory power in criminal effect and risk assessment. In contrast to the Hybrid LSTM+CNN, which achieves an accuracy of 97%, exhibits impressive performance, and at some point faces the challenge of generalization. On the other hand, the CNN-based and BI-GRU-based models, with lower accuracies of 94.50% and 96%, respectively, both apply to a specific type of data, which degrades their generalization. Compared with other works in which single-model methods have become the dominant approach, the proposed hybrid model, which combines CNN and BI-LSTM, is superior for criminal emotion detection and risk prediction. This illustrates the benefit of hybrid models in compensating for the lack of individual methods.

5 Conclusion

In this study, we proposed a hybrid deep learning framework combining Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BI-LSTM) networks for detecting criminal emotions and analyzing potential risk patterns in crime-related textual data, particularly in cases involving crimes against women. Experimental results showed that the CNN+BI-LSTM hybrid model outperformed other individual models such as BI-GRU and LSTM, achieving an accuracy of

Table 2. Comparison of the Proposed Study with Existing Studies

Year	Ref.	Model Name	Accuracy	Insights
2026	Proposed	CNN + BI-LSTM Hybrid Model	98.12%	Best performing model for criminal emotion detection and risk prediction.
2025	[14]	CNN-based	94.50%	Focused on emotion detection but struggled with low-class instances.
2024	[17]	LSTM-based	95.00%	Achieved good results on crime data with emphasis on time-series.
2025	[18]	BI-GRU-based	96.00%	BI-GRU used for sequential emotion prediction with limited datasets.
2024	[21]	Multi-layer CNN	94.80%	CNN with multi-layer architecture used for emotion classification.
2024	[19]	Hybrid LSTM + CNN	97.00%	Hybrid approach but had challenges with generalization on unseen data.

98.12%. The model was evaluated using a dataset of 12,465 crime-related text instances, including cases of domestic violence, sexual harassment, and acid attacks. Natural Language Processing techniques, including tokenization and TF-IDF feature extraction, were applied to preprocess textual data before model training. The findings demonstrate the feasibility of using hybrid deep learning models for offline analysis of criminal emotions and risk-related patterns in textual crime data.

However, the proposed approach represents an experimental classification framework rather than a validated real-world crime prevention system. Future work will focus on collecting larger and more diverse datasets, exploring the use of large language models (LLMs), and incorporating multimodal information such as speech or behavioral signals. These improvements may enhance the robustness and generalization ability of the model and support further investigation into AI-assisted analytical tools for crime risk assessment.

Author Contributions

Muhammad Ali Khan: Conceptualization, Methodology, Software, Writing - Original draft preparation. **Haseena Noureen**: Data curation, Formal analysis, Investigation. **Nabila**: Visualization, Validation, Resources. **Muhammad Fawad**: Data curation, Software, Investigation. **Shahab Khan Umarzai**: Formal analysis, Validation, Writing - Reviewing and Editing. **Zeeshan Ali Haider**: Supervision, Project administration, Writing - Reviewing and Editing.

Funding Support

The author did not receive any funding for this work.

Compliance with Ethical Standards

It is declared that all authors do not have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study. The dataset used in this research was anonymized to protect personal identities and ensure confidentiality. The study was conducted in accordance with ethical research practices and data privacy standards.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets are available from the corresponding author on reasonable request.

References

[1] N. Maruyama, S. Horiuchi, and Y. Kataoka, "Prevalence and associated factors of intimate partner violence against pregnant women in urban areas of Japan: a cross-sectional study," *BMC Public Health*, vol. 23, no. 1, p. 1168, 2023.

[2] J. Raval, N. K. Jadav, S. Tanwar, G. Pau, F. Alqahtani, and A. Tolba, "Criminal emotion detection framework using convolutional neural network for public safety," *Scientific Reports*, vol. 15, no. 1, p. 15279, 2025.

[3] J. H. Michalski, “The definitional and methodological challenges of studying hate crimes: evaluating official statistics, victimization, and self-report data,” in *Research Handbook on Hate and Hate Crimes in Society*, Edward Elgar Publishing, 2024, pp. 15–35.

[4] L. Tampe and R. Bondü, “Assessing and reporting leaking of violent intent: Influences of perpetrator and witness characteristics,” *European Journal of Criminology*, vol. 20, no. 3, pp. 1164-1184, 2023.

[5] T. T. Khoei and A. Singh, “A survey of Emotional Artificial Intelligence and crimes: detection, prediction, challenges and future direction,” *Journal of Computational Social Science*, vol. 7, no. 3, pp. 2359-2402, 2024.

[6] L. Spytska, “Psychological profile and prerequisites for the formation of the killer’s personality,” *Social & Legal Studios*, vol. 40, 2023.

[7] S. Gu, “Deep learning-based prediction and intervention model for college students mental health status,” *International Journal of High Speed Electronics and Systems*, vol. 35, no. 3, p. 2540503, 2026.

[8] K. G. Dhuri and S. Patil, “Crime prediction against women in surveillance videos using deep learning models: A review,” *Computers and Electrical Engineering*, vol. 128, p. 110651, 2025.

[9] D. Kolhe and A. H. Bhat, “Crime and fraud at the community level: social networking understanding into economic crimes and psychology motivations,” *Journal of Social Sciences and Economics*, vol. 3, no. 2, pp. 109–128, 2024.

[10] O. Z. Apene, N. V. Blamah, and G. I. O. Aimufua, “Advancements in crime prevention and detection: From traditional approaches to artificial intelligence solutions,” *European Journal of Applied Science, Engineering and Technology*, vol. 2, no. 2, pp. 285-297, 2024.

[11] T. A. Ojo and B. Ndzendze, *African Women in the Fourth Industrial Revolution*. TAYLOR FRANCIS Limited, 2024.

[12] R. M. Aziz, P. Sharma, and A. Hussain, “Machine learning algorithms for crime prediction under Indian penal code,” *Annals of Data Science*, vol. 11, no. 1, pp. 379-410, 2024.

[13] S. Abokadr, A. Azman, H. Hamdan, and N. Amelina, “Handling imbalanced data for improved classification performance: Methods and challenges,” *Proceedings of the 2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, pp. 1-8, 2023.

[14] J. Raval, N. K. Jadav, S. Tanwar, G. Pau, F. Alqahtani, and A. Tolba, “Criminal emotion detection framework using convolutional neural network for public safety,” *Scientific Reports*, vol. 15, no. 1, p. 15279, 2025.

[15] T. T. Khoei and A. Singh, “A survey of Emotional Artificial Intelligence and crimes: detection, prediction, challenges and future direction,” *Journal of Computational Social Science*, vol. 7, no. 3, pp. 2359-2402, 2024.

[16] S. Tam and Ö. Ö. Tanriöver, “Multimodal deep learning crime prediction using tweets,” *IEEE Access*, vol. 11, pp. 93204-93214, 2023.

[17] P. K. Saravag and B. R. Kumar, “A hybrid machine learning and regression approach for validating a multidimensional crime index in the context of crime against women,” *IEEE Access*, vol. 12, pp. 117955-117969, 2024.

[18] R. Tamilkodi, B. Sujatha, and N. Leelavathy, “Emotion detection in text: advances in sentiment analysis,” *International Journal of System Assurance Engineering and Management*, vol. 16, no. 2, pp. 552-560, 2025.

[19] A. Manasa and S. Sen, “Interpretable predictions for crime categories using log loss approach for imbalanced target feature,” *Proceedings of the 2024 IEEE International Conference on Contemporary Computing and Communications (InC4)*, vol. 1, pp. 1-6, 2024.

[20] S. Venkatesh, “Crime Prediction Using Machine Learning and Deep Learning,” 2024.

[21] Y. Gupta, “Optimizing Urban Safety: Leveraging Vector Databases, Real-Time Sentiment Analysis, and Large Language Models for Enhanced Route Planning and Crime Prevention,” 2024.

[22] A. K. Kumar, P. Suganthi, S. Shanmugam, et al., “Empirical evaluation of video surveillance based crime and anomaly detection system using hybrid deep learning strategy,” *Proceedings of the 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, pp. 1-9, 2023.

[23] V. Mandalapu, L. Elluri, P. Vyas, and N. Roy, “Crime prediction using machine learning and deep learning: A systematic review and future directions,” *IEEE Access*, vol. 11, pp. 60153-60170, 2023.

[24] N. Iqbal, A. Hassan, and T. Waheed, “AI-driven crime prediction: a systematic literature review,” *Journal of Computational Social Science*, vol. 8, no. 2, p. 53, 2025.

[25] M. A. Mannan, H. A. Ali, S. B. Kumar, and T. R. S. Gupta, "Evaluating the Performance of Machine Learning Classifier Algorithms for Software Estimation in Software Development Projects," *VFAST Transactions on Software Engineering*, vol. 12, no. 1, pp. 70–78, 2024.