International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 September 2023*

Identification of Crime using Multi Embedding BiLSTM

S. Jeya Selvakumari[1], Dr. V. Joseph Peter[2]
[1]Research Scholar, Research Department of Computer Science
Kamaraj College, Affiliated to Manonmanian Sundaranar University, Tirunelveli
Thoothukudi, Tamil Nadu India
jeyrobin@gmail.com
[2]Associate Professor, Research Department of Computer Science
Kamaraj College, Affiliated to Manonmanian Sundaranar University, Tirunelveli
Thoothukudi, Tamil Nadu India
peterchristy_in@yahoo.com

Abstract—Crimes pose significant societal challenges with implications for a nation's well-being, economic progress, and reputation. Precisely measuring crime rates, categories, and hotspots from historical patterns presents various computational complexities and opportunities. This study introduces and improves a deep learning approach for predicting crime types with high precision. The system can predict both crime categories and associated risk levels by analyzing concise summaries from criminal case reports. The predictive model is built on a neural network with LSTM and Bi-LSTM components, demonstrating remarkable accuracy in forecasting crime types despite limited training attributes. It is tested on a substantial real-world dataset containing historical urban crime data, offering a deep learning-based solution to enhance public safety in the face of criminal activities.

Keywords-crime; deep learning; LSTM; Bi-LSTM; classification.

I. INTRODUCTION

The daily tally of reported criminal incidents has experienced a substantial increase, though typically, only the most severe cases receive recognition. Nevertheless, it is imperative to scrutinize both major and minor incidents, as seemingly inconsequential events may have widespread consequences. Illustrative examples of such criminal incidents encompass activities like robbery, vandalism, assault, murder, and kidnapping, among others. Underscoring the importance of crime prevention is as critical as the focus on criminal investigation. Both functions must be efficiently executed within the Thoothukudi District. At present, the surveillance monitoring process heavily relies on manual labor, making it exceedingly impractical. The manual observation and data collection from video footage consumes a significant amount of time.

In the present landscape, Artificial Intelligence (AI) plays a pivotal role across a multitude of research domains. A fundamental component of AI, known as Natural Language Processing (NLP), proves exceptionally valuable when handling structured and unstructured data. An AI-based crime prediction system is proposed to enhance public safety and facilitate law enforcement by alleviating the labor-intensive tasks associated with crime recognition and optimizing resource allocation for prompt emergency responses. This system empowers police officers to remain continuously updated in real-time regarding ongoing incidents in their vicinity and foresee potential future criminal activities. This proactive approach aids law enforcement in devising and executing preventive measures to forestall similar incidents.

The primary objective of the proposed system is to predict crime types based on textual data using an innovative, fully connected Bidirectional Long Short Term Memory (BI-LSTM). BI-LSTM autonomously acquires features through multiple layers and transforms the essence of a crime into a vector format. LSTM trains the system to recognize crime-related terms by employing keyword matching and a rule-based Named Entity Recognition (NER) system. For instance, crimes involving children can be identified through keywords like "minor girl," "child pornography," "sexually assaulted," and "kidnapping." The system is trained to identify analogous terms by using iterative LSTM. By integrating BI-LSTM with GloVe embeddings, the system can make accurate predictions regarding the type of crime.

The paper's structure unfolds as follows: Section 2 provides an overview of related research about the proposed model. In Section 3, a comprehensive explanation of the system model is offered. Section 4 encompasses experiments involving all LSTM structures and a discussion of the outcomes. Finally, in Section 5, a comprehensive summary and a roadmap for future research endeavors are presented.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

II. RELATED WORKS

It is foreseeable that crime will persist and may even increase shortly due to the current structure of international economies and societies. Another noticeable trend is that urban areas experience higher crime rates compared to rural regions, primarily owing to their greater population density and income disparities, which have been linked to crime statistics [1]. Consequently, addressing urban crime calls for a thorough investigation and the implementation of preventive measures.

Following state reports, southern states are predicted to have the highest rates of severe violent crimes [2]. When identifying potential crime hotspots, it has been discovered that Deep Learning algorithms outperform conventional methods such as Random Forest and Naive Bayes [3]. An analysis of Vancouver's crime data over the past 15 years employed two different data processing approaches, utilizing machine learning prediction models like K-nearest neighbor and Boosted Decision Tree, resulting in an accuracy range of 39% to 44% [4]. To enhance crime prediction, data mining methods have played a significant role, contributing to both the prediction and prevention of criminal activities [9][11].

Machine learning and deep learning-based E-police systems aim to improve public safety and support law enforcement by preventing crime. E-Police serves as a real-time incident-reporting application that keeps law enforcement personnel informed about ongoing events and precautionary measures [6]. Leveraging advanced techniques, such as LSTM, ST-GCN, and fully-connected LSTM, have improved prediction accuracies, allowing for better pattern recognition and regularity extraction from historical crime data [7][12][14].

To enhance city safety and crime analytics, a new predictive system is under development, utilizing various machine learning algorithms applied to crime datasets from Chicago and Los Angeles to boost predictive accuracy for crimes. XGBoost has demonstrated the highest accuracy in Chicago datasets, while KNN performs best in Los Angeles datasets [13]. Data analysis using a variety of machine learning methods, including K-means, SVM, Apriori Algorithm, CART Algorithms, Fuzzy-C, and FP trees, has proven beneficial when removing anomalies and normalizing data [15]. An assemble-stacking-based crime prediction system (SBCPM) has been suggested to evaluate performance using advanced data mining techniques and a 5-fold cross-validation procedure [16]. Furthermore, the application of smart police techniques based on machine learning can forecast crime types and risk levels by examining text-based crime summaries [17].

In the realm of artificial intelligence, computer vision enables machines to interpret the visual world, fostering a heightened awareness of their surroundings [18]. Shah et al. introduced a concept of how computer vision, machine learning, and deep learning can collaboratively develop a more beneficial system for law enforcement, capable of tracking criminal hotspots and even identifying individuals through voice analysis [19]. Evidence-based research employing neural networks has led to the development of predictive crime models, including crime classification and the analysis of theft at various spatial and temporal scales, as well as the counting of crime and conflict points in urban areas [29]. Predictive analysis using RapidMiner and data processing through the Hadoop framework has aided police departments in saving time, money, and effort [20]. Wang et al. have provided a real-time spatial-temporal predictor for predicting crime intensity from start to finish, while a statistical approach has been utilized to establish connections between environmental context information and crime incidences [25].

III. SYSTEM MODEL

Figure 1 illustrates the application of deep learning in crime prediction. This study utilizes crime summaries gathered from the DCRB (District Crime Records Bureau) of Thoothukudi.

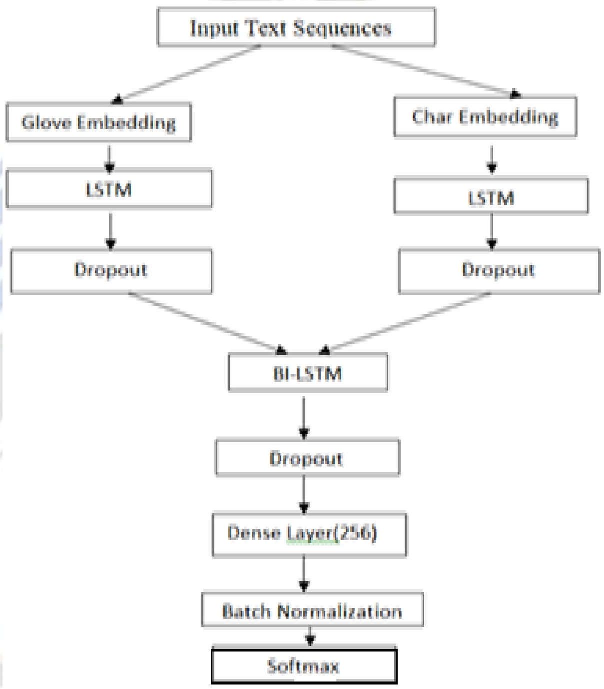

Figure 1. Architecture of the proposed method.

Let's imagine we have n input sequences, each consisting of w words. Let $m = m_1, m_1, \dots, m_n$ represents a set of words related to crime, and $R_n = r_1, r_2, \dots, r_w$ represent another set of words. The objective is to categorize crimes using R_n, and the outcome will be denoted as y_n, which signifies the type of crime that has taken place in Thoothukudi District. Figure 1 illustrates the outlined approach.

A. Preprocessing

The presented approach relies on real-time data obtained from the DCRB (District Crime Records Bureau) of Thoothukudi. This dataset encompasses various attributes, including Crime Number, section of law, date of occurrence,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

location, sub-division, reason, and a summary of the case known as the "gist of the case." Annotations are generated for each case to facilitate efficient processing. The gist of the case is further tokenized at the sentence level, and padding techniques are applied to ensure an equal number of tokens in each sentence. To enhance the efficiency of the proposed method, it is imperative to preprocess the raw dataset.

B. Embedding Layers

Embedding layers are employed to transform text data into a vector format, with the application of Glove and Char embeddings to this particular dataset.

C. GloVe Embeddings

This tool is employed for identifying distinct words and is a publicly accessible pre-trained word embedding method [23]. The package, named "glove.6B.300d.txt," has a size of 822 MB. Glove offers various embedding vector sizes, ranging from 50, 100, 200 to 300 dimensions. Glove operates by creating a co-occurrence matrix for the provided word [24]. The representation of the GloVe model for this co-occurrence matrix can be found in Equation (1).

$$w_i^T \widetilde{w}_j + b_i + b_j \approx \log(1 + X_i) \qquad (1)$$

This method generates two sets of word vectors w_i^T and $\widetilde{w}_j$. Typically, the left and right contexts are distinguished, so X_{ij} is asymmetric. These two-word vectors are different. A single-word vector is obtained as $w'_i = w_i + w'_i$. The co-occurrence matrix is sparse, containing only a few values with large probabilities and having most elements being very close to or equal to zero.

D. Character Embeddings

The character embedding model treats individual characters as distinct units and converts them into a fixed-length dense vector. Utilizing LSTM, the model generates a vector representation for a word based on the embeddings of its constituent characters. This approach enables character embedding models to generate embeddings for uncommon, misspelled, or out-of-vocabulary words (OOV). The challenge of OOV words is mitigated through character-level embedding, which leverages morphological information using an attention mechanism. The attention mechanism estimates the relevance of character positions by creating attention weights. Furthermore, the input text is encoded into a character embedding matrix and associated with these attention weights. Character-level input is derived from the enriched character embedding matrix, followed by the assignment of the corresponding label.

E. LSTM

The LSTM (Long Short-Term Memory) consists of numerous memory cells organized consecutively, along with four neural networks. A standard LSTM unit comprises a cell, an input gate, an output gate, and a forget gate. These three gates control the flow of information both into and out of the cell, allowing the cell to retain values for extended periods. The LSTM approach is particularly effective for classifying, analyzing, and predicting time series data of varying durations. Figure 2 provides a visual representation of the LSTM structure.

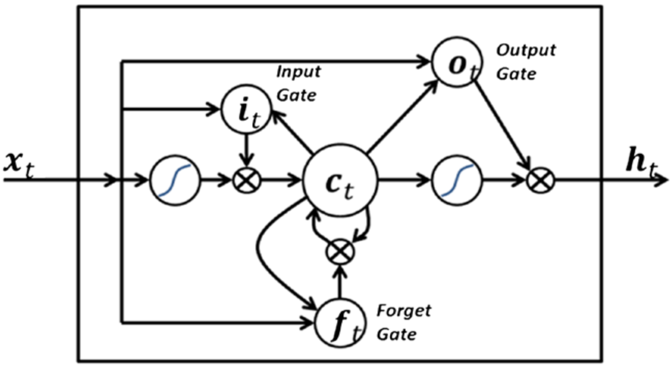

Figure 2. Block Diagram of LSTM.

The sigmoid units in a standard LSTM are represented in Eq. (2).

$$Y = S(\sum_{i=1}^{N} W_i X_i) \qquad (2)$$

The symbol σ denotes the squashing function. The memory input of the LSTM is supplied by the sigmoid unit on the left. The forget gate erases the value it had stored in memory and lets go of it. Meanwhile, the output gate determines whether the value retained in the LSTM's memory should be emitted or not. To ascertain the word's position in the sequence based on the spacing between words, the LSTM layer combines both GloVe and character embeddings. This layer generates an output consisting of 300 features.

F. Dropout

Overfitting poses a significant challenge in neural networks, and the use of large networks can be computationally slow. To combat overfitting, a solution involves combining the predictions of multiple large neural networks during testing. Figure 3 illustrates how dropout serves as a technique to tackle this issue.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

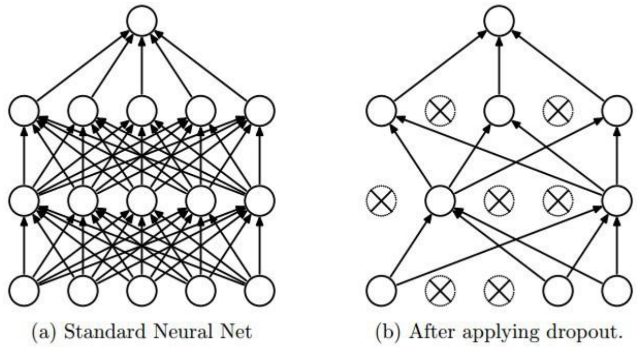

Figure 3. Dropout.

The fundamental idea is to introduce randomness by selectively deactivating units within the neural network during the training process. Dropout involves creating samples from an extensive range of distinct reduced networks throughout training, offering notable advantages over other regularization techniques and significantly mitigating overfitting. In the neural network architecture, the dropout layer follows the LSTM layer, and it essentially entails the random removal of units and connections during the training phase. Dropout samples are derived from an exponentially large pool of these "thinned" networks during training.

G. Bi-LSTM

For modeling word sequences, the bidirectional LSTM (Bi-LSTM) technique is applied. This approach efficiently incorporates both current and prospective attributes. In the initial phase of entity extraction, Bi-LSTM determines the most suitable representation for each word and assigns an entity state tag to each word in the input sequence. The core components of Bi-LSTM consist of the embedding layer, Bi-LSTM layer, and output layer, which collectively form the representation layers. Figure 4 illustrates the fundamental architecture of Bi-LSTM.

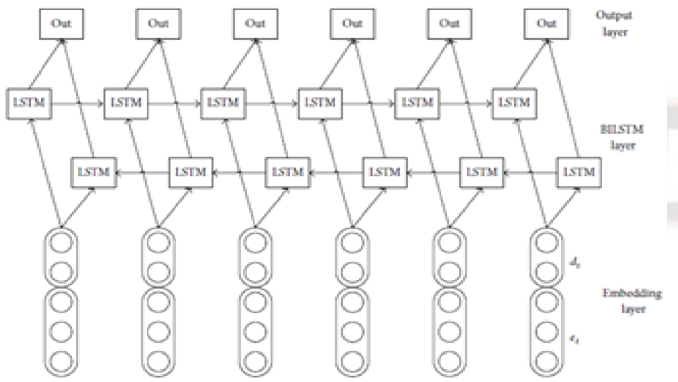

Figure 4. Structure of Bi-LSTM.

The output is fed into the Bi-LSTM model, which comprises 100 features. Bi-LSTM's role is to establish connections between the target words, condensing the initial 1024 features down to 100. To enhance the model's robustness, a Dropout layer is integrated at the end of the LSTM. All layers are interconnected and input into the Bi-LSTM layer. This Bi-LSTM layer identifies the most appropriate representation for each word and affixes an entity state tag to each word within the input sequence, facilitating the initial extraction of entities for crime prediction. Another Dropout layer is appended to the end of the Bi-LSTM to filter the features.

H. Dense Layer

A dense layer is essentially a matrix-vector multiplication operation. For simplicity, let's consider a batch size of 1. In this context, the values within the matrix are the learnable parameters that undergo updates during the backpropagation process, as illustrated in Equation (3).

$$u^T \times W, \qquad W \in R^{nm} \tag{3}$$

The output generated will be a vector with dimensions represented as "anm." The dense layer plays a role in altering the dimensions of each vector. The outcome of the Dropout operation is then supplied as input to the dense layer, which now consists of 256 components.

I. Batch Normalization

Certainly, neural networks, especially deep ones, demand meticulous adjustment of weight initialization and learning parameters. Batch normalization contributes to easing this process to some extent.

J. Softmax

The Softmax function is a different type of activation function employed in neural computing. Its purpose is to calculate a probability distribution from a set of real numbers. In the context of multi-class models, the Softmax function is utilized to provide probabilities for each crime type, with the target class receiving the highest probability.

IV. EXPERIMENTAL RESULTS

This section presents the findings from the data analysis. The dataset comprises information related to various criminal activities such as murder, robbery, kidnapping, and dowry. The initial dataset is split, allocating 80% for training and 20% for testing purposes.

A. Data Source and Selection

The actual dataset is gathered from the DCRB (District Crime Record Bureau) in Thoothukudi. This collected data encompasses factual reports related to various criminal incidents, including decoity, daytime house burglary, nighttime house burglary, murder, robbery, dowry, POSCO, rape, and theft, which are considered key elements. These crime records from the DCRB are prepared for convenient accessibility in further processing. The dataset encompasses all instances of violent crimes that have occurred within Thoothukudi District over the past 5 years, spanning from 2018 to 2022. In total,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

approximately 2556 cases of crimes have been documented, categorized into different divisions (comprising 8 divisions and 56 police stations) within Thoothukudi District. The proposed prediction model utilizes information regarding the nature of each crime. Each offense is associated with multiple attributes, including the crime number, section of the law, date of occurrence, location, sub-division, reason, and a summary of the case known as the "gist of the case." To streamline the operational process, the gist of the case is specifically considered as it provides a concise summary of the crime.

B. Dataset Used

The crime data used in this study were sourced from the DCRB's 'Crimes - 2018 to May 2022' dataset, which comprises real-world information about criminal cases occurring in Thoothukudi District from January 1, 2018, to May 2022, spanning six years. The data encompass all criminal cases within the district's 8 divisions during this period. The crime prediction is rooted in text analysis, and Table 1 provides details about specific crime types and sample keywords used for this prediction.

TABLE I. EXAMPLE KEY WORDS OF GIST

Crime Type	**Key Terms**
Murder	murdered, deceased, assaulted, enmity
Robbery	robbery, gold, booty, two-wheeler
Rape	sexually-assaulted, false-promise, to-marry
POSCO	sexual offence, against-children, minor-girl
Dowry	injuries, torture, demanding, dowries

C. Evaluation Metrics Used

We utilize a set of seven evaluation metrics to gauge the performance of the models. These metrics include the false positive rate, false negative rate, accuracy, precision, recall, specificity, and the F1 score.

A false positive in binary classification signifies an error where the test result erroneously suggests the presence of a condition, such as a specific crime type, when the crime name is not present. On the other hand, a false negative is a converse error, where the test result inaccurately fails to indicate the presence of a condition when it is, in fact, present.

The False Positive Rate (FPR), also referred to as the fall-out, quantifies the probability of a false alarm. It is calculated as the ratio of false positive values to the sum of false positives and true negatives, as demonstrated in Equation (4).

$$FPR = \frac{FP}{FP + TN} \tag{4}$$

The False Negative Rate (FNR), sometimes referred to as the miss rate, indicates the probability of false negatives. It is calculated by dividing the false negative value by the sum of false negatives and true positives, as outlined in Equation (5).

$$FNR = \frac{FN}{FN + TP} \tag{5}$$

Accuracy pertains to the quantity of accurately predicted data points concerning the entire set of data points, as presented in Equation (6). To be more precise, it can be defined as the number of true positives and true negatives divided by the sum of true positives, true negatives, false positives, and false negatives.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Precision, alternatively known as Positive Predictive Value (PPV), is calculated by dividing the count of correct positive results by the count of positive results predicted by the classifier, as outlined in Equation (7).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

Recall, also known as sensitivity or True Positive Rate (TPR), is determined by dividing the true positive values by the total number of actual positive cases. In essence, it represents the count of correct positive results divided by the count of all pertinent samples derived from data science, as depicted in Equation (8).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

The F1 score is the harmonic mean of precision and recall, as illustrated in Equation (9).

$$\text{F1} = 2 \times \frac{Precision \ \times Recall}{Precision + Recall} \tag{9}$$

D. Comparison with baseline models

The proposed model is compared to several baseline models. The Long Short Term Memory (LSTM), which incorporates multiple gates for information flow control, is employed. LSTM is augmented with embedding to focus on relevant words within lengthy sequences. Additionally, Bidirectional LSTM (BiLSTM) is employed to enhance accuracy performance. The proposed model is a combination of Embedding with LSTM and BiLSTM, featuring more layers than the other baseline models. Various dropout layers are also incorporated to prevent overfitting.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

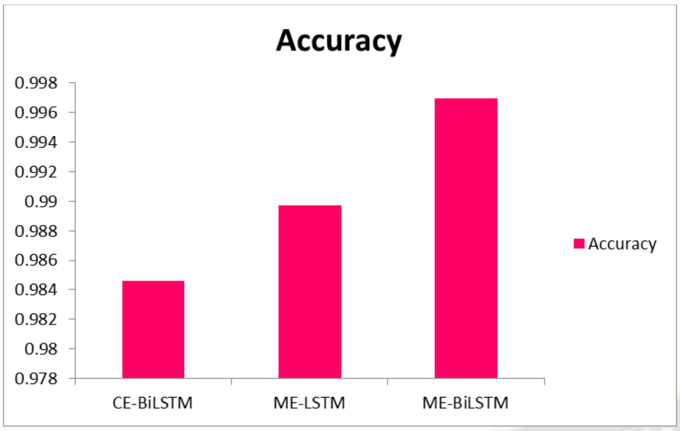

Figure 5. Accuracy of CE-BiLSTM, ME-LSTM and ME-BiLSTM.

TABLE II. COMPARISON BETWEEN CE-BILSTM, ME-LSTM- AND ME-BILSTM

	Accuracy	Precision	Recall	F1Score
CE-BiLSTM	0.9846	0.66514	0.6349	0.6447
ME-BiLSTM	**0.996923**	**0.910766**	**0.918681**	**0.913441**
ME-LSTM	0.99	0.66514	0.712	0.691

In the pursuit of improving safety and security in Thoothukudi City, as well as enhancing crime analytics, another predictive system is under development. This system employs various deep learning algorithms on crime datasets from DCRB's 'Crimes - 2018 to May 2022 present' to enhance the predictive accuracy for crimes involving children and women. When assessing performance on DCRB and Tuty_crime datasets, the multi-embedding with Bi-LSTM approach achieves the highest accuracy compared to the character embedding with Bi-LSTM. The comparative performance of these three models is detailed in Table 2, and Figure 5 illustrates that ME-BiLSTM outperforms ME-LSTM. This evaluation is conducted over 30 epochs.

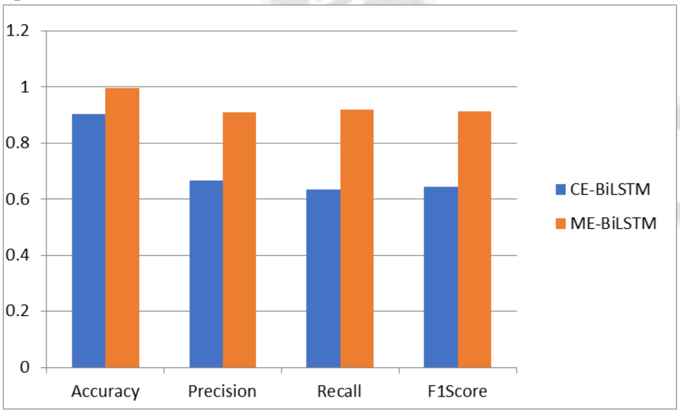

Figure 6. Comparison of CE-BiLSTM and ME-BiLSTM.

The performance of ME-LSTM and ME-BiLSTM is assessed based on metrics such as Accuracy, Precision, Recall, and F1 Score. Figure 6 demonstrates that ME-BiLSTM exhibits a notable improvement in overall accuracy, with a difference of 0.093 when compared to ME-LSTM and CE-BiLSTM. The achieved results are notably higher, with a substantial 0.1228 improvement over ME-LSTM.

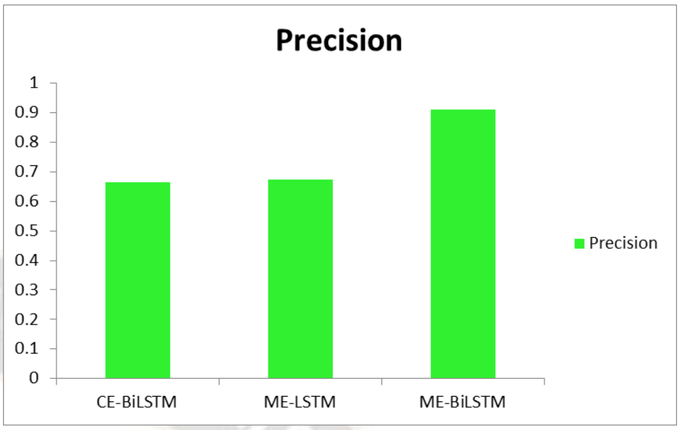

Figure 7. Precision of CE-BiLSTM, ME-LSTM and ME-BiLSTM.

In this case, ME-BiLSTM is trained for 30 epochs. Figure 7 illustrates that the overall Precision exhibits a notable improvement of 0.1746 when compared to ME-LSTM and CE-BiLSTM. The proposed approach has achieved a substantial improvement of 0.1718 over ME-LSTM.

As depicted in Figure 8, ME-BiLSTM is trained for 30 epochs. The overall Recall demonstrates a notable improvement of 0.2078 when compared to ME-LSTM and CE-BiLSTM. The proposed approach has achieved a significant 0.2141 improvement over ME-LSTM.

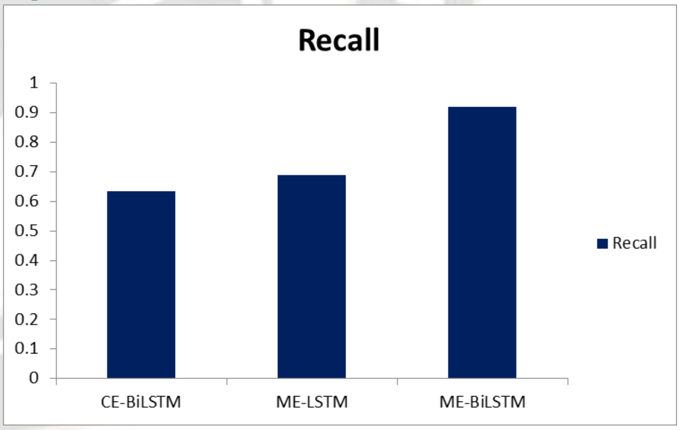

Figure 8. Recall of CE-BiLSTM, ME-LSTM and ME-BiLSTM.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

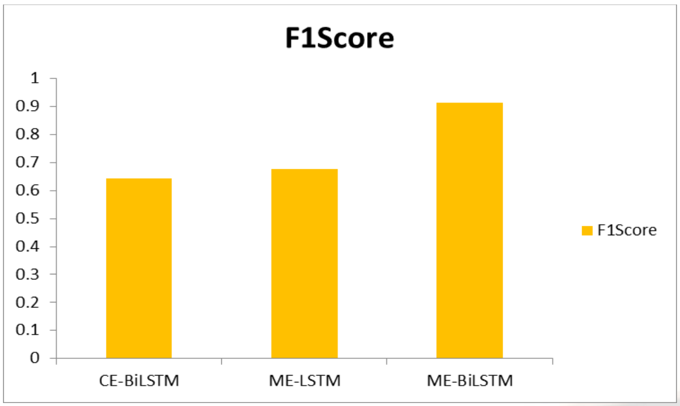

Figure 9. F1 Score of CE-BiLSTM, ME-LSTM and ME-BiLSTM.

ME-BiLSTM was trained for 30 epochs. The overall F1-Score has shown a substantial improvement of 0.2074 when compared to ME-LSTM and CE-BiLSTM. The proposed method has achieved a significant improvement of 0.2209 over ME-LSTM, as depicted in Figure 9.

V. CONCLUSION AND FUTURE WORK

A deep learning framework has been developed for the efficient prediction of crime types. To this end, a novel fully connected Bi-LSTM model with multiple embeddings is introduced. This study evaluates the performance of various embeddings on the DCRB 'Crimes - 2018 to May 2022 present' dataset. The results indicate that the combination of Bi-LSTM with GloVe and Character embeddings yields efficient results. It is evident that bidirectional LSTM outperforms LSTM, and the inclusion of GloVe embeddings further enhances recognition. The overall accuracy has seen a significant improvement of 0.093 compared to other methods.

While the proposed method demonstrates superior crime prediction performance, it is worth noting that the presence of a limited number of neutral sentences has an impact. Therefore, future testing on diverse crime datasets is planned. Additionally, the current crime prediction method does not provide information regarding the income and age of the offender.

REFERENCES

[1] Tahani Almanie, Rsha Mirza, Elizabeth Lor,” Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots", International Journal of Data Mining & Knowledge Management Process (IJDKP), July 2015, Boulder, USA.

[2] B. Cavadas and P. Branco, “Crime Prediction Using Regression and Resources Optimization Crime PredictionUsing Regression and Resources Optimization,” no.September 2015, 2017.

[3] Y. Lin, T. Chen and L. Yu, "Using Machine Learning to Assist Crime Prevention," 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), 2017, pp. 1029-1030, doi: 10.1109/IIAI-AAI.2017.46.

[4] S. Kim, P. Joshi, P. S. Kalsi and P. Taheri, "Crime Analysis Through Machine Learning," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 415-420, doi: 10.1109/IEMCON.2018.8614828.

[5] Soon Ae Chun, Venkata Avinash Paturu, Shengcheng Yuan, Rohit Pathak, Vijay Atluri and Nabil R. Adam. 2019. Crime Prediction Model using Deep Neural Networks. In Proceedings of dg. o 2019: 20th Annual International Conference on Digital Government Research (dg. o 2019), June 18-20, 2019, Dubai, United Arab Emirates. ACM, New York, NY, USA, https://doi.org/10.1145/3325112.3328221.

[6] C. Rajapakshe, S. Balasooriya, H. Dayarathna, N. Ranaweera, N. Walgampaya and N. Pemadasa, "Using CNNs RNNs and Machine Learning Algorithms for Real-time Crime Prediction," 2019 International Conference on Advancements in Computing (ICAC), 2019, pp. 310-316, doi: 10.1109/ICAC49085.2019.9103425.

[7] X. Zhang, L. Liu, L. Xiao, and J. Ji, "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots," in IEEE Access, vol. 8, pp. 181302-181310, 2020, doi: 10.1109/ACCESS.2020.3028420.

[8] X. Han, X. Hu, H. Wu, B. Shen, and J. Wu, "Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN," in IEEE Access, vol. 8, pp. 217222-217230, 2020, doi: 10.1109/ACCESS.2020.3041924.

[9] VO. Llaha, "Crime Analysis and Prediction using Machine Learning," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, pp. 496-501, doi: 10.23919/MIPRO48935.2020.9245120.

[10] A. Shukla, A. Katal, S. Raghuvanshi and S. Sharma, "Criminal Combat: Crime Analysis and Prediction Using Machine Learning," 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-5, doi: 10.1109/CONIT51480.2021.9498397.

[11] Stalidis P, Semertzidis T, Daras P.,” Examining Deep Learning Architectures for Crime Classification and Prediction,” Forecasting 2021, 3, 741–762. https://doi.org/10.3390/forecast3040046.

[12] X. Zhang, L. Liu, L. Xiao, and J. Ji, "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots," in IEEE Access, vol. 8, pp. 181302-181310, 2020, doi: 10.1109/ACCESS.2020.3028420.

[13] W. Safat, S. Asghar and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques," in IEEE Access, vol. 9, pp. 70080-70094, 2021, doi: 10.1109/ACCESS.2021.3078117.

[14] Muthamizharasan, M., Ponnusamy, R. (2022). Forecasting Crime Event Rate with a CNN-LSTM Model. In: Raj, J.S., Kamel, K., Lafata, P. (eds) Innovative Data Communication Technologies and Application. Lecture Notes on Data Engineering and Communications Technologies, vol 96. Springer, Singapore. https://doi.org/10.1007/978-981-16-7167-8_33.

[15] P. Kaur, G. Rani, T. Sharma, and A. Sharma, "A Comparative Study to analyze crime threats using data mining and machine learning approach," 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), 2021, pp. 1-4, doi: 10.1109/ICSCAN53069.2021.9526489.

[16] S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Qasim, and G. R. Sinha, "An Empirical Analysis of Machine Learning

Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach," in IEEE Access, vol. 9, pp. 67488-67500, 2021, doi: 10.1109/ACCESS.2021.3075140.

[17] M. -S. Baek, W. Park, J. Park, K. -H. Jang and Y. -T. Lee, "Smart Policing Technique With Crime Type and Risk Score Prediction Based on Machine Learning for Early Awareness of Risk Situation," in IEEE Access, vol. 9, pp. 131906-131915, 2021, doi: 10.1109/ACCESS.2021.3112682.

[18] Patel H, Prajapati D, Mahida D, Shah M (2020) Transforming petroleum downstream sector through big data: a holistic review. J Pet Explor Prod Technol 10(6):2601–2611. https://doi.org/10.1007/s13202-020-00889-2.

[19] Shah, N., Bhagat, N. & Shah, M. Crime forecasting: a machine learning and computer vision approach to crime prediction and prevention. Vis. Comput. Ind. Biomed. Art 4, 9 (2021). https://doi.org/10.1186/s42492-021-00075-z.

[20] Jha, Pranay, Raman Jha, and Ashok Sharma. "Behavior analysis and crime prediction using big data and machine learning." International Journal of Recent Technology and Engineering 8.1 (2019): 1.

[21] Wang, Bao, et al. "Deep learning for real-time crime forecasting and its internalization." Chinese Annals of Mathematics, Series B 40.6 (2019): 949-966.

[22] Sangeetha and Prabha, "Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM", Journal of Artificial Intelligent Human Computing, 2020.

[23] Shah, Bailey, and Williams, "Natural language processing for disease phenotyping in UK primary care records for research: a pilot study in myocardial infarction and death", Journal of Biomedical Semantics, Vol. 10, 2019. https://www.analyticsvidhya.com/blog/2022/03/an-overview-on-long-short-term-memory-lstm/

[24] Kang H-W, Kang H-B (2017) Prediction of crime occurrence from multi-modal data using deep learning. PLoS ONE 12(4): e0176244. https://doi.org/10.1371/journal.pone.0176244.

[25] Bates, Adam. "Stingray: A new frontier in police surveillance." Cato Institute Policy Analysis 809 (2017).

[26] Chen and Jiang, "Fully-connected LSTM–CRF on medical concept extraction", International Journal of Machine Learning & Cyber, Vol. 11, pp. 1971–1979, 2020.

[27] Muzammil Khan, Azmat Ali, Yasser Alharbi, and Gonzalo Farias. 2022. Predicting and Preventing Crime: A Crime Prediction Model Using San Francisco Crime Data by Classification Techniques. Complex. 2022 (2022). https://doi.org/10.1155/2022/4830411.

[28] Cortes, Ana & Fuentes-Silva, Carlos. (2020). Artificial Intelligence Models for Crime Prediction in Urban Spaces. Machine Learning and Applications: An International Journal. 8. 1-13. 10.5121/mlaij.2021.8101.