Integrating Long Short-Term Memory and Multilayer Perception for an Intelligent Public Affairs Distribution Model

Hong Fang[1], Minjing Peng[1*], Xiaotian Du[2], Baisheng Lin[2], Mingjun Jiang[2], Jieyi Hu[2], Zhenjiang Long[2], Qiaoxian Hu[3]

[1] Department of Economics and Management, Wuyi University, 529020 Jiangmen, China

[2] Department of Project Management Center, Jiangmen City Domain Social Wisdom Governance Technology Innovation Center, 529040 Jiangmen, China

[3] Jiangmen Jianghai District Branch, China Telecom Co. Ltd., 529030 Jiangmen, China

* Correspondence: Minjing Peng (peng.mj@wyu.edu.cn)

Abstract: In the realm of urban public affairs management, the necessity for accurate and intelligent distribution of resources has become increasingly imperative for effective social governance. This study, drawing on crime data from Chicago in 2022, introduces a novel approach to public affairs distribution by employing Long Short-Term Memory (LSTM), Multilayer Perceptron (MLP), and their integration. By extensively preprocessing textual, numerical, boolean, temporal, and geographical data, the proposed models were engineered to discern complex interrelations among multidimensional features, thereby enhancing their capability to classify and predict public affairs events. Comparative analysis reveals that the hybrid LSTM-MLP model exhibits superior prediction accuracy over the individual LSTM or MLP models, evidencing enhanced proficiency in capturing intricate event patterns and trends. The effectiveness of the model was further corroborated through a detailed examination of training and validation accuracies, loss trajectories, and confusion matrices. This study contributes a robust methodology to the field of intelligent public affairs prediction and resource allocation, demonstrating significant practical applicability and potential for widespread implementation.

Keywords: Long Short-Term Memory; Multilayer Perceptron; Public affairs prediction; Intelligent distribution

1 Introduction

In today's era of information explosion, the frequent occurrence and rapid spread of public events pose severe challenges to social stability and public safety. Public events typically include natural disasters (such as earthquakes, floods and hurricanes), public health incidents (such as epidemic outbreaks), social security incidents (such as crime and violence), and sudden public safety incidents (such as fires and explosions) [1]. Among these, the timely and accurate distribution of affairs is crucial for decision-making and emergency response, especially in sudden public events [2].

In recent years, the development of machine learning technology, particularly the widespread application of deep learning, has provided substantial support for acquiring and utilizing public affairs information. Neural Networks (NNs), a form of machine learning based on the functionality of the human brain and its cognitive processing, have become one of the most powerful and accurate clustering technologies available today [3]. To effectively address the complex and diverse data features in public affairs, this study employs the LSTM and MLP models within NN technology.

Relevant studies have demonstrated that LSTM outperforms other statistical and machine learning methods [4]. It is renowned for its superior time-series processing capabilities and effectively captures time dependencies in data. Public events often have significant sequential characteristics. LSTM, through its unique memory units and gating mechanisms, can retain and manage important information flows over long periods, making it particularly adept at handling sequential data. MLP is a classical feedforward NN suitable for handling high-dimensional feature data and capturing complex feature relationships. Public affairs data usually include multidimensional features such

as geographical location, event type, and related numerical indicators. MLP, through its multi-layer structure and nonlinear activation functions, can capture these complex relationships between features. Moreover, hybrid models are increasingly popular and widely applied across various fields [5]. For instance, the Convolutional LSTM (ConvLSTM) model, initially proposed for predicting rainfall intensity [6], has since been successfully applied in various fields such as predicting wind speed and direction [7], epidemic diagnosis [8] and crime prediction [9]. This demonstrates that the hybrid model can mitigate or eliminate the limitations of individual models by combining their advantages, further validating the broad applicability and practicality of hybrid models in addressing complex prediction tasks. Therefore, this study also proposes a hybrid model combining LSTM and MLP to fully utilize the strengths of both models, providing more comprehensive feature extraction and pattern recognition capabilities, thereby further enhancing model predictive performance and generalization capability.

This study details and compares the performance of the LSTM model, the MLP model, and the hybrid model combining LSTM and MLP in the task of public affairs distribution. Through experiments, the prediction accuracy and loss rates of each model were validated, analyzing their effectiveness and superiority in handling complex public affairs distribution tasks. Ultimately, this study aims to provide a novel, practically valuable solution for intelligent public affairs distribution, offering new ideas and technical support for public affairs management and optimization.

The structure of this study is as follows: the first part introduces the research background; the second part reviews related work, summarizing existing intelligent public affairs distribution technologies and methods; the third part describes the design and implementation of Word2Vec, LSTM, MLP, and the hybrid model; the fourth part elaborates on the experimental process; the fifth part conducts a result analysis, comparing the performance of each model; the sixth part summarizes this study; the seventh part discusses the shortcomings and improvements of the model in practical applications, as well as future research directions.

2 Related Work

2.1 Application of Deep Learning in Intelligent Public Affairs Distribution

Currently, an increasing number of scholars are utilizing deep learning to study intelligent public affairs distribution, including Convolutional Neural Networks (CNNs) [10] and Recurrent Neural Networks (RNNs) [11]. These technologies have shown immense potential in various fields.

In the field of natural disasters, deep learning is applied to predict and detect them in real time, aiding government departments in disaster management [12]. In public health, artificial intelligence (AI), machine learning, and deep learning technologies are widely used in epidemic prevention and control, enhancing the capability to respond to sudden health incidents [13, 14]. Additionally, in the realm of social security, integrated deep learning methods are used to enhance cybersecurity [15]. Deep learning models based on Swin Transformer are utilized to identify and analyze crowd behavior [16], ensuring social safety and stability. Lastly, in sudden public safety incidents, deep learning demonstrates its unique research value. Whether for traffic accidents [17], fires [18], or other emergencies, deep learning can rapidly and accurately analyze large amounts of data. It can conduct real-time monitoring and prediction, significantly improving the efficiency and effectiveness of emergency response.

2.2 Application of LSTM and MLP in Public Affairs

In public affairs management, the LSTM model, with its excellent sequential processing capabilities, can accurately predict the occurrence of public events, achieving timely warning through intelligent distribution [19, 20]. In disaster and emergency response, the combined application of CNN and Bidirectional LSTM (BiLSTM) can precisely identify key information from Twitter posts. This provides a critical base for emergency decision-making and assists relevant personnel in the reasonable allocation and dispatch of resources [21]. Furthermore, the LSTM model can effectively predict public sentiment changes during sudden public events. It improves prediction accuracy through the adaptive moment estimation (Adam) optimization algorithm, thereby guiding online public opinion and maintaining social stability [22]. The application of the LSTM model not only enhances the efficiency of public affairs management but also strengthens the capability to respond to sudden events.

The MLP model also demonstrates significant application value. For instance, by analyzing accident and weather data, the MLP model can accurately predict the severity of accidents, constructing automatic emergency braking systems to reduce traffic accidents [23]. Additionally, the MLP model shows high accuracy in predicting different types of crimes (such as theft and assault) [24]. In the field of epidemic prevention and control, the COVID-19 diagnosis model developed based on MLP-NN can effectively distinguish whether patients are infected with COVID-19, improving diagnostic accuracy and efficiency [25].

Meanwhile, the hybrid model combining LSTM and MLP is also applied in medical [26], transportation [27] and cybersecurity fields [28], but is still absent in the crime domain. Therefore, this study introduces the LSTM and MLP hybrid model into the crime prediction field. This study also explores and compares the performance of the LSTM model, MLP model, and LSTM+MLP hybrid model in crime prediction tasks, aiming to further improve and optimize the intelligent public affairs distribution system.

3 Methodology

To achieve intelligent prediction and resource allocation of crime events, this study employs Word2Vec to transform textual data into vector representations and designs three different models: the LSTM model, the MLP model, and the LSTM+MLP hybrid model. These models predict crime data by capturing data features and patterns through different structures. Below is the detailed design of these components.

3.1 Word2Vec

Current researchers have extensively studied Word2Vec and its applications in various fields, demonstrating its effectiveness in capturing semantic relationships between words. Recent studies have shown that Word2Vec can significantly improve the performance of natural language processing tasks by providing meaningful vector representations of words [29, 30]. Given its proven capability, Word2Vec has been widely adopted for tasks requiring detailed text analysis.

To effectively handle and represent the textual data in crime descriptions, Word2Vec was used in this study to convert words into dense vector representations. Word2Vec is a model used to map words into vector space, including two training strategies: Skip-gram and Continuous Bag of Words (CBOW). Skip-gram aims to predict the context words based on a given word, while CBOW predicts the current word based on its context words. Skip-gram is particularly useful for capturing detailed relationships between words, especially in situations where the context is sparse or varied. The Skip-gram model was adopted in this study because it is better suited for capturing the precise semantic relationships in the crime description texts, which often contain important nuances and context-specific information that need to be understood at a granular level.

In the Skip-gram model, given a word w_t, the model attempts to maximize the probability of the context words $w_{t-k}, ..., w_{t+k}$, where k is the window size. The Skip-gram model is trained through the following objective function:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-k\leq j\leq k,\, j\neq 0}\log P\left(w_{t+j}\mid w_t\right) \tag{1}$$

where, T represents the total number of words in the corpus, and $P\left(w_{t+j}\right)\mid w_t$ is the probability of word w_{t+j} occurring given the word w_t. By maximizing this objective function, the Word2Vec model can learn the vector representation of each word.

3.2 Architecture of the LSTM Model

LSTM is a special type of RNN proposed by Hochreiter and Schmidhuber [31] in 1997 to address the issues of gradient vanishing and explosion in traditional RNNs when processing long sequences. Recently, LSTM has achieved remarkable success in various time series data analysis fields, including speech recognition, natural language processing, financial market prediction and medical data analysis.
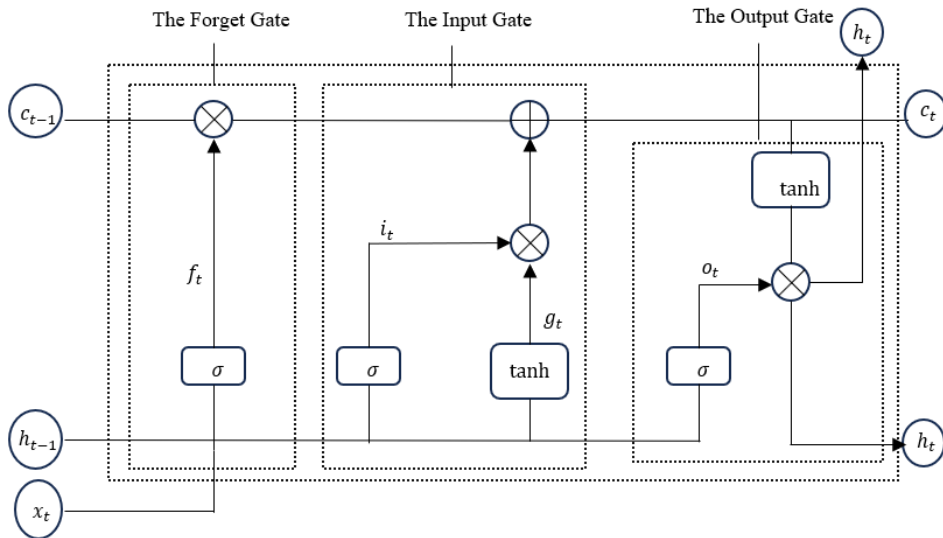

Figure 1. LSTM principle

The core components of the LSTM network include three gates: the forget gate, the input gate and the output gate. These gates control the forgetting, inflow, and output of information, respectively, thereby dynamically adjusting the state in the memory unit. The principle of LSTM is shown in Figure 1.

The function of the forget gate is to decide which information to forget from the memory unit. By using the sigmoid activation function, the output is a value between 0 and 1 , where 1 means "completely keep" and 0 means "completely discard." The forget gate f_t, weights W_f and bias b_f are used, with the hidden state h_{t-1} and current input x_t.

$$f_t = \sigma \left(W_f \cdot [h_{t-1}, x_t] + b_f \right) \tag{2}$$

The input gate determines which new information will be stored in the memory unit. It consists of two parts: the sigmoid activation function, which decides the update portion, and the tanh activation function, which generates candidate values. The input gate i_t and candidate memory unit $\widetilde{C}_t$ are calculated as follows:

$$i_t = \sigma \left(W_i \cdot [h_{t-1}, x_t] + b_i \right) \tag{3}$$

$$\widetilde{C}_t = \tanh \left(W_C \cdot [h_{t-1}, x_t] + b_c \right) \tag{4}$$

The memory unit, which is the core of LSTM, can retain information over long periods in a time series. It interacts through the forget gate and input gate, allowing it to selectively remember or forget information. The next cell state C_t is given as follows:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{5}$$

The output gate decides the next hidden state (i.e., the output at the next time step). Firstly, the output gate uses the sigmoid activation function to determine which parts of the memory unit will be output. Then, this value is multiplied by the tanh-activated memory unit value to get the final output. The output gate o_t and next hidden state h_t are given as follows:

$$o_t = \sigma \left(W_o \left[h_{t-1}, x_t \right] + b_o \right) \tag{6}$$

$$h_t = o_t * \tanh \left(C_t \right) \tag{7}$$

Through these well-designed gates and memory units, LSTM achieves precise control of information, enabling it to capture complex dependencies and long-term dependencies in sequences, significantly outperforming traditional RNNs. Therefore, this study uses the LSTM model to predict crime events, with the specific structure shown in Figure 2.
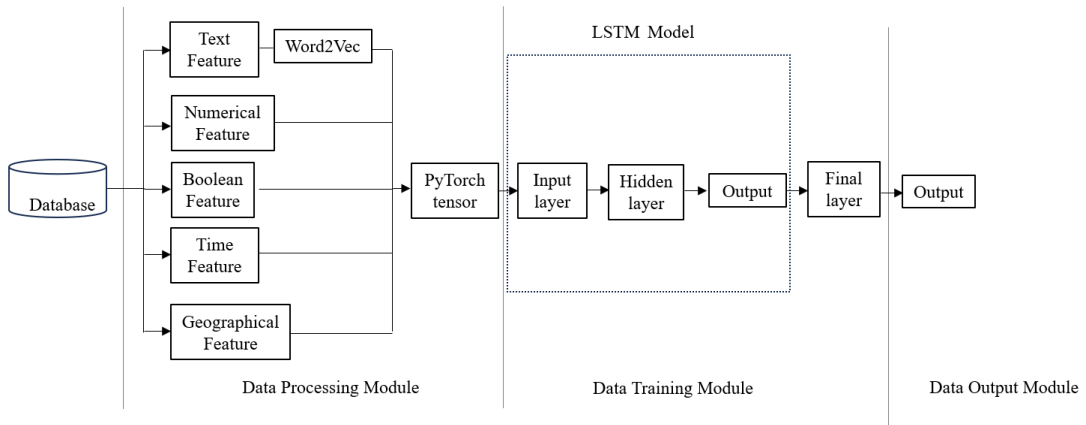

Figure 2. Structure of the LSTM model

The preprocessing involves converting textual data into dense vector representations using Word2Vec and integrating numerical, boolean, temporal, and geographical features into a PyTorch tensor. The processed data is fed into the LSTM model's input layer. The LSTM layer captures sequential dependencies and time-series features

through its memory cells and gates, effectively managing the flow of information. The hidden layer further refines these features, and the final output of the LSTM layer aggregates all sequential information. This output is passed through a final layer, mapping the aggregated information to the target prediction, thus forecasting crime events. This structure enables the LSTM to utilize diverse features and capture complex temporal patterns, making it suitable for intelligent crime prediction.
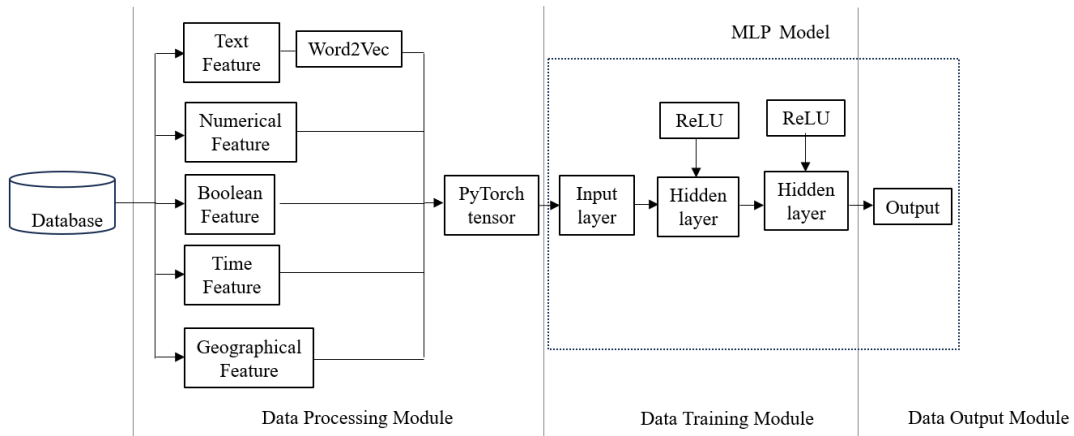
3.3 Architecture of the MLP Model

MLP is widely used in classification and regression tasks [32]. It consists of an input layer, multiple hidden layers, and an output layer, where each node in a layer is connected to all nodes in the previous layer, forming a fully connected layer. By applying nonlinear activation functions, MLP can capture complex feature relationships, making it suitable for handling high-dimensional static feature data. The principle of MLP is shown in Figure 3.
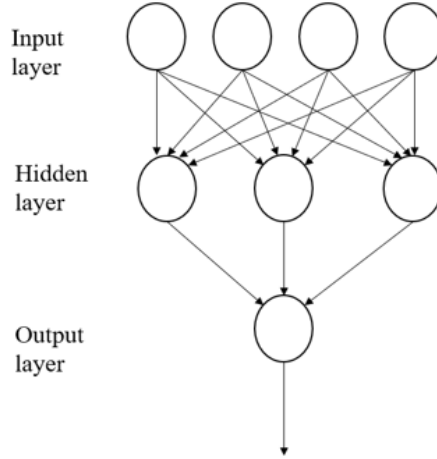
Figure 3. MLP principle

Assuming the input layer is represented by vector X, the output bf the hidden layer is $f\left(W_1 X + b_1\right)$, where W_1 is the weight (also called connection coefficient), b_1 is the bias, and the function f can be a commonly used sigmoid function or tanh function. In this study, the ReLU activation function was chosen because it has a larger gradient compared to the sigmoid and tanh functions, which helps to alleviate the gradient vanishing problem. Its formula is $f(x) = \max(0, x)$. Finally, the connection from the hidden layer to the output layer can be viewed as a multi-class logistic regression, also known as softmax regression. Therefore, the output layer applies the softmax function: softmax $\left(W_2 X_1 + b_2\right)$, where W_2 is the weight of the output layer, X_1 represents the output of the hidden layer, and b_2 is the bias of the output layer.

Figure 4. Structure of the MLP model

Based on the above principles, the specific structure of the MLP model used to predict crime events is shown in Figure 4. As with the LSTM model structure, the preprocessing step involves converting textual data into dense

vector representations using Word2Vec and integrating numerical, boolean, temporal, and geographical features into a unified PyTorch tensor. The processed data is then fed into the MLP model's input layer. The MLP model consists of an input layer, two hidden layers with ReLU activation functions, and an output layer. The input layer receives the combined features and passes them through the first hidden layer, where ReLU activation introduces non-linearity and helps capture complex patterns in the data. The data then flows into the second hidden layer, which also employs ReLU activation to further refine the learned features. The output from the second hidden layer is finally passed to the output layer, which generates the predictions for the crime events.

This structure allows the MLP model to effectively utilize the processed features, leveraging its multiple layers and ReLU activations to capture intricate relationships within the data, making it a robust tool for intelligent crime prediction.

3.4 Architecture of the LSTM+MLP Hybrid Model

The LSTM+MLP hybrid model combines the advantages of LSTM and MLP to handle two different types of input data: text data and other feature data. The model structure is shown in Figure 5.

Research has shown that utilizing Word2Vec and LSTM to process text features can significantly enhance the effectiveness of text processing tasks [33]. Consequently, this study employs LSTM to handle text data. The output of the last time step of the LSTM was extracted as a high-level feature representation of the text data, ensuring the comprehensiveness and temporal dependency of the information. Simultaneously, the MLP was used to process other feature data. The MLP consists of two hidden layers, and employs the ReLU activation function to enhance nonlinearity. The design of these two components remains consistent with their individual model structures, simplifying the process of hyperparameter tuning and enhancing the interpretability and comparability of the model. Finally, the feature representations from the LSTM and MLP were concatenated to form a comprehensive feature vector. This vector was then mapped to the output through a fully connected layer to achieve the final prediction.

By integrating the strengths of both LSTM and MLP, this hybrid model is able to effectively leverage sequential dependencies in the text data and complex patterns in the other feature data, resulting in a robust and accurate predictive model for intelligent crime prediction.
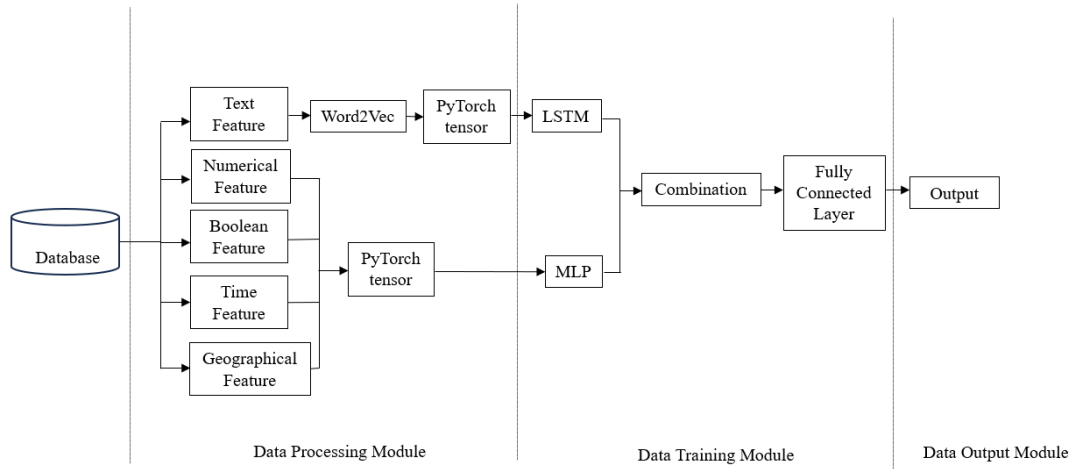
Figure 5. Structure of the LSTM+MLP model

4 Experiment

4.1 Data Processing

The data used in this study was sourced from the publicly available crime records database of the city of Chicago, focusing on crime records from 2022. The dataset contained 219,042 crime records, covering information such as the time, location, and type of crime. The specific field descriptions are provided in Table 1. To ensure data quality and meet model requirements, operations such as deduplication, missing value imputation, and standardization were performed on the data.

(a) Duplicate records were identified and removed to ensure each record is unique.

(b) Columns with many missing values were deleted. Columns with a few missing values were handled using interpolation.

(c) Numerical features (IUCR, Beat, District, Community_Area) were standardized to eliminate the influence of different feature value ranges on model training.

(d) Boolean features (Arrest, Domestic) were converted to integer values 0 and 1 to facilitate model processing and calculation.

(e) In this study, the label field is FBI_Code, representing different types of crimes. According to the conversion rules provided in Table 2, the 26 crime types in FBI_Code were mapped to integers 0 to 25. This conversion process not only simplifies the data format but also assists subsequent machine learning models in recognizing and handling labels. Additionally, Table 2 provides a clear understanding of the distribution of various crime types, which is useful for subsequent data analysis and model tuning.

Table 1. Data field description

Attribute	Description
ID	Unique identifier of the crime event
Case_Number	Case number
Date	Date and time the crime occurred
Block	Block where the crime occurred
IUCR	Crime report code
Primary_Type	The primary category of the crime
Description	A detailed description of the crime
Location_Description	Description of the crime location
Arrest	Whether there was an arrest
Domestic	Whether it was a domestic crime
Beat	Police patrol area number
District	Police district number
Ward	City council ward number
Community_Area	Community area number
FBI_Code	FBI-defined crime type
Latitude	Latitude of the crime location
Longitude	Longitude of the crime location
Location	Combined geographical location description

Table 2. FBI_Code data conversion rule

Original Data	Converted Data	Sample Size
6	0	50077
27	1	16476
17	2	1532
11	3	11662
28	4	31851
20	5	636
14	6	25237
5	7	7024
26	8	14423
30	9	7114
7	10	19269
2	11	1625
12	12	13
15	14	8461
22	15	195
29	16	6848
3	17	8299
9	18	393
110	19	670
18	20	3711
10	21	2143
24	22	1027
13	23	66
16	24	278
19	25	10
111	19	2

4.2 Feature Extraction

In this study, various features were extracted from the crime data to provide comprehensive information for model training. These features include time features, geographical features, and text features.

4.2.1 Time features

Temporal features can reflect the time distribution patterns of criminal behavior. The recorded crime timestamp (Date) was converted to date-time format, and four key temporal features were extracted: year, month, day, and hour. Finally, all extracted temporal features were standardized to ensure model validity.

4.2.2 Geographical features

Geographical features help identify crime hotspots and provide geographical information. This study extracted longitude and latitude information (Longitude, Latitude), community area number (Community_Area), police district number (District), and other relevant information to help the model recognize crime patterns and trends in specific areas. Additionally, to ensure consistent feature weights during model training, location features were standardized.

4.2.3 Text features

Text features provide information on crime types and specific details. This study extracted the primary category (Primary_Type) and detailed description (Description) from the dataset and used the Word2Vec model to convert them into vector representations, capturing semantic relationships between text features. During training, the vector dimension was set to 100, the window size to 5, the minimum word frequency to 1, and the number of parallel processing threads to 4, ensuring effective capture of semantic information in text features. After training, each text feature was converted into a vector representation, and these vectors were combined into the original feature set, generating text feature representations rich in semantic information.

4.3 Data Partitioning

The dataset was divided into a training set and a validation set, with 80% of the data allocated for training and the remaining 20% for validation. This partitioning ensures that the model can learn from the training data while using the validation data for performance evaluation and hyperparameter tuning, thereby verifying the model's generalization ability. The training set was primarily utilized for learning model parameters, whereas the validation set was used to evaluate the model's performance on unseen data. The results from the validation set were then employed to adjust the model's hyperparameters to optimize its performance.

4.4 Hyperparameter Tuning

This study selected appropriate hyperparameters for training the LSTM model, MLP model, and LSTM+MLP hybrid model to optimize their performance for intelligent crime prediction.

For the LSTM model, the hidden layer dimension was set to 64 units, balancing model complexity and computational efficiency, and capturing temporal dependencies in the data. The MLP model employs two hidden layers, each with 64 neurons, using ReLU activation functions to introduce non-linearity and capture complex patterns. The LSTM+MLP hybrid model has an LSTM component with an input dimension of 100 and a hidden layer dimension of 64 units, capturing temporal dependencies. The MLP component consists of two hidden layers, each with 64 neurons and ReLU activation functions, to process other features.

To ensure uniformity and facilitate performance comparison, the models were configured with consistent hyperparameters. The cross-entropy loss function was used as the loss function, a common loss function in classification problems, which effectively measures the difference between the model's predicted values and the true labels. The Adam optimizer was selected as the optimizer, which combines the advantages of momentum and adaptive learning rate, enabling quick convergence and avoiding local optima. The learning rate was set to 0.001, and the number of training epochs was set to 50, ensuring the model was sufficiently trained while avoiding overfitting and improving generalization ability.

By standardizing these hyperparameters, the study aims to create a controlled environment for evaluating the effectiveness of each model in leveraging the processed features for accurate crime prediction. This approach ensures that any differences in model performance are attributable to the model architectures themselves, rather than variations in hyperparameter settings.

4.5 Training Steps

During model training, lists for storing training and validation accuracy and loss were initialized, and the number of training epochs was set. In each training epoch, the model enters training mode, performing forward propagation and loss calculation by iterating through the training dataset, followed by backpropagation to update model parameters. Additionally, the loss and correct prediction count for each batch were accumulated to calculate the overall training loss and accuracy. Next, the model switches to evaluation mode, iterating through the validation dataset without

gradient calculation to perform forward propagation and loss calculation, similarly accumulating the loss and correct prediction count for each batch to calculate the average validation loss and accuracy. At the end of each training epoch, the training and validation loss and accuracy were recorded and output to monitor the model's training process and performance.

5 Experimental Results and Analysis

### 5.1 Analysis of Accuracy and Loss

The performance of the LSTM model, MLP model and LSTM+MLP hybrid model in crime prediction tasks in this experiment was compared. The analysis of accuracy and loss rates on the training and validation sets for each model is presented below.

Table 3. Accuracy and loss rates

Model	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
LSTM	0.9668	0.9647	0.0788	0.0870
MLP	0.9671	0.9666	0.0773	0.0851
LSTM + MLP	0.9719	0.9714	0.0584	0.0608
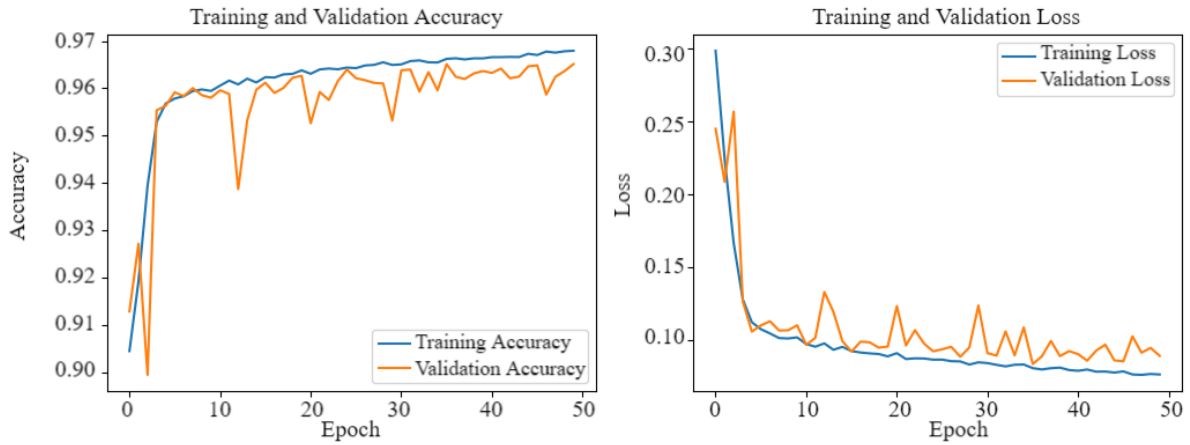
Figure 6. Training and validation accuracy and loss (LSTM)
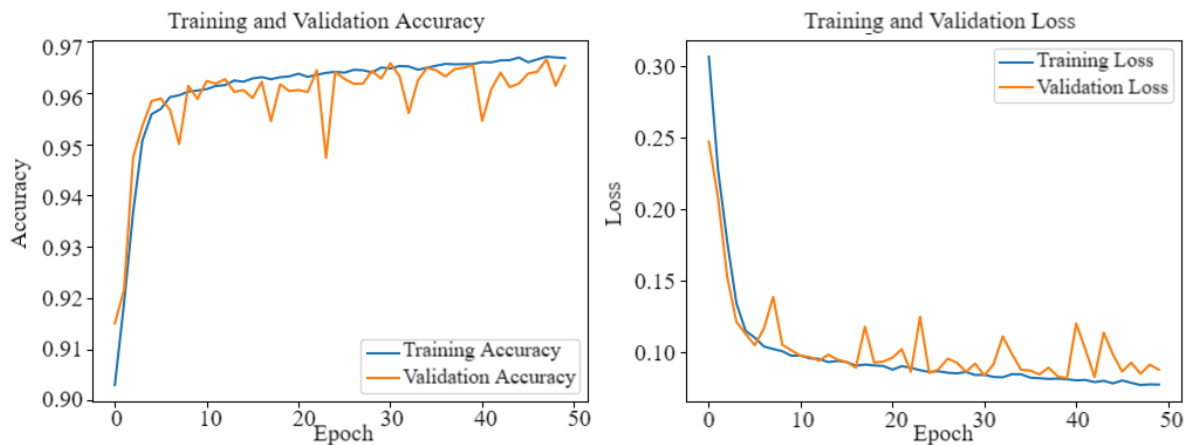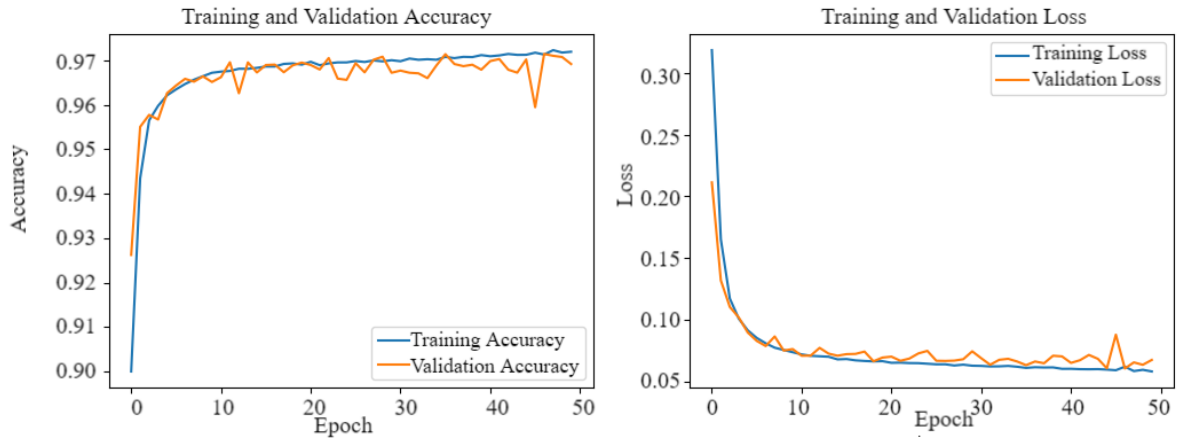

Figure 7. Training and validation accuracy and loss (MLP)

The LSTM model performs well on both the training and validation sets. As shown in Table 3, the highest training accuracy reaches 0.9668, with the lowest training loss of 0.0788. The highest validation accuracy is 0.9647, and the lowest validation loss is 0.0870. Figure 6 shows that the training accuracy and loss curves of the LSTM

model stabilize after the initial few epochs. However, the validation accuracy curve shows some fluctuations, and the validation loss curve fluctuates slightly more than the training loss, indicating slight overfitting.

The MLP model also performs well, with the highest training accuracy reaching 0.9671 and the lowest training loss being 0.0773. The highest validation accuracy is 0.9666, and the lowest validation loss is 0.0851. As shown in Figure 7, the training accuracy curve of the MLP model is similar to that of the LSTM model, stabilizing after a rapid increase. The validation accuracy and loss curves of the MLP model fluctuate slightly less than those of the LSTM model, indicating that the MLP model's simpler structure leads to more stable performance on the validation set.

The LSTM+MLP hybrid model performs best on both the training and validation sets. Its highest training accuracy reaches 0.9719, with the lowest training loss being 0.0584. The highest validation accuracy is 0.9714, and the lowest validation loss is 0.0608. As shown in Figure 8, the hybrid model achieves the highest accuracy and the lowest loss during training and validation, with very stable training and validation loss curves, indicating that the model not only has strong learning ability but also good generalization ability, effectively avoiding overfitting. The hybrid model combines the ability of LSTM to capture text data features and the ability of MLP to process structured data, making full use of data features and performing best in crime prediction tasks.

Figure 8. Training and validation accuracy and loss (LSTM+MLP)

Overall, the LSTM+MLP hybrid model performs best in this experiment, demonstrating its effectiveness and stability in handling complex data features. This model's comprehensive performance makes it suitable for intelligent public affairs distribution tasks requiring high accuracy and good generalization ability.

5.2 Confusion Matrix Analysis

In this study, the impact of different models on the task of crime type distribution was thoroughly compared by plotting confusion matrices for the top 10 most frequent crime types, as shown in Figure 9, Figure 10, and Figure 11. The results show that all three models exhibit fairly high classification accuracy overall but demonstrate certain differences in performance when dealing with specific categories.

Notably, the classification performance for categories 8 and 1 is less satisfactory across all models, exhibiting relatively lower accuracy compared to other categories. Specifically, the hybrid model, despite leveraging the strengths of LSTM and MLP, achieves only 0.84 and 0.85 in accuracy in categories 8 and 1, respectively. In comparison, the LSTM model performs slightly better in category 8, reaching an accuracy of 0.91, but drops to 0.81 in category 1. The MLP model achieves 0.85 and 0.82 in accuracy in these two categories, respectively. These results indicate that the data features of these two crime types may be more complex and overlapping, challenging the models' discriminative abilities.

Consequently, further exploration of the intrinsic characteristics of these two crime types is necessary to develop more effective feature representation methods. Innovations and optimizations in model architecture, learning algorithms, and data preprocessing are required to enhance the models' discriminative power and generalization performance in complex, overlapping feature spaces.

5.3 Summary of Experimental Results

In this experiment, a comparative analysis of the LSTM model, MLP model, and LSTM+MLP hybrid model reveals that the LSTM+MLP hybrid model performs best in handling complex time-series data and nonlinear mappings. It demonstrates higher training and validation accuracy, coupled with lower loss rates, indicating its overall superiority over single models. Additionally, the confusion matrix analysis shows that while all three models achieve high

classification accuracy across most categories, they exhibit lower accuracy in certain complex categories. This suggests that the data features of these categories may be more intricate or overlapping, thereby affecting the models' discriminative abilities. Overall, the LSTM+MLP hybrid model exhibits better performance, yet further optimization is necessary to enhance classification capabilities in specific complex categories. This could involve refining model architectures, improving learning algorithms, and developing advanced data preprocessing techniques to better address the challenges presented by complex, overlapping feature spaces.
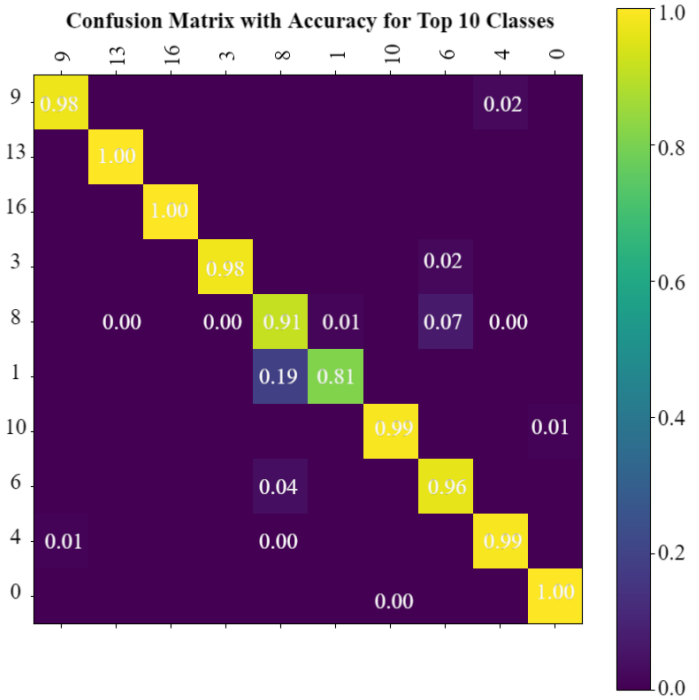

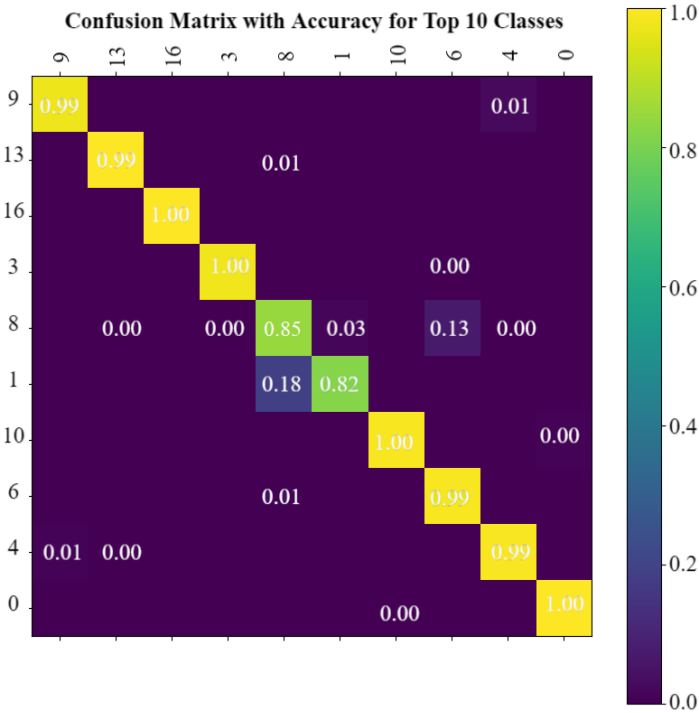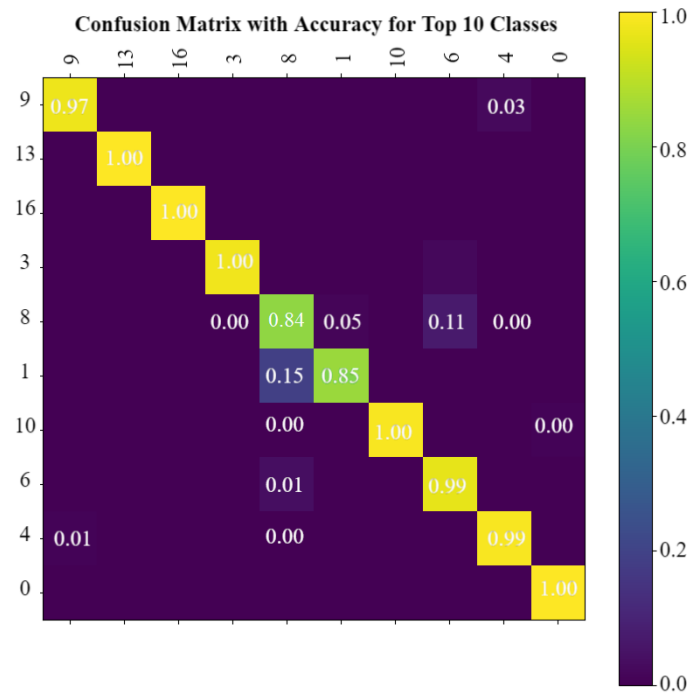**Figure 9.** Confusion matrix with accuracy for the top 10 classes (LSTM)

Figure 10. Confusion matrix with accuracy for the top 10 classes (MLP)

Figure 11. Confusion matrix with accuracy for the top 10 classes (LSTM+MLP)

6 Conclusion

This study aims to enhance the precision of intelligent public affairs distribution. Utilizing crime data from Chicago in 2022, three different machine learning models were proposed and compared in this study, i.e., the LSTM model, the MLP model, and a hybrid model combining both. Through comprehensive data preprocessing, the models effectively capture the complex relationships among textual, numerical, boolean, temporal, and geographical features, ultimately achieving efficient classification and prediction of public affairs events. Experimental results indicate that the LSTM+MLP hybrid model outperforms the individual models in terms of accuracy and loss rates, exhibiting greater stability during both training and validation phases. This suggests that the hybrid model has a significant advantage in capturing complex patterns and trends in public affairs. However, the confusion matrix analysis reveals that, despite overall good performance, the prediction accuracy for certain complex categories (e.g., categories 8 and 1) remains low, suggesting that the model needs improvement in handling specific categories.

7 Strategies for Enhancement and Future Directions

To address the challenges in practical applications, the primary issue of data imbalance must be resolved. This can be effectively managed through the implementation of oversampling techniques, employing Generative Adversarial Networks (GANs) to generate samples for minority classes, and adjusting class weights during model training to balance class distribution. It is also paramount to optimize feature engineering. Integrating contextual features and developing sophisticated feature combinations can significantly enhance the model's ability to capture and analyze critical information. Moreover, in terms of model architecture, adding more hidden layers and incorporating attention mechanisms can improve the overall predictive performance and flexibility of the model. During the model deployment phase, it is essential to establish robust performance monitoring mechanisms to ensure stability and accuracy in real-world applications. Iterative optimizations based on continuous feedback are necessary for maintaining and enhancing the model's performance. Furthermore, it is important to adhere to ethical and legal requirements, thereby ensuring privacy protection during data processing and enhancing the model's fairness to avoid bias against different groups.

Future research should focus on further optimizing the model and integrating more data sources. The model's performance in practical applications can be improved by introducing more advanced deep learning technologies and algorithms. Additionally, extensive testing on different types of public affairs datasets, such as natural disaster data and public health incident data, is necessary to verify the model's generalization ability and robustness. Continuous refinement of data preprocessing methods and model architectures, tailored to specific business needs, will enhance the model's practicality and reliability. This approach will ultimately lead to efficient management and precise prediction in intelligent public affairs distribution systems. Through these efforts, intelligent public affairs distribution

systems will be able to more accurately predict and allocate resources, improving social governance efficiency and achieving the goal of intelligent management.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] R. Williams, V. Kemp, K. Porter, T. Healing, and J. Drury, *Major Incidents, Pandemics and Mental Health: The Psychosocial Aspects of Health Emergencies, Incidents, Disasters and Disease Outbreaks*. Cambridge University Press, 2024.

[2] J. C. Pine, *Technology and Emergency Management*. John Wiley & Sons, 2017.

[3] J. Mena, *Machine Learning Forensics for Law Enforcement, Security, and Intelligence*. CRC Press, 2011.

[4] F. Zhang and Y. Xia, "Carbon price prediction models based on online news information analytics," *Financ. Res. Lett.*, vol. 46, p. 102809, 2022. https://doi.org/10.1016/j.frl.2022.102809

[5] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, "Hybrid approaches to optimization and machine learning methods: A systematic literature review," *Mach. Learn.*, vol. 113, pp. 4055–4097, 2024. https://doi.org/10.1007/s10994-023-06467-x

[6] X. J. Shi, Z. R. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *arXiv:1506.04214*, 2015. https://doi.org/10.48550/arXiv.1506.04214

[7] P. Sari, T. Yasuno, D. A. Prasetya, and M. M. Al Haromainy, "Forecasting system of wind speed and direction by neural network," in *2023 IEEE 9th Information Technology International Seminar*, Batu Malang, Indonesia, 2023, pp. 1–5. https://doi.org/10.1109/ITIS59651.2023.10420216

[8] Y. H. He, X. W. Zheng, and Q. Miao, "TFA-CLSTMNN: Novel convolutional network for sound-based diagnosis of COVID-19," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 21, no. 3, p. 2250058, 2023. https://doi.org/10.1142/S0219691322500588

[9] J. T. K, G. J, and P. S, "A survey on prediction of risk related to theft activities in municipal areas using deep learning," in *2023 Second International Conference on Electronics and Renewable Systems*, Tuticorin, India, 2023, pp. 1321–1326. https://doi.org/10.1109/ICEARS56392.2023.10085123

[10] H. L. Lei, H. Wang, L. L. Wang, Y. H. Dong, J. J. Cheng, and K. Cai, "An analysis of the evolution of online public opinion on public health emergencies by combining CNN-BiLSTM+attention and LDA," *J. Comput. Commun.*, vol. 11, no. 4, pp. 190–199, 2023. https://doi.org/10.4236/jcc.2023.114009

[11] F. Kadri and K. Abdennbi, "RNN-based deep-learning approach to forecasting hospital system demands: Application to an emergency department," *Int. J. Data Sci.*, vol. 5, no. 1, pp. 1–25, 2020. https://doi.org/10.1504/IJDS.2020.10031621

[12] V. Nunavath and M. Goodwin, "The use of artificial intelligence in disaster management-Asystematic literature review," in *2019 International Conference on Information and Communication Technologies for Disaster Management*, Paris, France, 2019, pp. 1–8. https://doi.org/10.1109/ICT-DM47966.2019.9032935

[13] J. Budd, B. S. Miller, E. M. Manning *et al.*, "Digital technologies in the public-health response to COVID-19," *Nat. Med.*, vol. 26, no. 8, pp. 1183–1192, 2020. https://doi.org/10.1038/s41591-020-1011-4

[14] C. Wen, W. Liu, Z. H. He, and C. Y. Liu, "Research on emergency management of global public health emergencies driven by digital technology: A bibliometric analysis," *Front. Public Health*, vol. 10, p. 1100401, 2023. https://doi.org/10.3389/fpubh.2022.1100401

[15] K. Yamarthy and C. Koteswararao, "MDepthNet based phishing attack detection using integrated deep learning methodologies for cyber security enhancement," *Cluster Comput.*, pp. 1–19, 2024. https://doi.org/10.1007/s10586-024-04313-w

[16] M. Qaraqe, A. Elzein, E. Basaran *et al.*, "PublicVision: A secure smart surveillance system for crowd behavior recognition," *IEEE Access*, vol. 12, pp. 26 474–26 491, 2024. https://doi.org/10.1109/ACCESS.2024.3366693

[17] V. Adewopo and N. Elsayed, "Smart city transportation: Deep learning ensemble approach for traffic accident detection," *arXiv:2310.10038*, 2023. https://doi.org/10.48550/arXiv.2310.10038

[18] K. Avazov, M. Mukhiddinov, F. Makhmudov, and Y. I. Cho, "Fire detection method in smart city environments using a deep-learning-based approach," *Electronics*, vol. 11, no. 1, p. 73, 2021. https://doi.org/10.3390/electronics11010073

[19] S. Nevo, E. Morin, A. G. Rosenthal *et al.*, "Flood forecasting with machine learning models in an operational framework," *Hydrol. Earth Syst. Sci.*, vol. 26, pp. 4013–4032, 2022. https://doi.org/10.5194/hess-26-4013-2022

[20] C. Y. Wang, T. C. Huang, and Y. M. Wu, "Using LSTM neural networks for onsite earthquake early warning," *Seismol. Soc. America*, vol. 93, no. 2A, pp. 814–826, 2022. https://doi.org/10.1785/0220210197

[21] R. Koshy and S. Elango, "Utilizing social media for emergency response: A Tweet classification system using attention-based BiLSTM and CNN for resource management," *Multimed. Tools Appl.*, vol. 83, pp. 41 405–41 439, 2024. https://doi.org/10.1007/s11042-023-16766-z

[22] M. Chen and W. H. Du, "The predicting public sentiment evolution on public emergencies under deep learning and Internet of Things," *J. Supercomput.*, vol. 79, pp. 6452–6470, 2023. https://doi.org/10.1007/s11227-022-04900-x

[23] L. Han, R. Y. Fang, H. Zhang, G. P. Liu, C. S. Zhu, and R. F. Chi, "Adaptive autonomous emergency braking model based on weather conditions," *Traffic Inj. Prev.*, vol. 24, no. 7, pp. 609–617, 2023. https://doi.org/10.1080/15389588.2023.2227304

[24] H. Alatrista-Salas, J. Morzán-Samamé, and M. Nunez-del Prado, "Crime alert! Crime typification in news based on text mining," in *Proceedings of the 2019 Future of Information and Communication Conference*, San Francisco, USA, 2020, pp. 725–741. https://doi.org/10.1007/978-3-030-12388-8_50

[25] M. M. Ahsan, T. E. Alam, T. Trafalis, and P. Huebner, "Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and non-COVID-19 patients," *Symmetry*, vol. 12, no. 9, p. 1526, 2020. https://doi.org/10.3390/sym12091526

[26] P. Kompunt, S. Yongjoh, P. Aimtongkham, P. Muneesawang, K. Faksri, and C. So-In, "A hybrid LSTM and MLP scheme for COVID-19 prediction: A case study in Thailand," *Trends Sci.*, vol. 20, no. 10, p. 6884, 2023. https://doi.org/10.48048/tis.2023.6884

[27] W. W. Zhu, J. L. Wu, T. Fu, J. H. Wang, J. Zhang, and Q. Q. Shangguan, "Dynamic prediction of traffic incident duration on urban expressways: A deep learning approach based on LSTM and MLP," *J. Intell. Connect. Veh.*, vol. 4, no. 2, pp. 80–91, 2021. https://doi.org/10.1108/JICV-03-2021-0004

[28] S. A. Alshaya, "IoT device identification and cybersecurity: Advancements, challenges, and an LSTM-MLP solution," *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 6, pp. 11 992–12 000, 2023. https://doi.org/10.48084/etasr.6295

[29] D. Jatnika, M. A. Bijaksana, and A. A. Suryani, "Word2vec model analysis for semantic similarities in English words," *Procedia Comput. Sci.*, vol. 157, pp. 160–167, 2019. https://doi.org/10.1016/j.procs.2019.08.153

[30] S. J. Johnson, M. R. Murty, and I. Navakanth, "A detailed review on word embedding techniques with emphasis on Word2Vec," *Multimed. Tools Appl.*, vol. 83, pp. 37 979–38 007, 2024. https://doi.org/10.1007/s11042-023-17007-z

[31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735

[32] H. Taud and J. F. Mas, "Multilayer perceptron (MLP)," in *Geomatic Approaches for Modeling Land Change Scenarios*. Springer, Cham, 2018, pp. 451–455. https://doi.org/10.1007/978-3-319-60801-3_27

[33] J. Z. Sun, X. D. Zhang, and S. J. Lei, "The evolution of public opinion and its emotion analysis in public health emergency based on weibo data," in *International Conference on Logistics, Informatics and Service Sciences*, Beijing, China, 2022, pp. 415–434. https://doi.org/10.1007/978-981-99-2625-1_33