IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 August 2024, accepted 25 August 2024, date of publication 29 August 2024, date of current version 6 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3451522

RESEARCH ARTICLE

Road Traffic Accident Risk Prediction and Key Factor Identification Framework Based on Explainable Deep Learning

YULONG PEI, YUHANG WEN, AND SHENG PAN
School of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, China

Corresponding author: Yuhang Wen (wen_yu_hang@163.com)

This work was supported in part by the Key Research and Development Projects in Heilongjiang Province under Grant JD22A014, and in part by the Natural Science Foundation of China under Grant 50778056.

ABSTRACT The prediction and identification of key factors in road traffic accidents are crucial for accident prevention, yet previous studies have often examined these aspects separately. To comprehensively assess the risk level of road traffic accidents and their key determinants, this paper proposes a comprehensive forecasting and analysis framework that offers a novel perspective for identifying key risk factors from a modeling standpoint compared to existing methods. The CNN-BiLSTM-Attention model was developed for predicting the risk value of road accidents, and DeepSHAP was employed to interpret the model and extract the key factors contributing to traffic accidents. This deep learning framework combines convolutional neural networks (CNN) and Bi-directional long short-term memory (BiLSTM), while incorporating a spatial-temporal local attention mechanism to enhance its capability in capturing spatiotemporal features. Through analysis and experimentation on real-world datasets, our model demonstrates superior accuracy in predicting traffic accident risk compared to the benchmark model, achieving a Mean Absolute Error (MAE) of 0.2475 on the UK dataset and 0.2683 on the US dataset. The results obtained from DeepSHAP were found to be more rational and informative in identifying key factors of different severity levels using four methods. To verify the rationality and stability of obtaining these key factors, the first 15 factors were reintegrated into the prediction model, resulting in almost unchanged accuracy and reduced model iteration time. By improving the influential factors, road traffic accidents can be effectively mitigated.

INDEX TERMS Traffic accidents, risk prediction, explainability, deep learning.

I. INTRODUCTION

Currently, approximately 1.27 million individuals perish in traffic accidents annually, with nearly 20 to 50 million sustaining injuries because of such incidents. Road traffic accidents represent a significant cause of death, injury, permanent disability and property loss [1]. They impact not only the economy but also the healthcare system. It is crucial to reduce accident likelihood by identifying key factors and road accident risk levels since accidents do not occur randomly; they can be predicted and prevented [2]. Predicting regional accident risk degree constitutes an essential aspect of accident management that enables rescue personnel to evaluate regional traffic accident risks and their potential impacts. Simultaneously, implementing effective accident management procedures through improving critical influencing factors plays a vital role in preventing accidents [3].

In order to mitigate the impact of traffic accidents on regional safety, it is imperative to conduct an analysis of traffic accident data in order to investigate the correlation between regional accident risk levels and associated risk factors, thereby developing a novel predictive model for regional risk assessment. The degree of traffic accident risk is often influenced by four primary factors, namely the driver, vehicle, road conditions, and environmental conditions [4]. By exploring the influence of these various factors on the

The associate editor coordinating the review of this manuscript and approving it for publication was Huan Zhou.

level of regional traffic accident risks, effective measures can be implemented to enhance safety. Currently, diverse research methodologies are employed for analyzing and forecasting the degree of regional traffic accident risks.

The first research method involves analyzing the key factors contributing to traffic accidents through data analysis and making statistical predictions [5]. While this traditional approach allows for a comprehensive examination of the interrelationships among various factors and enables the establishment of linear or parametric models, it falls short in accurately predicting accident risks. Additionally, this method heavily relies on extensive data support but lacks effectiveness in processing multidimensional data [6]. Consequently, machine learning methods have gradually replaced it and garnered significant attention from researchers in recent years.

The field of machine learning can be categorized into two research methodologies: traditional machine learning models and deep learning models. Due to their proficiency in effectively capturing the nonlinear relationship between input and output data, traditional machine learning methods can explore the contribution of various factors to traffic accident risk and predicting the degree of risk associated with such accidents [7]. Simultaneously, by uncovering the interrelationships among these factors, it becomes possible to determine their respective contributions towards the degree of risk in traffic accidents, thereby establishing their relative importance as influencing factors [8]. However, traditional machine learning models tend to overlook certain spatiotemporal correlations, resulting in weaker predictive capabilities.

Currently, deep learning has emerged as a cutting-edge analytical technology extensively employed in the analysis and prediction of traffic accident factors. These models can accurately assess regional-level traffic accident risks by utilizing activity data such as GPS, thereby achieving superior precision [9]. However, while deep learning models excel at predicting traffic accident risks with remarkable accuracy, they often only consider specific contributing factors to accidents, which may not fully capture the complexity of real-world scenarios.

The machine learning and deep learning methods mentioned above have distinct focuses on various aspects of predicting traffic accident risks; nevertheless, there exist several unresolved issues in the field of traffic accident prediction. Firstly, previous studies have rarely compared state-of-the-art machine learning models with hybrid models [10]. Secondly, machine learning, being a black box method, fails to provide explanations for the relationship between influencing factors and the degree of traffic accident risk [11]. The prediction of traffic accidents is often disconnected from the identification of key factors, with the key factors not being determined from a model's perspective. Ultimately, due to the consideration of individual factors, the predictive capability of the model becomes weakened when dealing with unbalanced data, thereby diminishing prediction accuracy.

The aim of this study is to address the issue of deep learning models in the field of traffic accident prediction by accurately forecasting the risk of traffic accidents and identifying key factors from a model perspective. To achieve this, a comprehensive framework is developed that integrates predictive analysis of traffic accident risk levels with identification of crucial factors. The framework proposes an innovative CNN-BiLSTM-Attention model, which integrates a convolutional neural network (CNN), Bi-directional long short-term memory (BiLSTM), and spatial-temporal local attention mechanism to construct a prediction model capable of capturing both temporal and spatial characteristics of traffic accidents.

The problem of data imbalance is addressed through the utilization of data cleaning and local attention mechanism, while the reliability of the model is validated by employing real-world data to predict traffic accident risk levels in different regions. To tackle the issue of black box models and identify key factors, DeepSHAP is introduced for model interpretation and assessing the contribution degree of influential factors. This approach enables us to identify crucial determinants, mitigate prediction errors caused by risk disparities, and provide a novel perspective for comprehending model predictions.

The main contributions of this paper are as follows:

- The present study proposes a novel framework for traffic accident analysis that integrates deep learning models and explainable algorithms to combine accident risk prediction with critical factor identification. This framework constructs a CNN-BiLSTM-Attention model, utilizing CNN to capture the spatial features of accidents, BiLSTM to capture the bidirectional temporal features of accidents, and incorporating spatial-temporal local attention mechanism to enhance the predictive capability of the model, thereby improving its accuracy.
- The DeepSHAP algorithm is specifically employed in this paper to interpret and analyze the influencing factors of deep learning models, aiming to provide a comprehensive explanation and analysis from the model's perspective. It identifies key factors at different levels of traffic accidents and compares them with those identified by different models. Furthermore, these key factors are re-input into the model to verify their robustness.
- The proposed analytical framework was validated using real data, encompassing a decade of traffic accidents in the UK, thereby establishing its validity. Additionally, to assess the model's applicability, we employed traffic accident data from the United States for forecasting purposes. The results demonstrate that our CNN-BiLSTM-Attention model achieves higher accuracy than any other benchmark model. The comparison of key factors calculated by different methods simultaneously validates the comprehensiveness of DeepSHAP.

II. RELATED WORKS

In recent years, predicting traffic accident risk has emerged as a pivotal aspect within the realm of safety. Consequently, comprehending and interpreting deep models have become focal points in algorithmic research. The subsequent section will analyze and summarize both aspects.

A. STUDY OF TRAFFIC ACCIDENT RISK PREDICTION

For the prediction of traffic accident risks, statistical research primarily involves constructing various parameter models. Xu et al. [12] analyzed and predicted the characteristics and models of collision accidents involving Connected and Autonomous Vehicles (CAV) using a logit model, thereby identifying key influencing factors. Kwak [13] analyzed key variables by constructing a multivariate logistic model based on Korean highway data set, resulting in the development of a real-time collision program prediction model. Ma et al. [14] developed a prediction model that combines cloud modeling and Markov chain analysis to forecast the number of traffic accidents based on accident characteristics. Liu et al. [15] employed grey correlation analysis to establish association rules for mining traffic accidents, which were then validated using data from 31 provinces. Reeves et al. [16] used partially constrained generalized logistic regression models to characterize traffic accidents in the United Kingdom while exploring combinations of key risk factors.

Traditional machine learning models primarily employ algorithms to analyze and predict the risk of traffic accidents. Taamneh et al. [17] utilized a decision tree model (DT) for analyzing traffic accidents in ABU Dhabi and predicting the extent of damage caused. Ma et al. [18] integrated evidence theory with Bayesian network to forecast the probability of accident occurrence. Tang et al. [19] compared the random forest model (RF) with the K-nearest neighbor model (KNN) for predicting the duration of traffic accidents. Gong et al. [20] proposed A quantum K-nearest neighbor algorithm, which has higher classification efficiency and accuracy in high-dimensional data and greatly reduces the time.

Silagyi and Liu [21] employed a support vector machine model (SVM) to predict accident severity and personal injuries, ranking 14 factors contributing to accidents. Some combination models achieve improved prediction accuracy by integrating different algorithms. Assi et al. [22] combined fuzzy c-means (FCM) with SVM, inputting 15 collision factors for predicting and analyzing traffic accident severity. Peng et al. [23] predicted taxi accidents through a combination of eXtreme Gradient Boosting tree model (XGBoost), highlighting job-related factors as more significant.

As an emerging machine learning technology, deep learning is gradually being applied in the prediction of traffic accident risks. Gong et al. [24] inspired by quantum computing, a quantum convolutional neural network (QCNN) is proposed, which greatly improves the convergence speed and classification accuracy compared with traditional models. Jiang et al. [25] employed Long Short-Term Memory (LSTM) with varying time resolutions to predict the pre- and post-collision situations. Huang [26] utilized a sophisticated deep learning model to forecast collision accidents and validated it using accident data from the United States. Liu and Ukkusuri [27] combined spatial-temporal convolution with LSTM to predict the impact of Manhattan accidents. Additionally, numerous novel models have been proposed and implemented for traffic accident risk prediction. Zhou et al. [28] developed a minute-level urban traffic accident prediction model based on Graph Neural Network (GNN). Ma et al. [29] constructed an accident prediction framework utilizing Semi-Supervised Autoencoders (SSAE) to safeguard vulnerable road users (VRU), ensuring their safety. In the latest study, Wang et al. [30] used distance Graphs to capture the correlation of accidents in unbalanced data and proposed an Adaptive graph with Self-Supervised Learning (AGSSL) traffic accident prediction method. Zhou et al. [31] studied feasible edge caching strategies by using the distributed Multi-Agent Reinforcement Learning (MARL). They developed a computationally efficient method called DeepDMRE, and experimental results demonstrate its enhanced efficiency.

B. STUDY OF MODEL EXPLAINABLE ALGORITHMS

The machine learning algorithm, being an exceptional data-driven model, demonstrates the predictability effectively. However, due to its opaque nature, the internal decision-making process of machine learning is challenging to visualize and comprehend [32].

To gain a comprehensive understanding of the intrinsic relationship between meaningful input features and output goals, as well as comprehend the decision-making process of the model, extensive research has been conducted by numerous scholars in this field. Lundberg [33] introduced Shapley Additive Explanations (SHAP), which is rooted in game theory's concept of Shapley value and offers explanations for model predictions through the calculation of each feature's contribution. To interpret traditional machine learning models, Apley and Zhu [34] proposed local effects plots to provide explanations. Ding [35] developed a neural network interpreter-segmentation recognition and interpretation (NNI-SRI) algorithm capable of interpreting CNN models.

Currently, the explanatory algorithm introduces the concept of local explanation and global explanation. Ribeiro et al. [36], [37] have proposed two successive explanatory algorithms, namely LIME and Anchor, to provide local explanations for models and identify key factors. Binder et al. [38] have introduced Layer-Wise Relevance Propagation (LPR) for neural networks to quantify the contribution of input features. Some explanatory algorithms are also being developed for specialized models. Wei et al. [39] have constructed NeuronMotif, a neural network interpretation model that deciphers gene codes layer-by-layer. Zhao and Hastie [40] have conducted a causal analysis of

the model's interpretation, starting from Pearl's back-door adjustment.

The prediction of traffic accident risks necessitates the consideration of multiple factors and an understanding of their respective degrees of influence. Consequently, this study aims to establish an interpretable analytical framework for predicting and identifying the degree of risk and key factors associated with traffic accidents.

III. FRAMEWORK FOR ANALYZING TRAFFIC ACCIDENTS

The proposed framework in this paper for predicting traffic accident risk consists of two components: risk prediction and key factor identification. This section initially outlines the fundamental concepts and model structures of CNN and BiLSTM models used for traffic accident risk prediction. And it discusses the incorporation of a spatial-temporal local attention mechanism. Subsequently it introduces the principles and calculation methods of DeepSHAP. Finally, it summarizes the traffic accident risk prediction framework based on an explainable deep learning model.

A. CNN-BiLSTM-ATTENTION MODEL

The occurrence of traffic accidents is influenced by complex spatial-temporal factors. To comprehensively study the influence of spatial correlation among various influencing factors on traffic accidents and to obtain the key factors of traffic accidents. The CNN model is employed to extract spatial features of samples, while BiLSTM captures temporal features of accidents to comprehensively investigate the impact of spatial correlation among various influencing factors on traffic accidents. Additionally, the spatial-temporal local attention mechanism is incorporated to mitigate noise interference and further enhance the predictive performance of the model.

CNN possesses the characteristics of local connectivity, weight sharing, and a multi-level pooling structure, which effectively capture local features such as road type and accident location. These attributes are pivotal in predicting traffic accident risks, while the CNN's processing model adeptly captures key spatial features related to traffic accidents. Within model, BiLSTM plays a primary role in capturing temporal information regarding accidents including weather fluctuations and changes in accident timing among other characteristics. The bidirectional structure of BiLSTM enables the model to learn from both past and future information, enhancing its feature-capturing capabilities and ultimately improving prediction accuracy.

Meanwhile, we introduce a novel spatiotemporal local attention mechanism into the model to further enhance its focus on crucial information. In contrast to the global attention mechanism, this approach enables the model to identify the spatial-temporal segments that have the greatest impact on traffic accident risk and concentrate on significant features, thereby enhancing the accuracy of traffic accident risk prediction. To comprehensively capture traffic accident characteristics and accurately predict their likelihood, we propose a CNN-BiLSTM-Attention model for prediction.

This paper utilizes a combination of convolutional layer and Batch Normalization (BN) layer to enhance the stability of the model [41]. When processing textual data, a 1-dimensional convolutional neural network (1DCNN) was employed to capture the salient information, followed by utilization of a fully connected layer (FC) for classification and feature extraction. The computational procedure is illustrated in equations (1)-(3).

$$F = \sigma_F(BN(X^* w_F + b_F)) \tag{1}$$

$$Q = \delta(F) + b_Q \tag{2}$$

$$P = \sigma_P(Q^* w_P + b_P) \tag{3}$$

where F, Q, P represent the outputs of the convolutional layer, pooling layer, and fully connected layer, respectively; BN normalizes to a standard; σ_F, σ_P represent the activation function of the convolutional layer and the fully connected layer; X is the matrix of the input data; δ is the pooling layer pooling mode; w_F, w_P are the weight matrix of convolutional layer and fully connected layer; b_F, b_Q, b_P are the bias vectors of convolutional layer, pooling layer, and fully connected layer.

In addition to spatial correlation, there is also a clear temporal correlation observed in traffic accident data, encompassing both short-term and long-term correlations. LSTM, as opposed to traditional RNN, effectively addresses the issues of gradient explosion and gradient disappearance through the utilization of gated units [42].

LSTM consists of oblivion gate, input gate and output gate. The structure of the model is shown in Figure 1.

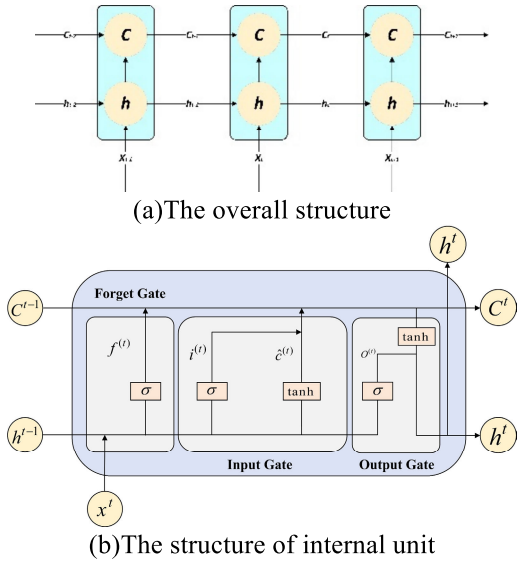

FIGURE 1. The structure of LSTM model.

Given its inherent memory function, this paper employs BiLSTM for extracting temporal features from samples. The BiLSTM model not only inherits the advantages of LSTM

but also effectively incorporates temporal information from both past and future directions, thereby enhancing prediction accuracy [43].

The calculation process of LSTM is illustrated in equations (4)-(9).

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{4}$$
$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{5}$$
$$g_t = \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \tag{6}$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{7}$$
$$c_t = g_t i_t + c_{t-1}f_t \tag{8}$$
$$h_t = \tanh(c_t)o_t \tag{9}$$

where x_t is the input at time t; W_{xf}, W_{xi}, W_{xg}, W_{xo} represents the weight matrix associated with x_t; h_{t-1} is the previous moment; W_{hf}, W_{hi}, W_{hg}, W_{ho} are the weight matrix associated with h_{t-1}; b_f, b_i, b_g, b_o are the bias vector.

The addition of an attention mechanism simultaneously enhances the spatial and temporal feature capturing abilities of CNN and BiLSTM models. This mechanism effectively reduces interruption loss during training, enabling the assignment of varying weights to input features for improved prediction accuracy and faster convergence speed [44]. Consequently, it facilitates the extraction of relationships between traffic accidents and different traffic indicators.

The local spatial attention mechanism in the encoder stage serves as a weighting mechanism for input data, enabling enhanced focus on the impact of regional traffic indicators. In each region, there exists a complex relationship between multiple local traffic flow data and traffic accident risk data with future traffic accident risk, which dynamically evolves over time [45]. The spatial attention mechanism is shown in equations (10)-(11).

$$a_t^j = w_1 \tanh(w_2 h_{t-1} + w_3 c_j^t + b_j) \tag{10}$$
$$\alpha_t^j = \frac{\exp(a_t^j)}{\sum_q^n (a_t^q)} \tag{11}$$

where a_t^j is the importance of the hidden layer at the previous moment for moment t; c_j^t is the j th indicator sequence at time t; α_t^j is the weight of attention; b_j is the bias vector.

The temporal attention mechanism is employed to adaptively capture the dynamic significance of historical time slices with respect to future time slices, thereby facilitating the model in acquiring a more comprehensive understanding of the impact of past moments on future instances. The hidden layer attention value at each moment is computed using equations (12) - (14).

$$C_t = \sum_{i=1}^{T} \theta_t^i h_i \tag{12}$$
$$\theta_t^i = \frac{\exp(r_t^i)}{\sum_{i=1}^{T} \exp(r_t^i)} \tag{13}$$
$$r_t^i = h_{t-1} W_r h_i + b_r \tag{14}$$

where r_t^i is the information at moment t;W_r is the associated weight matrix; b_r is the bias vector.

The final step involves utilizing a fully connected layer to establish connections between each node and the nodes in the preceding layer. This ensures that the extracted features from both before and after are interconnected, ultimately yielding an output value typically employed for multi-classification through Softmax logistic regression. The propagation formula for this fully connected layer is provided in equation (15).

$$x_{out} = \sigma(W_{out} + b_{out}) \tag{15}$$

where x_{out} is the final output value; σ is the sigmoid function; W_{out} is the weight between neurons in the last layer; b_{out} is the bias vector of the last layer.

The proposed CNN-BiLSTM-Attention model, as illustrated in Figure 2. Firstly, the processed data is initially fed into the spatial attention mechanism to enhance the spatial feature extraction capability of 1DCNN and mitigate the impact of imbalanced data. Subsequently, it undergoes normalization processing in the BN layer, followed by feature extraction through convolutional and pooling layers. Next, the information processed by 1DCNN is integrated into the temporal attention mechanism via an FC layer to reinforce time information extraction, which is further complemented by bidirectional BiLSTM for comprehensive time feature extraction. Ultimately, a prediction result indicating road traffic accident risk level is generated using an FC layer.

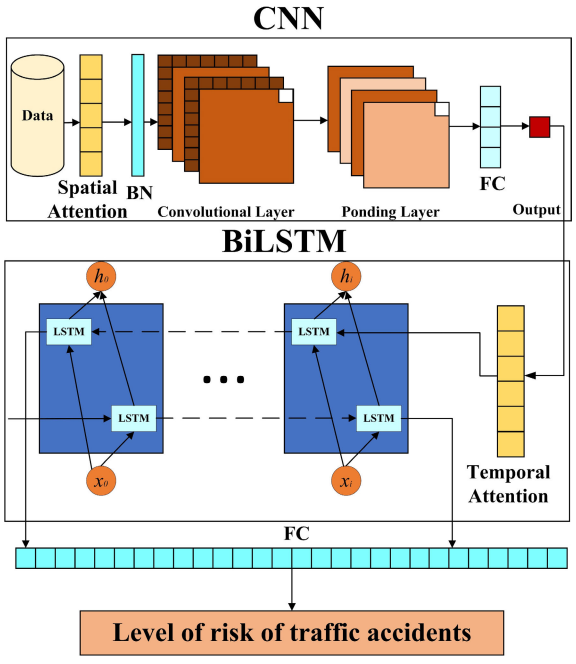

FIGURE 2. Structure of CNN-BiLSTM-Attention model.

According to the number of casualties in traffic accidents, traffic incidents are categorized into three levels: slight

incidents, serious incidents, and fatal incidents. The corresponding risk values for these traffic accidents are 1, 2, and 3. The prediction of the degree of traffic accident risk aims to forecast the cumulative sum Y_i^t of traffic accident risk values occurring within a given time period t for road section i based on input indicators.

When the accident sample contains numerous zero values, the marginal probability of "zero value" for each influencing factor tends to be overestimated, leading to erroneous conclusions [46]. To mitigate the issue of zero inflation, after computing the accident risk for each road section during any given period, we employ mean square error (MSE) as the loss function. During loss calculation, samples with higher accident risks are assigned greater weights based on a formulated calculation equation (16).

$$Loss(Y, Y') = \frac{1}{I}\sum_{i=1}^{I}(Y_i - Y_i')^2 \tag{16}$$

where Y is the accident risk value for all road sections; Y' is the predicted accident risk value for all road segments; I represents total number of accidents; Y_i is the actual risk value of the ith accident; Y_i' is the predicted risk value for the ith accident.

B. EXPLAINABLE ALGORITHMS AND FRAMEWORK BUILDING

The quality assessment of a model in algorithmic modeling typically relies on the precision and recall metrics of the test set. However, not all models conform to regular patterns, which poses challenges for comprehending and trusting their prediction outcomes. The feature contributions of traditional interpretable models are easily obtained due to their simplistic structure, whereas obtaining such contributions for deep learning models is challenging. Hence, explainable algorithms are employed to extract key features. To quantify the impact of features and provide a comprehensive explanation, we propose employing DeepSHAP [47] as an explainable tool to account for the model's predictions. This tool offers feature importance scores that elucidate the contribution of each input feature towards the model's predictions.

The DeepSHAP algorithm computes the Shapley value for each feature using the Deep Learning Important FeaTures (DeepLIFT), making it well-suited for deep learning models. equation (17) provides the formula to calculate the Shapley value.

$$\phi_j(v) = \sum_{S\subseteq(x_1,\ldots x_m)/x_j} \frac{|s|!(M-|S|-1)!\cdot(v(S\cup x_j)-v(S))}{M!} \tag{17}$$

where x_j is the jth feature; $\phi_j(v)$ is the characteristic contribution; S is the subset of features; M is the total number of features; $v(x)$ is the model predicted value for the study.

The Shapley values are approximated by DeepLIFT through the calculation of feature contributions using the multiplier principle and the chain rule. The theoretical formulation of DeepLIFT is provided in equation (18).

$$\sum_{i=1}^{n} C_{\Delta x_i \Delta G} = \Delta G \tag{18}$$

where x_i is the set of neural networks; x_0 is the reference value for x_i, $\Delta x_i = x_i - x_0$; G is the target output; ΔG is the sum of the contributions of the individual input features.

The explanations provided by DeepSHAP enable us to quantitatively assess the degree to which relevant factors impact accident risk. This information can then be utilized for the development of targeted policies and measures aimed at mitigating accident risk.

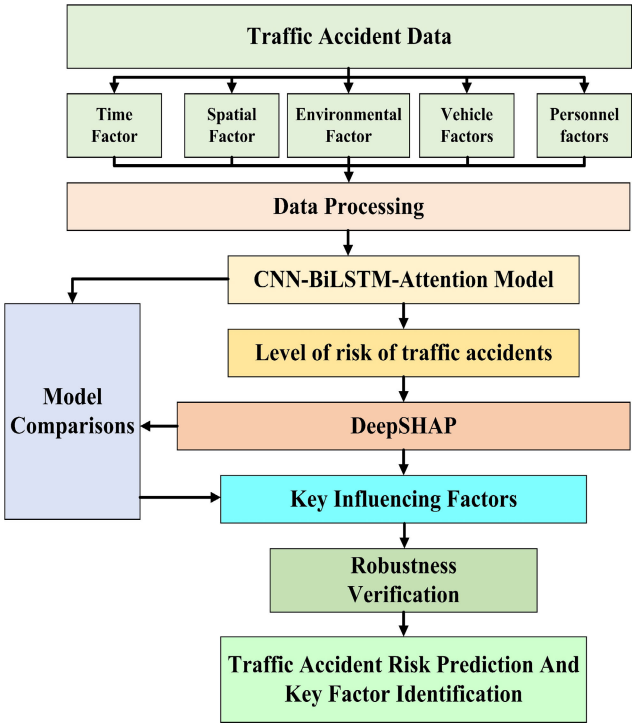

FIGURE 3. Framework process.

Therefore, the framework flow depicted in Figure 3 illustrates the integration of risk prediction and key factor identification. Initially, accident data is acquired and selected features are pre-processed into five categories for visualization. Subsequently, the pre-processed data is divided into training and test sets, which are then inputted into the CNN-BiLSTM-Attention model to predict traffic accident risk levels. Evaluation metrics are employed to assess the effectiveness of the model's predictions. Furthermore, a comparison between this paper's model, classical models, and novel models is conducted to demonstrate the superiority of our proposed approach.

The DeepSHAP technique is utilized to explain our model and determine each influencing factor's contribution extent. Additionally, different models are employed to obtain key factors associated with accidents of varying severity degrees for comparison purposes to validate the rationality of information obtained through DeepSHAP analysis. Through this

comparative analysis process, we select an optimal model that identifies key factors with maximum contribution degrees. Finally, these top 15 contributing factors are re-inputted into the CNN-BiLSTM-Attention model for prediction purposes to evaluate performance time and verify robustness while assessing rationality before outputting both risk prediction results and identified key factors.

IV. CASE STUDY

A. DATASET SOURCE

In order to validate the algorithm's predictive performance, this paper utilizes a publicly available dataset on road safety accidents and vehicles in the UK spanning from 2005 to 2018. The dataset consists of two components: accident-related personnel information and vehicle conditions, which are merged based on the accident index for experimental purposes in this study.

(https://www.kaggle.com/datasets/tsiaras/uk-road-safety-accidents-and-vehicles/code?resource=download)

The preprocessing of data can enhance computational performance. In this way, features missing more than 10% in the data are deleted, and to find suitable features, the proportion of discrepancy less than 0.02 is also deleted. Categorical variables in the processed data are imputed using mode-based filling for missing values, while temporal information is transformed into year, month, week, and hour formats to emphasize accident-related temporal characteristics. The original dataset comprises 2,058,408 accident record. The processed data set involved 2,755,286 injured persons. Of these, 1,734,548 (84.7%) were slightly injured, 286,339 (14%) were seriously injured, and 26,369 (1.3%) were killed.

The training set exhibits a significant disparity in the distribution of three types of traffic accident risk levels. To address this issue, this study employs SMOTE for oversampling and RandomUnderSampler for under sampling to achieve data balance. SMOTE is a widely used technique that generates new samples by interpolating existing ones, particularly useful in the domain of traffic accidents [48]. The RandomUnderSampler is an under-sampling technique that aims to balance the class distribution by reducing the sample size of the majority class. By combining these two methods, we can effectively address data imbalance and mitigate overfitting issues in forecasting tasks. To ensure robustness in data processing, 10-fold cross-validation is performed during the training process.

Additionally, 32 indicators with minimal missing data that are relevant to accident prediction were selected. Non-numerical indicators were encoded using the LabelEncoder algorithm from Sklearn package to establish one-to-one mapping for subsequent processing. The selected indicators are presented in Table 1.

The processed data were subjected to statistical analysis to observe the temporal and spatial patterns of the accidents. The trends depicting the three levels of accident severity are illustrated in Figure 4-6.

TABLE 1. UK data set traffic accident influencing factors.

Factor classification	Factor name	Factor classification	Factor name
Temporal Factor	Year	Vehicle factors	Road Surface Conditions
	month		Speed limit
	Day of Week		Propulsion Code
	hour		Age of Vehicle
Spatial Factor	Latitude & Longitude		Engine Capacity
Environmental Factor	Junction Control		Towing and Articulation
	Junction Detail		Skidding and Overturning
	Junction Location		Vehicle Manoeuvre
	Number of Vehicles		Vehicle Type
	Pedestrian Crossing Physical Facilities		Vehicle Reference
	Light Conditions	Driver factors	Age Band of Driver
	Weather Conditions		Driver Home Area Type
	Urban or Rural Area		Journey Purpose Of Driver
	Special Conditions at Site		Sex of Driver
	Carriageway Hazards		Was Vehicle Left Hand Drive
	Road Type		Accident Severity

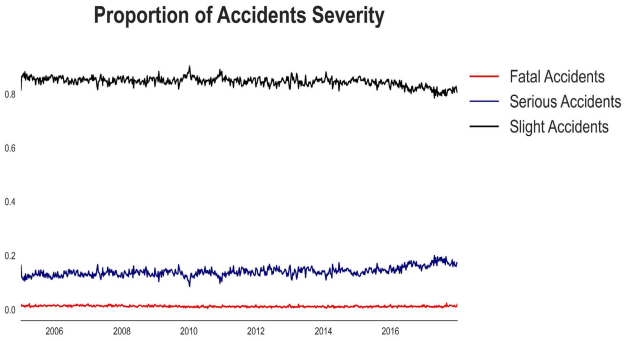

FIGURE 4. Annual change in severity of three types of accidents.

According to the annual variation of the data, there has been a slight decrease in the incidence of minor accidents, while the incidence of serious accidents has shown a gradual increase. The number of fatal accidents, however, has remained relatively stable. On a weekly basis,

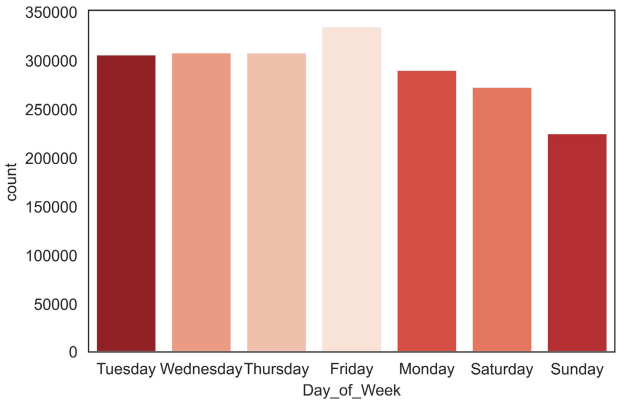

FIGURE 5. Weekly change in accidents.

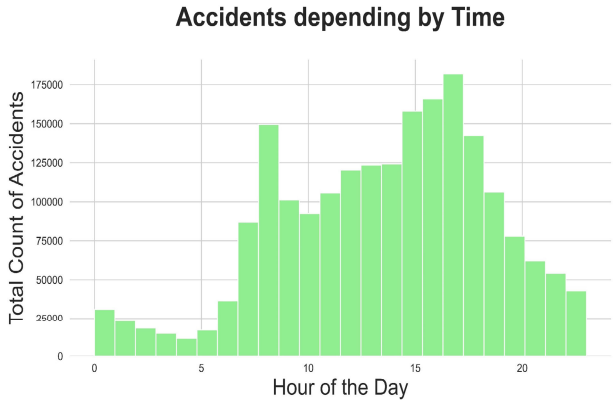

FIGURE 6. Daytime change in accidents.

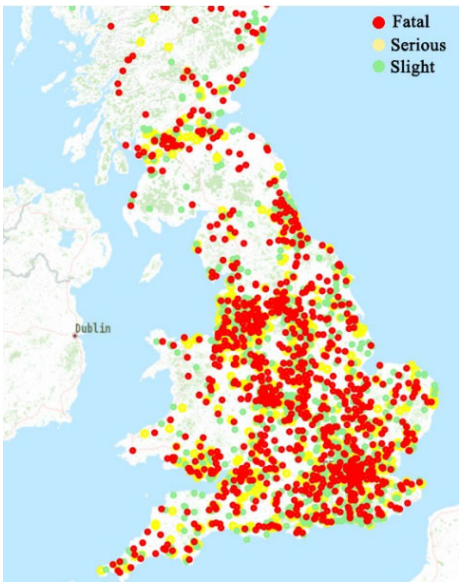

FIGURE 7. Spatial distribution of traffic accidents.

Fridays consistently record the highest number of traffic accidents, whereas Sundays consistently have the lowest number. This can be attributed to reduced travel during weekends. Analyzing daily changes reveals that peak hours in the morning and evening witness the highest occurrence of accidents due to high volumes of commuting trips.

The spatial distribution of traffic accidents is depicted in Figure. 7, wherein red indicates fatal accidents, yellow represents serious accidents, and green denotes minor accidents. Based on the spatial analysis, it can be observed that accident occurrences are concentrated along the road network and within urban areas such as London and Birmingham. Notably, there is a clear clustering pattern evident in the spatial characteristics of fatal accidents

At the same time, to verify the applicability of the model, road traffic accident data sets of 49 states in the United States from 2016 to 2020 were selected for comparison. (https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents) After the same pre-processing of the data, the selected features are shown in Table 2.

The analysis of Table 1 and Table 2 reveals that the features collected in the US data set are comparatively less comprehensive than those obtained from the UK data set, exhibiting significant disparities in environmental characteristics and vehicle attributes. Consequently, the third part of feature interpretation primarily focuses on analyzing and comparing the British data set's features.

TABLE 2. USA data set traffic accident influencing factors.

Factor classification	Factor name	Factor classification	Factor name
Temporal Factor	Year	Environmental Factor	VEHCOUNT
	Month		COLLISIONTYPE
	Day of week		JUNCTIONTYPE
	hour		WEATHER
Spatial Factor	Latitude & Longitude		ROADCOND
Environmental Factor	ADDRTYPE		JUNCTIONTYPE
	PEDCOUNT	Driver factors	SERIOUSINJURIES
	PERSONCOUNT		FATALITIES
	PEDCYLCOUNT		UNDERINFL
	LIGHTCOND		HITPARKEDCAR
	SPEEDING		INJURIES

B. PARAMETER SELECTION

The dataset in this study is divided into a training set, validation set, and test set in chronological order with a ratio of 6:2:2. The system configuration includes an AMD Ryzen9 7845HX CPU, GTX4060 GPU, and Windows 11,64-bit operating system. Python3.7 is used as the programming language, while Pytorch1.12.0 serves as the framework for building the model. After parameter tuning and inspection, the hyperparameter settings presented in Table 3 are adopted.

TABLE 3. Hyperparameter settings.

Parameter name	Parameter value
1DCN	Convolutional kernel 3:1, depth 32
BiLSTM	2 hidden layers
Dropout	1DCNN (rate=0.3) BiLSTM(rate=0.2)
Optimizers	Adam
Initial learning rate	0.001
regularization parameter	0.1113e-04
activation function	Relu
Number of iterations	100/200

As it assesses the risk of traffic accidents, it can be considered a regression model for predicting accident risks. It can also serve as a predictive model for assessing the degree of accident risks. Therefore, to evaluate the prediction model in this paper, four evaluation indicators are employed to measure its accuracy, including Mean Absolute Error (MAE), Precision, Recall, and F1 Score. MAE reflects the average magnitude of errors made by the prediction model. Also, Precision, Recall, and F1 Score enable evaluation of the model's performance in imbalanced datasets from different perspectives. The relevant formulas are provided below.

$$MAE = \frac{\sum_{t=1}^{N} |y(t) - \tilde{y}(t)|}{N}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ \text{score} = \frac{2Precision^{*}Recall}{Precision + Recall}$$

where N is the number of samples in the test set; $y(t)$ is the true value of risk for the t th sample; $\tilde{y}(t)$ is the predicted value of risk for the t th sample; TP is the number of samples that are actually true and predicted to be true; FN is the number of samples that were actually true but predicted to be false; FP is the number of samples that were actually false and predicted to be false.

The presence of a small MAE in a model indicates its strong fit. And if the model exhibits high Precision, Recall, and F1 score, it signifies its exceptional accuracy in predicting traffic accident risk.

C. ANALYSIS OF EXPERIMENTAL RESULTS

To validate the superiority of the proposed model, a module ablation experiment was concurrently conducted, wherein 10 benchmark models were selected for comparative analysis. These encompass the historical average model (HA), support vector machine model (SVM), back propagation neural network model (BP), multilayer Perceptron model (MLP), time series prediction model (GRU), stack denoising autoencoder (SDAE) [49], CNN model, LSTM model, BiLSTM model, CNN+BiLSTM model and CNN+BiLSTM+ Global attention model. The comparison model encompasses a broad spectrum of domains, ranging from conventional statistical models to state-of-the-art deep learning algorithms, to establish a comparative analysis with the model constructed in this paper.

The classical machine learning models employed in this paper include HA and SVM, while the most advanced models encompass BP, MLP, GRU, and SDAE. Currently, HA, SVM, BP, MLP, and GRU find extensive applications across various fields. On the other hand, SDAE represents the latest approach for constructing deep networks and acquiring hierarchical feature representations from inputs.

The role and effectiveness of each module in the model are verified through module ablation experiments using CNN, LSTM, BiLSTM, and CNN+BiLSTM. Additionally, a CNN+BiLSTM+Global attention model is designed to demonstrate the superiority of the proposed spatiotemporal local attention mechanism over the common global attention mechanism. Table 4 presents the optimal parameter settings for comparison methods. The parameter settings for the CNN model, BiLSTM model, CNN+BiLSTM model, and CNN+BiLSTM+Global attention model align with those of the CNN-BiLSTM-Attention model.

TABLE 4. Optimal parameter setting of the comparison method.

Models	Optimal parameter setting
SVM	regularization parameter：0.5 kernel type: rbf kernel coefficient: scale
BP	learning rate:0.001 L2 regularization term:0.001 batch size:300 Optimizer: Adam The number of hidden layer neurons:30
MLP	Number of fully connected layers:3 The number of Relu activated hidden units:128,64,32
GRU	Hidden size:100 Num layers:3 batch size:500 learning rate:0.001 Optimizer: Adam
SDAE	learning rate:0.001 L2 regularization term:0.001 number of denoising autoencoder layers :3 hidden dimension of each layer:30
LSTM	learning rate:0.001 L2 regularization term:0.001 number of LSTM layers:2 number of FC layers:2 LSTM unit hidden dimension:64 FC hidden dimension:40

Due to its complex nature as a voluminous data set, traffic accident data necessitates multiple iterations for effective feature acquisition by machine learning models to capture both temporal and spatial characteristics along with significant attributes associated with accidents. Nevertheless, excessive iteration runs may lead to overfitting issues within these models. Henceforth, each learning architecture undergoes 100/200 iterative processes utilizing identical datasets as a precautionary measure against incomplete execution of

TABLE 5. Model prediction performance on a UK dataset.

		Epoch=100				Epoch=200			
	Model	MAE	Precision	Recall	F1 Score	MAE	Precision	Recall	F1 Score
Classical model	HA	0.3397	0.6186	0.6521	0.6349	0.3153	0.6431	0.6724	0.6574
	SVM	0.3113	0.6618	0.6778	0.6697	0.3082	0.6721	0.6947	0.6832
State-of-the-art	BP	0.3049	0.6765	0.6913	0.6838	0.2924	0.6946	0.7016	0.6981
	MLP	0.2973	0.7176	0.7478	0.7324	0.2774	0.6894	0.7614	0.7236
	GRU	0.2989	0.7267	0.7596	0.7428	0.2767	0.7380	0.7816	0.7592
	SDAE	0.2752	0.7986	0.8328	0.8153	0.2498	0.7924	0.8477	0.8191
Modular ablation	CNN	0.2897	0.7485	0.7913	0.7693	0.2717	0.7378	0.7924	0.7641
	LSTM	0.2878	0.7707	0.7858	0.7782	0.2733	0.7683	0.8029	0.7852
	BiLSTM	0.2836	0.7889	0.8061	0.7974	0.2658	0.7749	0.8326	0.8027
	CNN-BiLSTM	0.2767	0.8003	0.8426	0.8209	0.2543	0.7958	0.8536	0.8237
	CNN+BiLSTM+ Global attention	0.2721	0.8095	0.8511	0.8298	0.2513	0.8017	0.8641	0.8317
Our model	CNN-BiLSTM-Attention	0.2654	0.8148	0.8641	0.8387	0.2475	0.8191	0.8782	0.8476

TABLE 6. Model prediction performance on a USA dataset.

		Epoch=100				Epoch=200			
	Model	MAE	Precision	Recall	F1 Score	MAE	Precision	Recall	F1 Score
Classical model	HA	0.3628	0.6048	0.6481	0.6257	0.3524	0.6138	0.6527	0.6327
	SVM	0.3425	0.6627	0.6628	0.6627	0.3386	0.6754	0.6754	0.6754
State-of-the-art	BP	0.3311	0.6824	0.6872	0.6848	0.3219	0.6928	0.6918	0.6923
	MLP	0.3218	0.7021	0.7326	0.7170	0.3127	0.7104	0.7452	0.7274
	GRU	0.3201	0.7128	0.7425	0.7273	0.3057	0.7221	0.7526	0.7371
	SDAE	0.2846	0.7483	0.8103	0.7781	0.2764	0.7612	0.8241	0.7914
Modular ablation	CNN	0.3126	0.7325	0.7829	0.7569	0.2931	0.7415	0.7886	0.7643
	LSTM	0.3108	0.7386	0.7814	0.7594	0.2917	0.7462	0.7901	0.7675
	BiLSTM	0.3084	0.7417	0.8122	0.7754	0.2856	0.7538	0.8216	0.7862
	CNN-BiLSTM	0.2957	0.7657	0.8236	0.7936	0.2754	0.7758	0.8452	0.8089
	CNN+BiLSTM+ Global attention	0.2858	0.7741	0.8339	0.8029	0.2711	0.7861	0.8516	0.8175
Our model	CNN-BiLSTM-Attention	0.2784	0.7882	0.8403	0.8134	0.2683	0.7962	0.8679	0.8305

100 cycles. This approach enables comprehensive comparison and analysis by selecting two representative iterations.

The experimental results of each comparison model on the UK traffic accident dataset are shown in Table 5. The experimental results of each comparison model on the US traffic accident data set are shown in Table 6.

Through comparative analysis of the models, it is evident that classical machine learning algorithms such as HA, and SVM exhibit relatively poor prediction performance, with their various indicators being inferior to those of the new model. Among these classical algorithms, SVM demonstrates exceptional capability in selecting significant features associated with traffic accidents while remaining resilient to input noise; thus, it outperforms other algorithms in terms of indicators.

The comparison of deep learning models reveals that the performance slightly improves with 200 iterations compared to 100 iterations. This highlights the capability of deep architectures in effectively modeling complex relationships. However, BP, MLP and GRU models exhibit better prediction effects than the basic machine learning model but fall short in terms of accuracy and precision when compared to the combined model. This discrepancy may arise from their limited focus on either spatial or temporal factors of traffic accidents without considering their spatial-temporal correlation. While SDAE performs closely to the combined model,

TABLE 7. Main factors affecting traffic accidents of different severities.

Slight				
Ranking of contributions	LR	RF	GCV-LIME	DeepSHAP
1	hour	Day of Week	Day of week	hour
2	Day of Week	Engine Capacity	Speed limit	Junction Control
3	Speed limit	Age of Vehicle	Latitude & Longitude	Engine Capacity
4	Road Type	Speed limit	Road Type	Latitude & Longitude
5	Number of Vehicles	Latitude & Longitude	Engine Capacity	Speed limit
Serious				
Ranking of contributions	LR	RF	GCV-LIME	DeepSHAP
1	Speed limit	Engine Capacity	Junction Control	Speed limit
2	Day of Week	Junction Control	Latitude & Longitude	Number of Vehicles
3	hour	Latitude & Longitude	Speed limit	Engine Capacity
4	Latitude & Longitude	Speed limit	Day of Week	hour
5	Number of Vehicles	hour	Number of Vehicles	Latitude & Longitude
Fatal				
Ranking of contributions	LR	RF	GCV-LIME	DeepSHAP
1	Number of Vehicles	hour	Speed limit	Speed limit
2	Speed limit	Speed limit	hour	Number of Vehicles
3	Day of Week	Latitude & Longitude	Age of Vehicle	Engine Capacity
4	Age of Vehicle	Number of Vehicles	Engine Capacity	hour
5	hour	Age of Vehicle	Latitude & Longitude	Latitude & Longitude

it is comparatively weaker due to its failure in accounting for dynamic influences of temporal features [50].

The accuracy of prediction in the module analysis is significantly diminished upon removal of the Attention module, BiLSTM module, and CNN module individually. It can be observed that each structure contributes uniquely to the model's performance. The CNN module captures spatial and related factors, the BiLSTM module captures temporal factors and retains them in memory, while the Attention module captures intricate information details to achieve optimal predictive efficacy. The prediction accuracy of the spatial-temporal local attention mechanism is higher compared to both the global attention mechanism and the spatial-temporal local attention mechanism, owing to its specific ability to capture appropriate features. In summary, CNN-BiLSTM-Attention surpasses all other methods across evaluation metrics due to its well-designed spatial-temporal information processing modules and utilization of attention mechanism for noise elimination and semantic representation learning from external information, rendering the model more robust.

According to Table 6, it can be observed that the model's trend is roughly like that of the UK data set. Models incorporating both temporal and spatial features tend to exhibit superior performance, and by incorporating spatial-temporal local attention mechanisms, they can effectively enhance the accuracy of accident risk prediction. Additionally, there is no evidence of overfitting, indicating that the inclusion of a BN layer and other techniques contribute to the stability of the model. However, due to fewer features in the US dataset compared to the UK dataset, the predictive accuracy on the US dataset is not as high as that on the UK dataset. Therefore, for identifying key factors, we selected the UK dataset for factor analysis.

D. MODEL EXPLANATION AND IDENTIFICATION OF KEY FACTORS

To enhance the model's predictive credibility and validate its interpretability, this paper employs DeepSHAP for explicating and analyzing the model, thereby yielding feature importance rankings. By identifying pivotal factors, effective policies can be formulated to mitigate traffic accident risk.

The DeepSHAP algorithm produces a feature score for each input, with positive scores indicating a favorable impact on the model's prediction and the magnitude of the score reflecting the extent of that impact. DeepSHAP is implemented using Python's SHAP library.The data set is divided into three levels based on the risk of traffic accidents, as different influencing factors are associated with each level. Subsequently, specific key factors are selected for each level to accurately identify the significant determinants.

The present study employed three methods, namely logistic regression (LR), random forest (RF), and GCV-LIME [51], to predict and analyze the same dataset, thereby validating the reliability of the interpretation algorithm. Comparative analysis was conducted by outputting feature importance. Random forest explanations calculated feature importance, while logistic regression explanations were based on feature coefficients. GCV-LIME is a text mining technique utilized to obtain the frequency of occurrence of accident factors through text mining and calculate scores for each factor in order to identify key factors. The top five factors with the highest impact on the three levels of predicted traffic accidents are presented in Table 7 using four different methods.

The results presented in Table 7 demonstrate that the primary influencing factors of minor accidents, as identified by the four methods, are all temporally related. Furthermore, intersection type and geographical location exert a significant impact, indicating a higher prevalence of minor accidents at intersections. Additionally, there is an observable correlation between speed limits and minor accident occurrence. The disparity in influencing factors between slight accidents and other types of accidents is significant, while the distinction between serious accidents and fatal accidents is minimal. The primary determinants in serious and fatal accidents remain to be vehicle speed limits, the number of vehicles, and Engine Capacity. Consequently, it can be concluded that influential factors for serious and fatal accidents often intertwine with vehicular and road constraints.

From a methodological perspective, in the case of slight accidents, RF and GCV-LIME identified the week in which the accident occurred as the primary influencing factor, while LR and DeepSHAP attributed importance to the time of day. This is because RF prioritize specificity over time, making it a key feature for them but less plausible as an influencing factor. Additionally, GCV-LIME's text mining approach tends to highlight the occurrence of weekdays more frequently, leading to this issue. In serious and fatal accidents, LR also identifies both weekday and time of day as significant factors with an overrepresentation of temporal characteristics.

The features provided by random forest tend to have a high incidence of missing values, rendering them less reliable. These missing values are imputed using the mode, resulting in a significant exacerbation of the problem. Consequently, the explanations offered by random forests become nonsensical when dealing with datasets containing substantial missing values. The GCV-LIME method, however, tends to emphasize the features of text occurrences and cross occurrences more prominently. This discrepancy between key judgment factors and model predictions makes it inappropriate to assess accident key factors solely from a model perspective.

In contrast, DeepSHAP employs a model proposed in this paper that effectively mitigates the impact of missing values. Furthermore, compared to LR, RF and GCV-LIME, DeepSHAP provides more extensive and detailed information. Therefore, combining the DeepSHAP algorithm with the CNN-BiLSTM-Attention model presented in this study yields a more rational and effective explanation and analysis.

The global contribution degree of influence factors is determined by integrating DeepSHAP with the proposed model, and Figure 8 illustrates the contribution size of the top 15 influence factors along with their respective contribution degrees.

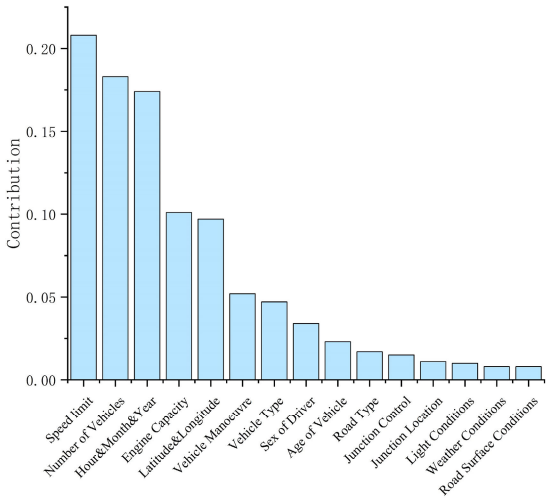

FIGURE 8. Degree of contribution of the top 15 influencing factors Degree of contribution of the top 15 influencing factors.

According to the given contribution degree of influencing factors, it can be observed that among the global factors, speed limit, number of vehicles, time, and vehicle performance exert the most significant impact on the risk level of traffic accidents. This implies that higher speeds and increased vehicular density are associated with a greater likelihood of traffic accidents occurring.

Additionally, both time and location play crucial roles in determining accident risks. Enhancing safety measures in high-risk locations and during critical time periods can effectively mitigate the risk of traffic accidents. However, factors such as vehicle type, road conditions, and weather conditions contribute relatively less to the overall degree of accident risk due to their lower occurrence probabilities.

The top 15 key factors at various levels are re-entered into the model for predicting traffic accident risk, and the robustness of both the model and the selected key factors is validated. The results are shown in Table 8.

The results in Table 5 demonstrate that the selection of the 15 key factors for predicting traffic accident risk leads to a slight decrease in all indicators, yet there is only a minimal increase in MAE. This suggests that the accuracy of the prediction remains largely unaffected.

The marginal impact on Precision, Recall, and F1 Score is limited to a mere 5%, thereby substantiating that the predictive accuracy remains largely unaffected even when considering only the top 15 key factors. This underscores the rationality and efficacy of utilizing these influential factors for predicting traffic accident risks.

TABLE 8. Predictive performance of models with selected key influencing factors.

Degree of risk	MAE	Precision	Recall	F1 Score
Slight	0.2524	0.8155	0.8584	0.8364
Serious	0.2563	0.8104	0.8428	0.8263
Fatal	0.2578	0.8102	0.8416	0.8256
Global	0.2517	0.8134	0.8629	0.8427
All factors	0.2475	0.8191	0.8782	0.8476

The findings in Table 8 reveal a greater reduction in the indicators of severe accidents and fatal accidents. This can be attributed to the limited amount of data available for serious and fatal accidents, which leads to decreased prediction accuracy when input factors are also reduced. The accuracy of model prediction remains largely unaffected, while the overall training time can be reduced by nearly one hour through a reduction in input features during the training process. This enhancement enables us to expedite model training under the same computational resources and significantly improves its efficiency.

Therefore, by carefully selecting the key influencing factors, it is possible to achieve comparable accuracy results within a shorter timeframe. This substantiates the rationality and effectiveness of identifying key factors for predicting traffic accident risks, thereby validating the model explanation provided by DeepSHAP.

By identifying these influential factors, it becomes evident that effective measures to mitigate traffic accident risks include enforcing vehicle speed limits, managing high-risk areas prone to accidents, and deploying personnel for supervision during peak accident periods. Additionally, timely vehicle inspections are crucial in reducing the likelihood of traffic accidents.

V. CONCLUSION

The present study proposes an analytical framework to forecast and identify crucial factors contributing to the risk of road traffic accidents, thereby addressing the issue of fragmented research. Within this framework, a CNN-BiLSTM-Attention model is established to predict traffic accident risk, while DeepSHAP is utilized to elucidate the model and identify the crucial influencing factors of such risks

The proposed model integrates a combination of CNN and BiLSTM to effectively capture both the temporal and spatial characteristics of the features. Moreover, it incorporates the spatial-temporal local attention mechanism to enhance model performance by capturing intricate spatiotemporal information and bolstering model robustness.

The UK's actual traffic accident dataset is utilized for prediction, and the efficacy and superiority of the traffic accident risk prediction model are validated through algorithmic comparison experiments and module ablation experiments. The model achieves an MAE of 0.2475, a Precision of 0.8191, a Recall of 0.8782, and an F1 Score of 0.8476, enabling accurate anticipation of traffic accident risks. The model's applicability was further verified by comparing it with USA data, which also yielded a high prediction accuracy, as evidenced by an MAE of 0.2683. This model can proactively alert potential hazards in advance and assist individuals in selecting safer travel routes.

After completing the prediction, DeepSHAP is utilized to elucidate the model and generate the feature contribution degrees for three categories of traffic accident severity. The top five features with the highest contribution degrees are selected for analysis, and upon comparison with results derived from logistic regression, random forest and GCV-LIME, it is observed that DeepSHAP provides more reasonable factors containing richer information.

Subsequently, the top 15 contributing features are employed for prediction. The experimental results demonstrate that the predictive performance of key factors remains largely unchanged, while there is a significant reduction in model iteration time. The identified key influencing factors will aid in mitigating traffic accident risks through policy adjustments and measures such as intersection modifications, thereby ensuring personal and property safety of road users.

REFERENCES

[1] J. J. Rolison, S. Regev, S. Moutari, and A. Feeney, "What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records," *Accident Anal. Prevention*, vol. 115, pp. 11–24, Jun. 2018.

[2] H. Manner and L. Wünsch-Ziegler, "Analyzing the severity of accidents on the German autobahn," *Accident Anal. Prevention*, vol. 57, pp. 40–48, Aug. 2013, doi: 10.1016/j.aap.2013.03.022.

[3] M. Riveiro, M. Lebram, and M. Elmer, "Anomaly detection for road traffic: A visual analytics framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 2260–2270, Aug. 2017.

[4] B. Wali, A. J. Khattak, and T. Karnowski, "The relationship between driving volatility in time to collision and crash-injury severity in a naturalistic driving environment," *Analytic Methods Accident Res.*, vol. 28, Dec. 2020, Art. no. 100136.

[5] A. Osama and T. Sayed, "Macro-spatial approach for evaluating the impact of socio-economics, land use, built environment, and road facility on pedestrian safety," *Can. J. Civil Eng.*, vol. 44, no. 12, pp. 1036–1044, Dec. 2017, doi: 10.1139/cjce-2017-0145.

[6] W.-H. Chen and P. P. Jovanis, "Method for identifying factors contributing to driver-injury severity in traffic crashes," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1717, no. 1, pp. 1–9, Jan. 2000, doi: 10.3141/1717-01.

[7] S. Sarkar, S. Vinay, R. Raj, J. Maiti, and P. Mitra, "Application of optimized machine learning techniques for prediction of occupational accidents," *Comput. Oper. Res.*, vol. 106, pp. 210–224, Jun. 2019, doi: 10.1016/j.cor.2018.02.021.

[8] Y. Zou, B. Lin, X. Yang, L. Wu, M. Muneeb Abid, and J. Tang, "Application of the Bayesian model averaging in analyzing freeway traffic incident clearance time for emergency management," *J. Adv. Transp.*, vol. 2021, pp. 1–9, Mar. 2021, doi: 10.1155/2021/6671983.

[9] Q. Chen, X. Song, H. Yamada, and R. Shibasaki, "Learning deep representation from big and heterogeneous data for traffic accident inference," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2016, vol. 30, no. 1, pp. 1–7, doi: 10.1609/aaai.v30i1.10011.

[10] M. Manzoor, M. Umer, S. Sadiq, A. Ishaq, S. Ullah, H. A. Madni, and C. Bisogni, "RFCNN: Traffic accident severity prediction based on decision level fusion of machine and deep learning model," *IEEE Access*, vol. 9, pp. 128359–128371, 2021, doi: 10.1109/ACCESS.2021.3112546.

[11] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, 2014, pp. 818–833, doi: 10.1007/978-3-319-10590-1_53.

[12] C. Xu, Z. Ding, C. Wang, and Z. Li, "Statistical analysis of the patterns and characteristics of connected and autonomous vehicle involved crashes," *J. Saf. Res.*, vol. 71, pp. 41–47, Dec. 2019.
[13] H.-C. Kwak and S. Kho, "Predicting crash risk and identifying crash precursors on Korean expressways using loop detector data," *Accident Anal. Prevention*, vol. 88, pp. 9–19, Mar. 2016.
[14] X. Ma, J. Zhang, P. Huang, H. Sang, G. Sun, and J. Chen, "Traffic accident prediction based on Markov chain cloud model," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 526, no. 1, Jun. 2020, Art. no. 012188, doi: 10.1088/1755-1315/526/1/012188.
[15] Y. Liu, X. Huang, J. Duan, and H. Zhang, "The assessment of traffic accident risk based on grey relational analysis and fuzzy comprehensive evaluation method," *Natural Hazards*, vol. 88, no. 3, pp. 1409–1422, Sep. 2017, doi: 10.1007/s11069-017-2923-2.
[16] K. Reeves, J. S. Chandan, and S. Bandyopadhyay, "Using statistical modelling to analyze risk factors for severe and fatal road traffic accidents," *Int. J. Injury Control Saf. Promotion*, vol. 26, no. 4, pp. 364–371, Oct. 2019, doi: 10.1080/17457300.2019.1635625.
[17] M. Taamneh, S. Alkheder, and S. Taamneh, "Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates," *J. Transp. Saf. Secur.*, vol. 9, no. 2, pp. 146–166, Apr. 2017, doi: 10.1080/19439962.2016.1152338.
[18] X. Ma, Y. Xing, and J. Lu, "Causation analysis of hazardous material road transportation accidents by Bayesian network using genie," *J. Adv. Transp.*, vol. 2018, pp. 1–12, Aug. 2018, doi: 10.1155/2018/6248105.
[19] J. Tang, L. Zheng, C. Han, W. Yin, Y. Zhang, Y. Zou, and H. Huang, "Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review," *Analytic Methods Accident Res.*, vol. 27, Sep. 2020, Art. no. 100123, doi: 10.1016/j.amar.2020.100123.
[20] L. Gong, W. Ding, Z. Li, Y. Wang, and N. Zhou, "Quantum K-nearest neighbor classification algorithm via a divide-and-conquer strategy," *Adv. Quantum Technol.*, vol. 7, no. 6, Mar. 2024, Art. no. 2300221, doi: 10.1002/QUTE.202300221.
[21] D. V. Silagyi and D. Liu, "Prediction of severity of aviation landing accidents using support vector machine models," *Accident Anal. Prevention*, vol. 187, Jul. 2023, Art. no. 107043, doi: 10.1016/j.aap.2023.107043.
[22] K. Assi, S. M. Rahman, U. Mansoor, and N. Ratrout, "Predicting crash injury severity with machine learning algorithm synergized with clustering technique: A promising protocol," *Int. J. Environ. Res. Public Health*, vol. 17, no. 15, p. 5497, Jul. 2020.
[23] Z. Peng, J. Zuo, H. Ji, Y. RengTeng, and Y. Wang, "A comparative analysis of risk factors in taxi-related crashes using XGBoost and SHAP," *Int. J. Injury Control Saf. Promotion*, vol. 31, no. 3, pp. 508–520, May 2024, doi: 10.1080/17457300.2024.2349555.
[24] L.-H. Gong, J.-J. Pei, T.-F. Zhang, and N.-R. Zhou, "Quantum convolutional neural network based on variational quantum circuits," *Opt. Commun.*, vol. 550, Jan. 2024, Art. no. 129993, doi: 10.1016/j.optcom.2023.129993.
[25] F. Jiang, K. K. R. Yuen, and E. W. M. Lee, "A long short-term memory-based framework for crash detection on freeways with traffic data of different temporal resolutions," *Accident Anal. Prevention*, vol. 141, Jun. 2020, Art. no. 105520.
[26] T. Huang, S. Wang, and A. Sharma, "Highway crash detection and risk estimation using deep learning," *Accident Anal. Prevention*, vol. 135, Feb. 2020, Art. no. 105392.
[27] J. Bao, P. Liu, and S. V. Ukkusuri, "A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data," *Accident Anal. Prevention*, vol. 122, pp. 239–254, Jan. 2019.
[28] Z. Zhou, Y. Wang, X. Xie, L. Chen, and H. Liu, "RiskOracle: A minute-level citywide traffic accident forecasting framework," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 1, pp. 1258–1265, doi: 10.1609/aaai.v34i01.5480.
[29] Z. Ma, G. Mei, and S. Cuomo, "An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors," *Accident Anal. Prevention*, vol. 160, Sep. 2021, Art. no. 106322, doi: 10.1016/j.aap.2021.106322.
[30] S. Wang, Y. Zhang, X. Piao, X. Lin, Y. Hu, and B. Yin, "Data-unbalanced traffic accident prediction via adaptive graph and self-supervised learning," *Appl. Soft Comput.*, vol. 157, May 2024, Art. no. 111512, doi: 10.1016/J.ASOC.2024.111512.
[31] H. Zhou, K. Jiang, S. He, G. Min, and J. Wu, "Distributed deep multi-agent reinforcement learning for cooperative edge caching in Internet-of-Vehicles," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 9595–9609, Dec. 2023, doi: 10.1109/TWC.2023.3272348.
[32] U. Bansal, "Explainable AI: To reveal the logic of black-box models," *New Gener. Comput.*, vol. 42, no. 1, pp. 53–87, Mar. 2024, doi: 10.1007/s00354-022-00201-2.
[33] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 4765–4774.
[34] D. W. Apley and J. Zhu, "Visualizing the effects of predictor variables in black box supervised learning models," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 82, no. 4, pp. 1059–1086, Sep. 2020, doi: 10.1111/rssb.12377.
[35] Y. Ding, "Visualizing deep networks using segmentation recognition and interpretation algorithm," *Inf. Sci.*, vol. 609, pp. 1381–1396, Sep. 2022.
[36] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.
[37] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2018, vol. 32, no. 1, pp. 1–9, doi: 10.1609/aaai.v32i1.11491.
[38] A. Binder, G. Montavon, S. Lapuschkin, K.-R. Müller, and W. Samek, "Layer-wise relevance propagation for neural networks with local renormalization layers," in *Artificial Neural Networks and Machine Learning—ICANN*. Cham, Switzerland: Springer, 2016, pp. 63–71, doi: 10.1007/978-3-319-44781-0_8.
[39] Z. Wei, K. Hua, L. Wei, S. Ma, R. Jiang, X. Zhang, Y. Li, W. H. Wong, and X. Wang, "NeuronMotif: Deciphering cis-regulatory codes by layer-wise demixing of deep neural networks," *Proc. Nat. Acad. Sci. USA*, vol. 120, no. 15, Apr. 2023, Art. no. e2216698120, doi: 10.1073/pnas.2216698120.
[40] Q. Zhao and T. Hastie, "Causal interpretations of black-box models," *J. Bus. Econ. Statist.*, vol. 39, no. 1, pp. 272–281, Jan. 2021, doi: 10.1080/07350015.2019.1624293.
[41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, Jun. 2015, pp. 448–456. [Online]. Available: https://proceedings.mlr.press/v37/ioffe15.html
[42] J. Wu and Z. Wang, "A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory," *Water*, vol. 14, no. 4, p. 610, Feb. 2022, doi: 10.3390/w14040610.
[43] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Dallas, TX, USA, Jun. 2017, pp. 88–95, doi: 10.1109/ICPHM.2017.7998311.
[44] Y. Guo, J. Ji, X. Lu, H. Huo, T. Fang, and D. Li, "Global-local attention network for aerial scene classification," *IEEE Access*, vol. 7, pp. 67200–67212, 2019, doi: 10.1109/ACCESS.2019.2918732.
[45] S. Zhao, R. Y. Zhong, Y. Jiang, S. Besklubova, J. Tao, and L. Yin, "Hierarchical spatial attention-based cross-scale detection network for digital works supervision system (DWSS)," *Comput. Ind. Eng.*, vol. 192, Jun. 2024, Art. no. 110220.
[46] Q. Ren, M. Xu, and X. Yan, "An investigation of heterogeneous impact, temporal stability, and aggregate shift in factors affecting the driver injury severity in single-vehicle rollover crashes," *Accident Anal. Prevention*, vol. 200, Jun. 2024, Art. no. 107562, doi: 10.1016/j.aap.2024.107562.
[47] Z. T. Fernando, J. Singh, and A. Anand, "A study on the interpretability of neural retrieval models using DeepSHAP," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.* New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 1005–1008, doi: 10.1145/3331184.3331312.
[48] A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, and A. Mohammadian, "Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis," *Accident Anal. Prevention*, vol. 136, Mar. 2020, Art. no. 105405, doi: 10.1016/j.aap.2019.105405.
[49] L. Xian and L. Tian, "Passenger flow prediction and management method of urban public transport based on SDAE model and improved bi-LSTM neural network," *J. Intell. Fuzzy Syst.*, vol. 45, no. 6, pp. 10563–10577, Dec. 2023.
[50] S. Kamoji and M. Kalla, "Effective flood prediction model based on Twitter text and image analysis using BMLP and SDAE-HHNN," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106365.

[51] C. Arteaga, A. Paz, and J. Park, "Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach," *Saf. Sci.*, vol. 132, Dec. 2020, Art. no. 104988, doi: 10.1016/j.ssci.2020.104988.

YULONG PEI received the B.S. degree from Harbin University of Civil Engineering and Architecture, Harbin, China, in 1982, the M.S. degree from Chang'an University, Xi'an, China, in 1988, and the Ph.D. degree in transportation planning and management from Southeast University, Nanjing, China, in 2002.

He has been involved in transportation engineering research for an extensive period, mainly focusing on urban transport planning, traffic safety, and transportation management and control. He is currently working as a Professor with Northeast Forestry University and the Director of the Transportation Research Centers, Northeast Forestry University, and Heilongjiang Provincial Key Laboratory of Cold Land Road Transportation Technology.

YUHANG WEN received the B.S. degree in transportation engineering from Nantong University, in 2022. He is currently pursuing the master's degree in transportation engineering with the School of Transportation, Northeast Forestry University, Harbin, China. His current research interests include land transportation corridor networks in urban agglomerations and traffic accident risk prediction.

SHENG PAN received the B.S. degree in transportation from Wuhan University of Technology, in 2021. He is currently pursuing the master's degree in transportation engineering with the School of Transportation, Northeast Forestry University, Harbin, China. His current research interests include lane-changing behavior and capacity of intelligent networked vehicles and reinforcement learning model prediction.