Research article

Traffic accident duration prediction using multi-mode data and ensemble deep learning

Jiaona Chen [a,*], Weijun Tao [a], Zhang Jing [a], Peng Wang [a], Yinli Jin [b]

[a] *Xi'an Shiyou University School of Electronic Engineering, Xi'an, 710065, China*
[b] *Chang'an University School of Electronic and Control Engineering, Xi'an, 710061, China*

A R T I C L E   I N F O

A B S T R A C T

Predicting the duration of traffic accidents is a critical component of traffic management and emergency response on expressways. Traffic accident information is inherently multi-mode data in terms of data types. However, most existing studies focus on single-mode data, and the influence of multi-mode data on the prediction performances of models has been the subject of only very limited quantitative analysis. The present work addresses these issues by proposing a heterogeneous deep learning architecture employing multi-modal features to improve the accuracy of predictions for traffic accident durations on expressways. Firstly, six unique data modes are obtained based on the structured data and the text data. Secondly, a hybrid deep learning approach is applied to build classification models with reduced prediction error. Finally, a rigorous analysis of the influence for multi-mode data on the accident duration prediction performances is conducted using a variety of deep learning models. The proposed method is evaluated using survey data collected from an expressway monitoring system in Shaanxi Province, China. The experimental results show that Word2Vec-BiGRU-CNN is a suitable and better model using text features for traffic accident duration prediction, as the F1-score is 0.3648. This study confirms that the newly established structured features extracted from text data substantially enhance the prediction effects of deep learning algorithms. However, these new features were a detriment to the prediction effects of conventional machine learning algorithms. Accordingly, these results demonstrate that the processing and extraction of text features is a complex issue in the field of traffic accident duration prediction.

1. Introduction

Traffic accidents have a highly negative impact on the efficiency and service quality of transportation networks. Among the many negative impacts of traffic accidents, delays caused by traffic accidents are most significant because they reduce the efficiency of the entire network while increasing fuel consumption and air pollution. In fact, the expense of delays caused by accidents is around 70% greater than the expense of material damage [1]. Hence, the period of time required to clear an accident and return traffic to normal conditions, which is typically denoted as the accident duration, is the dominant feature affecting the level of delay and its cost to transportation networks. As a part of the predictive maintenance, it is essential for measuring the deterioration level of the

transportation system in term of accident duration to achieve economic and environmental objectives [2]. Consequently, accurately predicting the durations of accidents from start to finish would be highly useful for guiding transportation agencies and drivers in mitigating the congestion arising from accidents.

Early studies focused on developing parametric prediction models based on an assumption that accident durations statistically follow some distribution. For example, Li proposed an accelerated failure time hazard-based model for predicting the durations of different stages of traffic incidents [3]. In a follow-up study, Li et al. investigated the influence of clearance methods and various covariates on the duration of traffic incidents. They verified that a competing risks mixture model obtained greater accuracy for predicting traffic accident duration than the conventional AFT (Accelerated Failure Time) model [4]. However, the factors influencing the occurrence of accidents are numerous and complex. Accordingly, parametric models are limited in practice regarding their assumptions pertaining to the overall distribution of data. In this regard, Wali et al. determined that random parameter models can obtain more accurate predictions than quantile regression and fixed parameter models based on more than 45,000 traffic incidents recorded in the state of Virginia, USA [5].

These issues can be effectively addressed by developing accident duration prediction approaches based on machine learning methods, such as support vector machine (SVM) [6], k-nearest neighbor (KNN) algorithms [7], and artificial neural networks (ANNs) [8]. Gao et al. conducted a study where they developed a random survival forest model to analyze accident durations. They found that this model outperformed the random forest model in terms of prediction accuracy. The results of their study indicate that the duration of traffic accidents is influenced by various factors including the type of traffic accident, the length of the road segment, the location of the accident, the number of remaining lanes, and the traffic intensity [9]. Li et al. conducted a comprehensive review of research on the analysis and prediction of traffic incident duration. The objective of their study was to discuss the evolution of research in this field, focusing on different phases of incident duration, data resources, and the various methods used in analyzing factors that influence traffic incident duration and predicting duration time [10]. Similarly, He et al. developed combined model of principal component analysis (PCA) and random forest (RF) to predict the duration of traffic accidents for highway tunnels. The results show that the average absolute error of PCA-RF combined model is 12.80 min, and the accuracy of error within 20 min is 89.15% [11]. Shang et al. introduced a new method for predicting the duration of traffic incidents. They utilized the Neighborhood Components Analysis (NCA) and the Bayesian Optimization Algorithm (BOA) to optimize the Random Forest (RF) model. The evaluation of the method using the confusion matrix showed that it achieved high accuracy and demonstrated excellent reliability and robustness [12]. Later, Hamad et al. established RF models for predicting traffic accident durations over both a wide duration range (1–1440 min) and a short range (5–120 min) based on 146,573 incident records and 52 variables [13]. A detailed study by Saracoglu et al. demonstrated that CHAID (Chi-square Automatic Interaction Detector), CART (Classification and Regression Tree), C4.5 (Cart 4.5), and LMT (Logistic Model Tree) models provided nearly equivalent prediction accuracies between 74% and 75.6% [14]. Ghosh et al. conducted an analysis on prediction errors obtained from Bayesian Support Vector Regression (BSVR) and Gaussian Process (GP) models for traffic incident duration predictions using two datasets, one from Singapore and the other from the Netherlands. They found that for incidents in Singapore, the sensitivity of BSVR and GP at 70% specificity were 67% and 78% respectively. For incidents in the Netherlands, the sensitivity of the two methods at the same level of specificity were 55% and 60% respectively. The study concluded that the BSVR model was more reliable [15]. Jia et al. compared the traffic incident duration prediction performances of a back propagation (BP) neural network, RF, SVM, and long short-term memory (LSTM) methods based on 550 traffic incidents from the Xi'an Ring expressway in China. The results demonstrated that the proposed LSTM model with its attention mechanism obtained superior performance [16]. Their conclusion is consistent with Zhu et al. [17].

In addition to the use of machine learning technologies for predicting the duration of traffic accidents, a growing body of literature has demonstrated that ensemble learning methods can obtain enhanced predictive performance by applying multiple learning algorithms. For example, Ghosh et al. proposed an adaptive ensemble model that can provide reliable predictions even under limited available information. They compared the prediction performance of various traditional regression methods and found that the Treebagger method outperformed the others. For incidents with durations ranging from 36 to 200 min, the mean absolute percentage error (MAPE) in predicting the duration ranged from 25% to 55% [18]. Zhao et al. obtained good prediction accuracy by applying data associated with millions of traffic incidents in the United States in conjunction with eXtreme Gradient Boosting (XGBoost), light gradient-boosting machine (LightGBM), and category boosting (CatBoost) gradient boosting frameworks, and a stacking and elastic network to establish a heterogeneous ensemble learning scheme. Their result shows that the MAPE, MAE, and MSE of heterogeneous ensemble learning are 35.6101%, 30.7432, and 4252.1728 [19]. In Zhao's study, they utilized an ensemble clustering-based approach that combined an artificial neural network (ANN) model and a random forest (RF) model to predict the durations of traffic incidents. Their dataset consisted of 18,462 incidents from Singapore. The results indicate that the ensemble model performed better than both the traditional model with fixed clusters and the classical model without clustering [20]. Grigorev et al. proposed a new intra-extra joint optimization machine learning (IEO-ML) approach and verified that 40–45 min is the best threshold for distinguishing between short-term versus long-term traffic incidents [21]. Lv et al. developed an optimized integrated learning algorithm to estimate the duration of traffic accidents. The temporal stability of the model was evaluated using traffic accident data from Guizhou province over different time periods. Additionally, the spatial stability of the model was assessed by comparing the results obtained from traffic accident data of Guizhou Province and Shandong province during the same period. The results indicate that while the model demonstrates temporal stability, it lacks spatial stability [22].

The above-discussed models developed to predict the duration of traffic incidents uniformly applied structured datasets that employ predefined data fields. While good prediction performance has been obtained, prediction models are inevitably influenced by the source, quantity, and quality of the data employed for model training and testing. We first note that research on the subject has been mostly restricted to an analysis of the impact of the heterogeneity of the predefined fields on the outcomes of prediction models.

Furthermore, we note that studies focused on this topic tend to design an increasing number of predefined data fields with increasingly limited applicability, which makes no obvious sense for practical applications. Moreover, the use of single-mode data involving a single data field or datatype as the object of analysis cannot meet the demands of increasing randomness and complexity. As a result, multi-mode analysis involving multiple data fields or multiple datatypes has generated increasing interest in the prediction of traffic incident durations. Through effective fusion, multi-mode data can provide richer information than single-mode data, and thereby improve the prediction accuracy of models. For example, Li et al. used a stacked restricted Boltzmann machine (RBM) to fuse the features extracted from both traffic accident data and traffic flow data. The results demonstrated that the fused model obtained improved prediction accuracy [23]. At the same time, the information expressed by different modes may complement each other, and thereby enhance model accuracy, or conflicting relations between some data modes may be a detriment to accuracy. This represents one of the main difficulties associated with multimodal data analysis. As a result, most studies have focused on single-mode data only, and the influence of multi-mode data on the accident duration prediction performance of models remains poorly understood due to a lack of quantitative analysis.

In fact, traffic incidence information is itself typical multi-mode data in terms of datatypes and time series, which include a wealth of structured data, text, and video owing to the rapid development of information technology. Hence, accident information is recorded many times at different intervals by an intelligent transportation information system as traffic events evolve.

As discussed, unstructured text data reflecting the status of a traffic incident is also collected by traffic information systems. In addition to many of the factors directly pertaining to traffic incidents, such as time, location, and event type, the evolution of the incident and its clearance status are described in text. Accordingly, text information integrates entire traffic incident cycles of discovery, response, and recovery. Hence, text data contains more comprehensive accident information than the conventional predefined fields in structured data. Moreover, the duration of traffic accidents is affected by a wide range of factors. Accordingly, it is not practical to pre-define all potentially important factors.

All these issues highlight the benefits of applying text data for accident duration prediction. However, as unstructured data, text cannot be directly used as model inputs. This is addressed in the field of natural language processing by transforming the text into numerical vectors through feature extraction. Therefore, the emergence of natural language processing technology has the potential to bring accident duration prediction research to a new level of accuracy. For example, the use of pre-trained models [24] and deep learning models [25,26] has recently provided a new foundation for implementing unstructured text data in the field of natural language processing. Wei conducted an analysis on 400 valid literatures from January 1985 to March 2023, focusing on the factors affecting road traffic accidents in China. The literatures were retrieved from the journal database of China National Knowledge Network (CNKI). The study utilized the visualization function of CiteSpace bibliometric software platform to provide insights into the current situation, future development trends, and hot spots in the field of road traffic accident influencing factors [27]. In another study, Yang et al. examined the current research on highway traffic accidents and the application of new technologies in highway traffic safety. They found differences between conventional traffic accidents and non-conventional accidents in terms of influencing factors mining, severity analysis, and accident prediction [28]. However, few studies have applied natural language technology for the purpose of predicting accident durations. Among the few such published studies, Zhang et al. analyzed and constructed traffic accident knowledge graphs based on four factors, including accident portraits, accident classifications, accident statistics, and accident correlation paths [29]. Later, Guo et al. proposed a study on the automatic text quality analysis of emergency response plans written in Chinese, focusing on process descriptions. The proposed approach extracted message sending tasks, message receiving tasks, and regular tasks. This was achieved through the use of Bi-LSTM-CRF networks, which combined a Conditional Random Fields network with a Bidirectional Long Short-Term Memory network [30].

However, despite the many benefits of applying text data in accident duration predictions, relatively little work has actually been published in this regard. For example, Ji et al. applied ensemble learning using the V-Fisher ordered clustering model to predict accident durations with the records of 4164 expressway accidents in China. The results show that the accuracy of their approach reaches 0.82 [31]. Similarly, Ji et al. predicted the duration of highway accidents based on the text data associated with 969 traffic accident records and social network information, and an average mean absolute percentage error (MAPE) of less than 22% was obtained for accident durations within the range of 0–180 min [32]. Shang et al. established a mixed deep learning model combining latent Dirichlet allocation-BiLSTM (LDA-BiLSTM) networks to predict the duration of traffic accidents based on an analysis of the text data associated with 78,317 traffic incidents [33]. However, the traffic accident durations considered in this work were limited to within 90 min, and far too little attention was given toward capturing multi-mode data via the fusion of structured data and text data. Finally, Chen et al. applied text data fusion and ensemble learning algorithms to predict accident duration based on specific feature variables of text data selected by an RF algorithm. The results show that the MAPE value is stable within 73–75% [34].

As discussed above, relatively little work has been published regarding the application of unstructured text data in accident duration prediction operations, and the influence of multi-mode data on duration prediction remains poorly understood due to a lack of quantitative analysis. Nonetheless, the above discussion clearly demonstrates the considerable benefits of applying text data, and that combinations of features can have a profound impact on the performance of prediction models. The present work addresses these issues by proposing a heterogeneous deep learning architecture employing multi-modal features to improve the accuracy of predictions for traffic accident durations on expressways. Accordingly, this study makes the following contributions.

(1) Unstructured text information pertinent to traffic accidents is extracted statistically to obtain new structured features, and typical pre-trained natural language processing models are also applied to the unstructured text to obtain text data in numerical form as an additional data mode.

(2) A hybrid deep learning approach is applied to build classification models with reduced prediction error. The influence of pre-trained models on the approach is summarized.

(3) Standard structured data is combined with the newly established structured data and numerical text data into five unique data modes to conduct a rigorous analysis of the influence of multi-mode data on the accident prediction performances of a variety of deep learning models.

The proposed method is evaluated using survey data collected from an expressway monitoring system in Shaanxi Province, China. This study provides a new insight into the consistency and differences between prediction models based on multimodal analysis.

The remainder of this paper is organized as follows. Section 2 provides a brief introduction to the prediction methodologies and measures applied herein for evaluating model performance. Section 3 discusses the data sources and the prediction results based on multi-mode data. Section 4 compares and summarizes the numerical results obtained by the proposed method. Finally, Section 5 concludes the paper and provides suggestions for future research.

2. Methodology

2.1. Construction of multi-mode data and prediction model training

This study utilizes text-based input features combined with traditional accident data to develop an ensemble deep learning based accident duration prediction model. The proposed prediction process is based on the heterogeneous ensemble prediction architecture illustrated in Fig. 1, which consists of three main components: (1) extraction of features from the dataset; (2) multimodal feature fusion conducted by combing the various extracted feature variables via simple concatenation; (3) prediction model training.

The standard structured features in the dataset were transferred directly into the multimodal dataset. The feature extraction process for text was divided into two sub-processes. In the first sub-process, text features were extracted using pre-trained models in natural language processing techniques. This is discussed in detail in Subsection 2.2. In the second sub-process, we constructed new structured features based on properties extracted from the text using a one-hot model. These newly established features in the dataset were then transferred directly into the multimodal dataset. The precise nature of the multimodal dataset constructed in the present work is a function of the database employed, and is therefore discussed in detail in Subsection 3.2 after introducing the database.

In the model training process, the numerical vectors of text data and structured features were concatenated in different combinations as six defined modes, which were then employed as inputs for training the proposed accident duration prediction model denoted as a bidirectional gated recurrent unit-convolutional neural network (BiGRU-CNN) model. The model training process was repeated several times, and the model parameters attaining the best results were adopted. The advantages of the proposed method were demonstrated by comparing the obtained prediction results with those obtained using currently well-established models, including an ensemble learning algorithm, deep learning based on LSTM, and deep learning based on gated recurrent units (GRUs). The BiGRU-CNN prediction model is presented in detail in Subsection 2.3.

2.2. Text vectorization

Text vectorization is a process employed for representing text semantics using numerical vectors, which ensures that the textual information in different texts is distinguishable, and facilitates the extraction of features. Hence, text vectorization is an important part of text representation.

Conventional methods applied to extract text vectors employ statistical models, such as one-hot models, bag of words (BOW) models, and the term frequency-inverse document frequency (TF-IDF) metric. However, the obtained text vectors are high-dimensional and sparse. Moreover, BOW models and the TF-IDF metric focus exclusively on the number of words appearing in the text, and therefore ignore contextual correlations. Accordingly, we applied the Word2Vec algorithm pre-trained on a large data corpus
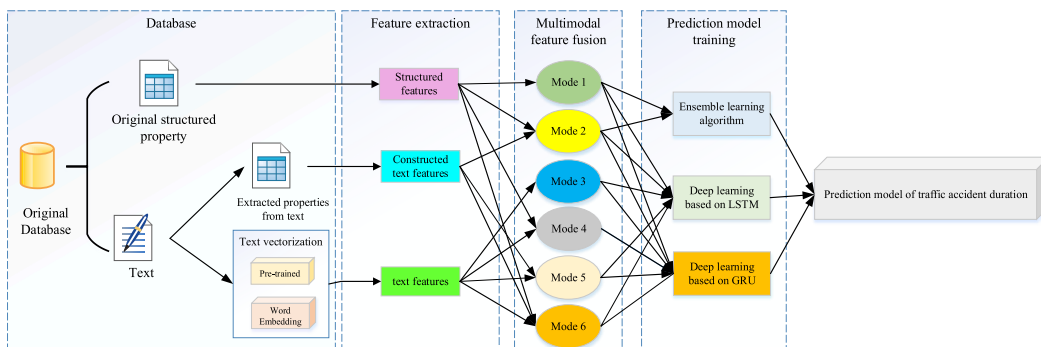

Fig. 1. Heterogeneous ensemble prediction architecture based on multi-mode data and its three main components: (1) feature extraction; (2) multimodal feature fusion; (3) prediction model training.

as a word embedding model for generating vectors of real numbers pertaining to the nature of the words in a text. In addition, Word2Vec can map high-dimensional sparse vectors to low-dimensional dense vectors. The Word2Vec algorithm comprises either continuous BOW (CBOW) or skip-gram network architectures. The CBOW structure has been demonstrated to work well with small databases, while the skip-gram structure works best with large corpora [24]. Therefore, the present work applied the skip-gram architecture for Word2Vec.

Nonetheless, no pre-trained model can effectively take into account reasonable model size, advanced small sample capability, and advanced fine-tuning capability. Hence, we compared the text vectorization results obtained using the Word2Vec algorithm with those obtained by two other pre-trained natural language processing models, including the bidirectional encoder representation from transformers (BERT) model, which uses a deeply bidirectional self-attention mechanism, and the robustly optimized BERT approach (RoBERTa), which represents an improved version of BERT obtained by applying a larger training corpus, using dynamic masking patterns, training with longer text sequences, and omitting the next sentence pre-training objective.

2.3. Prediction model based on BiGRU-CNN

The proposed multi-feature prediction model based on the BiGRU-CNN architecture is illustrated in Fig. 2. As can be seen, the BiGRU-CNN framework consists of four layers, including an input layer with structured data and text data output by the feature extraction module in Fig. 1, a BiGRU layer that obtains contextual semantics features, a CNN layer that obtains local features, and an output layer. As can be seen, the CNN layer consists of a convolution layer, pooling layer, dropout layer, and a fully connected layer, and a softmax activation function is applied in the output layer to the output of the CNN.

The GRU model is a variant of the LSTM model that was developed to solve the vanishing gradient problem that interfered in the training process of recursive neural network (RNN) models employing backpropagation in conjunction with gradient-based learning approaches. Moreover, the GRU network structure is simpler than that of an RNN and with fewer parameters. The structure of a GRU model is illustrated in Fig. 3, which includes a hidden state H_{t-1} at time $t - 1$, and an input X_t and a hidden state H_t at time t, along with sigmoid σ and tanh activation functions. The GRU model is mainly composed of an update gate and a reset gate with values of U_t and R_t at time t, respectively, which are defined as Eq. (1) and Eq. (2).
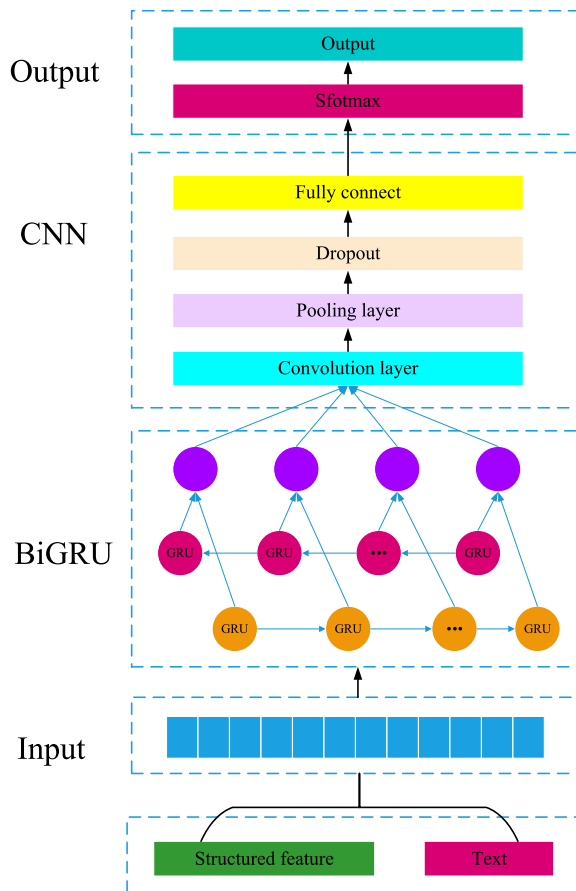

Fig. 2. Schematic illustrating the multi-feature fusion model based on a BiGRU-CNN architecture. BiGRU-CNN, bidirectional gated recurrent unit-convolutional neural network.

$$R_t = \sigma(w_r \cdot [H_{t-1}, X_t] + b_r) \tag{1}$$

$$U_t = \sigma(w_u \cdot [H_{t-1}, X_t] + b_u) \tag{2}$$

Here, the terms $[H_{t-1}, X_t]$ represent the concatenation of vectors H_{t-1} and X_t, w_r and w_u are weight matrices, and b_r and b_u are bias values. The process then yields the candidate set $\tilde{H}$ of the current state and the value of H_t as Eq. (3) and Eq. (4).

$$\tilde{H}_t = \tanh(w_h \cdot [R_t * H_{t-1}] + b_h) \tag{3}$$

$$H_t = U_t * H_{t-1} + (1 - U_t)\tilde{H}_t \tag{4}$$

where ω_h and b_h are the corresponding weight matrix and bias value, respectively. The weight matrices and bias values are determined during the training process.

The BiGRU model consists of forward GRUs and reverse GRUs, which are denoted in Fig. 2 by the orange and magenta circles, respectively. The input is passed to both the first forward GRU, which processes the forward information of the sequence, and the first reverse GRU, which processes the reverse information of the sequence, and the final results are assembled from the outputs of each GRU in both processes. This processing ensures more accurate results. Accordingly, the output $\overrightarrow{H}_t$ of a forward GRU and the output $\overleftarrow{H}_t$ of a reverse GRU are given as Eq. (5) and Eq. (6).

$$\overrightarrow{H}_t = GRU\left(X_t, \overrightarrow{H}_{t-1}\right) \tag{5}$$

$$\overleftarrow{H}_t = GRU\left(X_t, \overleftarrow{H}_{t-1}\right) \tag{6}$$

$$H_t = \left[\overrightarrow{H}, \overleftarrow{H}\right] \tag{7}$$

Here, $GRU(\bullet)$ represents the output of Eq. (4), and the final output is a combination of $\overrightarrow{H}_t$ and $\overleftarrow{H}_t$. As illustrated in Fig. 2, the output vector $H = [H_1, H_2, \cdots, H_t]$ of all GRU sets in the BiGRU model obtained from Eq. (7) serves as the input of the CNN. The final output Y obtained after applying the softmax activation function $soft(\bullet)$ is given as Eq. (8).

$$Y = sfot(W \cdot H + b) \tag{8}$$

where W is the weight matrix and b is the bias.

2.4. Performance evaluation

The prediction of traffic accident durations can be regarded as a classification task involving multiple categories. Based on previous research, we know that the accident samples considered herein have durations between 5 min and 200 min. Accordingly, we can conveniently divide the durations into four categories, including durations with intervals of 5–30 min, 31–60 min, 61–120 min, and 120–200 min. Therefore, the goal of the prediction model is to classify the accident durations into the four categories according to relevant traffic accident features.

We evaluated the performances of the prediction models according to the precision P_j, recall R_j, and F_j metric values of categories $j = 1, 2, 3, 4$, which are calculated as Eq. (9), Eq. (10), and Eq. (11).

$$P_j = \frac{TP_j}{TP_j + FP_j} \tag{9}$$
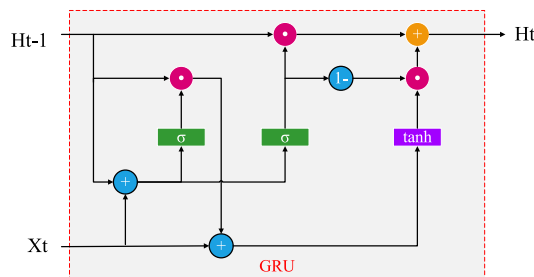

Fig. 3. GRU network structure, which includes a hidden state H_{t-1} at time $t-1$, and an input X_t and a hidden state H_t at time t, along with sigmoid σ and tanh activation functions. GRU, gated recurrent unit.

$$R_j = \frac{TP_j}{TP_j + FN_j} \tag{10}$$

$$F_j = \frac{2P_jR_j}{P_j + R_j} \tag{11}$$

Here, TP_j is the number of true positives, FP_j is the number of false positives, and FN_j is the number of false negatives observed for the j-th category. As can be seen, the prediction performance of an algorithm for a given category increases with increasing values of P_j, R_j, and F_j. However, the different categories suffer from unbalanced sample sizes. Therefore, we adopted a weighted F value defined as Eq. (12).

$$F_w = \sum_{j=1}^{4} w_jF_j \tag{12}$$

where w_j is the proportion of class j in the training dataset.

3. Results

3.1. Multimodal data collection

The survey data employed in this work consisted of traffic accident records recorded in Shaanxi Province from June 2021 to August 2022. As discussed above, the same accident is reported multiple times at each stage of the incident. Obtaining accurate predictions requires survey data collected from the information that was first reported. Then, the accident duration was defined as the difference between the time the accident was cleared and the time the accident was first reported. The original structured data describes information in the form of natural language associated with pre-defined fields, such as accident type, time, weather, and location. The text data is very sparse and heterogeneous, and includes a number of aspects, such as accident descriptions, treatment measures, and empirical predictions.

On the basis of literature research [35], the accident duration employed in the current work was limited to the range of 5–200 min. Therefore, removing outliers and empty values from the survey data resulted in a total of 3887 samples. In addition, statistical descriptions of the traffic accident durations are listed in Table 1. According to the proportions shown in Fig. 4 for accident duration samples with different time intervals observed in the dataset, about 83% of the accidents had durations less than 120 min. Accordingly, the dataset is characterized by considerable unbalance. A resampling method adopted to reduce the implications of class imbalance. After data balancing, five cross-validation techniques was used in model training. The data was divided into five parts. Each part of data was selected as the testing dataset and the remaining of the data was served as the training dataset. Finally, the mean of the five model training results is calculated as the model performance.

3.2. Feature extraction and fusion

The multimodal features observed or extracted from the dataset are listed in Table 2, along with the various modes, expressed as Mode 1, Mode 2, Mode 3, Mode 4, Mode 5, and Mode 6, that include select variables of the features considered, and were designed for various testing purposes. In this regard, the structured data obtained without data fusion in Mode 1 was employed as the baseline mode, Mode 2 fused these feature variables with the newly-established structured feature variables by simple concatenation, Mode 3 was formed of the text features extracted using one of the three pre-trained models, Mode 4 was the result of fusing Mode 3 with the feature variables in Mode 1 via simple concatenation, Mode 5 was the result of fusing Mode 3 with the new structured features via simple concatenation, and Mode 6 was designed to evaluate the benefits of including all feature variables. We note that the impact of the newly established structured features alone can be ascertained by comparing the results obtained using Modes 4 and 6.

3.3. Comparisons based on multimodal features

The benefits of the various feature variables considered on accident duration predictions were evaluated by comparing the weighted F1 scores obtained using the various modes in conjunction with various prediction models. We evaluated the benefits of the conventional feature variables in Modes 1 and 2 using various conventional and more recent methods, including decision tree [14], random forest [11], adaboost [19], gradient-boosted decision tree (GBDT) [34], catboost [19], extra tree [19], XGBoost [19], and the proposed BiGRU [36] and BiGRU-CNN [36] deep learning methods. In addition, the benefits of the feature variables in Modes 3 and 4 were evaluated using deep-learning-based BiLSTM [37], LSTM-CNN, BiLSTM-CNN [34], GRU-CNN, BiGRU, and BiGRU-CNN [36]

Table 1
Statistical descriptions of traffic accident durations in the dataset.

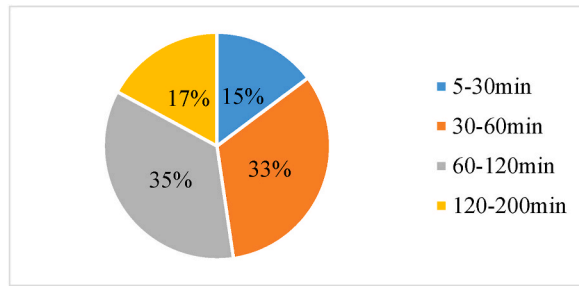| Maximum	Minimum	Average value	Standard deviation	Median	Coefficient of variation (cv)
200	5	73.634	45.376	62	0.616

Fig. 4. Proportion of accident duration samples with different time intervals in the dataset.

Table 2

Definitions of the multimodal features in the dataset.

Feature	Variable	Description	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
Original structured features	Hour	Discovery time in hours from 0 to 23	✓	✓		✓		✓
	Type	Accident type	✓	✓		✓		✓
	Scope	Spatial difference between the starting point and the ending point of the accident	✓	✓		✓		✓
	Injured	Number of people injured when the accident was discovered	✓	✓		✓		✓
	Death	Number of deaths when the accident was discovered	✓	✓		✓		✓
	Vehicles	Number of vehicles in the accident	✓	✓		✓		✓
	Location	Accident location type	✓	✓		✓		✓
	Weather	Weather at moment of accident	✓	✓		✓		✓
	Loss	Whether the road was damaged	✓	✓		✓		✓
New structured features	Dangerous	Vehicles transporting dangerous goods involved in accident, yes or no		✓			✓	✓
	Efftype	Whether the main line road was affected		✓			✓	✓
	Information	Source of information		✓			✓	✓
	Image	Location of accident includes traffic monitoring, yes or no		✓			✓	✓
	Struck	Small truck involved in accident, yes or no		✓			✓	✓
	Ltruck	Large truck involved in accident, yes or no		✓			✓	✓
	Estimatedtime	Estimated duration of accident at time of discovery		✓			✓	✓
	Control	Traffic control is involved in accident, yes or no		✓			✓	✓
	Diversion	Traffic diversion measures are involved in accident, yes or no		✓			✓	✓
	Rep_num	Number of accidents reported		✓			✓	✓
Text features	Pre-trained word vectors	Text vectorization matrix			✓	✓	✓	✓

Table 3

Weighted F1 scores obtained by different prediction models when employing only structured data (Mode 1) and both the structured data and the newly established structured features (Mode 2).

	Mode 2	Mode 1
Decision tree	0.315	0.371
Random forest	0.347	0.359
Adaboost	0.310	**0.426**
GBDT	0.318	0.385
Catboost	0.340	0.396
Extra tree	0.345	0.368
XGBoost	0.304	0.362
BiGRU	0.386	0.326
BiGRU-CNN	**0.527**	0.316

GBDT, gradient-boosted decision tree; XGBoost, eXtreme Gradient Boosting; BiGRU, bidirectional gated recurrent unit; BiGRU-CNN, bidirectional gated recurrent unit-convolutional neural network.

methods. These methods were applied because hybrid deep learning networks based on LSTM and GRU have been demonstrated to be suitable for use with text data, while conventional machine learning methods are not effective in processing high dimensional sparse text vectors. We note here that decision tree, random forest, adaboost, GBDT, catboost, extra tree, and XGBoost are not applicable to text data, and could therefore not be applied for evaluating Modes 3–4. Finally, the benefits of employing a fusion of all feature variables (Mode 5–6) were evaluated for the proposed BiGRU and BiGRU-CNN methods. Accordingly, the results also provided a detailed evaluation of the advantages of the various prediction models considered. In order to obtain convincing results, each model was trained for five times. The mean weighted F1 score was used as the model performance in case of experimental bias.

3.3.1. Predictions obtained using structured features

The weighted F1 scores obtained by the different prediction models (decision tree, random forest, adaboost, GBDT, catboost, extra tree, XGBoost, BiGRU, and BiGRU-CNN) when employing data in Mode 1 and Mode 2 are listed in Table 3. In addition, the results in Table 3 are presented graphically in Fig. 5. (a) and Fig. 5(b).

As can be seen in Fig. 5. (a), the gradient boosting models (adaboost, GBDT, catboost, and XGBoost) performed best for Mode 1, the BiGRU and BiGRU-CNN models exhibit similar and relatively poor performance, while the predictions of the tree and RF models (decision tree, random forest, and extra tree) were intermediate. These results are not surprising when restricting the data employed to structured data only. However, the conditions are altered significantly when combining the structured data with the newly established structured feature variables in Mode 2, as shown in Fig. 5. (b), where we note that the prediction performances of the BiGRU and
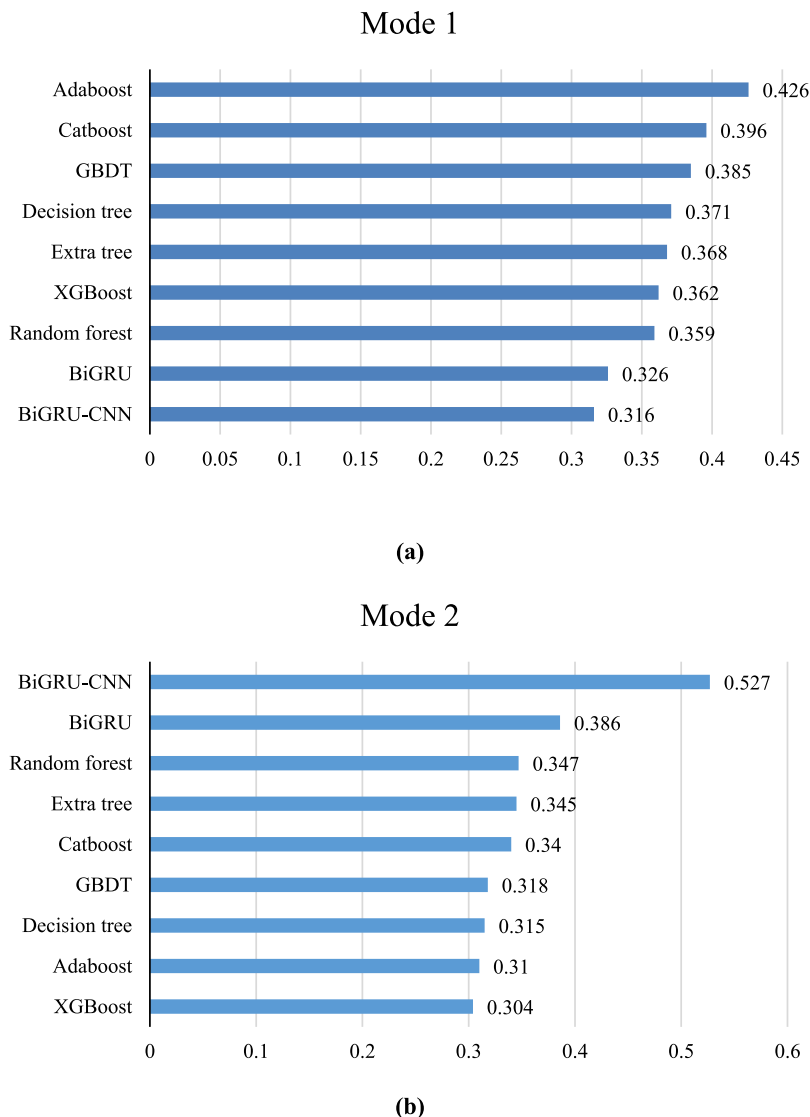

(a)

(b)

Fig. 5. Weighted F1 scores obtained by various prediction models when employing the data for **(a)** Mode 1 and **(b)** Mode 2.

BiGRU-CNN models were best overall, and that of the BiGRU-CNN model was 24% greater than the prediction performance obtained by adaboost for Mode 1. This benefit arises owing to the distinct advantages of the deep learning algorithms for addressing the complexity and higher dimensions arising when including the additional structured features in Mode 2.

3.3.2. Predictions obtained using text features

The weighted F1 scores obtained by the different deep-learning-based prediction models when employing only the text features extracted using each of the Word2Vec, RoBERTa, and BERT pre-trained models in Mode 3 are listed in Table 4. In addition, the results in Table 4 are presented graphically in Fig. 6.

These results present two distinct features of interest. Firstly, we note that using the Word2Vec pre-trained model resulted in very stable performance metrics for all deep-learning-based algorithms considered. In contrast, the RoBERTa and BERT pre-trained models provided wildly varying prediction performances depending on the prediction algorithm applied. Secondly, we note that all prediction algorithms provided their highest average weighted F1 values when extracting text features using the Word2Vec pre-trained model, and that the proposed BiGRU-CNN model obtained the highest average weighted F1 values of all algorithms considered. The result shows that Word2Vec-BiGRU-CNN is a suitable and better model using text features for traffic accident duration prediction.

3.3.3. Predictions obtained using fused features

Table 5 lists the weighted F1 scores obtained by the different deep-learning-based prediction models when fusing the text features extracted using each of the Word2Vec, RoBERTa, and BERT pre-trained models in Mode 3 with the standard structured features in Mode 1 and the newly established structured features. In addition, the results in Table 5 are presented graphically in Fig. 7.

As was observed in Table 5 with respect to Mode 3 data, we note that using the Word2Vec pre-trained model again resulted in much more stable performance metrics for all deep-learning-based algorithms considered than obtained when using the RoBERTa and BERT pre-trained models. We further note that BiGRU-CNN model provided their highest mean weighted F1 values when extracting text features using the pre-trained model. In fact, the BiGRU-CNN model was outperformed by the BiLSTM, LSTM-CNN, BiLSTM-CNN, BiGRU, and GRU-CNN models as well. This indicates that the relationship between the features of text vectors and structured labels can be recognized effectively by the BiGRU-CNN model.

Both Mode 4 and Mode 5 obtained the higher mean weighted F1 score than Mode 3. This indicates that the feature fusion is meaningful in prediction performance. In this case, we note that Mode 5 obtained the highest mean weighted F1 score of all algorithms considered and was about 5% greater than that obtained by Mode 3. In this case, we note that Mode 5 obtained the highest mean weighted F1 score of all algorithms considered and was about 5% greater than that obtained by Mode 3. This implies that the newly established structured features in Mode 5 exhibited significant differences. However, the lowest mean weighted F1 score was obtained in Mode 6. Accordingly, better results were not depend on more features.

3.3.4. BiGRU-CNN model predictions obtained using multimodal features

The weighted F1 scores obtained by the proposed BiGRU-CNN model for each of the six data modes are presented in Fig. 8, where the Word2Vec pre-trained model has been applied uniformly for extracting the data in Mode 3. As can be seen, the highest F1 value is 0.5270, which obtained from Mode 2 by the proposed prediction model. In other words, text feature extracted by the one-hot model provided the best prediction performances based on BiGRU-CNN. However, any inclusion of unstructured text features in either Mode 3, 4, 5 or 6 results in prediction performances that are more or less on par with those obtained when employing only standard structured feature variables in Mode 1. This suggests that the advantage of the BiGRU-CNN model for capturing local features of the experimental dataset is not exploited in the text vectors obtained by the pre-trained Word2Vec model because the one-hot model prediction performance is slightly better even though the pre-trained model provides a more scientific text vectorization method.

4. Discussion

As detailed in Table 3 and Fig. 5, the marked observation to emerge from the data comparison was deep learning algorithms have a greater advantage in prediction accuracy based on Mode 2. Therefore the BiGRU-CNN has the highest F1 value of 0.527. This finding is consistent with Shang et al. [33] who have reported the importance of deep learning algorithms in the duration prediction of traffic accidents by text data. In addition, this study confirms that the newly established structured features extracted from text data by the

Table 4
Weighted F1 scores obtained by different prediction models when employing only the text features extracted by text vectorization using the three pre-trained models (Mode 3).

Model	BiLSTM	LSTM-CNN	BiLSTM-CNN	BiGRU	GRU-CNN	BiGRU-CNN	Mean
Word2vec	0.3371	0.3285	0.3666	0.3332	0.3207	0.3648	**0.3418**
RoBERTa	0.2122	0.3556	0.3448	0.2660	0.3556	0.3795	0.3189
BERT	0.2122	0.3300	0.3349	0.3052	0.2578	0.3500	0.2984
Mean	0.2538	0.3380	0.3488	0.3015	0.3113	**0.3648**	–

LSTM, long short-term memory; BiLSTM, bidirectional LSTM; LSTM-CNN, long short-term memory convolutional neural network; GRU, gated recurrent unit; GRU-CNN, gated recurrent unit convolutional neural network; RoBERTa, robustly optimized BERT approach; BERT, bidirectional encoder representation from transformers.
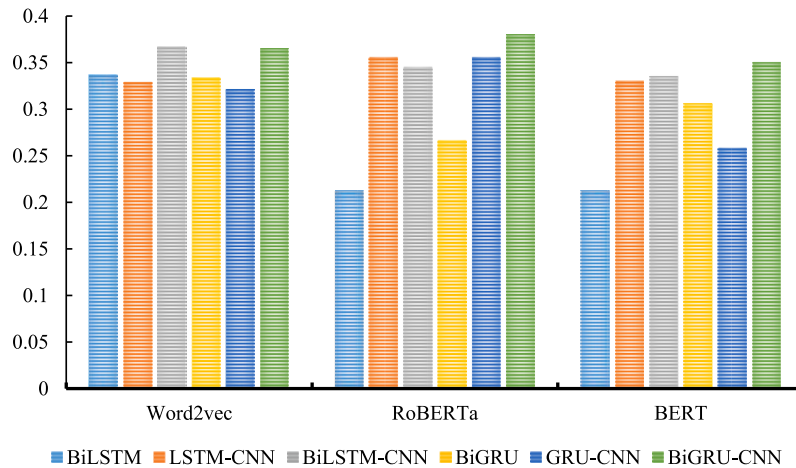
Fig. 6. Weighted F1 scores obtained by various prediction models when employing the data in Mode 3 extracted using different pre-trained models.

Table 5
Weighted F1 scores obtained by different prediction models using fused features.

|		BiLSTM	LSTM-CNN	BiLSTM-CNN	BiGRU	GRU-CNN	BiGRU-CNN	Mean
Mode 3	Word2vec	0.3371	0.3285	0.3666	0.3332	0.3207	0.3648	**0.3197**
	RoBERTa	0.2122	0.3556	0.3448	0.2660	0.3556	0.3795	
	BERT	0.2122	0.3300	0.3349	0.3052	0.2578	0.3500	
Mode 4	Word2vec	0.3501	0.3384	0.3562	0.3373	0.3630	0.3492	**0.3223**
	RoBERTa	0.3268	0.3957	0.3786	0.3178	0.3893	0.3768	
	BERT	0.3440	0.2202	0.2122	0.3224	0.2122	0.2122	
Mode 5	Word2vec	0.3597	0.3536	0.3611	0.3458	0.3230	0.3548	**0.3370**
	RoBERTa	0.3067	0.4094	0.4213	0.3037	0.4143	0.3868	
	BERT	0.2780	0.2122	0.2656	0.3194	0.3628	0.2879	
Mode 6	Word2vec	0.3513	0.3425	0.3727	0.3548	0.3639	0.3426	**0.2661**
	RoBERTa	0.3246	0.1893	0.1893	0.2968	0.1893	0.1893	
	BERT	0.1922	0.2570	0.1893	0.2673	0.1893	0.1893	
Mean		**0.2996**	**0.3110**	**0.3160**	**0.3141**	**0.3118**	**0.3153**	–
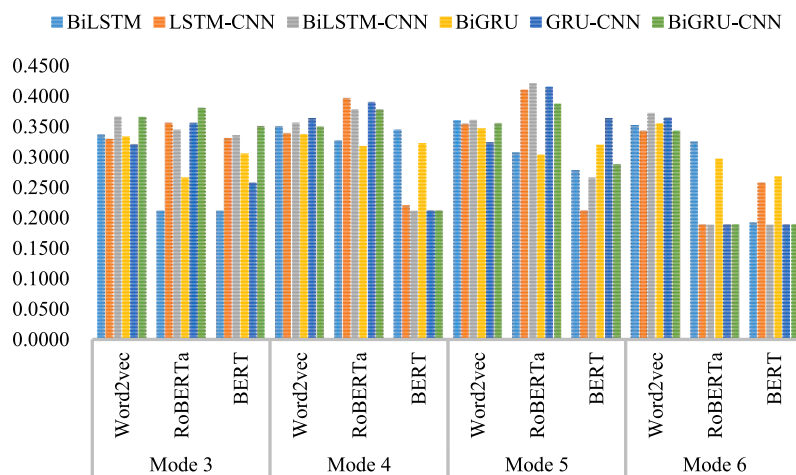
Fig. 7. Weighted F1 scores obtained by various prediction models when employing the data with text data extracted using different pre-trained models.
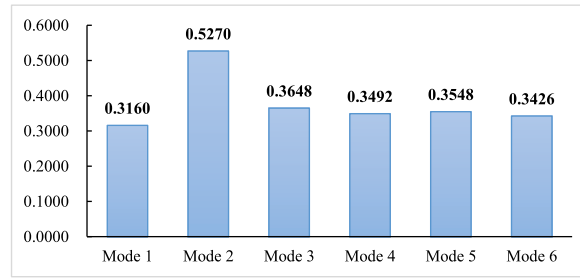
Fig. 8. Weighted F1 scores obtained by BiGRU-CNN model for different data modes, where the Word2Vec pre-trained model has been applied uniformly for extracting the data in Mode 3.

one-hot model substantially enhance the prediction effects of deep learning algorithms. However, these new features were a detriment to the prediction effects of conventional machine learning algorithms. These results corroborate that deep learning algorithms have been demonstrated to perform more effectively than conventional machine learning algorithms when working with complex, high-dimensional, and random data. Furthermore, meaningful features can be captured by deep learning models when working with multi-mode data.

As discussed, one of the important parameters of the current study was to assess the importance of text vectorization when working with unstructured text data. In Table 4, from the comparison of the mean F1 value, it can be seen that the Word2Vec performs better than the RoBERTa and BERT base on multiple deep learning models, which has highest F1 value of 0.3418. The most obvious finding to emerge from the analysis is that the Word2Vec pre-trained model provides text vectorization results that best support the performances of various deep-learning-based prediction algorithms operating downstream. This finding was also reported by Chen et al. [35]. In addition, as shown in Table 4, the highest mean F1 value for three pre-trained models is 0.3648, which obtained by the BiGRU-CNN. We can concluded that the proposed BiGRU-CNN model provided the best prediction performance when working with text data in term of Mode 3.

The results reported herein also broadly support the work of other studies demonstrating that the processing and extraction of text features is a complex issue in the field of traffic accident duration prediction [11,30,32]. This follows from the surprising results demonstrating that the prediction performances obtained by deep-learning-based models when employing data modes that included unstructured text data extracted using pre-trained models (i.e., Modes 3–6) were generally not better than the performances of conventional machine learning methods using only standard structured feature variables (i.e., Modes 1 and 2). Therefore, deep learning is more like a balance between accuracy, efficiency and interpretability. An inspection of the data in Table 5 reveals that more features do not benefit for prediction performance directly. Several factors could explain this observation. Firstly, the consistency between structured data and textual data has not been assessed. Thus it is necessary to evaluate the quality and consistency among multimodal data. Secondly, the introduction of too many input features will increase the complexity of the system. This F1 value is slightly lower than the value we expected and there is certainly room for improvement. However, this result has further strengthened our confidence in evaluating the prediction accuracy under different set of input features. Our study was successful in proving that the duration prediction of traffic accident can be improved using data fusion and deep learning models.

That the newly established structured features extracted from text data by the one-hot model substantially enhanced the prediction effects of deep learning algorithms may be explained by the fact that the textual descriptions of accidents given in the traffic management system are simple and regularized. As a result, the strengths of the more advanced RoBERTa and BERT pre-trained models were not utilized effectively in the vectorization of unstructured text data. This suggests that the text descriptions of traffic accidents employed by traffic management systems should be more diversified to record and capture potential influencing factors.

5. Conclusion

The present work proposed a heterogeneous deep learning architecture employing multi-modal features to improve the accuracy of predictions for traffic accident durations on expressways. Multi-modal features was created by combining standard structured data with unstructured text data related to traffic accidents. In addition, the influence of multi-mode data on the accident prediction performances of deep learning models was evaluated rigorously. The proposed method was evaluated using survey data collected from an expressway monitoring system in Shaanxi Province, China. The results confirmed that the newly established structured features extracted from text data substantially enhance the prediction effects of deep learning algorithms, but these new features were a detriment to the prediction effects of conventional machine learning algorithms. The results further demonstrated that the pre-trained Word2Vec model provides numerical text results that best support the performances of the deep-learning-based prediction algorithms considered. Moreover, the proposed BiGRU-CNN model provided the best prediction performance with the highest mean-F1 value when working with text data. Our results clearly augment the rapidly expanding field of traffic safety. Furthermore, the results demonstrated that the processing and extraction of text features is a complex issue in the field of traffic accident duration prediction. Finally, the present work considered only simple concatenation as the fusion method. Future work will investigate the impact of multi-mode data formed using additional multimodal data fusion methods on prediction accuracy.

Funding statement

Data availability statement

The data used in this study are available from the corresponding author upon request.

CRediT authorship contribution statement

Jiaona Chen: Writing – review & editing, Writing – original draft, Supervision. **Weijun Tao:** Methodology, Formal analysis, Data curation. **Zhang Jing:** Data curation, Formal analysis. **Peng Wang:** Data curation, Formal analysis. **Yinli Jin:** Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interestsChen jiaona reports financial support was provided by National Natural Science Fundation for China. Jin Yinli reports financial support was provided by National Key Research and Development Plan. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] K.G. Bardal, F. Jørgensen, Valuing the risk and social costs of road traffic accidents – seasonal variation and the significance of delay costs, Transport Pol. 57 (2017) 10–19, https://doi.org/10.1016/j.tranpol.2017.03.015.

[2] K. Koppiahraj, S. Bathrinath, S.A. Mithun, On sustainable predictive maintenance: exploration of key barriers using an integrated approach, Sustain. Prod. Consum. 2021 (27) (2021) 1537–1553.

[3] R.M. Li, Traffic incident duration analysis and prediction models based on the survival analysis approach, IET Intell. Transp. Syst. 9 (4) (2015) 351–358, https://doi.org/10.1049/iet-its.2014.0036.

[4] R.M. Li, F.C. Pereira, M.E. Ben-Akiva, Competing risks mixture model for traffic incident duration prediction, Accid. Anal. Prev. 75 (2015) 192–201, https://doi.org/10.1016/j.aap.2014.11.023.

[5] B. Wali, A.J. Khattak, J. Liu, Heterogeneity assessment in incident duration modelling: implications for development of practical strategies for small & large scale incidents, J. Intelligent Transportat. Syst. 26 (5) (2022) 586–601, https://doi.org/10.1080/15472450.2021.1944135.

[6] B. Yu, Y.T. Wang, J.B. Yao, J.Y. Wang, A comparison of the performance of ANN and SVM for the prediction of traffic accident duration, Neural Netw. World 26 (3) (2016) 271–287, https://doi.org/10.14311/nnw.2016.26.015.

[7] W. Yuan, S.C. Yan, Q.X. Yuan, R.B. Han, S.Y. Chen, Traffic incident duration prediction based on K-nearest neighbor, Appl. Mech. Mater. 253–255 (253–255) (2012) 1675–1681.

[8] Y. Lee, C.H. Wei, K.C. Chao, Non-parametric machine learning methods for evaluating the effects of traffic accident duration on freeways, Archiv. Transport 43 (3) (2017) 91–104, https://doi.org/10.5604/01.3001.0010.4228.

[9] Z. Gao, A.X. Ke, R.J. Yu, X.S. Wang, Urban expressway traffic incident duration prediction based on random survival forests, J. Tongji Univ. Nat. Sci. 45 (9) (2017) 1304–1310, https://doi.org/10.11908/j.issn.0253-374x.2017.09.008.

[10] R. Li, F.C. Pereira, M.E. Ben-Akiva, Overview of traffic incident duration analysis and prediction, European Transport Res. Rev. 10 (2) (2018) 1–13.

[11] K. He, S.X. Yang, Y.G. Gao, Prediction of traffic incident duration in tunnels based on a PCA-RF combined model, J. Transport Informat. Safe. 37 (5) (2019) 26–32, https://doi.org/10.3963/j.issn.1674-4861.2019.05.004.

[12] Q. Shang, D.R. Tan, S. Gao, L.L. Feng, A hybrid method for traffic incident duration prediction using BOA-optimized random forest combined with neighborhood components analysis, J. Adv. Transport. (2019) 4202735, https://doi.org/10.1155/2019/4202735, 2019.

[13] K. Hamad, R. Al-Ruzouq, W. Zeiada, S. Abu Dabous, M.A. Khalil, Predicting incident duration using random forests, Transportmetrica: Transport. Sci. 16 (3) (2020) 1269–1293, https://doi.org/10.1080/23249935.2020.1733132.

[14] A. Saracoglu, H. Ozen, Estimation of traffic incident duration: a comparative study of decision tree models, Arabian J. Sci. Eng. 45 (10) (2020) 8099–8110, https://doi.org/10.1007/s13369-020-04615-2.

[15] B. Ghosh, J. Dauwels, Comparison of different Bayesian methods for estimating error bars with incident duration prediction, J. Intelligent Transportat. Syst. 26 (4) (2021) 420–431, https://doi.org/10.1080/15472450.2021.1894936.

[16] X.L. Jia, S.Q. Li, H.Z. Yang, X.P. Chen, Prediction of the duration of freeway traffic incidents based on an ATT-LSTM model, J. Transport Informat. Safet. 40 (5) (2022) 61–69, https://doi.org/10.3963/j.issn.1674-4861.2022.05.007.

[17] W. Zhu, J. Wu, T. Fu, J. Wang, J. Zhang, Q. Shangguan, Dynamic prediction of traffic incident duration on urban expressways: a deep learning approach based on LSTM and MLP, J. Intelligent Connected Vehicles 4 (2) (2021) 80–91.

[18] B. Ghosh, M.T. Asif, J. Dauwels, U. Fastenrath, H. Guo, Dynamic prediction of the incident duration using adaptive feature set, IEEE Trans. Intell. Transport. Syst. 20 (11) (2019) 4019–4031, https://doi.org/10.1109/tits.2018.2878637.

[19] Y.X. Zhao, W. Deng, Prediction in traffic accident duration based on heterogeneous ensemble learning, Appl. Artif. Intell. 36 (1) (2022), https://doi.org/10.1080/08839514.2021.2018643.

[20] H. Zhao, W. Gunardi, Y. Liu, C. Kiew, T.H. Teng, X.B. Yang, Prediction of traffic incident duration using clustering-based ensemble learning method, J. Transport. Eng., Part A: Systems 148 (7) (2022) 04022044, https://doi.org/10.1061/jtepbs.0000688.

[21] A. Grigorev, A.-S. Mihaita, S. Lee, F. Chen, Incident duration prediction using a bi-level machine learning framework with outlier removal and intra–extra joint optimisation, Transport. Res. C Emerg. Technol. 141 (2022) 103721, https://doi.org/10.1016/j.trc.2022.103721.

[22] L. Lv, J. Li, Z.Y. Guo, Y. Yan, C. Gao, Study on calculation method of expressway accident duration, J. Highw. Transp. Res. Dev. 39 (12) (2022) 155–162.

[23] L.C. Li, X. Sheng, B.W. Du, Y.G. Wang, B. Ran, A deep fusion model based on restricted Boltzmann machines for traffic accident duration prediction, Eng. Appl. Artif. Intell. 93 (2020) 103686, https://doi.org/10.1016/j.engappai.2020.103686.

[24] X.P. Qiu, T.X. Sun, Y.G. Xu, Y.F. Shao, N. Dai, X.J. Huang, Pre-trained models for natural language processing: a survey, Sci. China Technol. Sci. 63 (10) (2020) 1872–1897, https://doi.org/10.1007/s11431-020-1647-3.

[25] D.W. Otter, J.R. Medina, J.K. Kalita, A survey of the usages of deep learning for natural language processing, IEEE Transact. Neural Networks Learn. Syst. 32 (2) (2020) 604–624, https://doi.org/10.1109/tnnls.2020.2979670.

[26] Z. Zhang, Q. He, J. Gao, M. Ni, A deep learning approach for detecting traffic accidents from social media data, Transport. Res. C Emerg. Technol. 86 (2018) 580–596.

[27] M. Wei, Bibliometric and visual knowledge garph analysis of road taffic accident influencing factors in cina based on CiteSpace, Saf. Environ. Eng. 1–16 (2023), https://doi.org/10.13578/j.cnki.issn.1671-1556.20230423 [2023-12-19].

[28] Y. Yang, W.H. Wang, X.Y. Wu, Y.P. Wang, Review of the research toward freeway unconventional traffic accidents, J. Basic Sci. Eng. 1–32 (2023) [2023-12-19], http://kns.cnki.net/kcms/detail/11.3242.TB.20231123.0950.002.html.

[29] L.Y. Zhang, M. Zhang, J.Z. Tang, J. Ma, X.K. Duan, J. Sun, X.F. Hu, S.C. Xu, Analysis of traffic accident based on knowledge graph, J. Adv. Transport. (2022) 3915467, https://doi.org/10.1155/2022/3915467, 2022.

[30] W.Y. Guo, Q.T. Zeng, H. Duan, W.J. Ni, T. Liu, C. Liu, N.F. Xie, Text quality analysis of emergency response plans, IEEE Access 8 (2020) 9441–9456, https://doi.org/10.1109/access.2020.2964710.

[31] K.K. Ji, J. Chen, S.Y. Xiao, X.Y. Wang, Y.X. Liu, Z.Y. Fu, A predictive model of highway accident duration driven by text data, J. Transport Informat. Safet. 38 (6) (2020) 9–16, https://doi.org/10.3963/j.jssn.1674-4861.2020.06.002.

[32] K.K. Ji, Z.Z. Li, J. Chen, G.Y. Wang, K.L. Liu, Y. Luo, Freeway accident duration prediction based on social network information, Neural Netw. World 32 (2) (2022) 93–112, https://doi.org/10.14311/nnw.2022.32.006.

[33] Q. Shang, T. Xie, Y. Yu, Prediction of duration of traffic incidents by hybrid deep learning based on multi-source incomplete data, Int. J. Environ. Res. Publ. Health 19 (17) (2022) 10903, https://doi.org/10.3390/ijerph191710903.

[34] J.N. Chen, W.J. Tao, Traffic accident duration prediction using text mining and ensemble learning on expressways, Sci. Rep. 12 (1) (2022) 21478, https://doi.org/10.1038/s41598-022-25988-4.

[35] J. Chen, W. Tao, Y. Jin, P. Wang, J. Zhang, Prediction of traffic incident duration on expressway based on multimodal text information, J. Safety Sci. Technol. 19 (6) (2023) 180–187.

[36] L. Duarte Soares, A. de Souza Queiroz, G.P. López, E.M. Carreño-Franco, J.M. López-Lezama, N. Muñoz-Galeano, BiGRU-CNN neural network applied to electric energy theft detection, Electronics 11 (5) (2022) 693, https://doi.org/10.3390/electronics11050693.

[37] G.L. Dai, J. Zhang, X. Han, A novel attention-based BiLSTM-CNN model in valence-arousal space, Int. J. Perform. Eng. 18 (12) (2022) 833–843, https://doi.org/10.23940/ijpe.22.12.p1.833843.