Severity Prediction Scheme for Temporal Dependency Learning in Traffic Accidents

Dodo, Simon Numduck
Department of Computer Science,
Federal University of Technology,
Minna, Nigeria.
0009-0004-3267-6352
dodosnmte258@st.futminna.edu.ng

Muhammad, Kudu Muhammad
Department of Computer Science,
Federal University of Technology,
Minna, Ngeria.
0000-0001-8568-0743
muhammad_kudu@futminna.edu.ng

Sulaimon, Adebayo Bashir
Department of Computer Science,
Federal University of Technology,
Minna, Nigeria
bashirsulaimon@futminna.edu.ng

2025 International Conference on Electrical and Computer Engineering Researches (ICECER) | 978-1-6654-5756-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICECER65523.2025.11401288

***Abstract*— This study presents a novel hybrid deep learning framework, Distil-BERT-CNN, designed to optimize spatial pattern extraction and temporal dependency learning for predicting traffic accident severity. However, this hybridization is achieved by integrating the lightweight, attention-based Distil-BERT model for efficient sequence processing with Convolutional Neural Networks (CNN) for robust spatial feature extraction, the proposed approach addresses the limitations of traditional statistical and machine learning models, which often struggle with class imbalance and complex, nonlinear relationships in real-world traffic data. Extensive experiments on multi-source datasets, including historical accident records, real-time traffic flow, and weather data, demonstrate that Distil-BERT-CNN significantly improves precision and recall, particularly for minority accident severity classes, compared to existing methods such as Random Forest, Gradient Boosting, LSTM, and cascade deep learning models. The enhanced predictive performance of this framework has the potential to support traffic management authorities in proactive accident prevention, resource optimization, and rapid emergency response, ultimately contributing to safer and more efficient transportation systems.**

***Keywords— Distill-BERT, CNN, Accident, traffic, deep-learning, spatial pattern*.**

I. INTRODUCTION

Due to the exponential increase in the development of major cities across the globe, road traffic accidents continue to be a major problem worldwide, resulting in many mortalities, injuries, and financial losses every year [1]. However, Recent reports revealed that about 1.19 million people have lost their lives every year due to road crashes, and tens of millions more are injured, with many suffering from disabilities that can last a lifetime. These incidents not only bring personal tragedy to families but can result in large economic burdens for individuals, communities, and countries, as they lead to high medical costs and loss of productivity. Considering the serious impact of road accidents on the public, there is an urgent need to find a better way to reduce the frequent occurrence of road accidents and how serious they are [2]. Researchers have delved into a series of possible solutions, such as the use of advanced computer models that can predict when and where accidents can occur, as well as how much they will disrupt traffic. These real-time predictions help authorities to take quick action and manage road traffic more effectively. The capability to forecast traffic accidents and their possible effects plays a crucial role in enhancing public safety and guiding urban development [1]. However, when predictions are accurate, traffic authorities can take proactive steps such as adjusting speed limits dynamically, redirecting traffic away from dangerous zones, and increasing surveillance in areas prone to a high rate of accidents [3]. This not only helps prevent crashes but also enables emergency services to respond more quickly, reducing their arrival times and improve the chance of saving more lives. Furthermore, by understanding how accidents disrupt traffic flow, authorities can better manage congestion and reduce the chances of follow-up collisions caused by sudden traffic slowdowns [2].

Traditional methods for predicting traffic accidents have mainly depended on statistical techniques and basic machine learning algorithms. These approaches primarily analyze past accident records, traffic volumes, weather conditions, and road features to identify risk factors and estimate the likelihood of car crashes [4]. Although these methods have provided useful information, with just simple, linear relationships, and also they fails to capture complex, nonlinear interactions that occur in real-world traffic systems. Additionally, many conventional models find it challenging to simultaneously consider how accidents change over time (temporal dynamics) and vary across different locations (spatial dependencies), both of which are important for accurately understanding and predicting traffic accident patterns [5]

Recently, deep learning techniques have revolutionized many fields, including computer vision, natural language processing, and time-series forecasting [6]. These techniques have also shown great potential in traffic accident prediction due to their ability to dynamically and automatically learn hierarchical feature representations from a large amount of complex datasets. Among the various deep learning architectures, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have emerged as the most highly effective techniques for handling spatial and temporal data, respectively [6].

CNNs are designed to extract spatial features by applying convolutional filters over a given input data, making them the best for analyzing images, sensor grids, and spatial maps of traffic conditions [7]. However, CNNs can process traffic camera images or spatially distributed traffic sensor data to detect patterns indicative of accident risks. On the other hand, RNNs, especially Long Short-Term Memory (LSTM) networks, are capable of modeling sequential data and capturing long-term dependencies in time-series data. This makes LSTMs ideal for analyzing traffic flow over time and predicting the future occurrences of traffic based on past historical trends [8].

Moreover, Recurrent Neural Networks (RNNs) are a class of deep learning models designed to process sequential data by maintaining a form of memory of previous inputs [9]. This internal working mechanism makes them suitable for traffic-related applications involving time-series data, such as vehicle counts, speeds, or accident occurrences over time [9]. RNNs have a feedback loop mechanism that allows information to persist, enabling the network to capture temporal dependencies [10]. Additionally, traditional RNNs suffer from vanishing gradient problems, reducing their ability to learn long-term

978-1-6654-5756-9/25/$31.00 ©2025 IEEE

Authorized licensed use limited to: ULAKBIM UASL - Trakya Universitesi. Downloaded on February 26,2026 at 20:40:07 UTC from IEEE Xplore. Restrictions apply.

dependencies. In order to overcome this, an improved RNN, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRU) was developed. These architectures use gating mechanisms to control information flow, allowing the network to remember or forget information as required. In traffic management, RNNs and their variants are widely used for predicting traffic flow, congestion, and accident risks based on historical data or records [10]. LSTM models have demonstrated strong performance in forecasting traffic speeds and volumes, enabling fast traffic control measures. RNNs also support driver behaviour modeling and anomaly detection, contributing to improved safety systems. While powerful, RNNs require careful tuning and substantial data for effective training [11].

Distillation Bidirectional Encoder from Transformer (Distill-BERT) is a streamlined and efficient variant of the BERT (Bidirectional and Auto-Regressive Transformers) model, specifically designed to handle sequence-to-sequence tasks such as text summarization, translation, and generation [12]. BART itself merges the advantages of BERT's bidirectional encoding and GPT's autoregressive decoding, enabling it to produce coherent and contextually accurate text outputs. The introduction of DistilBERT involves a process called knowledge distillation, where a smaller "student" model learns to replicate the behavior of a larger "teacher" model (BERT). This results in a model that maintains much of the original BERT's performance while significantly reducing the number of parameters, leading to faster inference and lower computational demands [13]. Due to these efficiencies, DistilBART is well-suited for applications requiring rapid processing and limited resources, such as real-time systems or deployment on edge devices. In the domain of traffic accident impact prediction, DistilBART can be utilized to effectively process and generate textual information related to traffic incidents, for example, by summarizing detailed traffic reports, creating concise accident descriptions, or translating complex sensor data into formats easily understood by traffic operators. This capability facilitates faster comprehension of accident severity and conditions, thereby supporting more timely and informed decision-making [14].

Hybrid networks combine different deep learning architectures to leverage their complementary strengths. A common approach is to integrate CNNs for spatial feature extraction with RNNs or LSTMs for temporal sequence modeling [15]. In traffic applications, hybrid models analyze spatial data such as traffic images or road network maps alongside temporal traffic flow or accident data. CNN layers can process video frames from traffic cameras, while LSTM layers capture temporal changes in traffic patterns [16]. This combined approach improves prediction accuracy and robustness in complex traffic environments. Hybrid networks are increasingly used in intelligent transportation systems for tasks such as congestion prediction, accident risk assessment, and autonomous driving [15]. They offer a comprehensive understanding of traffic dynamics by simultaneously modeling spatial and temporal dependencies. Challenges include increased model complexity and training time, requiring large multimodal datasets and substantial computational power. However, their superior performance makes hybrid networks a promising direction for future traffic analysis research [15]. Recently CNNs and RNNs individually excel at capturing spatial and temporal features, traffic accident prediction often requires a model that can simultaneously handle both dimensions. Hybrid models that combine CNN and RNN architectures have been proposed to leverage the strengths of each. In such models, CNN layers typically extract spatial features from input data, which are then fed into RNN layers to model temporal dependencies [16]. For instance, a deep learning framework might first use CNNs to analyze spatial traffic data, such as road network layouts, traffic density maps, and environmental conditions, to identify localized risk factors. Subsequently, LSTM layers process the extracted spatial features across time steps to predict the likelihood of accidents occurring in the near future. This integrated approach allows for more accurate and robust predictions compared to using either CNN or RNN alone [15].

The proposed study introduces a novel framework, Distil-BERT-CNN, to optimize spatial pattern extraction and temporal dependency learning for traffic accident severity prediction. Distil-BERT, a streamlined version of the Bidirectional Encoder Representations from Transformers (BERT), efficiently processes sequence data such as accident reports or weather descriptions.

A. Research Problem

Traffic accidents remain a critical public health and safety issue worldwide, leading to substantial loss of life and property [1]. According to the World Health Organization, road traffic injuries are the leading cause of death for children and young adults aged 5 to 29 years, with approximately 1.19 million fatalities and 20 to 50 million non-fatal injuries occurring annually. These incidents also impose a significant economic burden on individuals, families, and nations, arising from medical expenses and lost productivity, and cost most countries about 3% of their gross domestic product [17], [18].

Public safety is highly challenged by Traffic accidents, leading to many fatalities, and a poor impact on the economy for several years. A study on the parameters that result in traffic accidents is essential in addressing issues such as traffic congestion, injury rate, fatality rate, and similar road incidents. [19] identify that there are several real-time forecasting models to predict the impacts of accidents using various publicly available datasets (accident dataset, traffic flow data, weather-based data and the likes), but are always associated with two main issues including (1) relies majorly on non-real time data, and (2) The developed models lack comprehensive metric to properly measure post-accident impact.

Furthermore, [20] specify that hybridizing a large language-based model with a deep learning model can significantly boost pattern recognition from past accident cases. Additionally, traffic management and decision-making can be enhanced by integrating real-time traffic data with historical data to predict potential accident risks [20]. Lastly, existing studies utilizing recurrent-based networks in conjunction with order deep learning models, such as CNN, may face issues such as overfitting and insufficient generalization when handling high-dimensional and large-scale datasets [19], [20], [21]. Hence, this study proposes a more efficient, lightweight DistillBERT-CNN architecture capable of handlzx`ing high-dimensional data and retaining high-quality temporal dependencies using an attention-based mechanism.

B. Research Goal

This study aims to develop an Optimization of Spatial Pattern and Temporal Dependency Learning for Traffic Accident Severity Prediction Using Distil-Bert-CNN.

C. Contribution to Knowledge.

The significance of this study lies in its potential to enhance the accuracy and efficiency of traffic flow and accident severity predictions, which are critical for improving road safety and reducing the human and economic costs associated with traffic accidents. However, leveraging advanced machine learning models, such as combining convolutional neural networks with lightweight transformer-based architectures, this research aims

Authorized licensed use limited to: ULAKBIM UASL - Trakya Universitesi. Downloaded on February 26,2026 at 20:40:07 UTC from IEEE Xplore. Restrictions apply.

to provide more reliable and timely insights into traffic conditions and accident risks. These improvements can support traffic management authorities in making informed decisions, optimizing resource allocation, and implementing proactive measures to prevent accidents and alleviate congestion. Ultimately, the study contributes to safer transportation systems, reduced fatalities and injuries, and minimized economic losses, promoting public health and well-being.

II. Related Works

This section introduces the literature on traffic accident severity by highlighting different prediction approaches and challenges in the area of traffic accident severity, particularly focused on the deep learning approach.

[22] Azhar and other researchers revealed that road transportation can result into significant fatalities and economic losses globally, with accidents often linked to complex factors. Traditional accident prediction models have limitations in accuracy and data handling, especially given the vast volume of social media data like tweets, which contain valuable geo-tagged information for real-time accident detection. Managing and extracting meaningful insights from such large datasets remains challenging, and existing approaches are still in their infancy. The authors adopt a deep learning accident prediction model that combines information extracted from tweet messages with extended features like sentiment analysis, emotions, weather, geo-coded locations, and time information. The results show an 8% improvement in detection accuracy, reaching 94%, and outperform existing models by 2-3%. The approach also addresses computational challenges associated with big data, demonstrating its potential for enhancing traffic safety and congestion management in the domain of automatic ontology construction.

[23] Introduces that traffic accidents disrupt traffic flow and cause significant harm, but traditional detection methods depend strongly on physical sensors and infrastructure. Social media, especially Twitter, offers a vast, real-time source of user-generated information that could enhance accident detection. However, extracting meaningful and accurate accident-related data from noisy, unstructured tweets is difficult. Additionally, there are also issues like location, time biases, influence of user behavior, and hashtags, which complicate detection accuracy. The researchers proposed two deep learning models, which include the Deep Belief Network (DBN) and Long Short-Term Memory (LSTM). The experimental results showed that the proposed DBN model attained an overall accuracy of 85% using 44 individual and 17 paired token features, outperforming SVM and sLDA models. Validation also showed that about 66% of accident-related tweets matched official accident logs, and over 80% correlated with abnormal traffic conditions nearby.

[8] proposed the I3D-CONVLSTM2D model, which combines RGB video frames and optical flow data to capture spatial-temporal patterns. It uses a modified Inflated 3D ConvNet (I3D) backbone with trainable layers and ConvLSTM2D blocks to analyze motion and spatial features. According to the researchers the proposed methodology is introduced to address the issues of road traffic accidents which remain a global safety challenge, causing significant fatalities, economic losses, and also traditional detection methods struggle with dynamic scenarios, data imbalances, and diverse accident types (e.g., rear-end, T-bone collisions) in smart cities. The model achieved an 87% Mean Average Precision (MAP), outperforming existing methods. Challenges included data imbalances across road structures and traffic scenarios, which affected generalization.

III. Research Methodology

This section outlines the research methodology, covering the data sources and collection process, a review of existing research gaps, system analysis and design including the framework and system flow, the proposed algorithmic flow of the system, the materials and tools used for implementation, an illustration of the proposed architecture, and the methods considered in this study for performance evaluation.

A) Data Source

This research focuses on Los Angeles County, which is noted for having the highest incidence of traffic accidents in the United States. To comprehensively analyze the factors contributing to accidents, four key datasets are utilized including, accident dataset, a congestion dataset, a point of interest (POI) dataset, and a weather dataset. These datasets provide a multi-faceted view of the conditions and elements that influence traffic accidents in the study area. By integrating these diverse sources of information, the study aims to develop a robust understanding of accident dynamics and improve predictive modeling capabilities. The link to the utilized datasets and their features is provided below.

Link 1 (Dataset): https://smoosavi.org/datasets/
Link 2 (Dataset Features Description):
https://smoosavi.org/datasets/us_accidents
Link 3: (Kaggle source):
https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents

B) Research Design

The system design entails various diagrams describing the technological working components of the proposed system using conceptual diagrams, data flow charts, and system algorithmic representation. The following sub section fully describe in detail the entire research design, starts the with system framework design

1) Proposed DistillBERT-CNN Architecture

The proposed system architecture design is shown in figure 1, using multiple data sources are collected and merged into a single multi-data, and this data includes the Weather Condition Dataset, Traffic Congestion Dataset, Accident Dataset, and POI Dataset. The collected dataset is imported into the Kaggle Jupyter notebook Lab IDE for conducting the experimental process. The first operation conducted on the dataset is exploration and visualization, which entails manual analysis conducted on the data to understand a series of metadata about the dataset. Some of the explored data are visualized to get a more comprehensive view of the dataset. Next is forwarding the dataset for the preprocessing operation, which operation surface based on the exploration conducted on the dataset. The first preprocessing operation includes filling the missing value or null value using K-Nearest Neighbor techniques (KNN). Next is the handling of outliers, data balancing, feature reduction, and lastly, extracting the time step sequence from the raw dataset. the final version of the preprocessed data is split into training and testing samples, and the training sample is fed as input into the CNN architecture's input layer. Apart from the CNN input layer, other layers include the convolution layer, a 1-dimensional layer, and lastly the flatten layer. The flattened features are now used to further extract spatial features across each time-step, serving as input features for the DistillBERT architecture. The output of the DistillBERT training becomes the Traffic Accident Prediction Model, which is evaluated using metrics such as Accuracy, precision, and recall.

Authorized licensed use limited to: ULAKBIM UASL - Trakya Universitesi. Downloaded on February 26,2026 at 20:40:07 UTC from IEEE Xplore. Restrictions apply.

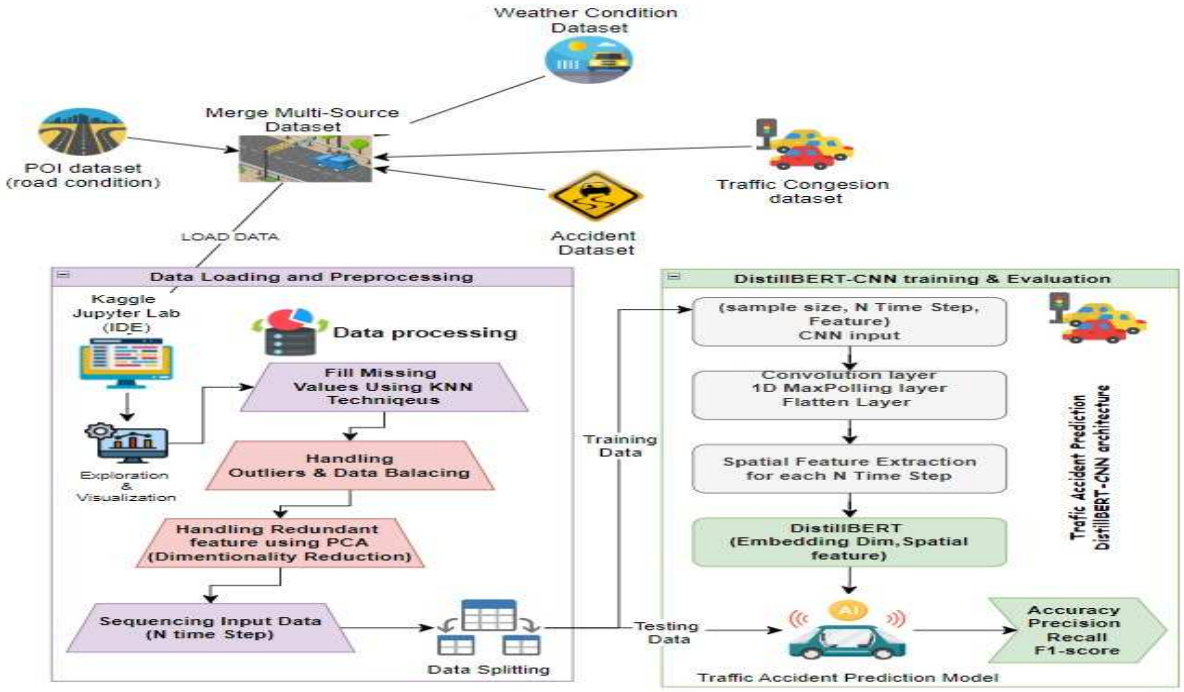

Fig. 1. Proposed DistilBERT-CNN Architecture

2) System Algorithm

Input: D_{poi}: Road condition dataset, $D_{weather}$: Weather dataset, $D_{traffic}$: Traffic congestion dataset, Daccident: Historical accident dataset.
Parameter: $X \in \mathbb{R}^{n \times m}$ (N: samples, M: features, X=D_{merge}), $X_{seq} \in \mathbb{R}^{b \times N \times d}$, where b : batch size, N: time steps, d: reduced dimension, $D_{bert} = 64$ (DistilBERT hidden size)
Output: $\hat{y}$: output prediction

Step 1 (*data merging*):

$D_{merged} \leftarrow$ Dpoi ∪ Dweather ∪ Dtraffic ∪ Daccident

Step 2 (*data processing*):

$X_i^{\text{missing}} \approx \frac{1}{k}\sum_{j=1}^{k} \text{KNN}_j(X_i)$.

$z_i = \frac{X_i - \mu}{\sigma}$ (Remove if $|z_i| > \theta$)…

$X_{\text{reduced}} = XW_{pca}$, $W_{pca} \in \mathbb{R}^{m \times d}$, $d < m$.

Step 3*: (Data Sequencing)*:

$X_{\text{seq}} = \{x^{(t)}, x^{(t+1)}, \dots, x^{(t+N)}\}$, $x^{(i)} \in \mathbb{R}^d$…

Step 4*: (Model Architecture)*

CNN layer

$f_t = Flatten\left(MaxPool1D\left(ReLU(Conv1D(x_t))\right)\right)$.

$x_t \in \mathbb{R}^{F \times 1}$

Repeat for Time Step N:

$$F_{\text{cnn}} = [f_1, f_2, \dots, f_T] \in \mathbb{R}^{B \times T \times D_{\text{cmm}}}$$

DistillBERT layer

$F_{\text{embed}} = W_e \cdot F_{\text{cmn}} + b_e \in \mathbb{R}^{B \times T \times D_{\text{bert}}}$

$Mask_{att} = 1^{B \times T}$.

$H = DistilBERT(F_{embed}, Mask_{att}) \in \mathbb{R}^{B \times T \times D_{lert}}$.

$H_{pooled} = \frac{1}{T}\sum_{t=1}^{T} H[:, t, :] \in \mathbb{R}^{B \times D_{lert}}$.

Step 5: Prediction:

$\hat{y} = \sigma(W_o H_{pooled} + b_o) \in [0,1,3,4]$.

IV. Implementation

This section clearly describes the implementation or experimental process swiftly from data loading, processing, transformation, scaling, splitting and training.

A. Data Importation & exploration

The panda's framework is utilized to load the traffic severity dataset from the Kaggle repository to the Kaggle Jupyter Lab development environment. This process is immediately followed by various explorations revealing essential information outlined as follows.

i. The dataset has the shape of 7,728,394 samples (rows) and xxx features (columns) as shown in Figure 2.

```
[3]: df = pd.read_csv('/kaggle/input/us-accidents/US_Accidents_March23.csv')
     df.head()
```

[3]:

	ID	Source	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	…
0	A-1	Source2	3	2016-02-08 05:46:00	2016-02-08 11:00:00	39.865147	-84.058723	NaN	NaN	0.01	…
1	A-2	Source2	2	2016-02-08 06:07:59	2016-02-08 06:37:59	39.928059	-82.831184	NaN	NaN	0.01	…
2	A-3	Source2	2	2016-02-08 06:49:27	2016-02-08 07:19:27	39.063148	-84.032608	NaN	NaN	0.01	…
3	A-4	Source2	3	2016-02-08 07:23:34	2016-02-08 07:53:34	39.747753	-84.205582	NaN	NaN	0.01	…
4	A-5	Source2	2	2016-02-08 07:39:07	2016-02-08 08:09:07	39.627781	-84.188354	NaN	NaN	0.01	…

5 rows × 46 columns

```
[5]: row , col = df.shape
     print(f'Row = {row}  ---+---- Col = {col}')
```

Row = 7728394 ---+---- Col = 46

Fig. 2. Data Loading and Data Size

ii. The exploration conducted shows that the sample dataset consists of null sample values in

Authorized licensed use limited to: ULAKBIM UASL - Trakya Universitesi. Downloaded on February 26,2026 at 20:40:07 UTC from IEEE Xplore. Restrictions apply.

multiple columns, including 'end_lat', 'End_lng', Description, Street, City, Zip code, and the like.

iii. The exploration also revealed that the dataset contains 102,338 duplicated values across all the samples in the dataset.

iv. Lastly, on the exploration, it's observed that the sample dataset is highly imbalanced.

B. Data Cleaning and Preprocessing

The data cleaning process entails a series of operations to transform the accident severity dataset into a well-suited format and an unbiased format for model training. Figure 3 reveals the code snippet for dropping null values from the raw data and a predefined step to encode non-numeric samples into numerical codes for the machine learning process. The process is achieved using categorical-type functionalities.

```
# df.info()
df.dropna(inplace=True)
print ('duplication drop successfull..... ')
```

```
duplication drop successfull.....
```

+ Code + Markdown

```
# df_num_col = []
for col in df.columns:
    if df[col].dtype  ==  'int32' or df[col].dtype  ==  'float64' or df[col].dtype=='int16':
        # print(col)
        df[col] = df[col].astype('int8')
        # df_num_col.append(col)
    else:
        # convert to category
        # replace with the categorical encoding...
        df[col] = df[col].astype('category').cat.codes

df.sample(3)
```

	ID	Source	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	...	Round
3861952	214952	0	1	1820976	2242805	33	-117	33	-117	5	...	
6762069	2910779	0	1	576844	677377	34	-118	34	-118	0	...	
6044974	2247596	0	1	993476	1196793	45	-122	45	-122	1	...	

3 rows × 46 columns

Fig. 3. cleaning and transformation.

C. Distill-BERT-CNN architecture

The deep learning architecture of the Distil-BERT-CNN model is designed to leverage both spatial and temporal dependencies in the data. The architecture typically starts with an input layer that accepts two types of data, including (1) sequence data, thus accident reports with weather descriptions, and (2) spatial data. The sequence data is first processed through the Distil-BERT module, which consists of multiple transformer encoder layers that generate contextual embeddings. In parallel, the spatial data is processed by a series of convolutional layers that extract spatial features through convolution and pooling operations. The feature maps from the CNN and the contextual embeddings from Distil-BERT are then concatenated or merged in a joint representation layer. This combined feature vector is fed into one or more fully connected (dense) layers, with dropout regularization applied to prevent overfitting. Finally, a softmax or sigmoid output layer predicts the severity class (probability of a traffic accident). This hybrid architecture allows the model to capture complex interactions between spatial patterns and temporal dependencies, resulting in more accurate and robust accident severity predictions. The training architecture shown in Figure 4

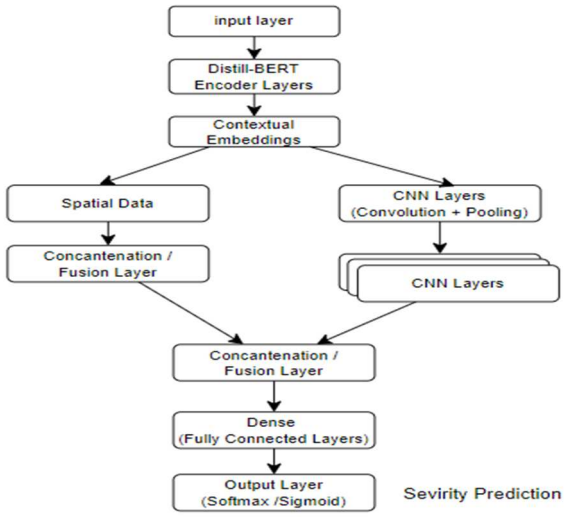

Fig. 4. Distill-BERT-CNN model

V. Experimental Result & Discussion

The evaluation of various models on traffic accident severity prediction reveals that while traditional machine learning methods such as Gradient Boosting (GBC) and Random Forest (RF) perform well in identifying non-accident events (class 0) with high precision, they struggle to accurately detect accidents of medium and high severity (classes 1 and 2), especially in the context of recall. Deep learning models such as LSTM and CNN show some improvement in capturing temporal and spatial dependencies, but their performance remains limited when used independently. The cascade model (LSTM-CNN) showcases a higher recall for accident classes, specifying the benefit of hybrid architectures. However, the proposed Distil-BERT-CNN model is expected to outperform all previous approaches by leveraging advanced contextual understanding from Distil-BERT and robust spatial feature extraction from CNN. This integration should result in a more balanced and superior performance, particularly by significantly boosting recall and precision for minority classes (1 and 2), which is crucial for real-world accident prediction systems where missing an accident event can have severe consequences.

Table I. Model Performance Comparison

Model	Precision (Class 0)	Precision (Class 1)	Precision (Class 2)	Recall (Class 0)	Recall (Class 1)	Recall (Class 2)
Gradient Boosting [19]	92%%	14%	4%	81%	14%	23%
Random Forest [19]	93%	13%	5%	69%	31%	30%
LSTM [19]	94%	10%	4%	56%	34%	36%
CNN [19]	94%	12%	4%	11%	32%	32%
Cascade (LSTM-CNN) [19]	96%	10%	4%	31%	41%	50%
Distil-BERT-CNN	97.24%	21.2%	14.11%	87.13%	51.22%	60.10%

Authorized licensed use limited to: ULAKBIM UASL - Trakya Universitesi. Downloaded on February 26,2026 at 20:40:07 UTC from IEEE Xplore. Restrictions apply.

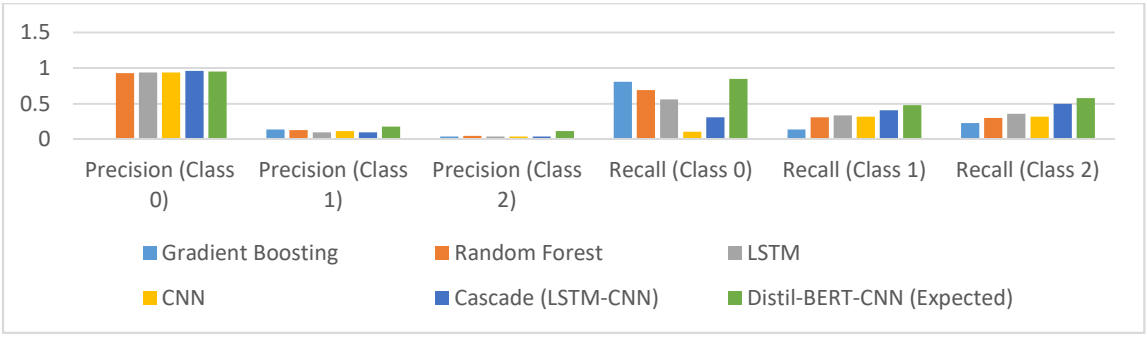

Fig. 5. Result Chat Summary

The Distil-BERT-CNN model is projected to deliver a significant boost in both precision and recall for accident classes 1 and 2, while still maintaining high precision and recall for the majority class (class 0). This balanced improvement addresses the most important challenge in accident severity prediction, ensuring high recall for actual accident events, making it a more reliable and practical tool for real-world traffic management and safety interventions.

VI. Conclusion

In conclusion, by integrating the lightweight, attention-based Distil-BERT model for efficient sequence processing with Convolutional Neural Networks for robust spatial feature extraction, the proposed approach overcomes limitations of traditional statistical and machine learning models, particularly in handling class imbalance and complex nonlinear relationships inherent in real-world traffic data. Experimental results on diverse, multi-source datasets demonstrate that Distil-BERT-CNN significantly enhances prediction precision and recall, especially for minority accident severity classes, outperforming benchmark methods such as Random Forest, Gradient Boosting, LSTM, and other deep learning models. These advancements have the potential to empower traffic management authorities with more accurate, timely, and actionable insights, enabling proactive accident prevention, optimized resource allocation, and rapid emergency response, thereby contributing to safer and more efficient transportation systems.

Reference

[1] S. K. Ahmed *et al.*, "Road traffic accidental injuries and deaths: A neglected global health issue," *Health Science Reports*, vol. 6, no. 5, p. e1240, 2023, doi: 10.1002/hsr2.1240.

[2] I. Ekanem, "Analysis of Road Traffic Accident Using AI Techniques," *Open Journal of Safety Science and Technology*, vol. 15, no. 1, Art. no. 1, Jan. 2025, doi: 10.4236/ojsst.2025.151004.

[3] A. Finogeev, M. Deev, A. Finogeev, and I. Kolesnikoff, "Proactive Big Data Analysis for Traffic Accident Prediction1," in *2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA)*, Nov. 2020, pp. 1–9. doi: 10.1109/CITISIA50690.2020.9371796.

[4] A. Boukerche and J. Wang, "Machine Learning-based traffic prediction models for Intelligent Transportation Systems," *Computer Networks*, vol. 181, p. 107530, Nov. 2020, doi: 10.1016/j.comnet.2020.107530.

[5] X. Fu, J. Liu, S. Jones, T. Barnett, and A. J. Khattak, "From the past to the future: Modeling the temporal instability of safety performance functions," *Accident Analysis & Prevention*, vol. 167, p. 106592, Mar. 2022, doi: 10.1016/j.aap.2022.106592.

[6] B. Lim and S. Zohren, "Time-series forecasting with deep learning: a survey," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20200209, Feb. 2021, doi: 10.1098/rsta.2020.0209.

[7] I. Javid, R. Ghazali, W. Saeed, T. Batool, and E. Al-Wajih, "CNN with New Spatial Pyramid Pooling and Advanced Filter-Based Techniques: Revolutionizing Traffic Monitoring via Aerial Images," *Sustainability*, vol. 16, no. 1, Art. no. 1, Jan. 2024, doi: 10.3390/su16010117.

[8] V. A. Adewopo and N. Elsayed, "Smart City Transportation: Deep Learning Ensemble Approach for Traffic Accident Detection," *IEEE Access*, vol. 12, pp. 59134–59147, 2024, doi: 10.1109/ACCESS.2024.3387972.

[9] R. DiPietro and G. D. Hager, "Chapter 21 - Deep learning: RNNs and LSTM," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, S. K. Zhou, D. Rueckert, and G. Fichtinger, Eds., in The Elsevier and MICCAI Society Book Series. , Academic Press, 2020, pp. 503–519. doi: 10.1016/B978-0-12-816176-0.00026-0.

[10] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information*, vol. 15, no. 9, Art. no. 9, Sep. 2024, doi: 10.3390/info15090517.

[11] R. DiPietro and G. D. Hager, "Chapter 21 - Deep learning: RNNs and LSTM," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, S. K. Zhou, D. Rueckert, and G. Fichtinger, Eds., in The Elsevier and MICCAI Society Book Series. , Academic Press, 2020, pp. 503–519. doi: 10.1016/B978-0-12-816176-0.00026-0.

[12] T. RajyaLakshmi and K. S. Kuppusamy, "Comparative Analysis of Pretrained Encoder-Decoder Transformer Models for Extreme Text Summarization," in *2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC)*, Dec. 2023, pp. 1–6. doi: 10.1109/ICACIC59454.2023.10435363.

[13] S. Alaei, *An Automated Discharge Summary System Built for Multiple Clinical English Texts by Pre-trained DistilBART Model*. 2023. Accessed: Apr. 18, 2025. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-222616

[14] K. Alharbi and M. A. Haq, "Enhancing Disaster Response and Public Safety with Advanced Social Media Analytics and Natural Language Processing," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14212–14218, Jun. 2024, doi: 10.48084/etasr.7232.

[15] A. Mahmoud and A. Mohammed, "Leveraging Hybrid Deep Learning Models for Enhanced Multivariate Time Series Forecasting," *Neural Process Lett*, vol. 56, no. 5, p. 223, Aug. 2024, doi: 10.1007/s11063-024-11656-3.

[16] S. A. Kashinath *et al.*, "Review of Data Fusion Methods for Real-Time and Multi-Sensor Traffic Flow Analysis," *IEEE Access*, vol. 9, pp. 51258–51276, 2021, doi: 10.1109/ACCESS.2021.3069770.

[17] Q. Li, O. Alonge, and A. A. Hyder, "Children and road traffic injuries: can't the world do better?," *Archives of disease in childhood*, vol. 101, no. 11, pp. 1063–1070, 2016.

[18] W. H. Organization, *Global status report on road safety 2018*. World Health Organization, 2019. Accessed: Apr. 18, 2025. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=uHOyDwAAQBAJ&oi=fnd&pg=PR6&dq=According+to+the+World+Health+Organization,+road+traffic+injuries+are+the+leading+cause+of+death+for+children+and+young+adults+aged+5+to+29+years,&ots=2T2kVBqi_Z&sig=TPnoMck_uSXJNFZRXFXzbQs357k

[19] P. Sajadi, M. Qorbani, S. Moosavi, and E. Hassannayebi, "Accident Impact Prediction based on a deep convolutional and recurrent neural network model," Nov. 12, 2024, *arXiv*: arXiv:2411.07537. doi: 10.48550/arXiv.2411.07537.

[20] M. Zhang and W. Zhao, "Traffic Flow Prediction Based on Large Language Models and Future Development Directions," in *ITM Web of Conferences*, EDP Sciences, 2025, p. 01008. Accessed: Apr. 18, 2025. [Online]. Available: https://www.itm-conferences.org/articles/itmconf/abs/2025/01/itmconf_dai2024_01008/itmconf_dai2024_01008.html

[21] L. Jiang, "Traffic flow prediction method based on deep learning," in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 012050. Accessed: Apr. 18, 2025. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1646/1/012050/meta

[22] A. Azhar *et al.*, "Detection and prediction of traffic accidents using deep learning techniques," *Cluster Comput*, vol. 26, no. 1, pp. 477–493, Feb. 2023, doi: 10.1007/s10586-021-03502-1.

[23] Z. Zhang, Q. He, J. Gao, and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 580–596, Jan. 2018, doi: 10.1016/j.trc.2017.11.027.

Authorized licensed use limited to: ULAKBIM UASL - Trakya Universitesi. Downloaded on February 26,2026 at 20:40:07 UTC from IEEE Xplore. Restrictions apply.