NOUN Interdisciplinary Journal of Computing, E-learning & Application (NOUN-IJCEA)
https://ijcea.nou.edu.ng
Volume 1 Number 1, April 2026 pp. 63-76

Spatial-Temporal CNNs: A Deep Learning Approach to Predictive Traffic Accident Hotspotting

Simon Numduck Dodo, Kudu Muhammad Muhammad, Adebayo Bashir Sulaimon, Femi Aminu Enesi

Department of Computer Science, Federal University of Technology, Minna, Niger State, Nigeria.
dodosnmte258@st.futminna.edu.ng, muhammad_kudu@futminna.edu.ng, bashirsulaimon@futminna.edu.ng, enesifa@futminna.edu.ng
*Corresponding Author: dodosnmte258@st.futminna.edu.ng

ABSTRACT

Road traffic accidents remain a serious global challenge, claiming millions of lives annually and straining economies worldwide. This study introduces a novel hybrid architecture that integrates Convolutional Neural Networks with DistillBERT transformer-based architecture to tackle the persistent problem of accident severity prediction. However, by merging different real-world data streams weather patterns, traffic congestion levels, road infrastructure details, and historical crash records from Los Angeles County. The model learns to read the complex language of urban risk. The approach transforms raw, noisy, and imbalanced data into meaningful structure sequence, enabling the CNN to map spatial pattern while DistillBERT's lightweight attention mechanism traces how these risks evolve over time. The experimental training and evaluation show that CNN-DistillBERT's models achieve the highest results of 94.66% accuracy in predicting accident severity. This study opens practical approach for real time road traffic analysis before and after it happened.

Keywords: Accident, CNN, Deep Learning, DistillBERT, Road Traffic, Special pattern, Transformer

1. INTRODUCTION

Recently, Road Traffic accidents are tagged as one of the major issues resulting into high rate of deaths, injuries and high cost of economic fund. Study conducted by Jaulkar and Parihar, (2025) revealed that about 1.19 million death rate are recorded on annual basic for road crashes only, and tens of millions of recorded injuries that can permanently cause damage in human body. Additionally, this can lead to economic burden due to high medical cost incur by an individual, communities or an entire country at large. Hence its essential to find a mean to limit the road traffic accident situation across the globe (Ekanem, 2025). Due to the advancement in technology, researcher claims that advance computer automation model can predict when and where a traffic accident might occur, and there level of traffic disruption (Berhanu *et al.,* 2023). Authorities, communities and individual can utilize such system to make quick decision and take appropriate action to prevent the occurrence of traffic incident in major roads. Public safety and urban development can only be enhance if traffic incident can be effectively control using automation system (Ahmed *et al.,* 2023).

Furthermore, accurate prediction helps traffic authorities to create a more proactive measure, which can includes dynamic speed limit enforcement, traffic rerouting or redirection from area with high possibility of accidents, and put in place real-time surveillance feeds in accident prone zone (Finogeev *et al.,* 2020). These actions put in place can reduce traffic accident situation while enhancing faster emergency responses, reduce arrival time and help save lives. More so, better congestion management can be improved, via enhance analysis on traffic accident data (Ekanem, 2025). Traditional traffic prediction model is majorly design using pure statistical model and conventional machine learning algorithm. This method analyze large historical amount of crash data, traffic volumes, weather pattern and road condition for identifying possibility of traffic accident (Boukerche and Wang, 2020). Technically, the approach generally struggles with capturing complex pattern nonlinear pattern inherent in real-world traffic conditions. Lastly, convention approach mostly fails to simultaneously account for temporal pattern and spatial dependencies which are both essential for traffic accident pattern prediction (Fu *et al.,* 2022).

DistillBERT (Distilled Bidirectional Encoder Representations from Transformers) is a lightweight, efficient adaptation of BERT, not BART, designed for bidirectional text encoding through knowledge distillation from its larger "teacher" model (RajyaLakshmi and Kuppusamy, 2023). This process reduces parameters by 40% while retaining approximately 97% of BERT's performance, enabling faster inference for resource-constrained applications (Alaei, 2023). Unlike BART which combines BERT's bidirectional encoding with GPT style autoregressive decoding for sequence-to-sequence tasks, DistillBERT focuses majorly on encoder efficiency. In the case of traffic accident prediction, DistillBERT excels at processing textual incident data, such as summarizing detailed reports, classifying severity from police logs, or generating operator friendly alerts from sensor descriptions (Alharbi and Haq, 2024).

Hybrid networks merge different deep learning architectures to improve their complementary strengths. A common method is to merge CNNs for spatial feature extraction with RNNs or LSTMs for temporal sequence modeling (Mahmoud and Mohammed, 2024). In traffic detection applications, hybrid models analyze spatial data such as traffic images or road network route alongside temporal traffic flow or accident data. study shows that CNN layers can process video frames from traffic cameras, while LSTM layers capture temporal changes in traffic patterns (Kashinath *et al.,* 2021). This combined approach enhances 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 (Mahmoud and Mohammed, 2024). 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 (Mahmoud and Mohammed, 2024). 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 (Kashinath *et al.,* 2021). 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 (Mahmoud and Mohammed, 2024).

Hence, this study proposes a hybrid CNN and DistillBERT framework to enhance the detection and prediction efficiency of traffic road accident severity by integrating convolutional neural networks for spatial feature extraction using multi-source data (weather, traffic congestion, POI data and accident) with DistillBERT, thus, a lightweight transformer model for capturing temporal dependencies in sequential patterns

1.1. Research Problem

Road Traffic Accidents still remain a serious public health and safety issue worldwide, resulting into a substantial loss of life and property (Ahmed *et al.,* 2023). In respect to the World Health Organization (WHO), road traffic injuries are the leading cause of death for children and young adults between the age of 5 to 29 years, with approximately 1.19 million fatalities and 20 to 50 million non-fatal injuries occurring on annual bases (Ahmed *et al.,* 2023). Additionally, these incidents also impose a high economic burden on individuals, families, and nations, based on medical expenses, loss of productivity, and cost most countries about 3% of their gross domestic product Li *et al.,* (2016).

However, there are several real-time forecasting models designed to predict and identified the impacts of accidents using various publicly available datasets (Sajadi *et al.,* 2024), but most of this approach faces issue because there relies majorly on non-real time data and models that lack comprehensive metric to properly measure post-accident associated with impact. Hybridizing large language-based model with a deep learning model can efficiently boost pattern recognition from past accident cases that lead to traffic management with decision-making that improved adoption of real-time traffic and accident historical data in predicting potential accident risks. The existing studies that utilized recurrent-based networks in combination with order deep learning models comes with issues such as overfitting, insufficient generalization, and large-scale datasets challenges. Hence, this study introduces and hybrid model that will address high-dimensional data issues while retaining high-quality temporal dependencies using attention-based mechanism for effective, efficient and lightweight using DistillBERT-CNN architecture as a predictive model.

1.2. Research Goal

This study primarily aims to develop a hybrid CNN-DistillBERT model for spatial pattern analysis and dependency analysis for road traffic accident prediction. In achieving this goal, the study first merge and optimized the four diverse road traffic dataset (traffic congestion, POI, weather condition and accident) using SMOTE-EEN and ADASYN data balancing techniques. The optimized dataset is then used to develop the road traffic prediction model using the combine CNN and DistillBERT. Next the study uses a fully customize hybrid CNN-DistillBERT Architecture to trained the Traffic Detection Model with the help of the extracted time series feature been extracted, then the performance of the model will be evaluated via Accuracy, precision, recall and F1-score performance metrics.

1.3. Signficiance of the study

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 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.

2. LITERATURE REVIEW

Based on the research work of Sajadi *et al.,* (2024), traffic accidents significantly threaten public safety, causing many fatalities, injuries, and economic losses annually. Existing predictive models for post-accident impact face two main challenges: they rely on either expensive or non-real-time data, or lack a comprehensive metric to accurately measure the impact of accidents on traffic flow. This study proposes a hybrid deep neural network called the cascade model, combining Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures. The LSTM captures temporal patterns in data, while the CNN extracts spatial patterns from a sparse accident dataset. The model uses readily available real-world data from Los Angeles County and integrates an external traffic congestion dataset to create a novel feature termed the "accident impact" factor, which quantifies how an accident affects surrounding traffic flow. Extensive experiments demonstrate that the proposed cascade model outperforms state-of-the-art baseline models. It achieves higher precision in predicting minimal impacts (cases with no reported accidents) and higher recall in predicting significant impacts (cases with reported accidents), indicating improved accuracy in forecasting post-accident traffic conditions.

Road traffic accidents present a growing global crisis, contributing to millions of fatalities annually and imposing substantial economic burdens. While prior research has framed accident prediction as a binary classification task, there remains a critical gap in analyzing the intricate relationships among contributing factors, particularly in understudied regions like New Zealand (NZ). Existing studies often overlook ensemble machine learning (ML) approaches and comprehensive model interpretability, limiting actionable insights for policy interventions. The authors (S. Ahmed *et al.,* 2023), evaluates six ensemble ML models Random Forest (RF), Decision Jungle (DJ), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (L-GBM), and Categorical Boosting (CatBoost) on NZ accident data (2016–2020). Shapley Additive exPlanations (SHAP) analysis is applied to interpret model outputs, enabling global feature importance ranking and local-level decision explanations through dependence plots. The experiment result show that RF outperformed other models with 81.45% accuracy, 81.68% precision, 81.42% recall, and 81.04% F1-score. SHAP analysis identified road category and number of vehicles involved as top severity determinants. Retraining models with SHAP-derived high-impact features improved performance: DJ (6% gain), AdaBoost (5%), and CatBoost (8%).

According to (Rahim and Hassan, 2021), it's been discussed that highway work zones are particularly susceptible to congestion and traffic collisions, making accurate and timely prediction of collision severity crucial for reducing emergency response times and improving traffic safety. Previous studies using statistical and machine learning models have shown limited performance, especially in predicting fatal and injury crashes, often focusing solely on accuracy. This study proposes a novel deep learning approach with a customized F1-loss function to predict traffic crash severity. The proposed methodology demonstrates improved performance in predicting crash severity, particularly for fatal and injury crashes, using the deep learning approach. This improvement can enhance traffic safety and reduce congestion at work zones and potentially other roadway segments.

Reviewing the work of (Azhar *et al.,* 2023), it is been reveal that road transportation causes 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.

Despite advances in traffic safety, highway accidents remain a leading cause of death, especially in developing nations where the societal and economic impacts are profound. Identifying key factors influencing accident severity is crucial for improving safety measures and decision-making. The challenge lies in accurately predicting accident severity using complex, multi-factor data, and improving upon existing models that often lack precision or interpretability. The study employs an ensemble approach combining Random Forest (RF) for feature importance analysis with a Convolutional Neural Network (CNN) for prediction, termed RFCNN. (Manzoor *et al.,* 2021) reveal that the proposed RFCNN model achieved high predictive performance, with an accuracy of 0.991, precision of 0.974, recall of 0.986, and F-score of 0.980, outperforming other models. The use of the top 20 significant features contributed to these results, demonstrating that the ensemble approach enhances decision-making and accident severity prediction reliability. The model's high accuracy indicates its potential utility in real-world traffic safety applications.

3. METHODOLOGY

The system design entails the methodological research design approach using Figure 1, revealing that 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 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 based on the exploration conducted on the dataset. The first preprocessing operation includes replacing the missing value or null value. Next is the handling of outliers, and data balancing using ADASYN and SMOTE-ENN. The second preprocessing state is conducted to extract sequential pattern and timestamp information which transform the dataset into three-dimensional input feature for CNN processing. The final version of the preprocessed data is split into training and testing samples for training and testing of the traffic accident prediction model. The output of the DistillBERT training becomes the Traffic Accident Prediction Model, which is evaluated using metrics such as Accuracy, precision, and recall.

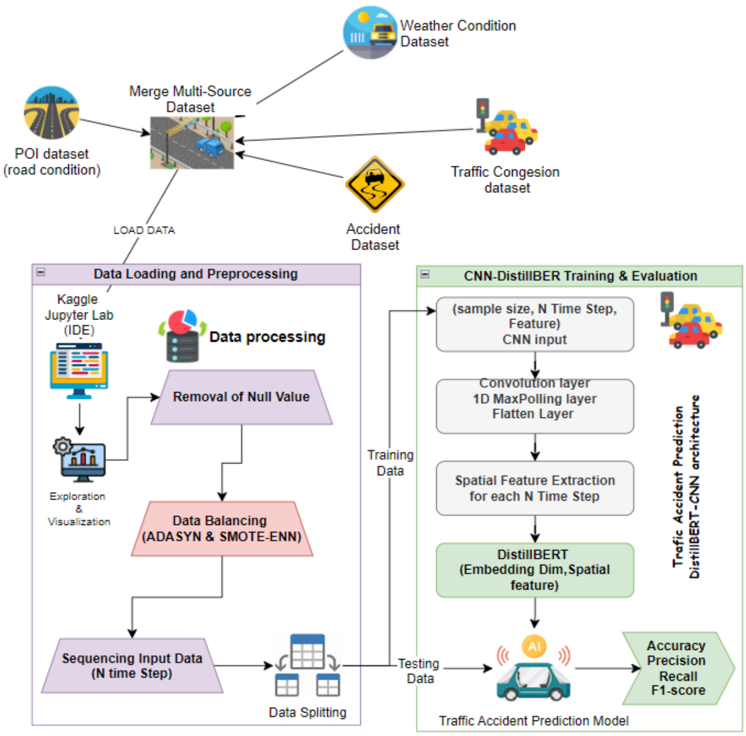

Figure 1: CNN-DistillBERT framework design

3.1. 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, which include the accident dataset, 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 as shown in table 1,2 and 3. Link 1 (Dataset): https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents

Table 1: Dataframe Sample (feature 1- 8)

	ID	Source	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng
0	A-6573928	Source1	8/3/2020 14:35	8/3/2020 15:20	32.20676	-110.981	32.20676	-110.981
1	A-6575538	Source1	9/22/2020 16:44	9/22/2020 17:29	32.38084	-110.964	32.38084	-110.964
2	A-6588337	Source1	8/3/2020 15:42	8/3/2020 15:57	41.37978	-81.9941	41.37978	-81.9941
3	A-6590489	Source1	9/22/2020 16:29	9/22/2020 17:14	33.6395	-112.134	33.6395	-112.134
4	A-6592050	Source1	8/3/2020 13:41	8/3/2020 13:56	38.0832	-122.115	38.0832	-122.115
5	A-6593108	Source1	7/31/2020 14:44	7/31/2020 15:29	36.11036	-86.7449	36.11036	-86.7449
6	A-6598867	Source1	8/3/2020 15:05	8/3/2020 17:05	35.59053	-82.5725	35.59053	-82.5725

Table 2: Dataframe Sample (feature 9-16) location-based feature

Distance (MI)	Description	Street	City	County	State	Zipcode	Country
0	At I-10/Freeway - Accident.	W Starr Pass Blvd	Tucson	Pima	AZ	85745	US
0	At N 1st Ave - Accident.	E Linda Vista Blvd	Tucson	Pima	AZ	85704	US
0	At I-80 - Accident.	I-80 W	North Ridgeville	Lorain	OH	44039	US
0	At N 35th Ave - Accident.	W Bell Rd	Phoenix	Maricopa	AZ	85053	US
0	At Lake Herman Rd - Accident. Hard shoulder blocked.	I-680 S	Benicia	Solano	CA	94510	US
0	At McCall St - Accident.	Nolensville Pike	Nashville	Davidson	TN	37211-2340	US

Table 3: Dataframe Sample (17-25) weather-based feature

Temperature(F)	Wind_Chill(F)	HUMIDITY (%)	Pressure(in)	Visibility(mi)	Wind_Direction	Wind_Speed(mph)	Precipitation(in)	Weather_Condition
103	103	22	27.15	10	ENE	10	0	Mostly Cloudy
95	95	19	27.16	10	WNW	10	0	Fair
68	68	93	29.14	9	E	5	0.02	Light Rain with Thunder
98	98	18	28.28	10	W	5	0	Fair
95	95	22	29.9	10	WNW	9	0	Fair
81	81	79	29.2	10	S	7	0.01	Mostly Cloudy

3.2. 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}), $X_{\text{seq}} \in \mathbb{R}^{b\times N\times d}$, where b : batch size, N: time steps, d: reduced dimension, $D_{\text{bert}} = 64$ (DistilBERT hidden size)

Output: $\hat{y}$: output prediction.

Step 1 (*data merging*):

D_{merged} 🡐 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
$f_t = Flatten\left(MaxPool1D\left(ReLU\left(Conv1D\left(x_t\right)\right)\right)\right)$
$x_t \in \mathbb{R}^{F \times 1}$ *:*
Repeat for Time Step N:
$F_{\text{cnn}} = [f_1, f_2, \dots, f_T] \in \mathbb{R}^{B \times T \times D_{\text{cmm}}}$
DistilBERT layer
$F_{\text{embed}} = W_e \cdot F_{\text{cmn}} + b_e \in \mathbb{R}^{B \times T \times D_{\text{bert}}}$
$Mask_{att} = 1^{B \times T}$
$H = DistilBERT\ (F_{embed}, Mask_{att}) \in \mathbb{R}^{B \times T \times D_{lert}}$
Step 5: Prediction:
$\hat{y} = \sigma\left(W_o H_{pooled} + b_o\right) \in [0,1,3,4]$

The proposed algorithm (algorithm 1), predicts traffic accidents by integrating spatial and temporal features from multi-source datasets using a hybrid CNN-DistilBERT architecture. First, each time step's features are treated as a 1D signal and passed through a 1D Convolutional Neural Network (CNN) to extract spatial characteristics, followed by max pooling and flattening. These spatial features across all time steps are stacked and linearly projected to match the embedding dimension expected by the DistilBERT model. Instead of textual input, the projected embeddings are directly fed into DistilBERT's transformer layers with an attention mask, allowing it to capture complex temporal dependencies. The transformer output is mean-pooled across time steps to generate a fixed-size representation, which is then passed through a sigmoid-activated dense layer for binary classification (accident or not). The model is trained using binary cross-entropy loss and evaluated using standard metrics like accuracy, precision, recall, and F1-score.

3.3. Performance Metrics

In this study the metrics assess the effectiveness of the proposed DistilBERT-CNN architecture in identifying and categorizing accident severity levels include, and this include the following.

3.3.1. Accuracy

Accuracy measures the overall correctness of the model across all severity classes. It is calculated as the ratio of correctly predicted instances to the total number of instances. Equation 3.1 shows the equation used in computing the performance accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) represent the four possible prediction severity outcomes.

3.3.2. Precision

Precision quantifies how many of the predicted severe accidents are severe, focusing on minimizing false alarms (FP). Equation 3.2 shows the equation for calculating the precision.

$$\text{Precision} = \frac{TP}{TP+FP} \dots \quad (2)$$

High precision is vital for traffic management systems to avoid unnecessary resource allocation (e.g., dispatching emergency services for non-critical cases).

3.3.3. Recall (Sensitivity)

Recall measures the model’s ability to identify all actual severe accidents, minimizing missed critical cases (FN). The equation is specified in equation 3.3.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

In accident severity prediction, high recall ensures that severe incidents are not overlooked, enabling timely interventions to mitigate risks.

3.3.4. F1-Score

The F1-score harmonizes precision and recall into a single metric, providing a balanced assessment of the model’s performance. Equation 3.4 shows the formula for calculating the F1-score metric.

$$\text{F1-Score} = 2 \text{ x } \frac{Precision \text{ x } Recall}{Precision \text{ x } Recall} \tag{4}$$

This metric is particularly useful when class imbalance exists, as it penalizes models that sacrifice one metric for the other.

4. RESULTS AND DISCUSSION

This study revealed the data sample class distribution for the traffic accident severity prediction classes, with severity 2 class having the highest sample and severity 1 having the lowest sample distribution. The severity ranges from 1 to 4, therefore the class 1 as the minimal accident impact and 4 with the maximum traffic accident impact. Additionally, the figure shows the Severity 2 contain 615,698,1 samples, while 3, 4, and 1 contain 129,933,7 sample, 204710 sample and 67,366 sample.

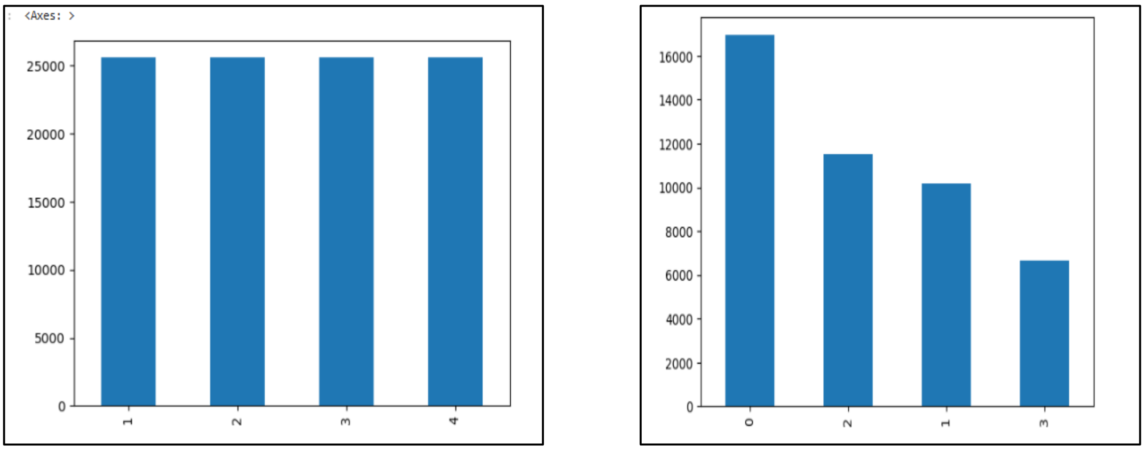

Figure 2: Balanced Data using ADASYN & SMOTEENN

However prior the implementation processes it’s essential to optimize the dataset by balancing the raw sample data using ADASYN and SMOTEENN data balancing techniques as visualized in figure 2. The ADASYN shows that each sample are equally balance, which could also be disadvantageous due to the introduction of too much noise. However, SMOTEENN show unequal samples across all class distribution due to is robustness in generating augmented data samples without introducing noisy samples via hybrid Synthetic Minority Over Sampling and Edited Nearest Neighbor Techniques (SMOTE-ENN).

Data transformation is also essential to convert the data samples into a numerically understandable format for machine learning or deep learning processing. Hence, this is achieved using categorical type datatype in pandas to convert objects or non-numerical type into numerical form (int, or float). The operational code uses a for loop code structure to check each feature with object or bool datatype,

then convert them into category type for code extraction. Furthermore, to reduce the computation cost, the dataset is compressed using min-max scaler approach.
This study conduct data splitting by partitioning the entire clean and preprocessed data into a training sample for training the deep learning traffic accident prediction model and a testing sample for validating the model performance using accuracy, recall, precision, and F1-score. Moreover, data feature extraction is conducted via sequence time-stamp extraction, which transform two-dimensional data into three-dimensional data containing sequence information for the CNN-DistillBERT road traffic detection model.

Layer (type)	Output Shape	Param #	Connected to
input_18 (InputLayer)	[(None, 10, 41)]	0	[]
tf.__operators__.add_10 (TFOpLambda)	(None, 10, 41)	0	['input_18[0][0]']
multi_head_attention_4 (MultiHeadAttention)	(None, 10, 41)	42793	['tf.__operators__.add_10[0][0]', 'tf.__operators__.add_10[0][0]']
dropout_16 (Dropout)	(None, 10, 41)	0	['multi_head_attention_4[0][0]']
tf.__operators__.add_11 (TFOpLambda)	(None, 10, 41)	0	['dropout_16[0][0]', 'tf.__operators__.add_10[0][0]']
layer_normalization_8 (LayerNormalization)	(None, 10, 41)	82	['tf.__operators__.add_11[0][0]']
dense_35 (Dense)	(None, 10, 128)	5376	['layer_normalization_8[0][0]']
dense_36 (Dense)	(None, 10, 41)	5289	['dense_35[0][0]']
dropout_17 (Dropout)	(None, 10, 41)	0	['dense_36[0][0]']
tf.__operators__.add_12 (TFOpLambda)	(None, 10, 41)	0	['layer_normalization_8[0][0]', 'dropout_17[0][0]']
layer_normalization_9 (LayerNormalization)	(None, 10, 41)	82	['tf.__operators__.add_12[0][0]']
multi_head_attention_5 (MultiHeadAttention)	(None, 10, 41)	42793	['layer_normalization_9[0][0]', 'layer_normalization_9[0][0]']
dropout_18 (Dropout)	(None, 10, 41)	0	['multi_head_attention_5[0][0]']
tf.__operators__.add_13 (TFOpLambda)	(None, 10, 41)	0	['dropout_18[0][0]', 'layer_normalization_9[0][0]']
layer_normalization_10 (LayerNormalization)	(None, 10, 41)	82	['tf.__operators__.add_13[0][0]']
dense_37 (Dense)	(None, 10, 128)	5376	['layer_normalization_10[0][0]']
dense_38 (Dense)	(None, 10, 41)	5289	['dense_37[0][0]']

Figure 3: The CNN + DistillBERT Network Layout.

Figure 3 illustrates the architecture of a hybrid CNN with DistillBERT model fictional network layers, which is designed for road traffic detection. Each block consists of a Conv2D layer (thus with 16 filters and kernel size 3), dropout for regularization, layer normalization, attentions mechanism dense layers, and additional dropout, maintaining progressive feature refinement across six, then batch normalization (BN) after most layer. Lastly, the model shows as a final dense layer and output, enabling efficient processing sequence data into predictions via convolutional operations followed by transformer-based DistillBERT integration for contextual understanding.

4.1. Training Summary

Figure 4-10 reveal the training epoch of the road accident severity prediction model using a hybrid CNN and Recurrent based Network such as LSTM, BI-LSTM, Attention Mechanism, and DistillBERT architecture. The Line charts reveal information such as the training accuracy and loss, with the validation accuracy and loss for 30 epochs (iterations). The training and validation curve shows that the hybrid version of Bi-LSTM with CNN are performing efficiently better than the LSTM version, but utilizing the SMOTE-EEN data sample with CNN-Bi-LSTM variant, a similar performance learning curve is observed revealing the immense contribution of SMOTE-ENN data re-sampling method. Similar observation is identified in attention variant and that of Bi-LSTM model.

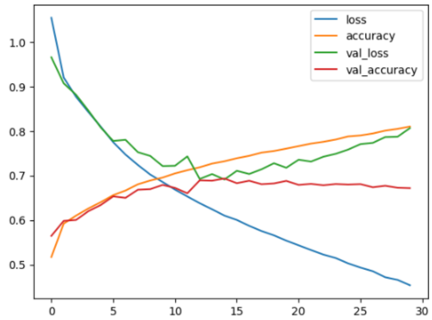

Figure 4: CNN + LSTM Training summary

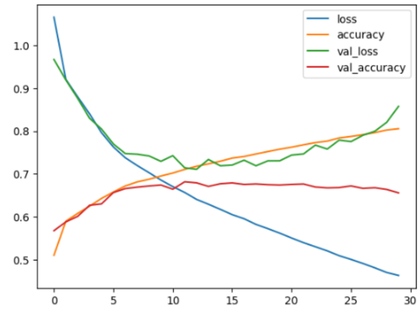

Figure 5: CNN + LSTM + ADASYN Training Summary

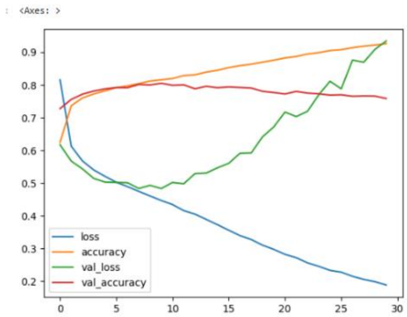

Figure 6: CNN + BI-LSTM + ADASYN Training Summary

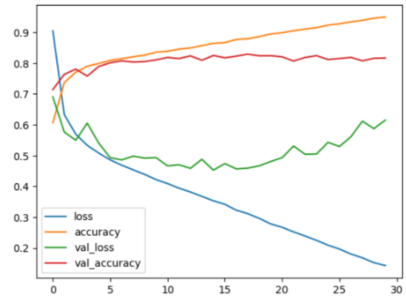

Figure 7: CNN + LSTM + SMOTE-ENN Training Summary

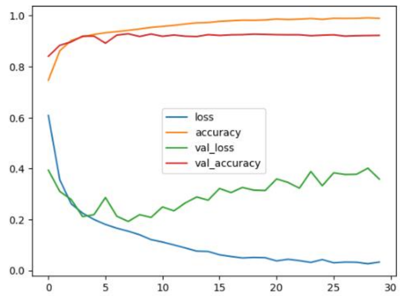

Figure 8: CNN + BI-LSTM + SMOTE-ENN Training Summary

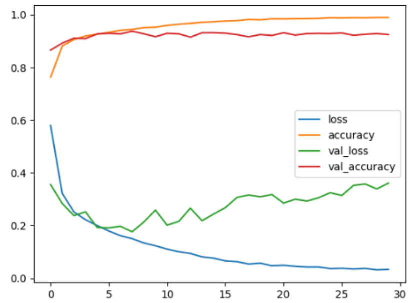

Figure 9: CNN + BI-LSTM + ATTENTION + SMOTE-ENN Training Summary

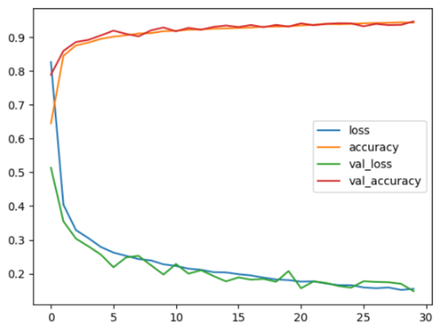

Figure 10: CNN + DistillBERT + SMOTE-ENN Training Summary

4.2. Experimental Results

This section introduces the result summary and result presentation via comprehensive classification report with accuracy, precision, recall, and F1-score. The classification report for CNN+LSTM, CNN + LSTM + ADASYN, CNN + BI-LSTM + ADASYN, CNN + LSTM + SMOTE-ENN, CNN + BI-LSTM + ATTENTION + SMOTE-ENN, and CNN + TRANSFORM + SMOTE-ENN are showcase in figures 11-15. each model, with a minimum accuracy of 65% observed using CNN + LSTM + ADASYN, while a maximum performance of 94.6% is observed with the hybridization of CNN + DistillBert + SMOTEENN.

```
Accuracy  : 0.6717094323628183
Precision : 0.6703304568808384
Recall    : 0.6717094323628183
F1-Score  : 0.6705789082255178
              precision    recall  f1-score   support

           0       0.71      0.76      0.74      7668
           1       0.66      0.66      0.66      7669
           2       0.68      0.67      0.68      7668
           3       0.63      0.59      0.61      7666

    accuracy                           0.67     30671
   macro avg       0.67      0.67      0.67     30671
weighted avg       0.67      0.67      0.67     30671
```

Figure 11: CNN + LSTM

```
959/959 [==============================] - 6s 6ms/step
Accuracy  : 0.6556682207948876
Precision : 0.6587414987998025
Recall    : 0.6556682207948876
F1-Score  : 0.6534593700998087
              precision    recall  f1-score   support

           0       0.70      0.74      0.72      7668
           1       0.69      0.53      0.60      7669
           2       0.63      0.72      0.67      7668
           3       0.60      0.64      0.62      7666

    accuracy                           0.66     30671
   macro avg       0.66      0.66      0.65     30671
weighted avg       0.66      0.66      0.65     30671
```

Figure 12: CNN + LSTM + ADASYN

```
425/425 [==============================] - 2s 5ms/step
Accuracy  : 0.8165954940362244
Precision : 0.8154614296971598
Recall    : 0.8165954940362244
F1-Score  : 0.8158672014618863
              precision    recall  f1-score   support

           0       0.86      0.87      0.87      5081
           1       0.81      0.84      0.83      3042
           2       0.81      0.80      0.81      3458
           3       0.71      0.67      0.69      2001

    accuracy                           0.82     13582
   macro avg       0.80      0.80      0.80     13582
weighted avg       0.82      0.82      0.82     13582
```

Figure 13: CNN + LSTM + SMOTE-ENN

```
425/425 [==============================] - 4s 7ms/step
Accuracy  : 0.9240907082903843
Precision : 0.9238535512104508
Recall    : 0.9240907082903843
F1-Score  : 0.9237792900087942
              precision    recall  f1-score   support

           0       0.95      0.96      0.95      5081
           1       0.90      0.93      0.92      3042
           2       0.92      0.92      0.92      3458
           3       0.88      0.83      0.86      2001

    accuracy                           0.92     13582
   macro avg       0.92      0.91      0.91     13582
weighted avg       0.92      0.92      0.92     13582
```

Figure 14: CNN + BI-LSTM + ATTENTION + SMOTE-ENN

```
425/425 [==============================] - 5s 10ms/step
Accuracy  : 0.946620527168311
Precision : 0.9464757610242441
Recall    : 0.946620527168311
F1-Score  : 0.9464554829163947
              precision    recall  f1-score   support

           0       0.97      0.97      0.97      5081
           1       0.93      0.95      0.94      3042
           2       0.94      0.95      0.95      3458
           3       0.92      0.88      0.90      2001

    accuracy                           0.95     13582
   macro avg       0.94      0.94      0.94     13582
weighted avg       0.95      0.95      0.95     13582
```

Figure 15: CNN + DistillBert + SMOTE-ENN

The results summary in Table 4 demonstrate a clear report on all the trained and evaluated accident severity model performance, starting with the baseline CNN + LSTM at metrics score of around 67% across accuracy, precision, recall, and F1-score, which slightly reduce with ADASYN oversampling (65%) due to potential class imbalance noise, but improves drastically with bidirectional LSTM (BI-LSTM) and advanced resampling like SMOTE-ENN, reaching 81-92% for CNN + BI-LSTM + SMOTE-ENN. The best performers model incorporates architectural enhancements including CNN + BI-LSTM + Attention + SMOTE-ENN edges out to 92.4%, while CNN + DistillBert + SMOTE-ENN achieves the highest at 94.66% accuracy 94.64%, revealing how combining convolutional feature extraction, bidirectional temporal modeling, attention mechanisms or transformers, and hybrid oversampling effectively addresses data imbalance and captures complex spatiotemporal patterns in accident data

Table 4: Accident Severity Model Result Summary.

S. N	MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE
1	CNN + LSTM	67.17%	67.03%	67.1%	67.05%
2	CNN + LSTM + ADASYN	65.56%	65.87%	65.56%	65.34%
3	CNN + BI-LSTM + ADASYN	75.79%	-	-	-
4	CNN + LSTM + SMOTE-ENN	81.65%	81.54%	81.65%	81.58%
5	CNN + BI-LSTM + SMOTE-ENN	92.23%	92.22%	92.23%	92.22%
6	CNN + BI-LSTM + ATTENTION + SMOTE-ENN	92.40%	92.38%	92.40%	92.37%
7	CNN + TRANSFORM + SMOTE-ENN	94.66%	94.64%	94.66%	94.64%

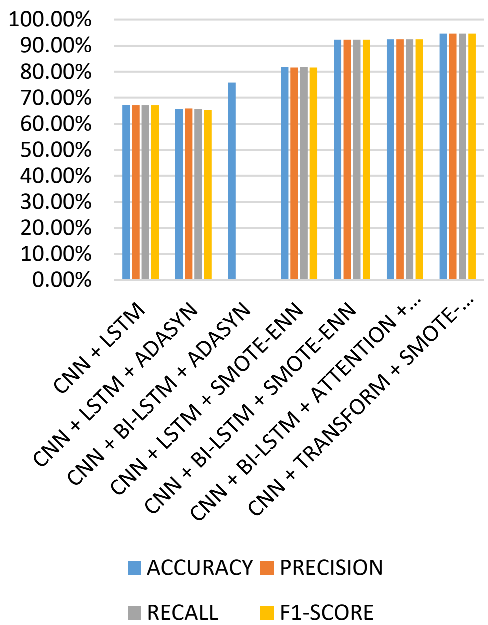

Figure 16: Model Performance Plot

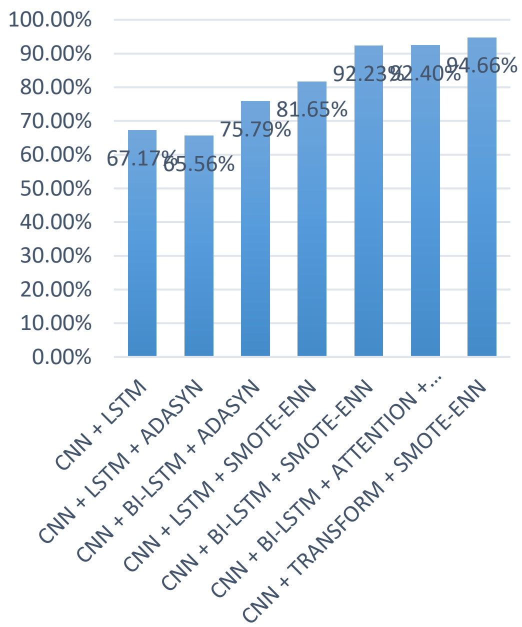

Figure 17: Model Accuracy Performance Plot

4.3. Discussion

The experimental result shows significant improvement in traffic accident severity prediction via the integration of CNN-DistillBERT architecture. The model achieved an efficient performance of 94.66% accuracy, 94.64% precision, recall, and F1-score when combined with SMOTE-ENN balancing. Moreover, baseline models such as CNN-LSTM achieved 67% across accuracy, precision and recall, while the bidirectional variants (CNN-BI-LSTM) and attention-based versions improved to 92.4%, revealing the importance of capturing spatio-temporal patterns from multi-source data. Data balancing approach utilized in this thesis proved important, as ADASYN alone achieved 65% due to synthetic noise, whereas SMOTE-ENN effectively mitigated class imbalance by enabling robust generalization post-feature scaling and 10 step timestamp sequencing.

These performances improve revealed the efficiency of transformer based DistillBERT over LSTM for temporal dependencies. DistillBERT strength is draw from the lightweight attention to process CNN-extracted spatial features from accidents, congestion, POI, and weather datasets. Figure 4.6 to 4.7 show the training curves (ranging from epochs 1-30) confirming the stable convergence with minimal validation loss, outperforming unidirectional models by processing sequences bidirectional and reducing overfitting on high-dimensional inputs.

Generally, the proposed model addresses the existing research gap identified in previous method (LSTM-CNN) by providing a real-time prediction efficiency of 94% above metrics, showcasing the potential for proactive traffic management and edge-deployable systems in urban safety applications.

5. CONCLUSION

In conclusion, the extensive study revealed that the developed hybrid-based CNN-DistillBERT framework for predicting the severity of road accident using multi-sources dataset, demonstrated and effective approach in capturing both spatial ban temporal dependency pattern due to the combination of convolutional feature extraction and transformer-based attention mechanism. Furthermore, experimental study revealed that the CNN-DistillBERT with SMOTE-ENN balancing technique achieved the best prediction performance with and accuracy, precision, recall and f1-socre of 94.66%, 94.64%, 94.66% and 94.64%, which outperform the benchmark CNN-LSTM, and Bi-LSTM. Findings further showcase that, with proper data balancing and good feature extraction (sequence-based feature) prediction can be drastically improved and model overfitting can me mitigated. Lastly, the lightweight nature of the model makes it suitable for real-time traffic accident severity prediction for supporting intelligent based transportation system and proactive traffic management in case there is an emergency.

REFERENCES

Ahmed, S., Hossain, M. A., Ray, S. K., Bhuiyan, M. M. I., & Sabuj, S. R. (2023). A study on road accident prediction and contributing factors using explainable machine learning models: Analysis and performance. *Transportation Research Interdisciplinary Perspectives*, *19*, 100814. https://doi.org/10.1016/j.trip.2023.100814

Ahmed, S. K., Mohammed, M. G., Abdulqadir, S. O., El-Kader, R. G. A., El-Shall, N. A., Chandran, D., Rehman, M. E. U., & Dhama, K. (2023). Road traffic accidental injuries and deaths: A neglected global health issue. *Health Science Reports*, *6*(5), e1240. https://doi.org/10.1002/hsr2.1240

Alaei, S. (2023). *An Automated Discharge Summary System Built for Multiple Clinical English Texts by Pre-trained DistilBART Model*. https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-222616

Alharbi, K., & Haq, M. A. (2024). Enhancing Disaster Response and Public Safety with Advanced Social Media Analytics and Natural Language Processing. *Engineering, Technology & Applied Science Research*, *14*(3), 14212–14218. https://doi.org/10.48084/etasr.7232

Azhar, A., Rubab, S., Khan, M. M., Bangash, Y. A., Alshehri, M. D., Illahi, F., & Bashir, A. K. (2023). Detection and prediction of traffic accidents using deep learning techniques. *Cluster Computing*, *26*(1), 477–493. https://doi.org/10.1007/s10586-021-03502-1

Berhanu, Y., Alemayehu, E., & Schröder, D. (2023). Examining Car Accident Prediction Techniques and Road Traffic Congestion: A Comparative Analysis of Road Safety and Prevention of World Challenges in Low-Income and High-Income Countries. *Journal of Advanced Transportation*, *2023*, 1–18. https://doi.org/10.1155/2023/6643412

Boukerche, A., & Wang, J. (2020). Machine Learning-based traffic prediction models for Intelligent Transportation Systems. *Computer Networks*, *181*, 107530. https://doi.org/10.1016/j.comnet.2020.107530

Ekanem, I. (2025). Analysis of Road Traffic Accident Using AI Techniques. *Open Journal of Safety Science and Technology*, *15*(1), Article 1. https://doi.org/10.4236/ojsst.2025.151004

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

Fu, X., Liu, J., Jones, S., Barnett, T., & Khattak, A. J. (2022). From the past to the future: Modeling the temporal instability of safety performance functions. *Accident Analysis & Prevention*, *167*, 106592. https://doi.org/10.1016/j.aap.2022.106592

Jaulkar, S., & Parihar, A. (2025). Different types of injury associated with road traffic accidents. *Multidisciplinary Reviews*, *8*(11). https://malque.pub/ojs/index.php/mr/article/view/5545

Kashinath, S. A., Mostafa, S. A., Mustapha, A., Mahdin, H., Lim, D., Mahmoud, M. A., Mohammed, M. A., Al-Rimy, B. A. S., Fudzee, M. F. M., & Yang, T. J. (2021). Review of Data Fusion Methods for Real-Time and Multi-Sensor Traffic Flow Analysis. *IEEE Access*, *9*, 51258–51276. https://doi.org/10.1109/ACCESS.2021.3069770

Mahmoud, A., & Mohammed, A. (2024). Leveraging Hybrid Deep Learning Models for Enhanced Multivariate Time Series Forecasting. *Neural Processing Letters*, *56*(5), 223. https://doi.org/10.1007/s11063-024-11656-3

Manzoor, M., Umer, M., Sadiq, S., Ishaq, A., Ullah, S., Madni, H. A., & Bisogni, C. (2021). RFCNN: Traffic Accident Severity Prediction Based on Decision Level Fusion of Machine and Deep Learning Model. *IEEE Access*, *9*, 128359–128371. https://doi.org/10.1109/ACCESS.2021.3112546

Rahim, M. A., & Hassan, H. M. (2021). A deep learning based traffic crash severity prediction framework. *Accident Analysis & Prevention*, *154*, 106090. https://doi.org/10.1016/j.aap.2021.106090

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

Sajadi, P., Qorbani, M., Moosavi, S., & Hassannayebi, E. (2024). *Accident Impact Prediction based on a deep convolutional and recurrent neural network model* (arXiv:2411.07537). arXiv. https://doi.org/10.48550/arXiv.2411.07537