International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(7s) 2026 pp. 575-588
https://doi.org/10.70917/ijcisim-2026-1839
Article

SKFA-LSTM: Spatial Keypoint Feature Aggregation with Long Short term Memory for Predicting Road Traffic Accident

Nithin Krishne Gowda[1]*, Raviprakash Madenur Lingaraju[2]

[1] Department of Computer Science Engineering, Kalpataru Institute of Technology, Tiptur, Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India.
Postal Address: Kalpataru Institute of Technology, B.H. Road, Tiptur, Tumakuru District, Karnataka 572201, India.
[2] Department of Artificial Intelligence and Machine Learning, Kalpataru Institute of Technology, Tiptur, Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India.
Postal Address: Kalpataru Institute of Technology, B.H. Road, Tiptur, Tumakuru District, Karnataka 572201, India.
***Corresponding Author**:Email: nithink666@gmail.com

Abstract: Road traffic accident prediction focuses on analyzing traffic patterns to identify potential risks and improve road safety. It considers important factors such as vehicle movement, environmental conditions, and road interactions to support proactive decision-making and reduce accident severity. However, accurately modeling traffic data is challenging because of its highly dynamic, non-linear, and spatio-temporal characteristics, which often result in lower prediction performance. To address this challenge, this study proposes a Spatial Keypoint Feature Aggregation with Long Short-Term Memory (SKFA-LSTM) model for accurate road traffic accident prediction. The SKFA module identifies the most informative spatial regions, such as vehicles and intersections, and provides enhanced feature representations to the LSTM network. This enables the model to effectively learn temporal dependencies from significant traffic patterns and improves prediction accuracy. Experimental results demonstrate that the proposed SKFA-LSTM achieves a mean Average Precision (mAP) of 93.00% on the DOTA dataset and 99.98% on the CCD dataset, outperforming existing methods such as TempoLearn. The proposed approach provides an effective solution for intelligent traffic monitoring and early accident prediction, contributing to safer and more reliable transportation systems.

Keywords: Mean Average Precision, Road traffic accident, Spatial Keypoint Feature Aggregation with Long Short Term Memory, Spatio-temporal dependencies, Vehicles

Introduction

Road traffic accidents constitute a persistent threat to public health worldwide and result in considerable economic losses. Based on World Health Organisation (WHO), nearly 1.35 million [1] people annually lose their lives because of road accidents, with another 20 to 50 million people suffering from non-fatal injuries, which leads to long-term disabilities. The scenario is particularly critical in low and middle-income countries, where inadequate emergency response systems, poor infrastructure, and high vehicular density exacerbate road safety challenges [2]. Meanwhile, a direct correlation exists between increasing traffic size and the maximum number of road accidents that cause deaths and injuries over each continent [3] [4]. High-frequency accidents lead to escalating social and economic challenges, including enhanced medical expenses and lower workforce productivity [5]. Predicting accident risk degree represents a significant aspect of accident management, which makes rescue personnel analyze traffic accident risks and their impacts [6] [7]. At the same time, developing an efficient accident management procedure via enhancing critical influencing factors plays a major role in avoiding accidents [8] [9]. This decrease probability and traffic accident losses [10] as well as increase overall road safety [11] [12].

Machine Learning (ML), statistical methods, and Deep Learning methods are the three primary approaches used to predict traffic accidents. Statistical methods determine historical accident data [13] to evaluate patterns depending on factors like weather, road conditions and traffic volume [14]. However, statistical methods based on linear assumptions struggle to capture nonlinear relationships in traffic data. To solve this issue, ML is utilised extensively to predict traffic, which allows for predicting crash occurrences and injury severity using data encompassing various factors from street and road events [15]. Moreover, ML [16] [17] provide more flexible modelling of intricate data but overlooks spatial dependencies, which leads to suboptimal performance in urban settings. Recent advances in DL [18] solve this issue through capturing non-linear [19] and spatial relationships more effectively, which enhance traffic accident prediction accuracy and minimise associated risks due to enabling timely interventions and more efficient traffic management [20].

Problem Statement and Objective

Road traffic accidents remain a primary global concern that causes injuries, significant loss of lives, and economic burden. However, existing methods for predicting accidents struggle to capture non-linear, highly dynamic, and spatio-temporal dependencies in traffic data, which results in inaccurate performance and less robustness in different accident scenarios. The objective of this research is to propose a robust accident prediction method using SKFA-LSTM to integrate spatial and temporal patterns in traffic data, which enables high accuracy and Mean Average Precision (mAP).

Contribution

- Median filter eliminates noise through conserving significant edges and object boundaries while Contrast Limited Adaptive Histogram Equalization (CLAHE) enhance local contrast and show clear visibility.
- VGG19 extracts features and capture both high-level semantic features and fine-grained information. This makes a robust representation of traffic scenes that enhances accuracy and reliability of accident prediction.
- SKFA-LSTM capture temporal modelling via selecting most relevant features from sequential traffic data, that decrease noise and redundancy. This leads to a more accurate prediction of road traffic accidents via emphasising critical patterns.

The research paper is structured as follows: Section 2 provides a literature survey; and Section 3 provides a proposed method. Section 4 analyses experimental results; and the conclusion of this research paper is given in Section 5.

Literature Survey

Soe Sandi Htun et al. [21] developed a TempoLearn that employs spatio-temporal learning for detecting traffic accidents. TempoLearn involves temporal convolutions, which identifies abnormalities and a dilation factor to attain high receptive fields. TempoLearn comprises two primary elements: accident localization, which predicts occurrence time of an accident within a video and accident classification, which classifies accident type depending on localization results. However, TempoLearn struggled with complex accident scenarios due to limited diversity in training data, which results in less robustness. Aonan Cheng et al. [22] established a K-CBST You Only Look Once (YOLO) to detect remote sensing objects. The K-CBST-YOLO combine the Swin transformer and Convolutional Block Attention Module (CBAM) to improve global semantic understanding of the feature map and enhance contextual information. Moreover, Non-Maximum Suppression (NMS) was used to minimise confidence of overlapping candidate boxes which decreases interference with subsequent detection outcomes. Nevertheless, K-CBST-YOLO performance degrades for densely packed objects due to the transformer's coarse feature representation. Farhan Mahmood et al. [23] presented a Graph Convolutional Recurrent Network (GCRN) to address visual biases for road traffic accidents. GCRN minimize bias and demonstrates better generalisation to unseen data by incorporating scene-agnostic geometric information during training. However, GCRN inherit structural bias from graph construction due to manually defined graph topology, which does not capture complex spatial-temporal relationships in accident scenarios. Shuang Li et al. [24] introduced a two-stage detection model for axis prediction and arbitrary-oriented traffic participant detection known as Axis Prediction Network. The heat-map prediction was used to fit oriented objects by employing a Gaussian model. This method utilises pixel-wise heatmap prediction to determine an object's long axis, which prevents the angle discontinuity problem in Oriented Bounding Boxes (OBB). Nevertheless, the two-stage detection model suffers from increased inference time due to sequential detection and axis prediction steps. Youcef Djenouri et al. [25] presented a Knowledge Guided DL for Near Crash Detection (KGDL-NCD) to enhance smart road safety. Initially, the data was

aggregated from each user to construct a global model that encapsulates the collective knowledge. Then, the model was aggregated from each cluster to generate local models. Users are generated with local and global models, which makes to choose the most appropriate model for crash detection. However, KGDL-NCD was limited by the quality and completeness of prior knowledge used due as insufficient knowledge biased the model's predictions.

From the overall analysis, the existing method had limitations like struggling with complex accident scenarios, inability to capture spatial-temporal relationships, and increased inference time. To solve this issue, SKFA-LSTM is proposed to accurately predict road traffic accidents. SKFA focus on the most relevant features from traffic data which minimise redundancy and noise. Then, LSTM effectively models temporal dependencies to capture spatiotemporal patterns, which enhance accuracy and minimise inference time for road traffic accidents.

Proposed Methodology

In this research, SKFA-LSTM is proposed for road traffic accident prediction. Initially, CCD and DOTA datasets are used to determine model performance. Median filter is applied to remove noise, whereas CLAHE enhance local contrast for accurate road traffic analysis. Then, VGG19 is employed to extract the features while the proposed SKFA-LSTM predicts road traffic accidents accurately. Figure 1 depicts the workflow of SKFA-LSTM method.

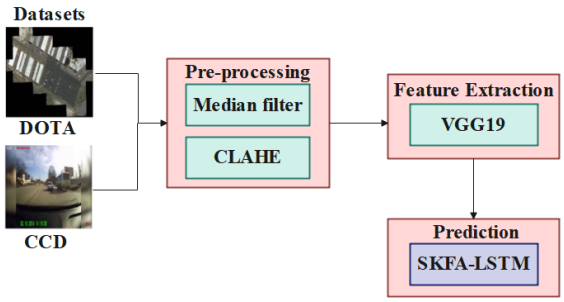


Fig. 1. Workflow of proposed method.

Datasets

This research employs DOTA and CCD datasets to analyze model performance for road traffic accident prediction. A detailed description of this dataset is mentioned below.

DOTA: This dataset is especially designed to detect aerial image objects. DOTA [26] has 2,806 aerial images and 188,282 object instances with image sizes varying from small pixels to large pixels. Images are collected from diverse sources and encompass a variety of scenes and backgrounds that exhibits a wide range of scales, object shapes, and orientations. Figure 2 shows a sample image from the DOTA dataset.

Fig. 2. Sample image from the DOTA dataset.

CCD: It comprises dashcam videos recorded under diverse environmental conditions. It has 4,500 videos, of which 1,500 are positive (accident) videos and 3000 are negative (normal driving) videos. Each video in CCD [27] is sampled at 10 frames per Second (FPS) and has 50 frames, i.e., a duration of 5 seconds per video. The dataset is split into 80% training and 20% testing, and Figure 3 shows sample frames extracted from the CCD dataset. The obtained images are pre-processed using a median filter and CLAHE to minimise noise and enhance quality.

Fig. 3. Sample frames converted from CCD datasets.

Pre-processing

After obtaining images, a median filter is used to minimise noise by preserving edges and CLAHE is applied to enhance local contrast. These methods enhance visibility and quality of features for accurate road traffic analysis. A detailed process for this method is explained as follows.

Median filter: CCD and DOTA have noise due to compression artifacts, sensor imperfections, varying environmental conditions, and motion blur. To solve this issue, median filter [28] eliminate noise by preserving edges and object boundaries, which are crucial for accurately identifying pedestrians, vehicles, and road markings. A primary benefit of using a median filter is preserving edges that are significant for image analysis. This method moves pixel by pixel across the image and replaces all values with the median value of nearby pixels. Mathematical formula of using median filter is represented in Equation (1),

$$y(t) = median((x(t - T/2), x(t - T + 1), ; x(t); x(t + T/2)) \quad (1)$$

Where $y(t)$ demonstrates output value of median filter at time , indicates input at time, and represents window size.

CLAHE: It enhance local contrast and avoids over-amplification of noise through limiting contrast enhancement in homogenous regions. In CLAHE [29], clip limit is set to 2.0 and tile grid size as which are the major parameters that manage improved image quality. Figure 4 depicts a sample image after applying CLAHE a) DOTA b) CCD datasets. The pre-processed images are fed as input into the feature extraction stage.

(a) DOTA

(b) CCD

Fig. 4. Sample image after applying CLAHE (a) DOTA (b) CCD datasets.

Feature Extraction

After pre-processing, VGG19 is used to extract the features and captures rich hierarchical features, which enhance model pe rformance. VGG19 [30] employs small convolutional kernels, which enables it to effectively identify fine-grained information like road markings, vehicle edges, and accident-relevant objects. For road traffic accident prediction, these detailed spatial features are important to understand scene context and distinguish between accident and normal scenes, which enhances the robustness and accuracy of prediction. VGG19 has 19 connection layers with 16 convolution layers, which extract the features. Moreover, max-pooling layer minimises the features and prevents overfitting. Each convolution layer output is indicated using Equation (2),

$$C_j^l = \varphi\ (\sum_{i=1}^{M^{l-1}} \quad C_i^{l-1} \times k_{ij}^l + b_j^{l-1}) \quad (2)$$

Where $\times$ indicates a convolutional function that represents the connection among weights of i^{th} and j^{th} features, b_j represents bias value, and φ denotes the activation function. Later, SKFA-LSTM is performed to predict road traffic accident.

Prediction

A extracted features are passed into SKFA-LSTM for predicting road traffic accidents. SKFA select most appropriate regions and ensures that features from significant areas like pedestrians and vehicles are emphasised. LSTM [31] is used to capture temporal dependencies among consecutive frames, which allows the model to understand patterns and event progression over time. Hence, it makes an accurate prediction of abnormal traffic events and improve model's ability to predict accidents. A detailed process of proposed SKFA-LSTM is elaborated below.

SKFA: It is applied to combine information from different regions within a frame into a unified representation. SKFA aggregates local key points and capture object locations, orientations, and interactions instead of relying on global features. This process ensure network to learn fine-grained spatial relationships among traffic participants, which are essential for prediction. By integrating these keypoints into a compact descriptor, the model obtains better scene understanding and minimize redundancy, which leads to more robust accident prediction. A simple k-Nearest Neighbor search is applied to sample local regions, which ensures that the number of extracted local regions matches number of generated spatial keypoints. Assume local neighbourhoods captured through spatial keypoints, which are grouped to create a tensor and given in Equation (3),

$$Cps_j = kNN(Cp_h|sk_j, h = 0,1,\ldots,H-1) \quad (3)$$

Where h represents number of captured points of spatial keypoint. Moreover, normalized spatial keypoints is used for capturing local spatial patterns. A normalized spatial keypoints Sk_j^{norm} is computed using Equation (4) by averaging location coordinates of all points.

$$Sk_j^{norm} = average(Cps_j) \quad (4)$$

Then, kNN search is used on normalised spatial keypoints for capturing local spatial patterns. The search outcomes are grouped to obtain $M \times K \times 3$ where K represents number of normalized spatial keypoints. The $M \times 256$ pattern feature is extracted through same method as acquiring detail feature. At last, pattern feature is concatenated to create $M \times 512$ intermediate representation, which is passed via a Fully Connected (FC) layer for final pattern. Then, the output of SKFA is fed into the LSTM layer to predict road traffic accidents.

LSTM: It captures long-term temporal dependencies and retains significant motion patterns to predict accidents. LSTM processes sequential input data via an LSTM layer that captures and retains temporal dependencies, followed by an output layer that produces final prediction outcomes. Each LSTM unit handles cell state to store information across a long sequence which enabling the model for learning long-term dependencies. A forward pass via each LSTM cell at time t is expressed in Equations (5) to (10)

$$i_t = \sigma\ (W_{xi}x_t + W_{hi}h_{t-1} + b_i) \quad (5)$$

$$f_t = \sigma\ (W_{xf}x_t + W_{hf}h_{t-1} + b_f) \quad (6)$$

$$g_t = tanh\ (W_{xg}x_t + W_{hg}h_{t-1} + b_g) \quad (7)$$

$$o_t = \sigma\ (W_{xo}x_t + W_{ho}h_{t-1} + b_o) \quad (8)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (9)$$

$$h_t = o_t \odot tanh(c_t) \quad (10)$$

Where x_t represents input at time t, h_t indicates hidden state at time t, i_t, o_t, f_t, denotes input, output gate, and forget cell, σ determines sigmoid activation function, and $\odot$ indicates element-wise multiplication. Hyperparameters are selected using Grid search to ensure optimal model performance. The learning rate of 0.01 with Adam optimizer is employed to ensure rapid and stable convergence, tanh activation function models non-linear patterns, 0.4 dropout to prevent overfitting, and binary cross-entropy loss function provides reliable performance for bianry accident prediction.

Algorithm 1 shows pseudo code of proposed method to ensure reproducibility.

Algorithm 1:

Input: Datasets {CCD, DOTA}
Output: Accident and Non-Accident prediction label
Start
Step 1: Datasets
i) Load CCD and DOTA datasets D
ii) Split datasets into 80% training ad 20% testing
Step 2: Pre-processing
For each dataset D in {CCD, DOTA} do
For each input frame in D do
i) Apply median filter and CLAHE
ii) Store pre-processed frame
End For
End For
Step 3: Feature Extraction
For each pre-processed frame I_i where I_i represents input frame in i^{th} sample
i) Pass I_i through VGG19 convolutional layers
ii) Extract deep feature vector F_i
iii) Store F_i
End For
Step 4: Prediction
For each feature vector F_i do
i)Detect spatial keypoints K_i
ii) For each keypoint k in K_i do
a)Perform KNN search to determine local neighborhood
b) Calculate normalized spatial keypoint
c) Extract local pattern features
End For
iv) Concatenate detail and pattern features
Generate and pass the aggregated spatial features to LSTM
For each time step t in feature sequence do
i) Compute input gate, forget data, output gate
ii) Update cell state and hidden state
iii) Pass final hidden state to FC
iv)Apply sigmoid for class prediction
v) Obtain predicted label y_i
Return y_i
End

Experimental Results

The SKFA-LSTM is simulated by utilizing a Python 3.4 environment with an Intel i7 environment, Windows 10 operating system, and 64 GB Random Access Memory (RAM), respectively. The performance metrics like accuracy, Area Under Curve (AUC), mean Time-To-Accident (mTTA), and mean Average Precision (mAP) are used to calculate model performance using Equations (11) to (14)

$$Accuracy = \frac{TP+TN}{TN+FP+FN+TP} \quad (11)$$

$$mTTA = \frac{1}{N}\sum_{i=1}^{N} \quad TTA_i \quad (12)$$

$$AUC = \int_0^1 \quad TPR(FPR^{-1}(x))dx \quad (13)$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} \quad AP_i \quad (14)$$

Where N determines number of accident classes, TP denotes True Positive, AP_i refers average precision of i^{th} class, FPrepresents False Positive, TPR illustrates True Positive Rate, FN indicates False Negative, and TN denotes True Negative.

Performance Analysis

Table 1 represents the evaluation of different prediction methods on the CCD and DOTA datasets. Compared to existing methods like Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), CNN-LSTM, and Vision Transformer (ViT), the proposed SKFA-LSTM obtains a high accuracy of 99.98% and 93.10% due to SKFA concentrating on the most informative regions, which minimises the impact of irrelevant background noise and enhances the quality of input data for the temporal model. Moreover, LSTM captures sequential dependencies and motion patterns over frames, which are important for detecting accidents. The integration of spatially focused features and learning temporal dependencies enhances the model's ability to determine subtle accidents. Hence, the proposed method attains more reliable predictions compared to existing methods on different datasets.

Table 1. Evaluation of different prediction methods on CCD and DOTA dataset

Methods	Datasets	mAP (%)	mTTA (s)	AUC (%)	Accuracy (%)
RNN	CCD	95.27	4.2	93.18	94.46
	DOTA	88.46	2.5	86.34	87.62
GRU	CCD	96.19	3.9	94.21	95.36
	DOTA	89.68	2.8	87.49	88.82
CNN–LSTM	CCD	97.83	4.1	95.47	96.94
	DOTA	90.57	2.95	88.39	89.68
ViT	CCD	98.64	4.55	96.26	97.52
	DOTA	91.46	2.36	89.28	90.59
SKFA-LSTM	CCD	99.98	4.8	99.99	99.98
	DOTA	93.00	3.2	92.52	93.10

Table 2 indicates the performance of the ablation study for individual components on the CCD and DOTA datasets. Compared to individual components like LSTM, ResNet+SKFA+LSTM, VGG+LSTM, the VGG+SKFA+LSTM achieves superior performance due to effectively extracting deep hierarchical features that capture both high-level and fine-grained spatial patterns for traffic scenes. SKFA refines these features through focusing on primary spatial regions which emphasise accident-relevant areas. Compared to other methods, VGG's simpler model solves overfitting and retains richer low-level details. This outperforms LSTM and VGG+LSTM by integrating spatial with temporal dependencies for more accurate prediction.

Table 2. Performance of ablation study on individual components

Methods	Datasets	mAP (%)	mTTA (s)	AUC (%)	Accuracy (%)
LSTM	CCD	95.27	4.6	93.66	94.58
	DOTA	88.46	3.4	86.71	87.69
ResNet+ SKFA+ LSTM	CCD	97.68	3.85	95.81	96.72
	DOTA	90.53	2.85	88.73	89.69
VGG+LSTM	CCD	96.38	4.66	94.69	95.56
	DOTA	89.48	3.0	87.64	88.59
VGG + SKFA+ LSTM	CCD	99.98	4.8	99.99	99.98
	DOTA	93.00	3.2	92.52	93.10

Table 3 provides an evaluation of k-fold validation for the proposed method on different datasets. The k=5 obtains high accuracy due to the balance between bias and variance. For k=2 and 3, model obtains less bias and high variance, leading to inaccurate performance. A k=5, variance is minimised without significantly increasing bias, which offers a more reliable and stable decision boundary. This balance leads to better generalization performance compared to other folds

Table 4 determines the analysis of computational complexity for different prediction methods on the CCD and DOTA datasets. Compared to existing methods, SKFA-LSTM obtains less training time of 20.8 ms and 23.22 ms on CCD and DOTA datasets due to SKFA minimize input feature dimensionality by concentrating only on the most relevant spatial keypoints. This decreases redundant computations and enhances data processing in traffic accidents with rapid convergence. The p-value identifies whether observed difference is statistically significant, whereas Confidence Interval (CI) provide a range of values within which a true population parameter is expected, that offers high reliability.

Table 3. Evaluation of k-fold validation for proposed method

k-Fold	Datasets	mAP (%)	Accuracy%
k=2	CCD	98.90	98.85
	DOTA	91.28	90.55
k=3	CCD	99.95	99.98
	DOTA	92.38	91.26
k=5	CCD	99.98	99.98
	DOTA	93.00	93.10

Table 4. Analysis of computational complexity on CCD and DOTA datasets

Datasets	Methods	Training Time (ms)	Execution Time (s)	Memory Consumption (MB)	Inference speed(ms)	p value from t-test	CI (%)
CCD	RNN	32.42	20.25	610	2.3	0.0041	93.01 – 95.71
	GRU	22.62	21.05	750	2.9	0.0038	85.46 – 89.55
	CNN–LSTM	28.45	23.10	595	2.8	0.0035	94.02 – 96.48
	ViT	25.31	36.47	610	3.6	0.0034	86.74 – 90.69
	SKFA-LSTM	20.8	15.8	520	1.58	0.0032	95.77 – 97.85
DOTA	RNN	34.75	19.12	621	2.21	0.0029	87.66 – 91.49
	GRU	47.88	23.21	772	2.94	0.0027	96.45 – 98.28
	CNN–LSTM	30.95	18.12	710	4.65	0.0025	88.61 – 92.34
	ViT	48.05	17.55	635	3.3	0.0023	99.44 – 100.00
	SKFA-LSTM	23.22	15.18	544	1.18	0.0021	91.26 – 94.68

Table 5. Cross-dataset validation in terms of training on the CCD and testing on DoTA dataset

Methods	mAP %	AUC %	Accuracy %
RNN	95.20	95.50	95.80
GRU	96.10	96.40	96.80
CNN–LSTM	97.00	97.20	97.50
ViT	97.50	97.60	97.90
SKFA-LSTM	97.98	98.00	98.48

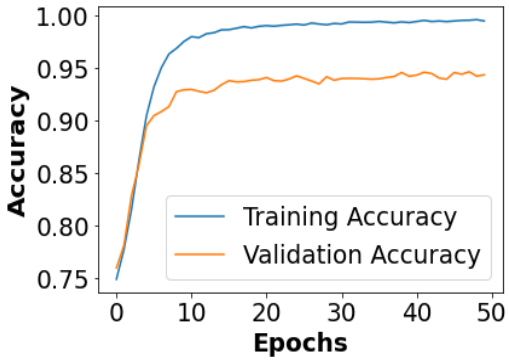
Table 6. Cross dataset validation in terms of training on DOTA and testing on CCD dataset

Methods	mAP %	AUC %	Accuracy %
RNN	87.91	88.73	87.54
GRU	88.68	89.42	88.21
CNN–LSTM	89.82	90.58	89.15
ViT	87.90	91.15	90.00
SKFA-LSTM	90.82	92.00	91.28

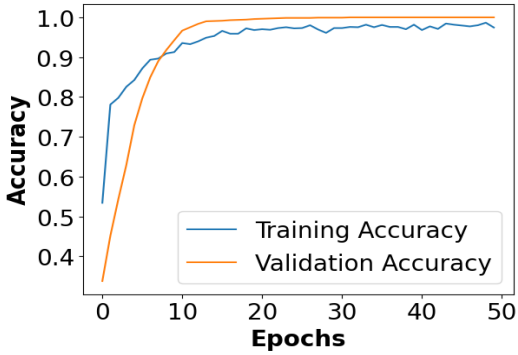
Table 5 shows cross-dataset validation in terms of training on the CCD and testing on the DOTA dataset. Table 6 demonstrates cross-dataset validation with respect to training on DOTA and testing on the CCD dataset to evaluate generalization ability on unseen data. Instead of testing on a subset of the same dataset, the model is trained on one

dataset and tested on another, which ensures a more realistic performance. This reveals the extent to which the model manages domain shifts, variations, and noise in data collection.

Figure 5 depicts training and validation accuracy over 50 epochs for the proposed method, a) DOTA b) CCD datasets. For DOTA, training accuracy reaches 100% whereas validation accuracy stabilises around 95% which shows better generalization with less overfitting. For CCD, both training and validation accuracy rapidly converge above 90% which represents better performance with effective learning. Overall, these outcomes confirm that the model performs well on different datasets with better reliability.



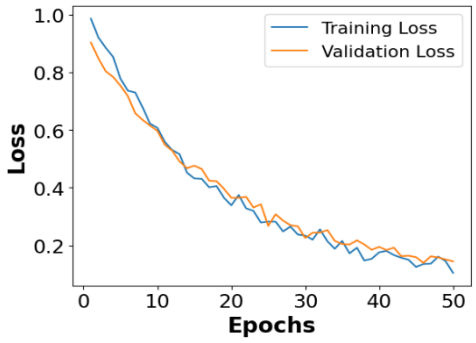
(a)



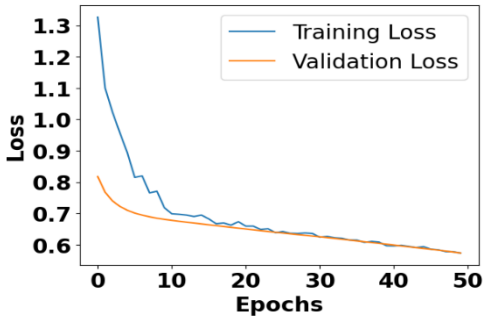
(b)

Fig.5. Evaluation of Epoch vs Accuracy for the proposed method (a) DOTA (b) CCD datasets.

Figure 6 represents training and validation loss for SKFA-LSTM a) DOTA b) CCD datasets. For DOTA, both validation and training loss minimize rapidly, which indicates consistent learning with less overfitting. For CCD, validation loss is slightly lower than training loss, which shows better robustness to unseen data. Therefore, these loss curves show effective performance and stable convergence across both datasets.

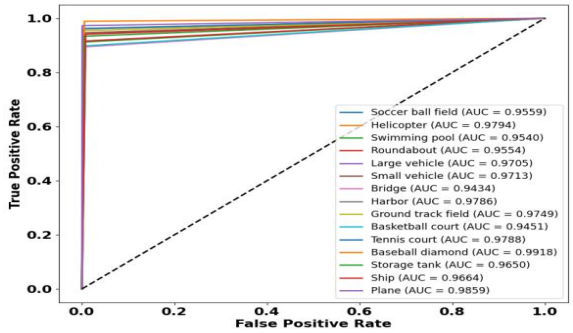


(a)

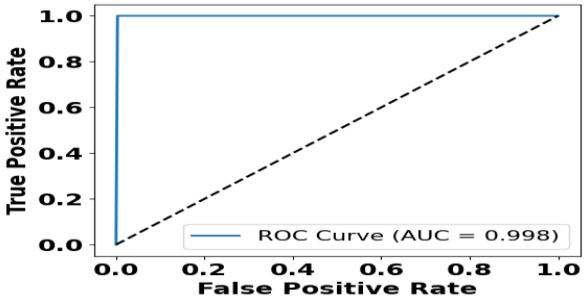


(b)

Fig. 6. Analysis of Epoch vs Loss for proposed method (a) DOTA (b) CCD datasets.



(a)



(b)

Fig. 7. Performance analysis of ROC curve for the proposed method (a) DOTA (b) CCD datasets.

Figure 7 demonstrates the Receiver Operating Characteristics (ROC) curve for the proposed method a) DOTA b) CCD. The AUC value is obtained above 0.93 in DOTA dataset, which indicates robust prediction performance across different classes. For CCD, the proposed method obtains 0.998 AUC, which shows high TPR with less FPR. Overall, these outcomes represent robustness and efficiency of the proposed method in both datasets.

Figure 8 illustrates a confusion matrix to determine model performance by analyzing FP, TP, TN, and FN a) DOTA b) CCD datasets. DOTA dataset obtains robust classification performance with less misclassification, whereas CCD odataset btains 299 normal and 600 accident cases classified correctly. These outcomes represent the model's reliability and robustness in managing both complex binary and multi-class classification tasks.

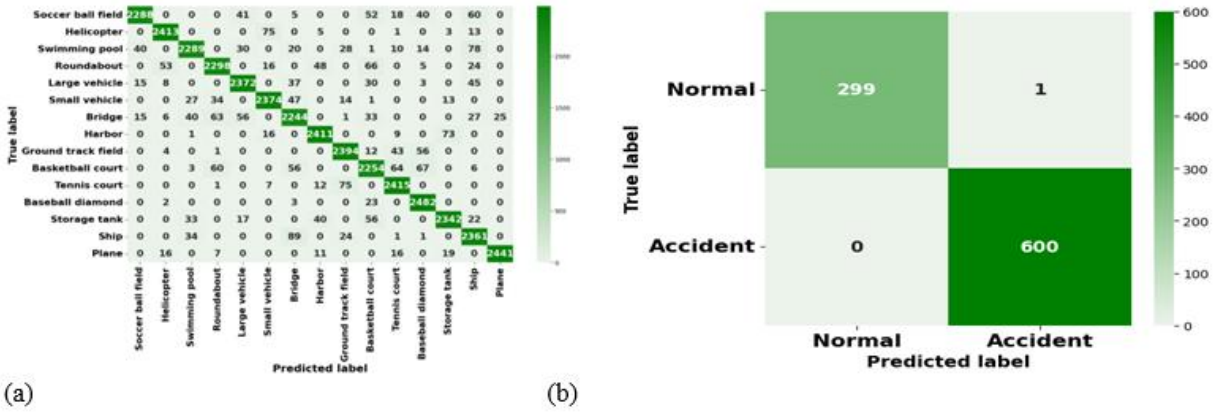


Fig. 8. Evaluation of confusion matrix for proposed method (a) DOTA (b) CCD datasets.

Comparative Analysis

Table 7 demonstrates a comparative analysis of existing methods on DOTA and CCD datasets. Compared to existing methods like [21] to [24], the proposed SKFA-LSTM obtains a high mAP of 93.00% and 99.98% on DOTA and CCD due to extracting the

most informative spatial key points, which minimize irrelevant features and enhance feature quality. Then, LSTM effectively captures temporal dependencies that enhance prediction consistency across sequences, which leads to more accurate performance.

Table 7. Comparative analysis of existing methods on the DOTA and CCD datasets.

Methods	Datasets	mAP (%)	AUC (%)	mTTA (s)
TempoLearn [21]	DOTA	16.5	92.4	N/A
	CCD	N/A	89.51	N/A
K-CBST YOLO [22]	DOTA	0.784	N/A	N/A
EGSTL [23]	DOTA	N/A	N/A	2.93
	CCD	N/A	N/A	4.79
ResNet101 [24]	DOTA	81.75	N/A	N/A
Proposed SKFA-LSTM	DOTA	93.00	92.52	3.2
	CCD	99.98	99.99	4.8

Discussion

A benefits of proposed method and disadvantage of existing methods are provided. The limitation of existing methods, like TempoLearn [21] struggles with complex accident scenarios due to limited diversity in training data, which results in less robustness. K-CBST-YOLO [22] performance degrades for densely packed objects due to the transformer's coarse feature representation. GCRN [23] inherits structural bias from graph construction due to a manually defined graph topology, which does not capture complex spatial-temporal relationships. Two-stage detection model [24] suffers from increased inference time, while KGDL-NCD [25] was constrained by the quality and completeness of its prior knowledge, leading to suboptimal performance. The SKFA-LSTM overcomes these existing method limitations due to its ability to capture both temporal and spatial patterns effectively. SKFA focus on spatial features that eliminate unwanted information and improve feature representation while LSTM captures temporal dependencies effectively. enhance prediction performance and decrease FP. Overall, the proposed SKFA-LSTM offers a robust and effective model for reliable accident prediction.

Conclusion

This research proposed SKFA-LSTM to accurately predict road traffic accidents. SKFA focus on the most appropriate spatio-temporal features from traffic data, whereas LSTM employs sequential learning ability for capturing temporal dependencies and dynamic traffic patterns efficiently. VGG19 is used to extract the features with small convolutional kernels, which capture both high-level semantic features and fine-grained information effectively.

This makes a robust representation of road scenes that enhances accident prediction accuracy. Experimental analysis of DOTA and CCD datasets shows that the proposed SKFA-LSTM attains superior performance in terms of accuracy, AUC, mAP compared to state-of-the-art methods with less training time and memory consumption. Comparative analysis shows that SKFA-LSTM obtains high mAP of 93.00% and 99.98% on DOTA and CCD, compared to existing methods like TempoLearn. In future, an improved optimization method will be employed to choose the most appropriate features that enhance model performance.

1. Acknowledgments The authors would like to thank the Department of Computer Science Engineering and Department of Artificial Intelligence and Machine Learning, Kalpataru Institute of Technology, Tiptur, for their support.

2. Funding This research received no external funding.

3. Author Contributions Nithin Krishne Gowda: Conceptualization, Methodology, Software, Writing – Original Draft. Raviprakash Madenur Lingaraju: Supervision, Validation, Writing – Review and Editing.

4. Conflict of Interest The authors declare no conflict of interest.

5. Data Availability Statement

The datasets used in this study are publicly available. The DOTA dataset is available at: https://www.kaggle.com/datasets/chandlertimm/dota-data. The Car Crash Dataset (CCD) is available at: https://www.kaggle.com/datasets/asefjamilajwad/car-crash-dataset-ccd.

Biographical Sketch

Nithin Krishne Gowda

Nithin Krishne Gowda is a researcher in the Department of Computer Science Engineering, Kalpataru Institute of Technology, Tiptur. His research interests include deep learning, computer vision, and intelligent transportation systems, with a focus on road traffic accident prediction. He has authored research work on applying spatio-temporal deep learning models to real-world safety problems.

Raviprakash Madenur Lingaraju

Raviprakash Madenur Lingaraju is with the Department of Artificial Intelligence and Machine Learning, Kalpataru Institute of Technology, Tiptur. His research interests include machine learning, deep learning applications, and AI-driven solutions for traffic safety and smart systems. He has guided and contributed to research on predictive modeling for real-world engineering problems

References

1. S. Pourroostaei Ardakani, X. Liang, K.T. Mengistu, R.S. So, X. Wei, B. He, A. Cheshmehzangi, “Road car accident prediction using a machine-learning-enabled data analysis,” Sustainability Vol. 15, Iss 7, p. 5939 (2023).
2. W. Sun, L.N. Abdullah, F.B. Khalid, P. Suhaiza Binti Sulaiman, “EAR-CCPM-Net: A Cross-Modal Collaborative Perception Network for Early Accident Risk Prediction,” Applied Sciences Vol. 15, Iss 17, p. 9299 (2025).
3. W.A. Syafei, A. Wibowo, I.A. Ashari, “A Novel SW-KMA-Bi-LSTM Approach for Improving Traffic Flow Prediction,” International Journal of Intelligent Engineering & Systems Vol. 18, Iss 7, pp. 82-97 (2025).
4. K.A. Sattar, I. Ishak, L.S. Affendey, S.N.B.M. Rum, “Road crash injury severity prediction using a graph neural network framework,” IEEE Access Vol. 12, pp. 37540-37556 (2024).
5. X. Wei, “Enhancing road safety in internet of vehicles using deep learning approach for real-time accident prediction and prevention,” International Journal of Intelligent Networks Vol. 5, pp. 212-223 (2024).
6. H. Abuzaid, R. Almashhour, G. Abu-Lebdeh, “Driving towards sustainability: a neural network-based prediction of the traffic-related effects on road users in the UAE,” Sustainability Vol. 16, Iss 3, p. 1092 (2024).
7. P. Sajadi, M. Qorbani, S. Moosavi, E. Hassannayebi, “Accident Impact Prediction based on a deep convolutional and recurrent neural network model,” Urban Science Vol. 9, Iss 8, p. 299 (2025).
8. Y. Pei, Y. Wen, S. Pan, “Road traffic accident risk prediction and key factor identification framework based on explainable deep learning,” IEEE Access Vol. 12, pp. 120597-120611 (2024).
9. Y. Berhanu, D. Schröder, B.T. Wodajo, E. Alemayehu, “Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing on accident and congestion-prone road networks,” Results in Engineering Vol. 23, p. 102737 (2024).
10. J. Tang, Y. Huang, D. Liu, L. Xiong, R. Bu, “Research on traffic accident severity level prediction model based on improved machine learning,” Systems Vol. 13, Iss 1, p. 31 (2025).

11. O.I. Aboulola, E.A. Alabdulqader, A.A. AlArfaj, S. Alsubai, T.H. Kim, "An automated approach for predicting road traffic accident severity using transformer learning and explainable AI technique," IEEE Access Vol. 12, pp. 61062-61072 (2024).
12. M. Wang, W.C. Lee, N. Liu, Q. Fu, F. Wan, G. Yu, "A data-driven deep learning framework for prediction of traffic crashes at road intersections," Applied Sciences Vol. 15, Iss 2, p. 752 (2025).
13. J. Liu, Z. Zhang, B. Lyu, R. Feng, Y. He, "A hybrid ARIMA-BP approach for superior accuracy in predicting traffic accident losses," Scientific Reports Vol. 15, Iss 1, p. 24328 (2025).
14. P. Zhang, W. Yi, Y. Song, P. Yan, P. Wu, A. Shemery, K. Hampson, A.P. Chan, "Dynamic spatiotemporal graph network for traffic accident risk prediction," GIScience & Remote Sensing Vol. 62, Iss 1, p. 2514330 (2025).
15. S.M.V. Reddy, A. Dammur, A., Narasimhamurthy, "Traffic Flow Prediction and Identifying Critical Node Networks Using Attention-based Spectral Temporal Graph Neural Network," International Journal of Intelligent Engineering & Systems, Vol. 16, Iss 6, pp. 946-955 (2023).
16. J. Chen, W. Tao, Z. Jing, P. Wang, Y. Jin, "Traffic accident duration prediction using multi-mode data and ensemble deep learning," Heliyon, Vol. 10, Iss 4, p. e25957 (2024).
17. G. Bahati, E. Masabo, "Optimizing ambulance location based on road accident data in Rwanda using machine learning algorithms," International Journal of Health Geographics Vol. 24, Iss 1, p. 23 (2025).
18. C. Chaoura, H. Lazar, Z. Jarir, "Enhancing Emergency Response in Road Accidents: A Severity Prediction Framework Using RF-RFE and Deep Learning Model," IEEE Access Vol. 13, pp. 136927-136944 (2025).
19. Y. Wang, T. Liu, Y. Lu, H. Wan, P. Huang, F. Deng, "Traffic Accident Risk Prediction of Tunnel Based on Multi-Source Heterogeneous Data Fusion," IEEE Access Vol. 12, pp. 18694-18702 (2024).
20. M. Molo, S. Changsan, L. Madares, R. Changkwanyeun, S. Wattanasoei, S. Vittaporn, P. Khamnuan, S. Pongpan, K. Pooseesod, S. Saita, "A decision tree model for traffic accident prediction among food delivery riders in Thailand," Epidemiology and Health Vol. 46, p. e2024095 (2024).
21. S.S. Htun, J.S. Park, K.W. Lee, J.H. Han, "TempoLearn network: Leveraging spatio-temporal learning for traffic accident detection," IEEE Access Vol. 11, pp. 142292-142303 (2023).
22. A. Cheng, J. Xiao, Y. Li, Y. Sun, Y. Ren, J. Liu, "Enhancing remote sensing object detection with K-CBST YOLO: integrating CBAM and swin-transformer," Remote Sensing Vol. 16, Iss 16, p. 2885 (2024).
23. F. Mahmood, D. Jeong, J. Ryu, "A new approach to traffic accident anticipation with geometric features for better generalizability," IEEE Access Vol. 11, pp. 29263-29274 (2023).
24. S. Li, C. Liu, L. Chen, "Arbitrary-oriented traffic participant detection and axis prediction for complex crossing road in aerial view," IET Intelligent Transport Systems Vol. 17, Iss 12, pp. 2419-2431 (2023).
25. Y. Djenouri, A.N. Belbachir, T. Michalak, A. Belhadi, G. Srivastava, "Enhancing smart road safety with federated learning for Near Crash Detection to advance the development of the Internet of Vehicles," Engineering Applications of Artificial Intelligence Vol. 133, p. 108350 (2024).
26. DOTA dataset link: https://www.kaggle.com/datasets/chandlertimm/dota-data (Accessed on September 18 2025)
27. CCD dataset link: https://www.kaggle.com/datasets/asefjamilajwad/car-crash-dataset-ccd (Accessed on September 18 2025)
28. A. Ijaz, B. Raza, I. Kiran, A. Waheed, A. Raza, H. Shah, S. Aftan, "Modality specific CBAM-VGGNet model for the classification of breast histopathology images via transfer learning," IEEE Access Vol. 11, pp. 15750-15762 (2023).
29. V.L. Narla, G. Suresh, C.S. Rao, M.A. Awadh, N. Hasan, "A multimodal approach with firefly based CLAHE and multiscale fusion for enhancing underwater images," Scientific Reports Vol. 14, Iss 1, p. 27588 (2024).
30. A.K. Rajendran, S.C. Sethuraman, "YogiCombineDeep: Enhanced yogic posture classification using combined deep fusion of VGG16 and VGG19 features," IEEE Access Vol. 12, pp. 139165-139180 (2024).
31. J. Zhao, J. Wu, S. Jiang, Z. Li, P. Liu, "LSTM-based classification of e-scooter trajectory features for single vs tandem riding detection," Traffic Injury Prevention (2025).