International Journal of Computer Techniques–IJCT Volume 13 Issue 3, May-June - 2026
Open Access and Peer Review Journal ISSN 2394-2231 https://ijctjournal.org/

Smart Traffic Accident Prediction System Using Machine Learning and Artificial Intelligence

POTTURU JAYA KRISHANA[1], SADDIKOOTI BHANU PRAKASH REDDY[2],DENDUKURI PRAVEEN VARMA[3],Ms.M. Vaishnavi[4].

123 UG Student, Department of Computer Science and Engineering, School of Engineering and Technology, Dhanalakshmi Srinivasan University, Trichy-621112-Tamilnadu. 4 Assistant Professor, Department of Computer Science and Engineering, Dhanalakshmi Srinivasan Institute of Technology, Trichy-621112-Tamil nadu

Email: mrkrishna8008@gmail.com, pprakash14366@gmail.com , praveenvarmadendu@gmail.com , vaishnavim.cse@dsuniversity.ac.in

Abstract — Traffic accidents continue to inflict devastating human and economic losses on a global scale, claiming approximately 1.35 million lives annually. This paper proposes STAPS-AI, an advanced Smart Traffic Accident Prediction System that tightly integrates classical machine learning (ML) algorithms with contemporary artificial intelligence (AI) architectures — including Long Short-Term Memory (LSTM) recurrent networks, Convolutional Neural Networks (CNN), and a Transformer-based attention model — to achieve robust, real-time accident risk prediction and severity classification. The system ingests heterogeneous data streams encompassing real-time traffic flow, historical accident records, meteorological readings, road geometry, driver-behavior telemetry, and connected-vehicle V2X communications, and fuses them within a unified preprocessing and feature-engineering pipeline. A comprehensive comparative evaluation is conducted across eight models — Logistic Regression, SVM, Random Forest, XGBoost, Multilayer Perceptron (MLP), LSTM, CNN-LSTM Hybrid, and a Transformer — on the UK Department for Transport Road Safety dataset comprising 1.83 million records (2015–2022). The Transformer model achieves the highest predictive performance with 96.1% accuracy, 95.8% precision, 95.2% recall, and an AUC-ROC of 0.983. A cost-sensitive four-tier severity classifier (Low / Moderate / High / Critical) provides granular, actionable risk intelligence directly consumable by traffic management centers, navigation platforms, and autonomous vehicle onboard systems. STAPS-AI outperforms all prior state-of-the-art baselines by margins of 4.8–17% in accuracy, and its modular REST-API architecture supports real-time deployment with median inference latency of 11 ms per 1,000 road segments.

Index Terms — Accident prediction; artificial intelligence; attention mechanism; convolutional neural network; deep learning; intelligent transportation systems; LSTM; machine learning; road safety; Transformer; V2X; XGBoost.

I. INTRODUCTION

Road traffic accidents remain among the most severe and persistent public-health crises worldwide. The World Health Organization (WHO) reports that roughly 1.35 million fatalities and up to 50 million non-fatal injuries occur annually on public roads, imposing an estimated economic burden equivalent to 3% of each nation's gross domestic product (GDP) [1]. Despite decades of countermeasures — including road redesign, speed enforcement, seat-belt mandates, and drunk-driving penalties — accident frequency has plateaued or increased in many low- and middle-income countries, where motorisation is accelerating faster than safety infrastructure can adapt.

Traditional road-safety engineering is inherently reactive: infrastructure is improved only after accident clusters are statistically identified, often requiring years of data accumulation before interventions are approved. Such inertia is incompatible with modern smart-city aspirations, which demand proactive, data-driven risk mitigation in real time. The convergence of four enabling technologies now makes genuinely predictive traffic safety feasible: (1) pervasive IoT sensor networks that stream high-resolution traffic and environmental data; (2) connected vehicle and V2X communication protocols delivering sub-second situational awareness; (3) big-data platforms capable of ingesting and storing terabytes of multimodal transportation data; and (4) advances in artificial intelligence — particularly deep learning — that can extract complex spatiotemporal patterns far beyond the reach of classical statistical models.

While machine learning methods such as Random Forest and Gradient Boosting have demonstrated competitive predictive accuracy on tabular accident datasets, they remain inherently limited by their inability to exploit sequential temporal dependencies or long-range spatial correlations in traffic dynamics. Artificial intelligence architectures — recurrent networks (LSTM), convolutional networks (CNN), and attention-based Transformers — are uniquely suited to capture these dynamics, yet have rarely been benchmarked against one another within a unified, reproducible experimental framework on real-world accident data at national scale.

This paper addresses these gaps by presenting STAPS-AI, an end-to-end Smart Traffic Accident Prediction System that unifies classical ML and modern AI in a single deployable pipeline. The principal contributions are:

1) A unified, multi-source data fusion architecture that integrates traffic sensors, meteorological APIs, road geometry databases, vehicle telematics, and V2X feeds into a consistent feature representation.
2) A systematic benchmark of eight ML/AI models under identical preprocessing, cross-validation, and evaluation protocols — the most comprehensive comparison reported on the UK DfT dataset to date.
3) A Transformer-based accident prediction model achieving state-of-the-art accuracy of 96.1% and AUC-ROC of 0.983 on 1.83 million real-world records.
4) A cost-sensitive four-tier severity classifier (Low / Moderate / High / Critical) with class-weighted threshold optimisation.
5) A deployable REST-API service architecture with demonstrated real-time inference at 11 ms median latency per 1,000 road segments.

The remainder of the paper is structured as follows. Section II surveys related work. Section III details the STAPS-AI architecture. Section IV describes the dataset and feature pipeline. Section V presents experimental results. Section VI discusses deployment and ethics. Section VII concludes.

II. RELATED WORK

The literature on accident prediction has evolved along two parallel trajectories: statistical-econometric models and data-driven ML/AI models. We survey each in turn, then identify the research gaps addressed by STAPS-AI.

A. Statistical and Econometric Models

Early studies relied on Poisson and negative binomial regression to model accident-count data as functions of road geometry and traffic volume [2]. Abdel-Aty and Radwan [3] extended this to ordinal logistic regression for severity modelling, achieving moderate explanatory power but highlighting strong multi-collinearity among covariates. Miaou and Lum [4] established the inadequacy of simple Poisson models due to overdispersion in real crash data, motivating the adoption of zero-inflated and hurdle variants. Though interpretable, these approaches assume linearity and independence — assumptions routinely violated in traffic environments.

B. Classical Machine Learning

Ensemble methods transformed the field. Yuan et al. [5] applied Random Forest to real-time freeway crash prediction and reported a 15% accuracy improvement over logistic regression by exploiting non-linear feature interactions. Sheykhfard and Haghighi [6] combined KNN with genetic-algorithm feature selection, reducing dimensionality while maintaining predictive power. Gradient boosting — particularly XGBoost [7] — became the dominant benchmark for tabular accident data due to its built-in handling of missing values, categorical features, and class imbalance. Several studies report XGBoost accuracies in the 91–95% range on standardised accident datasets, consistent with our own findings.

C. Deep Learning and AI Approaches

The shift toward deep learning accelerated with the availability of large-scale temporal traffic datasets. Zhang et al. [8] applied LSTM to sequential accident data and demonstrated that temporal context substantially improves short-horizon risk prediction. Spatiotemporal CNN models were applied by Yu et al. [9] to grid-based accident maps, capturing spatial autocorrelation that purely tabular models ignore. Hybrid CNN-LSTM architectures subsequently proved capable of simultaneously extracting spatial and sequential features, achieving accuracy gains of 2–4% over either architecture alone. Most recently, Moradi and Vagnoli [10] demonstrated that Transformer-based models — adapted from natural language processing — outperform LSTM on variable-length traffic event sequences, owing to multi-head attention's ability to directly model long-range temporal dependencies without vanishing-gradient limitations.

D. *Research Gaps*

Three limitations pervade the existing literature: (i) studies rarely compare ML and AI approaches under identical experimental conditions, making benchmarking unreliable; (ii) most produce binary (accident / no-accident) outputs rather than actionable severity gradations; and (iii) few present a deployable end-to-end system architecture. STAPS-AI is designed to close all three gaps simultaneously.

III. STAPS-AI SYSTEM ARCHITECTURE

STAPS-AI is built as a five-layer modular pipeline: Data Ingestion, Preprocessing and Feature Engineering, Model Ensemble, Severity Classification, and Output and Alerting. Each layer is independently deployable as a microservice, enabling horizontal scaling under high data volumes.

A. Data Ingestion Layer

Heterogeneous data streams are consolidated in a unified data lake. Five source categories are ingested: (1) Historical accident records from the UK Department for Transport (DfT), updated quarterly; (2) Real-time traffic flow parameters — speed, density, and volume at 5-minute granularity — from inductive loop detectors and radar sensors; (3) Meteorological conditions from the UK Met Office Datapoint API, providing precipitation type and intensity, visibility, temperature, wind speed, and road surface state at 1-hour resolution; (4) Road geometry and infrastructure attributes from Ordnance Survey OpenRoads — lane count, curvature, gradient, speed limit, junction classification, and road class; (5) Connected vehicle telemetry and V2X event messages where available, including harsh-braking events, near-miss detections, and hazard lights activations.

All streams are spatially indexed to standardised 500-metre road segments using a geospatial hash and temporally aligned to 5-minute epochs via forward-fill interpolation for lower-frequency sources.

B. Preprocessing and Feature Engineering

The preprocessing module executes six sequential operations: (i) Missing-value imputation using predictive mean matching for numerical features and mode imputation for categoricals; (ii) Outlier detection and Winsorisation at the 1st and 99th percentiles to suppress sensor noise; (iii) Categorical encoding via target encoding with 5-fold cross-validation to prevent target leakage; (iv) Cyclical temporal feature encoding — hour-of-day, day-of-week, and month are mapped to sine/cosine pairs to preserve circular continuity; (v) Spatial interaction feature construction — four neighbourhood-segment statistics (mean speed variance, accident density, near-miss rate, junction proximity index) are appended; (vi) Class-imbalance correction via Borderline-SMOTE, which generates synthetic minority-class instances near the decision boundary rather than throughout the minority class manifold.

C. Model Ensemble Layer

Eight candidate models are trained in parallel: Logistic Regression (baseline), SVM with RBF kernel, Random Forest (500 trees), XGBoost, Multilayer Perceptron (MLP), LSTM (2 layers, 256 hidden units), CNN-LSTM Hybrid, and a Transformer with 8 attention heads and 4 encoder layers. For the AI architectures, input sequences of 12 consecutive 5-minute epochs are used, representing one hour of prior traffic context. All models are optimised via 5-fold stratified cross-validation with Bayesian hyperparameter search (150 iterations).

A stacking meta-learner combining the three best-performing base models (Transformer, CNN-LSTM, XGBoost) via Ridge regression is evaluated as an optional ensemble stage, yielding marginal additional gains at the cost of increased inference complexity.

D. Severity Classification Module

Calibrated probability outputs from the best model (Transformer) are mapped to a four-tier risk taxonomy using cost-sensitive thresholds determined by minimising a weighted misclassification cost matrix in which Critical misclassification carries a penalty weight of 10, High carries 5, Moderate carries 2, and Low carries 1, reflecting the asymmetric real-world consequences of under-predicting severe accidents.

E. Output and Alerting Layer

Risk predictions and severity labels are published to a RESTful API consumed by: (i) Traffic Management Center (TMC) dashboards displaying heatmap overlays on geographic information systems; (ii) Variable Message Signs (VMS) controllers that automatically reduce speed limits on High/Critical segments; (iii) Navigation platforms (OpenStreetMap, future integration with TomTom and HERE) that reroute drivers away from Critical zones; (iv) Emergency dispatch systems that pre-position ambulances near Critical-flagged corridors; and (v) Autonomous vehicle OBD/V2X modules receiving structured JSON risk packets in real time.

IV. DATASET AND FEATURE ENGINEERING

A. Dataset Description

The UK Road Safety Data published by the UK Department for Transport covers all personal-injury road accidents reported to police between 2015 and 2022. After merging accident, vehicle, and casualty tables and joining with weather and road-geometry supplementary tables, the consolidated dataset comprised 1,832,441 samples. The positive-class (accident) prevalence was 4.6%, yielding a class ratio of approximately 1:21 — a severe imbalance that necessitates both oversampling and cost-sensitive learning.

B. Feature Selection

Recursive Feature Elimination with 10-fold cross-validation (RFECV) reduced the candidate feature set from 47 attributes to 22 retained features. Table I presents the eight highest-importance features ranked by both XGBoost gain scores and Transformer attention weights (averaged across all heads and layers). Strong concordance between the two ranking methods (Spearman's $\rho = 0.91$) validates feature relevance across both model families.

TABLE I. TOP-8 FEATURE IMPORTANCE: XGBOOST VS. TRANSFORMER

Feature	XGBoost Score	AI (Transformer) Score	Rank
Road Surface Condition	0.187	0.192	1
Weather Condition	0.163	0.171	2
Speed Limit (km/h)	0.154	0.148	3
Time of Day & Day of Week	0.143	0.139	4
Junction Type	0.112	0.117	5
Light Conditions	0.098	0.096	6
Traffic Volume	0.083	0.087	7
Urban / Rural Classification	0.060	0.050	8

Road surface condition ranked first in both models, reflecting the dominant influence of wet, icy, or damaged surfaces on friction and vehicle controllability. Weather condition and speed limit jointly accounted for approximately 31–33% of cumulative importance, consistent with prior literature attributing over one-third of all fatal accidents to adverse environmental conditions.

C. *Temporal Sequence Construction*

For the AI sequence models (LSTM, CNN-LSTM, Transformer), each sample is extended into a 12-step lookback window at 5-minute resolution, capturing one hour of traffic history. Samples are padded with zero vectors where fewer than 12 prior time steps exist (e.g., at the start of daily records). This temporal representation enables the AI models to detect pre-accident signatures — such as speed variance escalation and rapid density drops — that are invisible to single-snapshot models.

D. *Data Partitioning*

The dataset was split chronologically — 70% for training (2015–2020), 15% for validation (2021), and 15% for testing (2022) — to simulate realistic deployment conditions where models are evaluated on future, unseen data. All preprocessing transformers were fitted exclusively on the training partition.

V. EXPERIMENTAL RESULTS

All experiments were executed on an Ubuntu 22.04 server configured with dual Intel Xeon Platinum 8380 processors (80 logical cores, 3.3 GHz base), 256 GB DDR4 ECC RAM, and four NVIDIA A100 GPUs (80 GB HBM2 each). The software stack comprised Python 3.11, Scikit-learn 1.4, XGBoost 2.0.3, PyTorch 2.2.1, HuggingFace Transformers 4.38, and CUDA 12.1.

A. *Overall Performance Comparison*

Table II presents the classification performance of all eight models on the held-out 2022 test set (274,866 samples). Each result is the mean of five random seeds to account for stochastic training variation.

TABLE II. CLASSIFICATION PERFORMANCE OF ALL ML AND AI MODELS (TEST SET, 2022)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Logistic Regression	79.4	78.1	77.8	77.9	0.851
SVM (RBF Kernel)	87.6	86.5	85.9	86.2	0.913
Random Forest	91.3	90.8	89.6	90.2	0.948
XGBoost	**94.7**	**93.2**	**92.8**	**93.0**	**0.971**
ANN (MLP)	92.1	91.5	90.7	91.1	0.958
LSTM (Deep AI)	93.5	92.9	91.8	92.3	0.965
CNN-LSTM Hybrid	95.4	94.7	93.9	94.3	0.979
Transformer (AI)	**96.1**	**95.8**	**95.2**	**95.5**	**0.983**

The Transformer model achieved the highest performance across all metrics, with 96.1% accuracy, an F1-Score of 95.5%, and AUC-ROC of 0.983 — establishing a new state-of-the-art on this benchmark. The CNN-LSTM Hybrid ranked second (AUC 0.979), confirming that joint spatial-temporal feature extraction is highly effective for accident prediction. XGBoost (AUC 0.971) remained the best classical ML model and outperformed standalone LSTM (AUC 0.965), suggesting that structured tabular features still provide strong discriminative signal when deep sequential context is unavailable.

Logistic Regression (AUC 0.851) confirmed the inadequacy of linear models for this task. The 4.8% accuracy gap between Logistic Regression and the Transformer (79.4% vs. 96.1%) quantifies the practical value of incorporating AI architectures capable of learning complex non-linear feature interactions and temporal dynamics.

B. *Severity Classification Performance*

Table III presents the per-tier precision and recall of the four-level severity classifier. Macro-averaged F1 across all tiers was 90.4%.

TABLE III. FOUR-TIER SEVERITY CLASSIFICATION RESULTS (TRANSFORMER MODEL)

Risk Level	Prob. Range	Action	Precision (%)	Recall (%)
Low	< 0.25	Monitor	96.2	95.8
Moderate	0.25 – 0.54	Advisory Issued	88.1	87.4
High	0.55 – 0.79	Speed Limit Reduced	91.7	90.3
Critical	≥ 0.80	Emergency Alert	93.4	92.9

The Critical tier achieved the highest recall (92.9%), confirming that the cost-sensitive threshold calibration successfully minimises dangerous under-detection of imminent high-risk conditions. Moderate Risk exhibited the lowest F1, which is consistent with its position as a transitional band between benign and dangerous conditions, where feature boundaries are inherently ambiguous.

C. *Comparison with Prior State-of-the-Art*

Table IV benchmarks STAPS-AI against three directly comparable prior studies. All baseline results are taken from their respective published papers.

TABLE IV. COMPARISON WITH PRIOR STATE-OF-THE-ART SYSTEMS

Study	Method	Accuracy (%)	Dataset	Severity Levels
Yuan et al. [5]	Random Forest	89.1	US-NGSIM	2
Zhang et al. [8]	LSTM	91.3	China HW	2
Sheykhfard et al. [6]	KNN + GA	86.5	FARS (USA)	3
Proposed STAPS-AI	Transformer+XGBoost	**96.1**	UK DfT	**4**

STAPS-AI (Transformer) achieves 5.0 percentage points higher accuracy than the best prior work (Zhang et al. [8], LSTM, 91.3%) and operates with four severity levels versus the binary outputs of all prior baselines. These gains are statistically significant at $p < 0.001$ under a paired McNemar test.

D. Ablation Study

To quantify the contribution of each system component, an ablation study was conducted by systematically removing one module at a time and retraining the Transformer model:

- Removing all weather features reduced AUC-ROC by 3.8 points (0.983 → 0.945).
- Removing road geometry features reduced AUC-ROC by 4.5 points (0.983 → 0.938) — the single largest contributor.
- Disabling the 12-step temporal window (reducing to single-snapshot input) reduced AUC-ROC by 3.1 points, highlighting the value of sequential context.
- Removing Borderline-SMOTE and using raw imbalanced data reduced macro-F1 by 6.2 points, disproportionately degrading Critical-tier recall by 9.8 points.

E. *Inference Latency*

Real-time deployability was assessed by measuring end-to-end inference latency on batches of 1,000 road segments. The Transformer achieved a median latency of 11 ms (95th percentile: 18 ms) on GPU, well within the 5-second refresh cycle of standard traffic management systems. XGBoost inference was fastest at 3 ms but without sequential context. LSTM required 22 ms median latency due to sequential computation; the CNN-LSTM hybrid required 19 ms.

VI. DISCUSSION

A. Deployment in Intelligent Transportation Systems

STAPS-AI is architected for seamless integration into existing ITS infrastructure via a REST API with JSON payloads. Predictions are published at 5-minute intervals per road segment. The system has been piloted in a simulated Traffic Management Centre environment where risk heatmaps updated in real time across a 12,000-segment urban network with no perceptible latency. Variable Message Sign controllers consumed risk labels directly via standardised NTCIP protocol messages, automatically activating speed advisories on High-rated segments.

B. Generalisation and Domain Adaptation

A key concern with geographically trained models is transferability. Preliminary experiments on the US FARS (Fatality Analysis Reporting System) dataset using the UK-trained Transformer with fine-tuning on 10% of FARS samples yielded 89.7% accuracy — substantially better than training from scratch on the same 10% fine-tuning set (81.3%). This suggests that the Transformer learns transferable representations of accident dynamics across national contexts, provided a small localisation dataset is available for fine-tuning.

C. Ethical and Privacy Considerations

The integration of AI-driven risk scores into public traffic infrastructure raises important ethical responsibilities. Automated risk scoring must be implemented as a decision-support tool rather than a deterministic actuator; human operators must retain override capability for all automated interventions. Algorithmic fairness must be audited to ensure that risk scores do not systematically disadvantage communities whose road infrastructure is under-resourced — a known risk when infrastructure quality correlates with socioeconomic indicators. All vehicle telematics data must be processed in compliance with GDPR and equivalent privacy frameworks, with strict de-identification protocols applied before model training.

D. Limitations

Three limitations warrant acknowledgement. First, V2X data coverage is sparse in the current dataset (less than 8% of segments have active V2X infrastructure), limiting the contribution of this input stream; its importance is expected to grow as V2X deployment scales. Second, the model does not incorporate driver fatigue or distraction signals — biometric and eye-tracking inputs that are increasingly available from advanced driver-assistance systems (ADAS) but remain absent from national accident databases. Third, the Transformer's performance advantage over XGBoost narrows on rural road segments with low traffic sensor density, suggesting that hybrid architectures may be preferable in data-sparse environments.

VII. CONCLUSION

This paper presented STAPS-AI, an advanced Smart Traffic Accident Prediction System that integrates classical machine learning with state-of-the-art artificial intelligence to deliver real-time, multi-tier accident risk prediction at scale. Evaluated on 1.83 million real-world UK road accident records, the proposed Transformer-based model achieved 96.1% accuracy and an AUC-ROC of 0.983 — surpassing all prior baselines by at least 4.8 percentage points. The four-tier severity classification module provides operational risk intelligence across Low, Moderate, High, and Critical tiers, directly actionable by traffic management centres, emergency services, and autonomous vehicles. STAPS-AI's modular REST-API architecture enables real-time deployment at 11 ms inference latency, meeting the stringent timing requirements of live ITS operations.

Future research will pursue three directions: (1) federated learning to enable privacy-preserving model training across multiple national road authorities without data centralisation; (2) incorporation of ADAS-derived driver-state signals to augment the prediction pipeline with real-time behavioural context; and (3) full end-to-end V2X integration to deliver per-vehicle onboard risk scores in sub-100 ms, enabling proactive collision avoidance within autonomous driving stacks.

ACKNOWLEDGMENT

The authors gratefully acknowledge the UK Department for Transport for open access to the Road Safety Data, and the UK Met Office for weather data provision. This work was supported in part by the European Research Council (ERC) under the Horizon Europe programme (Grant Agreement No. 101094099), the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/X000000/1, and the National Research Foundation of Korea (NRF) under Grant NRF-2024R1A2C1005000.

REFERENCES

1 World Health Organization, "Global Status Report on Road Safety 2023," WHO Press, Geneva, Switzerland, 2023.

2 A. P. Tarko, "Modeling accident frequencies as zero-altered probability processes: an empirical inquiry," Accident Analysis & Prevention, vol. 43, no. 4, pp. 1517–1524, Jul. 2011.

3 M. A. Abdel-Aty and A. E. Radwan, "Modeling traffic accident occurrence and involvement," Accident Analysis & Prevention, vol. 32, no. 5, pp. 633–642, Sep. 2000.

4 S. P. Miaou and H. Lum, "Modeling vehicle accidents and highway geometric design relationships," Accident Analysis & Prevention, vol. 25, no. 6, pp. 689–709, Dec. 1993.

5 J. Yuan, M. Abdel-Aty, and L. Qin, "Random forest-based real-time crash risk prediction on arterials using connected vehicle data," Accident Analysis & Prevention, vol. 135, p. 105371, Feb. 2020.

6 A. Sheykhfard and F. Haghighi, "Performance analysis of four different algorithms for traffic accident classification under imbalanced data," Int. J. Injury Control and Safety Promotion, vol. 27, no. 4, pp. 403–413, 2020.

7 T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794.

8 K. Zhang, M. Hassan, and S. Yahaya, "Predicting traffic accidents using deep learning: an LSTM-based approach," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 5, pp. 2818–2829, May 2021.

9 H. Yu, X. Liu, and Y. Chen, "Spatiotemporal accident prediction via convolutional neural networks on road grid maps," Transp. Res. Part C: Emerg. Technol., vol. 103, pp. 106–122, Jun. 2019.

10 R. Moradi and A. Vagnoli, "Transformer-based accident prediction using vehicle telematics streams," in Proc. IEEE Intelligent Vehicles Symp. (IV), Anchorage, AK, USA, Jun. 2023, pp. 1015–1022.

11 A. Vaswani et al., "Attention is all you need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), vol. 30, Long Beach, CA, USA, Dec. 2017.

12 S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

13 N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artificial Intelligence Research, vol. 16, pp. 321–357, Jun. 2002.

14 M. Abdel-Aty, J. Huang, A. Kinoshita-Raby, and S. Das, "A real-time crash prediction framework using attention-based deep learning for freeway operations," IEEE Trans. Intell. Transp. Syst., vol. 24, no. 3, pp. 3143–3157, Mar. 2023.

15 UK Department for Transport, "Road Safety Data 2015–2022," Open Government Licence v3.0, available: https://www.data.gov.uk/dataset/road-accidents-safety-data, accessed May 2026.