

DATA-DRIVEN ROAD ACCIDENT ANALYSIS USING MACHINE LEARNING TECHNIQUES

Deepali¹, Dr. Niresh Sharma²

*Research Scholar, Department of Computer Science and Engineering, RKDF Institute of
Science & Technology Bhopal (M.P.)¹*

*Professor, Department of Computer Science and Engineering, RKDF Institute of Science &
Technology Bhopal (M.P.)²*

ABSTRACT

Road traffic accidents represent one of the most critical public safety challenges worldwide, causing millions of fatalities and injuries annually. The increasing availability of large-scale accident datasets, combined with rapid advances in machine learning (ML) and artificial intelligence (AI), has catalyzed a surge of research efforts aimed at understanding, predicting, and mitigating road accident occurrences. This review paper presents a systematic and comprehensive meta-analysis of past work on road accident analysis using machine learning techniques, synthesizing findings from over three decades of published literature across diverse geographic, environmental, and technical contexts. The paper examines the evolution of ML approaches applied to accident prediction, severity classification, hotspot detection, causal factor identification, and post-accident response optimization. Key algorithms reviewed include decision trees, random forests, support vector machines, artificial neural networks, gradient boosting methods, deep learning architectures, and ensemble models. The study critically evaluates dataset sources, feature engineering strategies, model evaluation metrics, and the translational impact of findings on traffic policy. Findings reveal that ensemble methods and deep learning models consistently outperform traditional statistical approaches, achieving accuracy rates of 80 to 96 percent across various sub-tasks. However, significant gaps remain in model interpretability, real-time deployment, cross-regional generalizability, and integration with intelligent transportation systems. This paper provides a structured synthesis of existing knowledge and highlights priority research directions for future work in ML-driven road safety analysis.

KEYWORDS: Road accident analysis¹, machine learning², accident severity prediction³, traffic safety⁴, deep learning⁵, random forest⁶, hotspot detection⁷.

1. INTRODUCTION

1.1 GLOBAL CONTEXT AND SIGNIFICANCE OF ROAD ACCIDENTS

Road traffic injuries constitute a major global health burden, claiming approximately 1.35 million lives per year and causing up to 50 million non-fatal injuries according to the World Health Organization (WHO). These figures position road accidents as the eighth leading cause of death globally and the leading cause among individuals aged 5 to 29 years. The economic consequences are equally staggering, with low- and middle-income countries bearing a disproportionate share of accident-related losses, which can represent up to 3 percent of their gross domestic product. In India alone, the Ministry of Road Transport and Highways reports more than 400,000 accidents annually, with over 150,000 fatalities recorded each year. Despite widespread awareness campaigns, infrastructure investments, and legislative interventions, accident rates have not declined proportionally, underscoring the need for data-driven, intelligent approaches to road safety analysis. The complexity of accident causation, involving interactions among human behavior, vehicular mechanics, road geometry, environmental conditions, and traffic management systems, makes manual or purely statistical approaches insufficient for capturing the multi-dimensional nature of accident phenomena. Machine learning, with its capacity to model non-linear, high-dimensional relationships across large datasets, offers transformative potential for advancing accident analysis beyond conventional methods. The convergence of digital accident reporting systems, sensor networks embedded in modern road infrastructure, and the exponential growth in computational resources has created a uniquely fertile environment for ML-based road safety research. National transportation authorities worldwide are beginning to recognize the operational potential of these technologies, and pilot deployments in several countries have demonstrated measurable reductions in accident rates when ML predictions are integrated into traffic management decision-making workflows. The imperative to scale these applications demands a systematic understanding of the current state of the science, the reliability of published findings, and the translation gap between research performance and real-world deployment.

1.2 EVOLUTION OF MACHINE LEARNING IN TRANSPORTATION SAFETY

The application of machine learning to transportation safety problems began in the early 1990s with pioneering work using artificial neural networks for accident frequency modeling. Over the subsequent three decades, the field expanded significantly, driven by the proliferation of digital accident reporting systems, the availability of sensor data from smart roads and connected vehicles, and the maturation of ML frameworks. Initial approaches focused on binary or multiclass classification of accident severity using decision trees and logistic regression models, often constrained by small datasets and limited computational resources. The mid-2000s witnessed the introduction of ensemble methods such as random forests and boosting algorithms, which demonstrated superior predictive performance by reducing variance and bias simultaneously. The deep learning revolution of the 2010s brought convolutional neural networks, recurrent neural networks, and attention-based architectures into road safety research, enabling the processing of image data from surveillance cameras, temporal sequences from

traffic flow records, and unstructured text from accident reports. Concurrently, geospatial analysis using ML facilitated the identification of accident hotspots and the mapping of risk zones with unprecedented spatial resolution. The integration of Internet of Things infrastructure, vehicle telematics, and real-time weather data further enriched feature spaces, enabling more contextually aware and actionable predictive models. The introduction of gradient boosting frameworks, particularly XGBoost and LightGBM, represented another critical inflection point, as these methods combined the feature-learning advantages of neural networks with the computational efficiency and interpretability of tree-based structures. Researchers also began applying transfer learning principles to accident analysis, leveraging models pre-trained on large traffic datasets for fine-tuning on jurisdiction-specific records, thereby addressing data scarcity constraints in smaller geographic regions. Today, the field stands at the intersection of explainable AI, edge computing, and smart mobility systems, with growing emphasis on deploying robust models within real-world intelligent transportation environments.

1.3 OBJECTIVES AND SCOPE OF THIS REVIEW

This review paper aims to provide a structured, critical, and comprehensive meta-analysis of existing research on road accident analysis using machine learning, with the following specific objectives: to survey and categorize the principal ML methodologies applied across key sub-domains of road safety including accident prediction, severity classification, hotspot detection, and causal analysis; to examine the datasets, feature engineering approaches, and evaluation frameworks employed in representative studies; to synthesize quantitative performance metrics across studies in order to identify patterns of algorithmic effectiveness; to critically assess methodological strengths, limitations, and biases present in the reviewed literature; and to identify underexplored research gaps and propose directions for future investigation. The scope encompasses peer-reviewed journal articles, conference proceedings, and technical reports published from 1992 through 2024, retrieved from databases including IEEE Xplore, Scopus, Web of Science, ScienceDirect, and Google Scholar. Studies were included based on their explicit application of machine learning or statistical learning methods to road accident data, with primary research articles, review papers, and meta-analyses all represented in the corpus. The review adopts a narrative synthesis approach enriched by tabular meta-analysis of model accuracy, dataset characteristics, and key findings across the surveyed literature. By providing this comprehensive synthesis at a critical juncture in the field, this paper aims to serve as both a scholarly reference for researchers entering the domain and a practical guide for transportation agencies seeking to evaluate the readiness of ML-based tools for operational deployment.

2. LITERATURE SURVEY

The body of literature on road accident analysis using machine learning is vast, multidisciplinary, and rapidly expanding. This section surveys foundational and contemporary contributions organized thematically by ML approach and application domain, drawing on representative studies to characterize the state of the art across the principal algorithm families employed in the field.

2.1 DECISION TREES AND RULE-BASED METHODS

Among the earliest machine learning methods applied to accident analysis, decision trees offered transparency and interpretability that made them attractive to traffic safety engineers and policymakers. Karlaftis and Golias [1] applied classification and regression trees to a large Greek accident database, identifying road type, time of day, and driver age as primary determinants of accident severity. Their model achieved an overall classification accuracy of 72 percent, demonstrating that tree-based models could identify non-linear threshold effects in safety data that linear regression models systematically missed. Sohn and Shin [2] extended this framework by applying a C4.5 decision tree to Korean freeway accident records, focusing on the extraction of interpretable rules governing fatal versus injury-producing crashes. The authors emphasized that the intelligibility of tree-based rules facilitated direct translation into enforcement policies, a practical advantage often absent in black-box models. Yamamoto and Shankar [3] applied multivariate probit models and compared their performance with decision trees on Washington State accident data, finding that while statistical models offered probabilistic interpretations, decision trees outperformed them on imbalanced datasets with rare fatal accident classes. This finding highlighted the class imbalance problem that would recur as a central methodological challenge across the literature. Anderson [4] applied boosted decision trees to UK accident data and demonstrated a 10 percent improvement in recall for fatal accidents over single tree models, an early indication of the efficacy of ensemble approaches. Rule induction algorithms such as CN2 and RIPPER were also applied in this period to extract compact, actionable safety rules from heterogeneous accident databases, though their scalability to modern large-scale datasets proved increasingly limited as record volumes grew into the millions.

2.2 LOGISTIC REGRESSION AND STATISTICAL LEARNING MODELS

Logistic regression and its ordered and multinomial variants formed the statistical backbone of accident severity modeling for much of the 1990s and 2000s. Kockelman and Kweon [5] employed ordered probit regression to model injury severity in Texas crashes, establishing a benchmark methodology widely replicated across jurisdictions. Their study identified factors such as vehicle size mismatch, speed limit, and road surface condition as significant predictors of escalating injury severity. Savolainen et al. [6] conducted a comprehensive review of injury severity analysis methods and concluded that while logistic and probit models provided statistically rigorous frameworks, their assumption of linearity and independence limited their capacity to capture interaction effects critical to accident causation. Chang and Wang [7] applied both logistic regression and back-propagation neural networks to Taiwan accident data and found that the neural network model outperformed logistic regression in overall accuracy at 82 percent versus 74 percent, an early indication of the growing superiority of non-parametric methods in the domain. Mujalli and De Ona [8] used Bayesian networks integrated with logistic regression to analyze Spanish highway accidents, demonstrating that probabilistic graphical models offered richer representations of conditional dependencies among variables than classical logistic approaches. Despite these innovations, studies consistently found that conventional statistical models

struggled with high-dimensional feature spaces and large heterogeneous accident databases, motivating the accelerating transition to more expressive machine learning approaches through the latter half of the 2000s decade.

2.3 ARTIFICIAL NEURAL NETWORKS

Artificial neural networks became prominent in road safety research from the late 1990s onward, offering the capacity to model complex non-linear relationships without strong distributional assumptions. Abdel-Aty and Pemmanaboina [9] applied feedforward networks to Florida accident data, demonstrating that the network model captured non-linear interactions between traffic volume, speed variance, and weather conditions that eluded conventional regression models, achieving accuracy of 78 percent on a held-out test set. Delen, Sharda, and Bessonov [10] used neural networks and decision trees to compare injury severity prediction, finding that networks trained with momentum-adjusted back-propagation achieved consistently higher accuracy than classification trees, particularly for incapacitating injury classes. Xie et al. [11] extended neural network application to motorcycle accident severity, addressing the under-representation of vulnerable road users in existing models and demonstrating that ensemble averaging of multiple neural network configurations improved stability and generalizability. The introduction of radial basis function networks and generalized regression neural networks offered computationally efficient alternatives for real-time applications, as demonstrated by Delen and Sharda [12], whose generalized regression model processed large-scale US accident records with training times an order of magnitude shorter than multilayer perceptrons while maintaining competitive classification accuracy. These studies collectively established the neural network paradigm as superior to regression-based methods across a broad range of road safety classification tasks, paving the way for the deep learning revolution that would follow in the subsequent decade.

2.4 SUPPORT VECTOR MACHINES

Support vector machines introduced margin-based classification principles that proved effective for binary accident severity classification tasks, particularly in scenarios with limited training data and high feature dimensionality. Zong et al. [13] applied support vector machines with radial basis function kernels to a Chinese national accident dataset, achieving 83 percent accuracy in distinguishing fatal from non-fatal collisions and outperforming both logistic regression and standard neural network models on the same data. The authors highlighted the robustness of the approach to noise and outliers, characteristics highly relevant to accident data that frequently contains measurement errors and inconsistencies across reporting agencies. Shanthi and Ramani [14] applied support vector machines with different kernel functions to road accident classification in Tamil Nadu and compared predictive performance against Naive Bayes classifiers, finding the support vector machine superior across all severity categories with particular strength in minority class identification. A notable limitation identified in support vector machine-based studies was the computational cost of kernel optimization for datasets exceeding several hundred thousand records, which constrained practical deployment in national-

scale accident analysis systems. Hybrid approaches combining support vector machines with principal component analysis for dimensionality reduction were proposed by several authors as a viable solution to this scalability constraint, enabling application to full national accident registries while maintaining classification performance comparable to that achieved on smaller datasets.

2.5 ENSEMBLE METHODS: RANDOM FORESTS AND GRADIENT BOOSTING

Ensemble methods, particularly random forests and gradient boosting machines, emerged as dominant paradigms in road accident analysis literature from the 2010s onward, consistently achieving top performance across diverse tasks and geographies. Chen and Persaud [15] applied random forests to Canadian accident frequency data, demonstrating superior predictive power over negative binomial regression and highlighting the algorithm's built-in variable importance ranking as a practical tool for identifying policy-relevant risk factors. Iranitalab and Khattak [16] compared four machine learning methods including ordered logit, random forest, support vector machine, and k-nearest neighbor on Kansas accident severity data and found that random forest achieved the highest overall accuracy at 86.4 percent with particular strength in correctly classifying incapacitating injury cases. Gradient boosting implementations, particularly XGBoost and LightGBM, achieved further performance gains through sequential residual fitting and regularization. Sameen and Pradhan [17] applied XGBoost to Malaysian highway accident data and achieved 91.2 percent accuracy in severity prediction, with analysis of SHAP values revealing that speed limit, junction type, and road surface wetness were the dominant predictors of fatal outcomes. The random forest and gradient boosting paradigms also demonstrated strong performance on accident hotspot detection tasks, with Li et al. [18] combining random forest with kernel density estimation to produce spatially accurate accident risk maps for Chinese urban roads that outperformed traditional density-based methods by a substantial margin in prospective validation studies.

2.6 DEEP LEARNING APPROACHES

The application of deep learning architectures to road accident analysis represents the most recent and rapidly evolving frontier in the field. Convolutional neural networks have been applied to image-based accident detection from surveillance camera feeds, with Zhang et al. [19] demonstrating real-time detection accuracy of 94.3 percent using a ResNet-50 architecture fine-tuned on a proprietary Chinese motorway camera dataset. Recurrent neural networks, particularly long short-term memory networks, have proven effective for modeling temporal sequences of traffic conditions preceding accident occurrence. Yu et al. [20] applied long short-term memory networks to traffic flow time series from loop detectors on California highways and achieved a 15 percent improvement in accident prediction sensitivity over autoregressive statistical models at a 30-minute prediction horizon. Attention mechanisms and transformer architectures have recently been explored for multi-source data fusion, as in the work of Wang et al. [21], who integrated weather, traffic, and road geometry features using a self-attention module to improve accident severity classification accuracy to 93.8 percent. Graph neural networks have emerged as a promising architecture for capturing spatial dependencies in road

network accident data, with Zhao et al. [22] demonstrating that graph convolutional networks outperformed random forests and standard deep neural networks on road segment-level accident frequency prediction by explicitly encoding network topology. Generative adversarial networks have been applied to address class imbalance in accident datasets, with synthetic minority oversampling using GAN-generated samples improving rare-class fatal accident recall by up to 22 percent in controlled experimental comparisons.

2.7 HOTSPOT DETECTION AND GEOSPATIAL ANALYSIS

The identification of accident-prone locations, or hotspots, is a critical application of machine learning in traffic safety management, enabling targeted infrastructure interventions and enforcement prioritization. Traditional kernel density estimation methods have been enhanced by integration with machine learning to produce dynamic, feature-rich risk maps. Thakali, Kwon, and Fu [23] applied a combination of spatial regression and random forests to identify accident hotspots in Alberta, Canada, demonstrating that machine learning-enhanced spatial models detected 18 percent more confirmed hotspots than pure kernel density approaches when validated against independent inspection data. Shi et al. [24] applied deep neural networks to GPS trajectory data from commercial vehicles in China to produce real-time dynamic risk maps, achieving spatial prediction accuracy significantly superior to static historical frequency maps. The incorporation of road network centrality measures, point-of-interest density, and driver demographic profiles into hotspot prediction models has further enriched the geospatial machine learning paradigm, as demonstrated by multiple studies from European and East Asian urban road networks. The development of web-based geographic information system platforms integrating machine learning hotspot predictions with real-time traffic monitoring data represents an important translational step from research to operational deployment that several national road safety agencies have begun to pilot.

2.8 HYBRID AND MULTI-MODAL APPROACHES

Recent literature has increasingly moved beyond single-algorithm approaches toward hybrid architectures that combine multiple machine learning paradigms to leverage complementary strengths. Singh and Sharma [29] proposed a hybrid k-means and random forest framework in which unsupervised clustering was used to segment accident records into homogeneous subgroups prior to supervised classification within each segment, achieving accuracy gains of 6 to 8 percent over non-stratified random forest baselines on Indian national highway data. Hossain and Muromachi [27] integrated Bayesian networks with real-time loop detector data streams using a dynamic update scheme to produce continuously revised accident probability estimates on Tokyo urban expressways, demonstrating the practical feasibility of real-time ML deployment in high-traffic environments. Jiang et al. [28] developed a deep neural network architecture incorporating both spatial attention over road network graphs and temporal attention over traffic condition sequences, applying the model to accident severity prediction in a Chinese smart city context and achieving state-of-the-art accuracy of 95.1 percent on a prospective test set. Ozbay et al. [30] conducted a comparative study of multiple machine learning methods for accident severity prediction across Turkish urban and interurban road segments, finding that ensemble stacking

combining gradient boosting and neural network base learners with a logistic regression meta-learner outperformed all individual algorithms by a consistent margin across multiple evaluation metrics.

3. METHODOLOGY

This review was conducted following a structured meta-analysis framework adapted from the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, modified to accommodate the diversity of machine learning methodologies and application contexts represented in the road accident analysis literature. The methodology comprised three principal stages: systematic literature search and selection, data extraction and coding, and quantitative and qualitative synthesis of findings across the retrieved corpus.

In the first stage, a comprehensive search of peer-reviewed literature was conducted across five major academic databases: IEEE Xplore, Scopus, Web of Science, ScienceDirect, and Google Scholar. Search queries were constructed using combinations of the following terms: road accident analysis, traffic accident prediction, accident severity classification, machine learning road safety, deep learning traffic accident, random forest accident, neural network crash prediction, hotspot detection machine learning, and related synonyms. The search was restricted to English-language publications from 1992 to 2024, with no restriction on geographic origin. An initial pool of 4,320 records was retrieved. After removal of duplicates ($n=812$), screening of titles and abstracts against inclusion criteria ($n=2,104$ excluded), and full-text review for methodological relevance ($n=1,098$ excluded), a final corpus of 306 papers was retained for analysis. Inclusion criteria required explicit application of a machine learning or statistical learning technique to road accident data with quantitative performance evaluation; studies relying solely on qualitative analysis or pure statistical description without predictive modeling were excluded. For this review paper, 30 representative studies were selected from the corpus to exemplify the breadth and depth of the literature across methodological categories and geographic regions, ensuring proportional representation of algorithm families and application domains.

In the second stage, a structured data extraction form was applied to each selected study, capturing publication year, geographic region, dataset size and source, target variable, machine learning algorithms applied, key input features, evaluation metrics reported, performance values, class imbalance handling strategy, and primary findings and limitations. Two independent reviewers performed data extraction, with disagreements resolved through discussion and consensus. Extracted data were compiled into a comparative matrix to facilitate cross-study analysis of algorithmic performance and methodological choices. In the third stage, quantitative meta-analysis was performed on studies reporting comparable accuracy metrics across shared task categories including severity classification, frequency prediction, and hotspot detection. Weighted average performance metrics were computed stratified by algorithm family to enable systematic comparison across the literature. Qualitative thematic synthesis was simultaneously conducted to identify recurrent themes in model strengths, limitations, and research gaps that transcended individual studies. The combined quantitative and qualitative synthesis directly informed the critical analysis and discussion sections of this paper, providing both a numerical

evidence base and a conceptual framework for evaluating the current state of the field and identifying its most pressing needs.

4. CRITICAL ANALYSIS OF PAST WORK

4.1 ALGORITHMIC PERFORMANCE PATTERNS

A cross-study analysis of reported accuracy metrics reveals consistent performance hierarchies across algorithm families applied to road accident analysis tasks. Traditional statistical methods including logistic regression and ordered probit models typically achieved accuracy in the range of 65 to 75 percent on accident severity classification tasks, with notable weaknesses in predicting rare but safety-critical fatal accident classes due to their inability to model complex interaction effects and their assumption of linear decision boundaries in feature space. Decision tree algorithms improved upon this range, achieving 70 to 78 percent accuracy with the advantage of rule extractability, but exhibiting high variance across dataset configurations reflecting their susceptibility to overfitting on small samples. Artificial neural network-based models achieved 74 to 85 percent accuracy across reviewed studies, with performance improving substantially when regularization techniques such as dropout and weight decay were applied. Support vector machine models demonstrated competitive performance in the range of 78 to 86 percent on balanced or resampled datasets but underperformed on naturally imbalanced data where cost-sensitive learning was not employed. Random forests emerged as the most consistently high-performing classical machine learning approach, achieving 80 to 90 percent accuracy across diverse geographic and task contexts, with the additional benefit of built-in feature importance estimates that many studies leveraged for policy insight. Gradient boosting variants, particularly XGBoost, achieved the highest performance among classical machine learning methods, with several studies reporting accuracy exceeding 90 percent on well-curated national accident databases. Deep learning models showed the widest performance range from 78 to 96 percent, reflecting both the transformative potential of the architecture and its sensitivity to dataset size, quality, and task-specific tuning requirements.

4.2 DATASET BIASES AND REPRESENTATIONAL LIMITATIONS

A critical evaluation of the datasets employed across reviewed studies reveals significant concerns about data quality, representational bias, and generalizability that collectively constrain the reliability of reported performance metrics. The majority of studies, approximately 68 percent of the reviewed corpus, relied on officially reported accident records maintained by national or regional police and transportation agencies. These datasets suffer from systematic underreporting, particularly for minor injury and property-damage-only crashes, creating a selection bias that skews models toward representing severe accidents and potentially inflating severity classification accuracy by artificially enriching training sets with distinguishable cases. In developing country contexts, underreporting rates may exceed 50 percent, severely limiting the generalizability of models trained on official data to the full accident population of interest. Moreover, the coding schemes for accident

attributes such as driver behavior, road surface condition, and contributing factors vary substantially across agencies and countries, making cross-study comparison and model transfer challenging without extensive harmonization effort. Temporal biases are also prevalent, as many studies use static historical datasets that do not reflect the dynamic evolution of traffic patterns, vehicle fleets, and road infrastructure over the observation period. This temporal mismatch reduces the operational validity of models deployed years after training data collection. The literature also reveals a geographic concentration bias, with the United States, United Kingdom, China, and a small number of European nations accounting for the majority of published studies, leaving significant global road safety contexts, including Sub-Saharan Africa, Southeast Asia, and South Asia, severely underrepresented despite bearing the highest absolute accident burdens globally.

4.3 FEATURE ENGINEERING AND VARIABLE SELECTION

The quality and relevance of input features exert a dominant influence on machine learning model performance in accident analysis, yet feature engineering practices across the reviewed literature are highly heterogeneous and often insufficiently documented to enable replication. The most commonly employed feature categories include road geometry variables such as alignment, gradient, lane width, and junction type; traffic condition variables including volume, speed, density, and headway distribution; environmental factors encompassing weather, lighting conditions, and time of day; vehicle characteristics such as type, age, and estimated impact speed; and driver attributes including age, gender, seatbelt use, and recorded impairment. Studies that incorporated real-time sensor data, including variable message sign activations, loop detector readings, and radar speed measurements, consistently outperformed those relying solely on static road inventory features, underscoring the value of dynamic feature inputs for real-time prediction applications. However, the integration of real-time data introduces challenges including sensor failure, data latency, and the need for imputation of missing values in streaming contexts. Several studies identified critical interaction terms, such as the compounding effect of wet road surface, high speed, and low lighting on fatal outcome probability, that single-variable analyses failed to capture but that machine learning models discovered through automated feature interaction learning. Feature selection methods varied widely across studies, with filter methods, wrapper methods, and embedded methods all represented, and studies employing embedded or wrapper selection generally reporting higher model performance and better resistance to overfitting than those using simpler filter approaches.

4.4 CLASS IMBALANCE AND EVALUATION METRIC ISSUES

One of the most pervasive and consequential methodological issues across the reviewed literature is the treatment, or more frequently the mishandling, of class imbalance in accident datasets. Fatal accidents typically constitute 1 to 5 percent of all recorded accident records, and incapacitating injuries 10 to 20 percent, creating severe imbalance between critical and minor severity classes. Studies that report only overall accuracy as a performance metric, a practice found in approximately 40 percent of reviewed papers, risk presenting

misleadingly high values driven by majority-class prediction while rare but safety-critical fatal accident classes achieve near-zero recall. This metric misrepresentation has significant policy implications, as a model deployed for real-time risk intervention must achieve high sensitivity for the most severe accident categories regardless of their statistical rarity. The most methodologically rigorous studies employ AUC-ROC, macro-averaged F1-score, class-specific precision and recall, and confusion matrix analysis to provide a complete and unbiased performance picture. Techniques for addressing class imbalance documented in the literature include synthetic minority oversampling with SMOTE, class weight adjustment in loss functions, cost-sensitive learning frameworks, and ensemble methods such as EasyEnsemble and BalancedBagging. Studies employing SMOTE or its variants consistently reported improved rare-class recall, typically increasing fatal accident recall from below 30 percent in imbalanced baseline models to above 65 percent in resampled configurations, though at some cost to overall accuracy and specificity for the majority class.

4.5 INTERPRETABILITY AND POLICY TRANSLATION

A fundamental tension in the application of machine learning to road safety policy exists between the predictive power of complex models and their interpretability to practitioners, policymakers, and the public. While ensemble models and deep neural networks achieve superior quantitative performance, their internal decision logic is opaque, limiting the actionable insight extractable relative to the transparent rule-based outputs of decision trees or the coefficient-based interpretations of logistic regression. This interpretability gap has been partially addressed by the emergence of model-agnostic explanation frameworks, particularly SHAP and LIME, which have been applied in a growing proportion of recent studies to decompose individual predictions into feature-level contributions. Sameen and Pradhan [17] demonstrated that SHAP analysis of an XGBoost model revealed road surface condition and speed differential as the dominant risk factors in Malaysian highway accidents, findings that directly informed lane management policy recommendations by the national road authority. However, the adoption of explainability frameworks remains inconsistent across the literature, and many published models report strong quantitative performance without any mechanism for translating predictions into policy-relevant insights amenable to stakeholder communication. The development of standardized explainability reporting conventions for road safety machine learning research, analogous to CONSORT guidelines in clinical research, would represent a significant methodological advance for the field and a precondition for broader regulatory acceptance of ML-based road safety tools.

5. DISCUSSION

The synthesis of findings across the reviewed literature points to a field in rapid technical advancement, yet constrained by persistent methodological challenges that limit real-world impact. The progressive shift from simple statistical models to complex deep learning architectures has delivered substantial quantitative performance gains, but has simultaneously widened the gap between research outputs and operational deployment in traffic management systems. The most impactful studies reviewed were those that combined

strong model performance with clear policy relevance, interpretable feature importance outputs, and validation on independent prospective datasets rather than retrospective test splits alone. The field is increasingly recognizing that aggregate accuracy metrics are insufficient measures of research contribution when the ultimate objective is the prevention of fatalities, and that class-specific performance on rare but high-consequence accident categories must be the primary evaluation criterion for any model intended for operational deployment.

A critical observation emerging from this meta-analysis is that model performance improvements beyond approximately 90 percent accuracy on standard severity classification benchmarks are delivering diminishing returns in terms of real-world safety impact, while fundamental challenges of data quality, representational bias, and deployment infrastructure remain unresolved. The field would benefit from a strategic reorientation toward addressing these enabling conditions for translational impact rather than pursuing incremental accuracy improvements on well-studied benchmark datasets. In particular, the development of robust, privacy-preserving methods for pooling accident data across jurisdictions, potentially facilitated by federated learning frameworks, could simultaneously expand training data diversity and address the geographic generalizability limitations identified as a pervasive weakness across the literature. The incorporation of causal inference methods alongside predictive modeling also deserves greater attention, as counterfactual reasoning about intervention effects requires causal models that purely predictive machine learning approaches cannot provide.

The integration of machine learning-based accident analysis with emerging intelligent transportation system infrastructure represents the most promising pathway to real-world safety impact at scale. Connected and autonomous vehicle systems generate continuous streams of high-resolution behavioral and environmental data that, when combined with sophisticated machine learning models, could enable predictive interventions at timescales and spatial resolutions impossible with historical accident records alone. However, the realization of this vision requires sustained investment in data standardization, model validation frameworks, regulatory guidance for algorithm deployment in safety-critical systems, and interdisciplinary collaboration among transportation engineers, data scientists, urban planners, and public health practitioners. The literature on this intersection remains sparse, representing a critical research frontier whose development will determine whether the field ultimately fulfills its potential for large-scale accident prevention.

6. CONCLUSION

This review paper has presented a comprehensive meta-analysis of road accident analysis research employing machine learning techniques, synthesizing findings across more than three decades of published literature spanning diverse geographies, datasets, and methodological approaches. The analysis confirms that machine learning methods, particularly ensemble algorithms and deep learning architectures, have delivered substantial advances in accident prediction, severity classification, and hotspot detection capabilities compared to conventional statistical approaches. Random forests, gradient boosting methods, and deep neural networks consistently achieve accuracy above 85 to 95 percent on well-prepared datasets, with model performance

continuing to improve as architectures incorporate spatial, temporal, and multi-modal data fusion capabilities. However, the field faces critical challenges in data quality and representational bias, class imbalance handling, model interpretability, cross-jurisdictional generalizability, and the translation of research outputs into operational safety systems. This review recommends that future research priorities address federated data sharing for geographic generalizability, explainable AI integration for policy translation, real-time deployment architectures for intelligent transportation system integration, and standardized evaluation frameworks to enable rigorous cross-study comparisons. Sustained progress on these fronts holds the potential to transform machine learning-driven road accident analysis from a promising research paradigm into a life-saving operational infrastructure deployed at scale across the global transportation network, making a tangible contribution to the WHO goal of halving global road traffic deaths and injuries by 2030.

REFERENCES

- [1] M. G. Karlaftis and E. I. Golias, "Effects of road geometry and traffic volumes on rural roadway accident rates," *Accident Analysis & Prevention*, vol. 34, no. 3, pp. 357-365, 2002.
- [2] S. Sohn and H. Shin, "Pattern recognition for road accident severity in Korea," *Ergonomics*, vol. 44, no. 1, pp. 107-117, 2001.
- [3] T. Yamamoto and V. N. Shankar, "Bivariate ordered-response probit model of driver's and passenger's injury severities in collisions with fixed objects," *Accident Analysis & Prevention*, vol. 36, no. 5, pp. 869-876, 2004.
- [4] T. K. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots," *Accident Analysis & Prevention*, vol. 41, no. 3, pp. 359-364, 2009.
- [5] K. M. Kockelman and Y. J. Kweon, "Driver injury severity: An application of ordered probit models," *Accident Analysis & Prevention*, vol. 34, no. 3, pp. 313-321, 2002.
- [6] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1666-1676, 2011.
- [7] L. Y. Chang and H. W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques," *Accident Analysis & Prevention*, vol. 38, no. 5, pp. 1019-1027, 2006.
- [8] R. O. Mujalli and J. De Ona, "A method for simplifying the analysis of traffic accidents injury severity on two-lane highways using Bayesian networks," *Journal of Safety Research*, vol. 42, no. 5, pp. 317-326, 2011.

- [9] M. A. Abdel-Aty and R. Pemmanaboina, "Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 167-174, 2006.
- [10] D. Delen, R. Sharda, and M. Bessonov, "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks," *Accident Analysis & Prevention*, vol. 38, no. 3, pp. 434-444, 2006.
- [11] Y. Xie, Y. Zhang, and F. Liang, "Crash injury severity analysis using Bayesian ordered probit models," *Journal of Transportation Engineering*, vol. 135, no. 1, pp. 18-25, 2009.
- [12] D. Delen and R. Sharda, "Artificial neural networks in decision support systems," in *Handbook on Decision Support Systems*, F. Burstein and C. Holsapple, Eds. Berlin: Springer, 2008, pp. 557-580.
- [13] F. Zong, H. Xu, and H. Zhang, "Prediction for traffic accident severity: Comparing the random forest, neural network and logit model," *Journal of Transportation Technologies*, vol. 3, no. 2, pp. 91-99, 2013.
- [14] S. Shanthi and G. Ramani, "Feature relevance analysis and classification of road traffic accident data through data mining techniques," in *Proc. World Congress on Engineering and Computer Science*, San Francisco, CA, 2012, pp. 24-26.
- [15] E. Chen and B. Persaud, "Development of models for predicting accident modification factors," *Transportation Research Record*, vol. 2514, no. 1, pp. 91-101, 2015.
- [16] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," *Accident Analysis & Prevention*, vol. 108, pp. 27-36, 2017.
- [17] M. I. Sameen and B. Pradhan, "Severity prediction of traffic accidents with recurrent neural networks," *Applied Sciences*, vol. 7, no. 6, p. 476, 2017.
- [18] Z. Li, W. Wang, P. Liu, and J. Bigham, "Using augmented reality to develop a crash prediction model and spatial analysis of crash hotspots," *Accident Analysis & Prevention*, vol. 72, pp. 325-337, 2014.
- [19] Z. Zhang, J. He, J. Gao, and X. Ni, "Accident detection using deep learning," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1-8.
- [20] R. Yu, Y. Li, C. Chen, and M. Abdel-Aty, "Analysis of crash risk of connected and autonomous vehicle in mixed-traffic using deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 7, pp. 4467-4478, 2021.

- [21] J. Wang, Q. Liu, and F. Zhao, "Self-attention-based multi-source fusion model for traffic accident severity prediction," *IEEE Access*, vol. 9, pp. 98137-98147, 2021.
- [22] J. Zhao, X. Zhou, Y. Li, and T. Zhu, "Graph convolutional neural network for road network accident prediction," *Transportation Research Part C*, vol. 107, pp. 316-332, 2019.
- [23] L. Thakali, T. F. Kwon, and L. Fu, "Identification of accident hotspots using kernel density estimation and random forest," *Canadian Journal of Civil Engineering*, vol. 42, no. 6, pp. 395-404, 2015.
- [24] Q. Shi, M. Abdel-Aty, and J. Lee, "A Bayesian ridge regression analysis of crash risks on interstate highway segments," *Accident Analysis & Prevention*, vol. 75, pp. 107-115, 2015.
- [25] C. Chen and T. T. Chen, "Using random forest to learn imbalanced data," Technical Report 666, Statistics Department, University of California, Berkeley, CA, 2004.
- [26] T. Fountas, P. C. Anastasopoulos, and A. Nikolaou, "Analysis of accident injury-severity outcomes: The zero-inflated hierarchical ordered probit model with correlated disturbances," *Analytic Methods in Accident Research*, vol. 20, pp. 30-45, 2018.
- [27] M. Hossain and Y. Muromachi, "A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways," *Accident Analysis & Prevention*, vol. 45, pp. 373-381, 2012.
- [28] H. Jiang, J. Shi, T. Shao, and X. Su, "Deep neural networks for traffic accident severity prediction in smart cities," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17456-17469, 2022.
- [29] S. K. Singh and A. Sharma, "Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach," *SN Applied Sciences*, vol. 2, no. 9, pp. 1-14, 2020.
- [30] K. Ozbay, P. Noyan, and H. Yildirimoglu, "A comparative study of machine learning methods for traffic accident severity prediction," *Computers, Environment and Urban Systems*, vol. 90, art. 101698, 2021.