International Journal of Science and Engineering Applications
Volume 11-Issue 12, 469 – 480, 2022, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1112.1068

Crash Severity Prediction Through Machine Learning Algorithms Analyzing Roadway Geometry, Driver Behavior, and Environmental Conditions in Multimodal Transport Networks

Adewumi Augustine Adepitan
Transportation Research
Institute (IMOB)
Hasselt University, Belgium

Abstract: Improving road safety in increasingly complex and multimodal transport networks requires analytical tools capable of capturing the multifaceted interactions between roadway geometry, driver behavior, and environmental conditions. Traditional statistical approaches though valuable often struggle to model nonlinear relationships and hidden dependencies that influence crash likelihood and severity across diverse transportation contexts. Recent advancements in machine learning offer a more adaptive and data-rich pathway for predicting crash severity by learning from large, heterogeneous datasets that integrate roadway design attributes, behavioral indicators, traffic flow characteristics, and real-time environmental factors. Machine learning algorithms such as gradient boosting machines, random forests, support vector machines, and deep neural networks excel in uncovering nonlinear patterns and high-order interactions that traditional models may overlook. When enriched with high-resolution roadway geometry data, including curvature, lane width, grade, and intersection design, these models can better capture how infrastructure directly shapes collision risk. Integrating driver behavior indicators, such as speeding patterns, distraction proxies, acceleration variability, and compliance with traffic control devices, further strengthens predictive accuracy by reflecting human factors that often precede severe crashes. Environmental elements weather, visibility, lighting conditions, and seasonal effects introduce additional variability that machine learning techniques can incorporate dynamically. In multimodal networks, where cars, cyclists, pedestrians, and transit vehicles converge, machine learning provides nuanced insights into exposure risk and conflict points by fusing sensor data, geospatial information, and historical crash records. These predictive capabilities enable transportation agencies to prioritize high-risk corridors, evaluate countermeasure effectiveness, and design targeted interventions that enhance safety across all modes of travel. By bridging infrastructure, behavior, and environmental analytics, machine learning-based crash severity prediction supports proactive, evidence-driven road safety management and more resilient urban mobility planning**.**

Keywords: Crash severity prediction; machine learning; roadway geometry; driver behavior; environmental conditions; multimodal transport networks

1. INTRODUCTION

1.1 Evolution of Road Safety Analytics in Modern Multimodal Networks

The evolution of road safety analytics in modern multimodal networks reflects a fundamental shift from isolated roadway assessments toward integrated views of mobility risk informed by diverse, high-resolution datasets [1]. Traditional safety programs depended on manual reports and periodic engineering studies, which provided only partial insight into how collisions occurred across increasingly complex networks [2]. As mobility expanded to include cyclists, pedestrians, transit vehicles, and emerging micro-mobility modes, analysts recognized that earlier methods could not adequately capture the interactions shaping network risk, especially in dense urban environments [3]. Rapid urban growth intensified exposure to congestion and conflict points, creating new safety demands.

With the rise of intelligent transportation systems, multimodal analytics benefited from high-frequency data streams such as vehicle telemetry, video detection, and environmental sensors, helping researchers monitor near-miss patterns and conflict dynamics rather than relying solely on historical crash outcomes [4]. Geographic information systems further transformed the field by enabling sophisticated spatial comparisons across intersections, corridors, and regional settings [3]. These tools supported risk modeling approaches aligned with Safe System and Vision Zero frameworks, emphasizing that human errors are inevitable and must be mitigated through systemic design [5].

The integration of pedestrian exposure metrics, cyclist trajectory mapping, and transit boarding information introduced a more holistic understanding of safety performance, revealing nuanced risk variations that earlier vehicle-centric studies overlooked [6]. Modern analytics also incorporate lighting conditions, micro-climate data, and roadway geometry captured with unprecedented precision [7]. By layering behavioral indicators such as speed variability and sudden braking onto infrastructure and environmental data, analysts can uncover emerging risk patterns not yet reflected in crash statistics [8]. As multimodal networks become increasingly interconnected through shared mobility platforms and connected infrastructure, road safety analytics continues evolving into a data-rich discipline that captures the complex interplay of human behavior, roadway design, and environmental context [9].

1.2 Traditional Crash Prediction Shortcomings: Linear Models, Limited Features, Low Resolution

Traditional crash prediction models, dominated by linear regression, Poisson, and negative binomial approaches, face persistent limitations because their structures assume simplified relationships between predictors and crash outcomes [5]. These models depend on assumptions of linearity, independence, and fixed variable interactions, which are rarely consistent with real-world crash dynamics shaped by nonlinear behavioral, geometric, and temporal factors [9]. As a result, they often fail to capture the compounded effects of geometry, traffic turbulence, and environmental conditions that elevate crash likelihood under specific combinations of circumstances [10].

Legacy models also relied on narrow feature sets derived from crash reports, low-frequency traffic counts, and broad roadway characteristics, excluding critical behavioral indicators such as hard-braking events, lateral acceleration patterns, or distraction-related anomalies [4]. Without these elements, models cannot reflect the underlying risk environment that persists even in locations with low recorded crash frequency. Furthermore, traditional methods treat crashes as isolated outcomes rather than manifestations of evolving exposure patterns, limiting their sensitivity to emerging hazards [8].

Resolution constraints further weaken predictive accuracy. Many earlier systems used annual crash totals aggregated at coarse spatial scales, masking variations within intersections or complex multimodal conflict zones [7]. These aggregations obscure temporal fluctuations like peak-hour congestion effects or the influence of transitional lighting at dawn and dusk [2]. Linear structures also struggle to incorporate heterogeneous data types video streams, real-time sensor feeds, or GPS-based trajectories which are now central to safety assessment.

Another challenge arises from imbalanced datasets: severe crashes represent a small fraction of total incidents, causing traditional models to gravitate toward majority outcomes and underrepresent high-severity events [10]. Combined, these limitations restrict the ability of legacy tools to represent dynamic, multimodal networks, reinforcing the need for analytical methods capable of capturing nonlinear patterns, integrating diverse data sources, and adapting to fine-grained spatial and temporal variation [1].

1.3 Why Machine Learning Enables a New Paradigm for Crash Severity Modeling

Machine learning enables a transformative paradigm in crash severity modeling by overcoming the structural and data limitations embedded in traditional prediction frameworks. Unlike linear models that impose predefined relationships, machine learning algorithms learn flexible, nonlinear interactions directly from data, capturing the complex mechanisms through which crashes escalate in severity across multimodal environments [6]. This is particularly important where pedestrian, cyclist, and vehicle trajectories intersect unpredictably, generating conflict patterns that vary across time, geometry, and environmental context [3].

Machine learning methods such as gradient boosting, random forests, and neural networks can identify subtle, multivariate effects for example, how curvature combined with surface conditions influences severity without requiring rigid functional assumptions [7]. They also support the integration of high-dimensional data, including pavement images, connected-vehicle telemetry, roadside sensor readings, and LiDAR-derived object detections, enabling richer contextual representations than those used in earlier models [4]. Surrogate safety measures, such as time-to-collision or trajectory divergence, further enhance predictive sensitivity by allowing models to infer severity risk in areas with low crash counts [9].

Another major advantage lies in handling heavily imbalanced datasets, where severe crashes are rare. Machine learning models can incorporate cost-sensitive loss functions, class-balancing techniques, and synthetic sampling approaches to improve sensitivity to high-injury outcomes [1]. Deep learning architectures, including recurrent neural networks, can also detect temporal dependencies capturing how driver behavior, environmental conditions, or traffic turbulence evolve immediately before severe incidents [5].

Additionally, machine learning frameworks offer scalability, allowing continuous retraining as new sensor data, roadway changes, or behavioral patterns emerge. By learning from real-time conditions rather than historical aggregates alone, these models provide transportation agencies with more actionable, proactive severity predictions. Collectively, these capabilities demonstrate why machine learning has become central to next-generation crash severity modeling, enabling a more precise understanding of risk across modern multimodal networks [10].

2. THEORETICAL FOUNDATIONS OF CRASH SEVERITY DETERMINANTS

2.1 Roadway Geometry and Infrastructure Design as Crash Severity Drivers

2.1.1 Horizontal Alignment, Curvature, Grade, and Intersection Control

Roadway geometry establishes the foundational risk landscape that governs how quickly minor driver deviations can escalate into severe crashes. Horizontal alignment, especially in areas with sharp or hidden curvature, elevates lateral acceleration demands, forcing drivers to make rapid stabilizing corrections in environments where traction margins may already be limited [14]. These curvatures interact with sight distance restrictions, reducing the time available to perceive lane position drift or oncoming hazards [9]. Grades intensify these issues: steep downhill segments shorten the effective braking zone, increasing the likelihood that misjudged speeds culminate in lane departures under high kinetic energy [16]. Even mild gradients can create asymmetric load transfers that destabilize vehicles during evasive maneuvers [11]. Intersections introduce additional high-severity exposure because conflict points multiply; inadequate signal timing or limited visibility translates into abrupt angle-impact scenarios, which typically involve greater injury forces [7]. Roundabouts mitigate such severity by redistributing conflict geometries and reducing approach speeds [15]. Ultimately, the interplay of curvature, grade, and control form a structural blueprint that shapes whether minor miscalculations become high-impact crash events [12].

2.1.2 Lane Width, Shoulder Design, Medians, and Traffic-Calming Elements

Lane width influences both operational stability and perceived roadway forgiveness. Narrow lanes restrict lateral maneuvering space, making drivers more vulnerable to drift-related conflicts when distracted or overcorrecting steering inputs [10]. Conversely, excessively wide lanes may unintentionally encourage higher speeds, thereby increasing crash severity even when crash frequency appears stable [7]. Shoulders play a critical buffering role. Wide paved shoulders offer a recovery zone that prevents collisions with rigid objects, whereas soft or narrow shoulders destabilize vehicles, particularly during emergency corrections at high speed [16]. Median design, especially the presence of barrier systems, significantly reduces the likelihood of catastrophic head-on collisions, though rebound impacts remain a possible consequence when vehicles strike rigid dividers at oblique angles [15]. Depressed medians, when adequately sized, prevent cross-over events by providing both spatial and frictional separation [13]. Traffic-calming elements curb extensions, mini-roundabouts, and chicanes reshape driver expectations by signaling the need to reduce speed before conflict areas emerge [9]. These engineering choices collectively regulate maneuverability, spatial tolerance, and impact geometry, forming multilayered defenses that directly modulate crash severity outcomes [14].

2.2 Driver Behavior, Cognitive Load, and Human Error Pathways

Driver behavior remains the most unpredictable determinant of crash severity because human attention and cognitive processing are inherently limited. Drivers continuously synthesize visual cues, environmental signals, and vehicle feedback, yet this processing becomes fragile under conditions of high workload, complex road layouts, or inconsistent signage [8]. When cognitive demand exceeds available capacity, perception delays and decision slowdowns occur, making it more likely that a potential hazard becomes an unavoidable crash event [15]. These delays are particularly dangerous at high speeds, where even a fraction of a second of delayed braking multiplies severity due to exponential increases in kinetic energy [12].

Distraction whether visual, manual, or cognitive reduces situational awareness by redirecting attention away from critical roadway changes [9]. Fatigue compounds this by shrinking the perceptual field and increasing the occurrence of microsleeps, during which drivers unintentionally relinquish control of the vehicle, often resulting in uncorrected departures or high-impact collisions [14]. Impairment from alcohol or psychoactive substances further distorts judgment, slows reaction time, and reduces lane-keeping precision, creating conditions in which crash severity rises sharply because evasive maneuvers are either delayed or executed inefficiently [11].

Experience moderates many of these risks. Novice drivers typically exhibit narrower scanning patterns and slower hazard anticipation, making them more susceptible to late detection of conflicts at intersections or on curved alignments [7]. Experienced drivers, although generally more efficient in perceptual processing, can still succumb to inattentional blindness when their expectations are violated by rare or unconventional hazards such as slow-moving agricultural vehicles or unexpected pedestrian crossings at mid-block locations [13].

Risk perception also influences severity. When drivers perceive a roadway as “safe” due to wide shoulders or gentle curvature, they may increase their speeds neutralizing the intended safety benefits of engineering upgrades [10]. Conversely, visually complex environments can trigger overcautious behavior that paradoxically increases rear-end crash severity when following drivers fail to anticipate sudden speed reductions [16]. Overall, human behavior interacts dynamically with geometric and environmental conditions to determine how pre-crash scenarios evolve into varying levels of severity [15].

2.3 Environmental and Situational Conditions Affecting Crash Outcomes

Environmental conditions modify the interaction between human behavior and roadway geometry, often transforming otherwise manageable situations into high-risk scenarios. Rain reduces pavement friction and elevates hydroplaning probability, impairing driver control during braking or steering adjustments [9]. Snow, ice, and frost magnify these effects by eliminating traction altogether, producing chain-reaction crashes when multiple drivers misjudge stopping distances simultaneously [14]. Fog, smoke, and nighttime lighting limitations reduce visual contrast and depth perception, delaying hazard recognition and shrinking the functional reaction window [11].

Situational traffic conditions further shape severity. Free-flowing high-speed segments allow drivers greater maneuvering freedom but produce extremely severe crashes when conflicts occur due to elevated impact energy [12]. In contrast, congested traffic lowers speeds but increases the risk of multi-vehicle interactions, particularly when speed variance between lanes widens unexpectedly [7]. Sudden deceleration waves, often occurring near work zones, provoke rear-end collisions that become severe when heavy vehicles are involved or when following distances are insufficient [16].

Surface quality and roadside environment also influence crash severity. Potholes, shoulder drop-offs, and debris cause abrupt vehicle destabilization, especially for motorcycles and small cars, triggering rollover or loss-of-control events [15]. Roadside objects trees, poles, steep embankments determine whether off-road excursions result in recoverable maneuvers or high-energy collisions with rigid obstacles [13].

Severity increases sharply when multiple adverse conditions coincide. A tight curve navigated at night during heavy rain combines impaired visibility, reduced friction, and trajectory instability into a single high-risk scenario where even slight steering errors lead to severe outcomes [10]. Likewise, construction zones often merge narrow lanes, temporary barriers, and inconsistent pavement textures, compounding environmental stressors that heighten both crash likelihood and severity [8]. These combined situational factors show that environmental conditions function as powerful amplifiers that shape the ultimate magnitude of crash consequences [14].

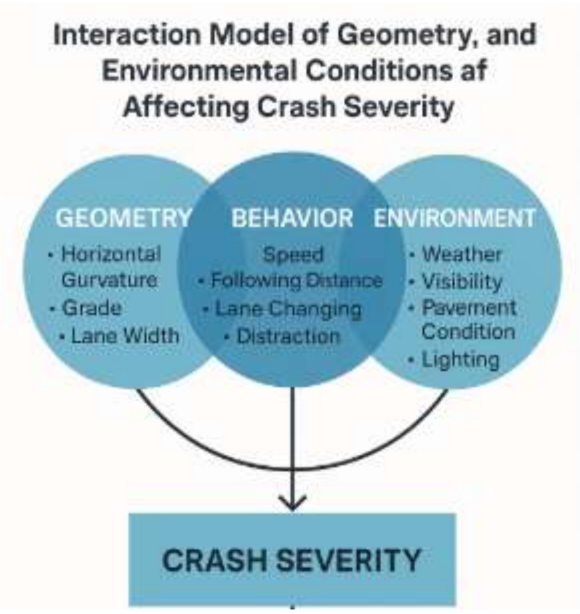

Figure 1. *Interaction Model of Geometry, Behavior, and Environmental Conditions Affecting Crash Severity.*

3. DATA ARCHITECTURE AND GIS-BASED FEATURE ENGINEERING

3.1 Data Sources for Crash Severity Prediction

3.1.1 Crash Records, Exposure Data, and Injury Severity Coding

Crash severity prediction depends fundamentally on the completeness and accuracy of crash records, which provide the essential ground-truth labels required for model training and validation. These datasets typically include detailed information on collision type, vehicle movements, injury levels, roadway context, occupant characteristics, and temporal factors. Police crash reports remain one of the most widely used data sources, containing structured variables and narrative fields that describe pre-crash events and contributing conditions [17]. Hospital trauma registries and emergency medical services introduce further granularity by capturing physiological metrics and confirmed injury severity, enabling cross-validation of field-reported severity levels [19]. Exposure data such as vehicle miles traveled, intersection turning counts, and pedestrian or cyclist volumes provide the denominators necessary to contextualize crash risk rather than focusing solely on outcomes [15]. Injury severity coding frameworks, including AIS, MAIS, and KABCO, standardize these records so that ML models can differentiate between fatal, serious, minor, and property-damage crashes [22]. Because reporting inconsistencies and underreporting bias often distort severity classification, integrating multiple data sources produces more reliable severity labels that enhance predictive performance and reduce model noise [16].

3.1.2 Roadway Geometry, Sensor Streams, Weather, and Traffic Data

Roadway geometry datasets supply critical structural variables that influence both crash likelihood and severity. GIS-based inventories include curvature radii, grade profiles, lane configurations, control types, and roadside clear zones, each of which contributes distinct geometric risk factors [24]. Increasingly, agencies rely on mobile LiDAR and high-resolution roadway imaging to capture finer roadway attributes such as barrier offsets, cross-slopes, and pavement textures [14]. Complementing static geometry, sensor streams from radar detectors, inductive loops, and Bluetooth tracking provide real-time measurements of traffic flow, headway distributions, and speed variance factors that strongly affect severity outcomes [20]. Weather data, obtained from mesonets and roadway environmental sensor stations, capture rainfall intensity, surface temperature, visibility, and freezing conditions that amplify crash severity during adverse events [18]. Traffic management centers provide dynamic data such as incident logs, congestion patterns, lane closures, and work-zone adjustments that influence crash contexts [21]. By integrating geometry, traffic flow, environmental conditions, and operational variables, ML pipelines gain access to rich multimodal datasets that reflect the layered determinants shaping crash severity [23].

3.2 GIS-Based Spatial Feature Engineering for ML Pipelines

Spatial feature engineering forms the backbone of modern crash severity modeling because crash risk is deeply tied to spatial context. GIS enables precise geolocation of crash points and supports the extraction of high-resolution roadway and environmental attributes from spatial layers. Kernel density estimation, for example, identifies spatial clusters of severe collisions, allowing models to incorporate hotspot intensity as a predictive variable [19]. Buffer-based feature extraction generates spatially linked variables such as proximity to schools, bars, high-speed corridors, or signalized intersections, all of which influence driver behavior and conflict likelihood [22]. Similarly, spatial joins integrate crash points with roadway geometry layers to quantify curve sharpness, intersection complexity, and roadside hazard density at the crash location [16]. Terrain layers introduce slope, elevation, and line-of-sight measures that shape pre-crash trajectories, particularly on rural networks [24].

GIS further supports temporal-spatial fusion by aligning weather rasters, traffic sensor feeds, and land-use maps with crash coordinates. This integration allows ML models to learn how environmental and operational variations interact with spatial features under specific conditions [18]. Road network topology is another critical spatial variable. Measures such as centrality, connectivity, and link hierarchy reveal whether a segment serves as a local access road or a high-speed through corridor, influencing the magnitude of forces experienced in crashes [20].

Advanced spatial encodings including graph-based embeddings, road-segment fingerprinting, and polyline convolution features allow neural networks to model the geometric continuity of the roadway environment rather than treating crash points as isolated observations [14]. When spatial features are properly engineered, ML pipelines gain the ability to detect nonlinear interactions across geometry, context, and behavior. Thus, GIS-based feature engineering transforms raw coordinates into structured, multidimensional representations that dramatically strengthen crash severity prediction performance [23].

3.3 Multimodal Network Integration: Pedestrian, Cyclist, and Transit Interactions

Crash severity is not solely a function of vehicle-to-vehicle dynamics; urban multimodal systems introduce additional interactions that shape severity patterns. Pedestrian and cyclist movements create exposure zones where vulnerable road users intersect with motorized traffic, often at locations where vehicle speeds remain high or sight distance is limited [21]. Integrating pedestrian counts, mid-block crossing density, and sidewalk continuity into ML models helps quantify these risk concentrations [16]. Cyclist route choice, derived from GPS traces, bike-share trip logs, and protected-lane inventories, reveals conflict lines where vehicle trajectories and cyclist movements intersect at acute angles, increasing severity risk even when crash frequency is low [19].

Transit systems also contribute crucial contextual data. Bus stop locations, boarding volumes, curbside dwell times, and transit signal priority phases alter local traffic dynamics by increasing deceleration events and lane merges near stops [24]. These micro-interactions elevate the probability of severe conflicts when drivers make sudden maneuvers to overtake or avoid stopped transit vehicles [15]. Rail crossings introduce additional multimodal hazards, where warning compliance behaviors, queue spillbacks, and gate timing influence peak crash severity [18].

Multimodal network integration requires harmonizing datasets with different spatial resolutions and temporal frequencies. Pedestrian counts may be intermittent, cyclist traces sparse, and transit logs minute-by-minute. GIS harmonization aligns these multimodal layers with crash coordinates, enabling ML models to learn how modal interactions reshape severity outcomes [23]. Such integration is essential for cities transitioning toward complete-streets frameworks, where modal density increases the complexity of crash dynamics [17]. Ultimately, multimodal datasets enrich ML pipelines by supplying nuanced behavioral and operational variables that traditional vehicle-centric datasets often overlook [20].

Table 1. Summary of Key Input Variables for ML Crash Severity Models

Category	Key Variables
Geometry	Curvature radius; Grade/slope; Lane width; Shoulder width; Median type; Intersection control; Sight distance
Behavior	Speed variation; Headway; Lane-changing frequency; Distraction indicators; Impairment/fatigue; Risk-taking behavior
Environment	Weather (rain/snow/fog); Lighting conditions; Pavement condition; Visibility; Temperature
Operational	Traffic volume; Speed profiles; Transit activity; Pedestrian/cyclist volumes; Work-zone conditions; Sensor-based flow metrics

4. MACHINE LEARNING FRAMEWORK FOR CRASH SEVERITY PREDICTION

4.1 Supervised Learning Approaches for Injury Severity Classification

Supervised learning remains the foundational approach for injury severity classification because labeled crash datasets enable direct mapping between input features and severity outcomes. In these models, injury levels often coded using KABCO or MAIS systems serve as targets that algorithms learn to differentiate based on geometric, behavioral, environmental, and operational predictors [27]. Logistic regression has traditionally served as a baseline classifier due to its interpretability, but its linear structure struggles to capture nonlinear interactions commonly present in crash dynamics [23]. Support Vector Machines provide improved boundary separation in high-dimensional spaces, yet their sensitivity to kernel selection and limited scalability in large datasets constrain their utility in broader roadway networks [29].

Decision trees offer intuitive rule-based segmentation and can model conditional interactions between speed, geometry, and environmental conditions, but they are prone to instability when small changes in data shift the structure of the tree [22]. Naïve Bayes classifiers perform efficiently with large datasets but assume conditional independence between predictors—a simplification that rarely holds in multimodal crash environments [24]. In contrast, k-Nearest Neighbors offers a distance-based classification perspective but becomes computationally intensive with expanding datasets and offers limited insight into the underlying relationships [28].

Supervised models rely heavily on high-quality labeled data, making consistent injury coding and accurate crash reporting essential. When inputs include harmonized geometry, traffic flow, and environmental variables, these classifiers begin capturing nuanced severity determinants that reflect real-world conditions [25]. However, because severe crashes are relatively rare events, supervised classifiers often struggle with imbalanced distributions that skew predictions toward frequent mild outcomes [30]. Therefore, supervised learning frequently acts as the foundation upon which more sophisticated ensemble, deep learning, and hybrid frameworks are constructed to capture the complexity of multimodal crash severity patterns [26].

4.2 Ensemble and Advanced ML Algorithms

4.2.1 Random Forest, Gradient Boosting, XGBoost, CatBoost

Ensemble tree-based algorithms have become dominant in crash severity modeling because they overcome many limitations inherent in single-model approaches. Random Forest, which aggregates predictions from multiple decorrelated decision trees, enhances generalization by combining diverse feature splits, making it robust against noise and overfitting [30]. Through random feature selection and bootstrap sampling, the model captures nonlinear relationships between speed variability, geometric features, environmental conditions, and multimodal traffic dynamics [23]. Gradient Boosting frameworks extend this logic by sequentially generating trees that correct residual errors from prior iterations, allowing the model to refine severity prediction patterns that single trees might overlook [29].

XGBoost further optimizes boosting through regularization, weighted loss functions, and efficient parallel computation, enabling it to learn subtle feature interactions, especially those involving curve sharpness, nighttime visibility, and pedestrian exposure [24]. Its ability to incorporate custom objective functions makes it valuable in addressing skewed severity distributions. CatBoost extends gradient boosting to handle categorical crash variables such as intersection types, traffic-control devices, and land-use categories without requiring extensive preprocessing [26]. This proves advantageous in crash datasets that mix continuous and categorical spatial attributes.

Collectively, these ensemble models deliver superior accuracy, accommodate heterogeneous data, and support feature-importance extraction that reveals critical severity determinants. Their flexibility and strong predictive performance make them widely adopted in injury severity prediction pipelines across multimodal networks [25].

4.2.2 Deep Neural Networks and Hybrid Spatiotemporal Models

Deep learning methods offer an advanced alternative capable of learning hierarchical relationships from complex multimodal datasets. Deep Neural Networks (DNNs) capture nonlinear interactions between roadway geometry, driver behavior indicators, weather sequences, and temporal traffic flow patterns, enabling a richer representation of the crash environment [27]. With sufficient data, DNNs detect latent severity patterns not easily identifiable through shallow algorithms, particularly when interactions span multiple layers of spatial context [22].

Hybrid spatiotemporal architectures build on this by combining convolutional and recurrent components. Convolutional Neural Networks (CNNs) extract spatial features from roadway segments, curvature grids, and heat-map encodings of traffic density, while Recurrent Neural Networks (RNNs) model temporal trends such as precipitation changes, peak-hour congestion, and sensor-based speed fluctuations [30]. Long Short-Term Memory (LSTM) variants are especially effective in retaining sequential dependencies that influence pre-crash conditions [28]. Graph Neural Networks (GNNs) extend this capability by modeling the roadway network as a graph structure, allowing algorithms to learn severity propagation patterns across connected links and multimodal corridors [24].

Hybrid designs allow deep learning models to incorporate geometry, environment, and behavior simultaneously, generating predictions that reflect real-world crash dynamics more accurately than single-input architectures. These systems are increasingly used in urban analytics where pedestrian exposure, cyclist flow, and transit operations introduce complex multimodal interactions [29].

4.3 Handling Imbalanced Crash Data and Rare High-Severity Events

Crash severity datasets are often highly imbalanced because severe and fatal crashes occur far less frequently than minor collisions. This imbalance poses a major challenge for machine learning algorithms, which tend to optimize toward the dominant class, undermining the model’s ability to detect rare but consequential high-severity events [25]. Traditional classifiers frequently mislabel severe crashes as minor because the cost of misclassification is lower under standard loss functions [22]. Therefore, specialized techniques are required to correct the imbalance and maintain reliability.

Resampling strategies include oversampling minority severe cases and undersampling majority mild cases. Synthetic Minority Oversampling Technique (SMOTE) generates synthetic injury-severity examples by interpolating minority instances, helping models learn more stable boundaries between classes [30]. Cluster-based undersampling removes redundant mild-severity examples while preserving representative variability, improving class balance without discarding key spatial patterns [27].

Algorithmic approaches incorporate cost-sensitive learning, where misclassifying a severe crash incurs a higher penalty, reshaping decision boundaries toward safety-critical predictions [24]. Ensemble variants, such as Balanced Random Forest or XGBoost with customized severity-weighted losses, offer further improvements by integrating sampling and cost adjustments within the training process [29]. Deep learning frameworks also address imbalance by using focal loss, which emphasizes harder-to-classify severe crashes and down-weights frequent minor cases [23].

Another challenge lies in the noise and uncertainty surrounding severe crash reports. Because high-severity crashes often occur under extreme or rare environmental conditions, models must avoid learning spurious correlations. Hybrid pipelines combine resampling, weighting, and regularization to ensure that rare severe events are adequately represented without inflating false positives [28]. Effectively managing imbalanced data improves not only classification accuracy but also real-world utility, given that identifying severe crashes is central to safety planning and resource allocation across multimodal networks [26].

4.4 Model Evaluation, Feature Importance, and Explainable AI Tools

Model evaluation requires metrics sensitive to class imbalance and reflective of real-world safety priorities. Accuracy alone is insufficient because it masks poor performance on rare severe cases. Metrics such as F1-score, weighted recall, ROC-AUC, and precision-recall curves provide more meaningful assessments of severity prediction quality [29]. Confusion matrices further reveal misclassification patterns across severity levels, allowing targeted adjustments to sampling or cost functions [22].

Feature-importance tools including permutation importance, SHAP values, and partial dependence plots enable analysts to interpret how geometric, behavioral, and environmental variables influence predictions [27]. These tools are essential for validating that models learn meaningful relationships rather than artifacts of imbalanced data [25]. Explainable AI techniques make complex ensemble and deep learning models more transparent by highlighting variable interactions and identifying key contributors in multimodal settings [30]. This interpretability strengthens trust and supports safety-oriented decision-making by transportation agencies [24].

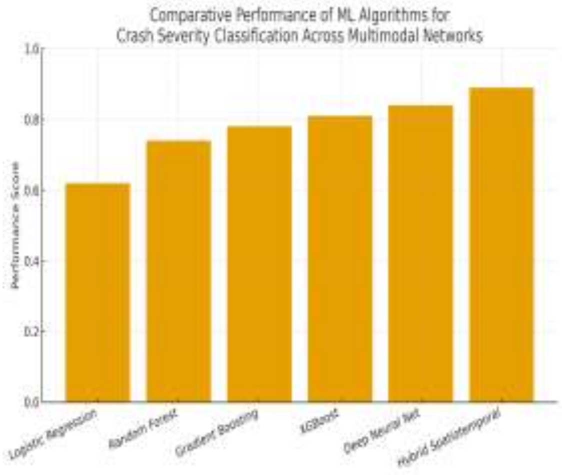

Figure 2. *Comparative Performance of ML Algorithms for Crash Severity Classification Across Multimodal Networks.*

5. MULTIMODAL SAFETY EVALUATION AND RISK INTERPRETATION

5.1 Severity Prediction Across Cars, Cyclists, Pedestrians, and Transit Modes

5.1.1 Vehicle Occupant Risk and High-Speed Geometry Factors

Crash severity prediction across vehicle occupants depends heavily on the interactions between vehicle mass, restraint systems, and roadway geometry. Motor-vehicle occupants experience a disproportionate escalation in severity when crashes occur at segments with long tangent sections, high approach speeds, or insufficient curvature warning, as kinetic energy compounds the forces applied during impact [34]. Severity increases further on multilane arterials where lane widths encourage higher operating speeds, reducing reaction time and magnifying crash intensity when lane departures occur [29]. Vehicle occupants are also highly sensitive to grade transitions; steep downhill segments decrease braking capacity, increase velocity at impact, and reduce the margin for evasive maneuvers [31]. Occupant injury profiles are shaped by seatbelt use, seating position, and vehicle structural characteristics, yet environmental context such as nighttime conditions or wet pavement remains equally influential because degraded visibility or traction amplifies crash forces experienced inside the vehicle [28]. Intersections with complex phasing, wide turning radii, or inconsistent signage frequently yield near-side or head-on collisions, which are inherently high severity due to direct energy transfer [33]. Predictive models integrating speed variance, geometry, and vehicle-specific parameters capture these layered determinants with greater fidelity [30].

5.1.2 Vulnerable Road User (VRU) Safety Under Mixed Traffic Conditions

Pedestrians and cyclists experience crash outcomes that differ substantially from vehicle occupants because they lack protective structures and are exposed directly to impact forces. Severity prediction for VRUs must therefore incorporate multimodal traffic flow, crossing density, bicycle lane connectivity, and conflict patterns generated by turning vehicles [35]. In mixed traffic environments, pedestrian risk spikes near unsignalized mid-block crossings, loading zones, and transit stops where movement patterns are highly unpredictable [30]. Cyclists encounter elevated severity risk at intersections where right-turning vehicles cross designated bike lanes or where drivers misjudge cyclist approach speeds under limited visibility [28]. Even at low speeds, VRU collisions result in severe injuries because impact forces are absorbed directly by the body, particularly head and lower-limb regions [33]. Environmental variability plays an even larger role for VRUs; wet pavement, poor lighting, and glare from oncoming headlights intensify the consequences of driver errors, especially in constrained urban corridors [29]. Predictive algorithms incorporating pedestrian density, cyclist routing, lateral clearance, and transit dwell-time patterns allow ML models to differentiate high-risk VRU zones from vehicular hotspots, generating more inclusive severity assessments [34].

5.2 Environmental and Situational Variability in Crash Severity

5.2.1 Weather, Visibility, Pavement Condition, and Lighting Effects

Environmental variability introduces significant fluctuations in crash severity because weather, visibility, and pavement condition directly alter driver perception, surface friction, and control stability. Rain reduces tire–road adhesion and increases braking distances, elevating the probability that minor steering errors escalate into high-energy crashes [28]. Snow and ice generate even greater severity potential, as vehicles lose traction entirely and collide with limited pre-impact deceleration, resulting in greater energy transfer at the moment of impact [35]. Fog and smoke impair depth perception and reduce contrast, delaying hazard recognition in conditions where reaction windows are already constrained by roadway geometry [30]. Nighttime conditions remain one of the strongest severity predictors because low illumination impairs driver scanning and increases the risk of striking pedestrians or roadside obstacles at high speed [33]. Pavement deterioration such as rutting or potholes forces abrupt maneuvers that destabilize vehicle alignment, intensifying crash severity on curves or narrow corridors [29]. When environmental hazards occur concurrently, such as rain at night or fog on high-speed curves, severity probability rises nonlinearly. Machine learning models incorporating spatiotemporal weather feeds and surface condition indices capture these compounding effects more accurately than traditional statistical methods [31].

5.2.2 Temporal Variations: Peak Traffic, Weekend Effects, and Seasonal Patterns

Severity levels fluctuate substantially across temporal dimensions, as traffic density, travel purpose, and behavioral patterns shift throughout the day and year. During peak hours, stop-and-go traffic reduces average speeds but increases rear-end collision likelihood; however, the lower kinetic energy often moderates severity outcomes [34]. Conversely, off-peak periods, especially late at night, produce fewer crashes but much higher severity due to elevated travel speeds and increased incidence of impairment or fatigue [28]. Weekend patterns introduce distinct risks as recreational travel, alcohol involvement, and rural roadway exposure combine to elevate fatality rates in low-density environments [32]. Seasonal variations also contribute: winter months present severe traction challenges, while summer increases pedestrian and cyclist activity, shifting severity patterns across modes [30]. Holiday periods, characterized by longer trips, driver fatigue, and higher traffic volumes, often yield spikes in high-severity crashes that reflect both behavioral and environmental stressors [29]. Incorporating hourly traffic flows, weekend seasonality, event-based surge patterns, and daylight cycles into ML severity classification models enables more accurate temporal risk forecasting and enhances proactive safety interventions [33].

5.3 Spatial Hotspot Analysis and Risk Mapping Using ML Outputs

Spatial hotspot analysis plays a central role in interpreting ML-based crash severity predictions because severity clusters often reflect systemic design vulnerabilities rather than isolated events. Hotspot mapping using GIS allows analysts to identify spatial concentrations of high-severity predictions along corridors with adverse geometry, large speed differentials, or multimodal conflict density [28]. Kernel density estimation and spatial autocorrelation metrics such as Moran’s I highlight severity-prone zones that may not be evident from raw crash counts alone, especially when multilayer factors such as weather sensitivity or visibility constraints interact spatially [35]. ML outputs enrich these analyses by incorporating learned nonlinear relationships between roadway context and severity, enabling identification of complex spatial risk signatures that reflect multimodal interactions rather than single-factor anomalies [31].

High-severity clusters frequently coincide with curved rural highways, multilane urban arterials, and locations where pedestrian or cyclist flows intersect high-speed traffic streams [29]. Integrating predicted severity probabilities with land-use layers reveals hotspots around nightlife districts, industrial loading zones, schools, and transit hubs contexts where exposure profiles shift rapidly across time and user groups [34]. Network-based hotspot detection methods further extend spatial analysis by examining severity across connected roadway links rather than isolated points. These methods identify corridors with propagation risk, such as segments where poor sight distance or inconsistent lane widths create sequential high-severity events [30].

ML-based severity surface maps also enable environmental overlay analysis. When weather sensitivity layers, lighting grids, or pavement-condition indices are combined with predicted severity, the resulting composite surfaces reveal zones prone to multi-factor severity amplification [33]. These risk maps guide targeted engineering interventions such as improved illumination, speed-calming features, signal timing adjustments, and protected VRU crossings. By translating model outputs into spatial decision-support tools, agencies can prioritize investments where they yield the highest safety benefit [32].

Table 2. Multimodal Severity Risk Indicators Derived From ML-Based Classification Models

Mode	Key Severity Risk Indicators
Vehicles	High speed variance; Curve-related instability; Nighttime visibility loss; Lane departures; Sudden braking clusters
Pedestrians	Mid-block crossing density; Poor lighting; Transit stop proximity; Long crossing distances; High turning-vehicle volume
Cyclists	Conflicts at right-turn lanes; Lack of bike-lane protection; Narrow lateral clearance; Surface defects; Mixed-traffic exposure
Transit	Bus stop dwell-time conflicts; Lane blocking; Pedestrian surges; Signal priority interactions; Queue spillback zones
Network-Level	Multimodal conflict hotspots; High-risk geometric segments; Weather-sensitive corridors; Temporal severity peaks

6. POLICY, INFRASTRUCTURE, AND OPERATIONAL APPLICATIONS

6.1 Data-Driven Infrastructure Design and Safety Countermeasures

Data-driven infrastructure design increasingly relies on machine learning (ML) outputs to identify where and why severe crashes occur. By analyzing predicted severity probabilities across large spatial networks, engineers can prioritize corridors exhibiting high-risk geometric and behavioral patterns, enabling more precise deployment of countermeasures [36]. ML-derived features such as speed variance, curvature thresholds, lighting deficits, and multimodal conflict densities highlight conditions that amplify severity and guide infrastructure redesign decisions [34]. For example, models frequently identify multilane arterials with limited median protection or inconsistent lane widths as high-severity corridors, prompting agencies to introduce raised medians, refuge islands, or road diets to reduce kinetic energy at conflict points [40].

Crash severity prediction also improves the selection of curve-specific safety interventions. Predictive outputs showing elevated severity at sharp curves or downhill segments inform the placement of chevron signs, high-friction surface treatments, and dynamic advisory speed systems [35]. When integrated with pedestrian and cyclist exposure maps, ML predictions guide the installation of protected intersections, widened sidewalks, or grade-separated crossings that reduce conflict between vulnerable road users and high-speed vehicles [38].

Spatial overlays combining predicted severity with land-use, transit activity, or nighttime illumination patterns help identify emerging hotspots before crashes accumulate. These insights support proactive deployment of lighting upgrades, signal timing modifications, speed humps, and automated speed enforcement where risks are highest rather than where incidents have already occurred [37].

Ultimately, ML-enabled infrastructure planning shifts the focus from reactive crash mitigation to preventive safety engineering. By translating high-risk patterns into countermeasure designs, agencies align roadway geometry, traffic control, and multimodal facilities with Safe-System-aligned severity reduction goals [39].

6.2 Real-Time Decision Support for Traffic Management and Enforcement

Real-time ML systems enhance traffic management by generating continuous risk assessments derived from live sensor streams, weather feeds, and connected-vehicle telemetry. These dynamic severity alerts identify segments where conditions such as rain-induced speed dispersion, fog-related visibility loss, or sudden congestion waves elevate the probability of severe crashes within minutes [34]. Control centers use this information to activate variable speed limits, reroute traffic, or deploy rapid response teams before hazardous conditions escalate into crashes [38].

Automated enforcement and patrol allocation also benefit directly from ML outputs. Severity-weighted heatmaps allow agencies to assign enforcement units strategically to corridors where risky driver behaviors such as excessive speeding or aggressive lane changes coincide with poor geometry or nighttime conditions [40]. This shifts enforcement efforts from broad area coverage to precision-focused intervention [35].

When integrated into adaptive signal control systems, ML predictions trigger proactive phasing adjustments at intersections experiencing high-severity potential due to pedestrian surges, bus dwell times, or reduced visibility [37]. Together, these real-time decision-support functions transform ML outputs into operational tools that reduce both crash likelihood and severity across multimodal networks [36].

International Journal of Science and Engineering Applications
Volume 11-Issue 12, 469 – 480, 2022, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1112.1068

6.3 Integrating Machine Learning Insights into Vision Zero and Safe Systems Frameworks

Machine learning insights align closely with Vision Zero and Safe Systems principles because predictive patterns illustrate where systemic risks accumulate and how the transportation network can absorb human error before severe injury occurs. ML models reveal nonlinear interactions between geometry, behavior, and environmental stressors, enabling agencies to identify failure points where crash forces exceed human biomechanical limits [39]. Vision Zero prioritizes speed management, and ML-derived severity predictions consistently show where small speed reductions achieve disproportionate safety benefits particularly at conflict-rich intersections and multimodal corridors [34].

Safe Systems frameworks emphasize protecting vulnerable road users. ML outputs incorporate pedestrian density, cyclist routing, and transit activity, enabling cities to engineer safer crosswalks, protected lanes, and transit-priority zones based on predicted severity gradients rather than historic crash records alone [38].

Equally important is the policy integration of ML-based equity assessment. Severity surfaces linked with socioeconomic and land-use layers identify neighborhoods disproportionately exposed to high-risk roadway environments, supporting targeted improvements in lighting, crossing accessibility, and traffic calming [36]. Through these applications, ML does more than predict crashes it reshapes how agencies prioritize safety resources in ways consistent with Vision Zero's commitment to eliminating severe injuries and fatalities [41].

6.4 Future Innovations: Digital Twins, Connected Vehicles, and Edge-Based Safety Analytics

Emerging technologies will deepen the role of machine learning in crash severity prevention. Digital twins virtual replicas of roadway networks enable simulation of crash-risk scenarios under changing geometry, traffic loads, or weather conditions. These platforms use ML-driven severity models to evaluate countermeasures before real-world deployment, reducing design uncertainty and accelerating decision-making [35].

Connected vehicles introduce real-time behavioral and environmental telemetry, including hard-braking events, traction loss, and near-miss patterns that enhance prediction accuracy for rare severe events [34]. As more vehicles share sensor data, ML models will capture risk signatures at unprecedented granularity, supporting hyper-localized safety interventions [37].

Edge computing brings severity prediction directly to roadside units, enabling low-latency alerts that notify drivers, cyclists, or pedestrians of imminent hazards including slippery surfaces, sudden congestion, or visibility deterioration without relying on centralized servers [40]. Integrating these capabilities with smart signals, dynamic speed advisories, and adaptive lighting systems allows cities to deploy automated, context-aware safety responses [36].

Together, digital twins, connected mobility, and edge analytics represent the next evolution of data-driven safety infrastructure one in which ML becomes an embedded, real-time component of roadway systems rather than a post-hoc analytical tool [38].

Figure 3. *Operational Pipeline for Real-Time Crash Severity Risk Alerts in Multimodal Transport Systems.*

7. CONCLUSION

7.1 Summary of Advancements in Spatial–Machine Learning Crash Modeling

Recent advancements in spatial–machine learning crash modeling have transformed the field from a historically reactive discipline into a forward-looking predictive science. The integration of high-resolution roadway geometry, multimodal flow data, and environmental context has enabled models to capture nonlinear severity relationships that traditional statistical approaches often overlooked. GIS-enabled feature engineering now allows crash locations to be enriched with spatial buffers, network topology, and land-use indicators, resulting in more context-sensitive predictions.

Ensemble algorithms and deep neural networks have further improved the capacity to detect complex risk signatures associated with speed variation, visibility loss, multimodal conflicts, and geometric discontinuities. Importantly, ML models now support real-time risk assessment through the fusion of connected-vehicle streams, weather sensors, and traffic monitoring systems. These advancements collectively shift safety management toward proactive identification of emerging hazards, allowing agencies to target countermeasures more efficiently and design infrastructure that better absorbs human error.

7.2 Future Research Agenda for Data Fusion, Ethics, and Multimodal Safety

Future research must prioritize deeper multimodal data integration to better capture the interactions among vehicles, cyclists, pedestrians, and transit systems. As connected-vehicle and IoT sensor networks expand, opportunities will grow for fusing continuous telemetry, roadway imaging, and environmental feeds to support more accurate real-time severity prediction. Ethical considerations require equal attention, including responsible data governance, mitigation of algorithmic bias, and transparency in safety-critical decision systems. Ensuring that ML models do not reinforce disparities in enforcement or infrastructure investment will be essential. Advancements in digital twins, edge-based analytics, and simulated human behavior modeling can unlock more dynamic risk assessments that reflect evolving traffic ecosystems. Research should also focus on harmonizing privacy-preserving technologies with safety analytics, enabling detailed insights without compromising user confidentiality. By addressing data fusion, ethical safeguards, and multimodal inclusion, the next generation of crash modeling can deliver safer, more equitable transportation networks.

8. REFERENCE

1. Wen X, Xie Y, Jiang L, Pu Z, Ge T. Applications of machine learning methods in traffic crash severity modelling: current status and future directions. Transport reviews. 2021 Nov 2;41(6):855-79.
2. Daniel ONI. TOURISM INNOVATION IN THE U.S. THRIVES THROUGH GOVERNMENTBACKED HOSPITALITY PROGRAMS EMPHASIZING CULTURAL PRESERVATION, ECONOMIC GROWTH, AND INCLUSIVITY. International Journal Of Engineering Technology Research & Management (IJETRM). 2022Dec21;06(12):132–45.
3. Eze Dan-Ekeh. DEVELOPING ENTERPRISE-SCALE MARKET EXPANSION STRATEGIES COMBINING TECHNICAL PROBLEM-SOLVING AND EXECUTIVE-LEVEL NEGOTIATIONS TO SECURE TRANSFORMATIVE INTERNATIONAL ENERGY PARTNERSHIPS. International Journal Of Engineering Technology Research & Management (IJETRM). 2018Dec21;02(12):165–77.
4. Alabede LA, Maimako SM. Optimizing autonomous drone deployment strategies to improve geological structural mapping accuracy in complex mines. *International Journal of Computer Applications Technology and Research*. 2019;8(12):634-646.
5. Alabede LA. Enhancing underground drone endurance by optimizing battery efficiency, communication robustness, and navigation autonomy. *International Journal of Computer Applications Technology and Research*. 2016;5(12):831-843
6. During D. Machine-driven quantitative modeling approaches for measuring liquidity risk propagation across interconnected financial systems. International Journal of Computer Applications Technology and Research. 2022;11(12):753–764. doi:10.7753/IJCATR1112.1033.
7. Olanlokun Y, Taiwo M. Quantifying economic impact of COVID-19-induced disruptions in Nigeria's medicine supply chains: prices, availability and out-of-pocket burdens. World J Adv Res Rev. 2020;7(3):357-371. doi:10.30574/wjarr.2020.7.3.0346
8. Rumbidzai Derera. HOW FORENSIC ACCOUNTING TECHNIQUES CAN DETECT EARNINGS MANIPULATION TO PREVENT MISPRICED CREDIT DEFAULT SWAPS AND BOND UNDERWRITING FAILURES. International Journal of Engineering Technology Research & Management (IJETRM). 2017Dec21;01(12):112–27.
9. During D. Applied statistical methods for stress-testing credit portfolios and forecasting default probabilities in volatile markets. Int J Sci Res Arch. 2022;7(2):924-937. doi:10.30574/ijsra.2022.7.2.036
10. Lukman A Alabede, Samuel Mohammed Maimako. Improving mine ventilation planning using drone-enabled atmospheric tracking, particulate mapping, and hazardous gas detection. Int J Comput Artif Intell 2022;3(2):113-123. DOI: 10.33545/27076571.2022.v3.i2a.212
11. Kolawole Oloke. CLOUD-ACCELERATED PREDICTIVE TREASURY MANAGEMENT USING DEEP REINFORCEMENT LEARNING AND FINANCIAL DIGITAL TWINS. International Journal Of Engineering Technology Research & Management (IJETRM). 2022Dec21;06(12):190–204.
12. Komol MM, Hasan MM, Elhenawy M, Yasmin S, Masoud M, Rakotonirainy A. Crash severity analysis of vulnerable road users using machine learning. PLoS one. 2021 Aug 5;16(8):e0255828.
13. Mannering FL, Bhat CR. Analytic methods in accident research: Methodological frontier and future directions. Analytic methods in accident research. 2014 Jan 1;1:1-22.
14. Mussone L, Bassani M, Masci P. Analysis of factors affecting the severity of crashes in urban road intersections. Accident Analysis & Prevention. 2017 Jun 1;103:112-22.
15. Atanda ED. EXAMINING HOW ILLIQUIDITY PREMIUM IN PRIVATE CREDIT COMPENSATES

ABSENCE OF MARK-TO-MARKET OPPORTUNITIES UNDER NEUTRAL INTEREST RATE ENVIRONMENTS. International Journal Of Engineering Technology Research & Management (IJETRM). 2018Dec21.;2(12):151-64.

16. Nwenekama Charles-Udeh. Leveraging financial innovation and stakeholder alignment to execute high-impact growth strategies across diverse market environments. Int J Res Finance Manage 2019;2(2):138-146. DOI: 10.33545/26175754.2019.v2.i2a.617
17. Lukman A. Alabede, Samuel Mohammed Maimak, Joel Mintah Opoku. LEVERAGING MULTISPECTRAL DRONE IMAGING TECHNOLOGIES FOR MONITORING OPEN-PIT MINES AND IMPROVING PRODUCTION EFFICIENCY. International Journal Of Engineering Technology Research & Management (IJETRM). 2020Dec21;04(12):199–212.
18. Eze Dan-Ekeh. Engineering high-value commercialization frameworks integrating technical innovation with strategic sales leadership to drive multimillion-dollar growth in global energy markets. World J Adv Res Rev. 2019;4(2):256-268. doi:10.30574/wjarr.2019.4.2.0152
19. Theofilatos A, Chen C, Antoniou C. Comparing machine learning and deep learning methods for real-time crash prediction. Transportation research record. 2019 Aug;2673(8):169-78.
20. Fiorentini N, Losa M. Handling imbalanced data in road crash severity prediction by machine learning algorithms. Infrastructures. 2020 Jul 20;5(7):61.
21. Al-Mistarehi BW, Alomari AH, Imam R, Mashaqba M. Using machine learning models to forecast severity level of traffic crashes by R Studio and ArcGIS. Frontiers in built environment. 2022 Apr 20;8:860805.
22. Jiang L, Xie Y, Wen X, Ren T. Modeling highly imbalanced crash severity data by ensemble methods and global sensitivity analysis. Journal of Transportation Safety & Security. 2022 Apr 3;14(4):562-84.
23. Elamrani Abou Elassad Z, Mousannif H, Al Moatassime H. Class-imbalanced crash prediction based on real-time traffic and weather data: A driving simulator study. Traffic injury prevention. 2020 Apr 2;21(3):201-8.
24. Oloke K. Designing Cloud-Native Risk Orchestration Layers for Real-Time Fraud Detection in Digital Banking Ecosystems. *International Journal of Computer Applications Technology and Research*. 2019;8(12):647-658.
25. Mafi S, AbdelRazig Y, Doczy R. Machine learning methods to analyze injury severity of drivers from different age and gender groups. Transportation research record. 2018 Dec;2672(38):171-83.
26. Wen X, Xie Y, Jiang L, Li Y, Ge T. On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development. Accident Analysis & Prevention. 2022 Apr 1;168:106617.
27. Udeh NC. *Building sustainable SME banking strategies that expand market access, boost client retention, and support economic inclusion*. International Journal of Financial Management and Economics. 2018;1(1):126-135. doi:10.33545/26179210.2018.v1.i1.674.
28. Zhang J, Li Z, Pu Z, Xu C. Comparing prediction performance for crash injury severity among various machine learning and statistical methods. IEEe Access. 2018 Oct 9;6:60079-87.
29. Man CK, Quddus M, Theofilatos A. Transfer learning for spatio-temporal transferability of real-time crash prediction models. Accident Analysis & Prevention. 2022 Feb 1;165:106511.
30. Islam MK, Reza I, Gazder U, Akter R, Arifuzzaman M, Rahman MM. Predicting road crash severity using classifier models and crash hotspots. Applied Sciences. 2022 Nov 9;12(22):11354.
31. Santos K, Dias JP, Amado C. A literature review of machine learning algorithms for crash injury severity prediction. Journal of safety research. 2022 Feb 1;80:254-69.
32. Dong C, Shao C, Li J, Xiong Z. An improved deep learning model for traffic crash prediction. Journal of Advanced Transportation. 2018;2018(1):3869106.
33. Morris C, Yang JJ. Effectiveness of resampling methods in coping with imbalanced crash data: Crash type analysis and predictive modeling. Accident Analysis & Prevention. 2021 Sep 1;159:106240.
34. Kolawole Oloke. Next-generation credit scoring systems enabled by explainable AI and cloud-orchestrated data pipelines. Int J Comput Programming Database Manage 2020;1(1):61-71.
DOI: 10.33545/27076636.2020.v1.i1a.133
35. Iranitalab A, Khattak A. Comparison of four statistical and machine learning methods for crash severity prediction. Accident Analysis & Prevention. 2017 Nov 1;108:27-36.
36. Yu R, Abdel-Aty M. Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis & Prevention. 2013 Mar 1;51:252-9.
37. Silva PB, Andrade M, Ferreira S. Machine learning applied to road safety modeling: A systematic literature review. Journal of traffic and transportation engineering (English edition). 2020 Dec 1;7(6):775-90.
38. Wang X, Kim SH. Prediction and factor identification for crash severity: Comparison of discrete choice and tree-based models. Transportation research record. 2019 Sep;2673(9):640-53.
39. Zhang X, Waller ST, Jiang P. An ensemble machine learning-based modeling framework for analysis of traffic crash frequency. Computer-Aided Civil and Infrastructure Engineering. 2020 Mar;35(3):258-76.
40. Akin D, Sisiopiku VP, Alateah AH, Almonbhi AO, Al-Tholaia MM, Al-Sodani KA. Identifying Causes of traffic crashes associated with driver behavior using supervised machine learning methods: Case of highway 15 in Saudi Arabia. Sustainability. 2022 Dec 12;14(24):16654.