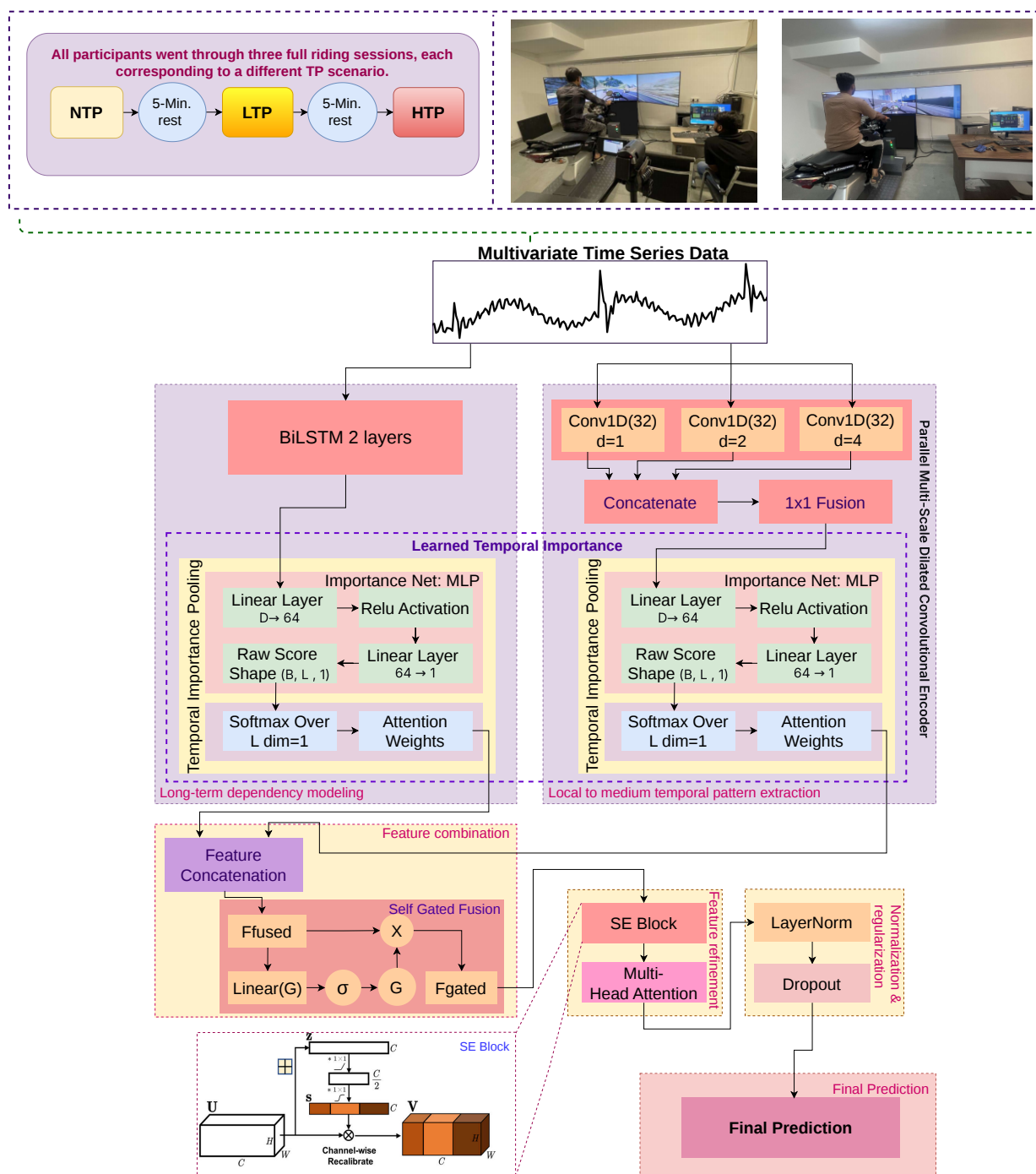


Graphical Abstract

MotoSafety: Edge-AI with Learned Temporal Importance for Two-Wheeler Collision Risk Assessment Under Time Pressure

Sumit S. Shevtekar¹, Chandresh K. Maurya², Gourab Sil³, Subasish Das⁴



MotoSafety: Edge-AI with Learned Temporal Importance for Two-Wheeler Collision Risk Assessment Under Time Pressure

Sumit S. Shevtekar^{a,1,*}, Chandresh K. Maurya^a, Gourab Sil^b, Subasish Das^c

^a*Department of Computer Science and Engineering, Indian Institute of Technology Indore, Khandwa Road, Simrol, Indore, 453552, Madhya Pradesh, India*

^b*Department of Civil Engineering, Indian Institute of Technology Indore, Khandwa Road, Simrol, Indore, 453552, Madhya Pradesh, India*

^c*Civil Engineering Program, Ingram School of Engineering, Texas State University, RFM 5202, San Marcos, 78666, TX, USA*

Abstract

Powered two-wheeler riders face critical safety challenges in low- and middle-income countries, yet limited studies exist on how cognitive stressors such as Time Pressure influence collision risk. We address this gap by introducing a comprehensive dataset consisting of over 129,000 labeled multivariate time-series samples, gathered across 153 simulator rides from 51 participants under No, Low, and High TP scenarios. Across each sequence, we capture 64 distinct attributes covering vehicle motion, rider control actions, spatial proximity, and rule compliance indicators. Using this dataset, we introduce MotoSafety, a new edge-AI framework built on the Learned Temporal Importance (LTI) concept. MotoSafety achieves 94.97% accuracy and 99.33% ROC AUC, outperforming ten baselines, including TimesNet and LLM4TS, and achieves 0.039 MSE and 0.094 MAE for forecasting ($4.4\times$ lower error than Time-LLM and iTransformer). With only 1.15M parameters and 0.135 ms latency, it is suitable for edge deployment on low-cost CPU hardware. Using ground truth TP as an inductive bias improves accuracy from 94.09% to 94.97%, while predicted TP achieves 94.82%. Using only 21 IMU+GPS features, it achieves 93.91% accuracy, indicating practical deployment. Beyond PTW safety, the architecture shows better transferability to human activity (97.66%) and clinical (99.65%) domains. This lightweight framework advances PTW

*Corresponding author. E-mail: sumit.shevtekar@gmail.com

collision risk assessment, supporting the Safe System Approach for Intelligent Transportation Systems.

Keywords: Powered two-wheeler safety, Time pressure, Collision risk assessment, Deep learning, Intelligent transportation systems

1. Introduction

Road traffic crashes are a major global health concern, particularly affecting Low- and Middle-Income Countries (LMICs), where the mortality burden is high [46]. Powered two-wheeler (PTW) riders represent one of the most vulnerable groups, facing disproportionately high fatality rates due to limited protection, complex traffic, and socio-economic pressures to meet mobility and livelihood deadlines. Globally, PTWs account for approximately 21% of road traffic fatalities, a figure that rises to nearly 38% in LMICs, where PTWs serve as an affordable and widely adopted mode of transport [46]. In India alone, PTWs constituted 74.4% of registered vehicles as of 2022 shown in Fig. 1 [29]. As of 2023, two-wheelers constitute the largest segment of the India’s vehicle population with more than 263 million registered units and were involved in 44.5% of total road traffic fatalities, as shown in Fig. 2 [30, 27]. Official MoRTH data from 2019 to 2023 consistently indicates that male riders account for the vast majority of road fatalities in India, with proportions varying between 85.2% and 87.3% across these years [23, 24, 25, 26, 28]. The highest concentration of these fatalities is observed within the 18–45 age demographic. These patterns underscore the elevated vulnerability of male PTW riders in India’s traffic environment. Despite advances in infrastructure design and traffic enforcement, human error remains the dominant contributor to PTW crashes, with risky behaviors such as overspeeding, abrupt maneuvering, and inadequate hazard response accounting for a large share of fatalities [43]. Naturalistic and simulator-based studies consistently show that riding speed, control variability, and situational context strongly influence crash risk [17]. The operating speeds of PTW riders are considerably higher than those of four-wheeler drivers, often reaching up to 2.3 times the speeds observed in cars [35].

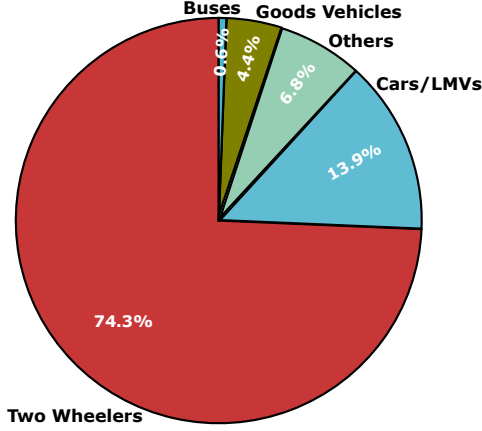


Figure 1: Registered Vehicles Share.

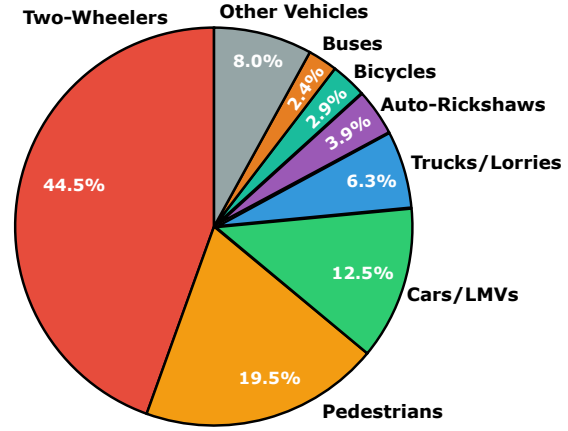


Figure 2: Road Crash Fatalities in India.

Time Pressure (TP) refers to the subjective perception of insufficient time to complete a riding task, commonly arising from risk-prone behaviors, situational urgency, or externally imposed constraints [35, 44]. This is especially common for app-based delivery riders in LMICs such as India. App-generated delivery deadlines, performance incentives, and intense customer demands for fast, flawless service combine to create persistent and often extreme TP [44, 38, 39]. Elevated TP affects cognitive processing and motor control, narrowing attentional bandwidth and accelerating decision-making. This in turn promotes risk-prone behaviors such as overspeeding, inconsistent braking, and delayed hazard response [35, 11, 43]. Simulator studies on car drivers report increased accident risk of up to 181% under high time pressure (HTP) compared to no time pressure (NTP) [35]. Related work [34] demonstrates that cognitive and emotional stress amplifies speed variability and control instability. However, a key cause of these behaviors, TP-induced cognitive stress, is still not well understood within the PTW-driving situation. Recent AI-driven traffic crash prediction leverages large-scale datasets, classical ML/DL architectures to improve safety [31, 33, 6, 1, 15]. Although effective in predicting four-wheeler crashes, these approaches inadequately model the unique dynamics and highly unstable control characteristics of PTWs. For example, [33] achieved $\sim 90\%$ accuracy using deep neural networks on highway data, [6] improved driver behavior classification with feature selection on large datasets, [1] used Random Forest to predict

injury severity (92% accuracy). However, these studies focus primarily on *four-wheelers or general traffic datasets and do not incorporate fine-grained temporal modeling of PTW riders or cognitive stress factors such as TP*.

To our knowledge, this is the first study to focus on predicting collision risk for two-wheeler riders under TP conditions. To address this gap, we introduce the **MotoSafety** specifically designed for multivariate time series data and to quantify and predict collision risk by modeling the interplay between a riders cognitive state under TP and their operational maneuvers. Our work addresses a critical public safety exigency: the protection of vulnerable PTW riders operating under chronic TP—a phenomenon increasingly prevalent in the rapidly urbanizing and dense traffic environments of emerging economies like India. The proposed system can be deployed as an edge-based alert system, where a handlebar-mounted device or smart helmet provides haptic/audio alerts when collision risk exceeds a threshold, enabling proactive safety interventions for riders under TP (e.g., emergency commuting, delivery deadlines, or urgent travel).

This methodology was co-developed with road safety authorities and transportation experts to address PTW fatalities in LMICs. By integrating stakeholder input on crash-prone scenarios and demographic vulnerability, we ensure that our edge-deployable model satisfies real-world operational requirements. Consequently, our approach transitions from theoretical inference to a practitioner-validated solution for tangible public safety impact. The major contributions of this research are summarized as:

1.1. Key Contributions

We make the following key contributions.

1. **Large-Scale PTW Simulator Data Under TP:** To study the effect of TP on PTW riders, we collected 129,000 labeled samples from 51 participants under No/Low/High TP. Each input sequence comprises 64 features categorized into vehicle dynamics, control inputs, headway/tailway distances, and behavioral violations.

2. **Novel Architecture for Collision Risk Forecasting and Classification:** We propose MotoSafety, a novel architecture grounded in the Learned Temporal Importance (LTI) principle. The architecture integrates: (i) parallel multi-scale dilated convolutions for local-to-medium pattern extraction; (ii) Bidirectional Long Short-Term Memory (Bi-LSTM) for long-range dependencies; (iii) Temporal Importance Pooling (TIP) for content-aware temporal collapse, applied independently to both CNN and BiLSTM branches, which reduces downstream fusion, SE-recalibration, and attention stages to $\mathcal{O}(1)$ relative to the sequence length, while the front-end encoder operates at $\mathcal{O}(L)$ complexity, resulting in an overall $\mathcal{O}(L)$ framework (Table 4); and (iv) a self-gated fusion mechanism that leverages squeeze-and-excitation (SE) blocks and multi-head attention (MHA) for adaptive representation refinement.
3. **SOTA Performance and Deployability:** MotoSafety achieves 94.97% accuracy and 99.33% ROC AUC, showing improved performance over ten baselines including TimesNet, Time-LLM and LLM4TS. For long-term forecasting, it achieves 0.039 MSE and 0.094 MAE ($4.4\times$ lower error than Time-LLM and iTransformer). Notably, MotoSafety requires only $\sim 1.15\text{M}$ parameters and achieves an inference latency of 0.135 ms, making it $21.9\times$ and $71.9\times$ faster than TimesNet and LLM4TS, respectively, and makes it suitable for edge devices.
4. **TP Inductive Bias, Real-World Feasibility, and Architecture Transferability:** Explicit TP prediction improves collision accuracy from 94.09% to 94.97% (ground truth) and 94.82% (predicted TP). Using only 21 real-world features (IMU+GPS), MotoSafety achieves 93.91% accuracy, indicating practical deployment potential; hardware validation remains future work. Beyond PTW safety, the architecture demonstrates strong transferability to human activity (97.66%) and clinical exercise (99.65%) domains when retrained from scratch.

2. Related Work

2.1. Time Pressure as a Latent Risk Factor in Driving Safety

Time Pressure (TP) is widely recognized as a cognitive stressor that degrades driver and rider decision-making and elevates crash risk. Under TP, individuals experience a perceived urgency to complete the driving task within insufficient time, which narrows attentional focus, accelerates cognitive processing, and increases reliance on risk-oriented heuristics [35, 11]. Empirical studies consistently show that TP leads to higher speeds, shorter accepted gaps, reduced safety margins, and increased crash likelihood across diverse traffic contexts [36, 43, 44]. Experimental evidence from simulator-based studies indicates that TP significantly alters longitudinal and lateral control behavior, including increased acceleration variability, unstable steering, and abrupt braking [37].

These behavioral deviations are strongly associated with near-crash and collision events, suggesting that TP acts as an upstream cognitive trigger for unsafe driving outcomes rather than merely a contextual modifier. PTWs are particularly susceptible to TP-induced risk due to unstable dynamics and high dependence on rider motor control compared to four-wheelers [36, 43, 9]. At unsignalized intersections, drivers experiencing low time pressure (LTP) and high time pressure (HTP) tend to accept substantially shorter gaps, leading to crash probability increases of 127% under LTP and 181% under HTP compared to no time pressure (NTP) conditions [35]. TP increases cognitive load and unintentional attentional lapses under stress [10, 20], reducing situational awareness and leading to errors that precede collisions. These findings highlight TP as a critical contributor to risk in dense, heterogeneous traffic flows common in LMICs.

2.2. Simulator-Based Collision and Risk Analysis

Driving simulators offer a secure and regulated setting for examining collision risk under cognitively demanding scenarios, including TP. They eliminate the hazards associated with real-world driving while facilitating high-resolution data acquisition for both collision risk

classification and kinematic state prediction [37, 3, 21]. Prior work demonstrates strong relative validity between simulated and real-world behavior, particularly for risk-related metrics including speed selection, braking intensity, and control variability [37, 3, 21]. Although the magnitude of observed behaviors may differ between simulator and naturalistic conditions, the core behavioral responses to TP are preserved: both environments exhibit higher speeds, more frequent lane changes, stronger braking, and greater variability in vehicle control [37]. This supports the ecological validity of using simulators to investigate TP-induced behavioral changes in PTW rider behavior, enabling the current study to train and evaluate MotoSafety’s classification and forecasting tasks under controlled yet realistic conditions.

2.3. Machine Learning for Collision and Risk Prediction

The application of ML and DL methods in intelligent transportation systems (ITS) has grown considerably, particularly for predicting risky events, collisions, and hazardous driving behaviors. Traditional techniques such as Decision Tree, Support Vector Machines, and feed-forward neural networks have shown moderate effectiveness in detecting high-risk maneuvers [5, 42]. In recent years, sequence-based architectures like Temporal Transformers, Informer, and TimesNet have demonstrated considerable effectiveness in capturing long-range temporal patterns in driving behavior data [53, 51]. However, most existing models focus on detecting externally observable outcomes (e.g., collisions or near-crashes) after risk has already manifested. They largely overlook latent cognitive precursors such as TP that shape behavior well before unsafe actions occur. Furthermore, few studies address PTW-specific dynamics, which are characterized by inherent instability and high dependence on rider motor control.

2.4. Research Gap and Motivation

Although TP is a critical contributor to unsafe driving behavior, it is rarely integrated into collision risk assessment frameworks as a latent risk factor. This gap is particularly pronounced for PTWs, where safety margins are narrow and early intervention is essential.

Motivated by these limitations, this work advances collision risk assessment under TP by leveraging high-resolution multivariate time-series data to infer risk directly from behavioral dynamics. By focusing on collision risk in cognitively stressed riding conditions, this study aligns with the Safe System Approach (SSA) [47] and the goal of Vision Zero [18], enabling the detection of high-risk states and supporting the development of intelligent rider-assistance systems and enhanced safety for PTW users.

3. Methodology

In this section, we provide details of our data collection protocol, followed by MotoSafety architecture.

3.1. Experimental Design and Data Collection

3.1.1. Simulator Setup

We carried out all experimental trials on a fixed-base PTW simulator at our institute (Fig. 3). The simulator features a Honda motorcycle frame fitted with working throttle, brake, clutch, and gear controls, designed to replicate the riding experience of a real motorcycle. Riders are surrounded by a three-screen setup that delivers a wide viewing angle and lifelike road imagery. High-precision industrial-grade sensors continuously record rider inputs, vehicle dynamics, and environmental parameters through a real-time data acquisition interface, complying with IEC 393 accuracy standards [13]. While the frame is static, a 4-actuator motion system provides vibration and acceleration. Complete technical specifications, including sensor accuracy ratings, appear in Table 1. This configuration offers a realistic yet controlled setting, allowing us to systematically study how riders behave across various traffic and road conditions.

This ISO-compliant simulator provides a controlled, scalable, and safe environment to investigate rider behavior, cognitive load, and collision risk under realistic operational conditions.

Table 1: Two-wheeler simulator: key technical parameters

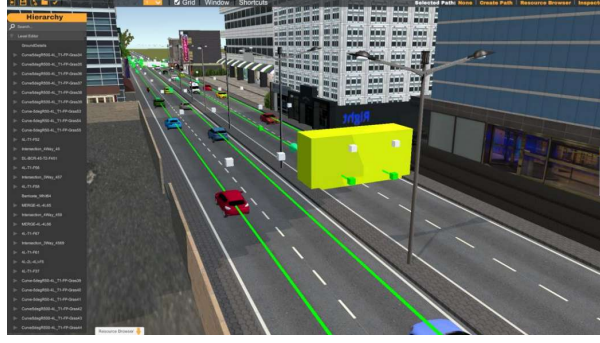
Component	Details
Platform	Fixed-base PTW frame with ISO certification
Controls	Acceleration, clutch, hydraulic brake unit, 5-speed shifting mechanism, handlebar steering, parking brake
Sensors	Wire-wound servo potentiometers (type 50WW): Resistance: 5 k Ω ($\pm 10\%$); Linearity error: $\pm 0.5\%$ (IEC 393 compliant), Angular range: $355^\circ \pm 3^\circ$; Full mechanical rotation: 360° Maximum power: 3 W (at 70°C); Expected operational life: 2 million cycles Safe operating limits: -40°C to $+105^\circ\text{C}$ Force transducers: Brake force sensing Rotary encoders: Gear position detection; SPDT microswitches: Gear shift sensing
Visual System	Three 50-inch LED panels with 180° horizontal coverage
Motion System	Four 3-DOF electric actuators capable of ± 1 G max acceleration, 0–100 Hz response Motion ranges: Pitch $\pm 1.75^\circ$, Roll $\pm 3.5^\circ$, Heave 38.1 mm; Actuator stroke: 1.5 inches Per-actuator load rating: 114 kg
Computer Machine	Rider station: Intel i7, 32 GB RAM, GTX 1650 graphics card Operator station: Intel i5 processor, 16GB RAM, GTX 1650 GPU
Software	TechnoSim (AI-powered traffic, environment, and scenario simulation)
Data Logging	Continuous data recording with sampling frequency $f_s \geq 100$ Hz



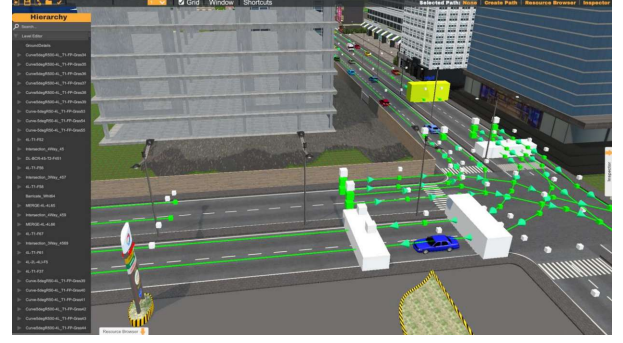
Figure 3: Participants performing the simulator test.

3.1.2. Simulated Scenarios

We design a 4.8 km urban scenario on a high-fidelity two-wheeler simulator to investigate rider behavior under cognitive stress (TP). The route comprises a combination of two-lane undivided and four-lane road segments, with a maximum speed of 50 km/h, representing common road conditions found in Indian cities. The scenario embeds diverse events—including pedestrian crossings, obstacle overtaking, intersections, bike and car following segments to elicit real-time decision-making, adaptive control, and risk-taking under varying TP. The key elements of the scenario are illustrated in Fig. 4: Fig. 4a: primary route with intersections and conflict zones, Fig. 4b: AI vehicle triggers simulating dynamic interactions,



(a) Primary route with intersections and conflict zones



(b) AI vehicle triggers simulating dynamic interactions



(c) Route direction changes and detour prompts



(d) Speed change triggers near sensitive areas

Figure 4: Sample snapshots of the simulated riding environment.

Table 2: Demographic and riding profile of participants (N = 51)

Attribute	Category / Range	Mean (SD) or %
Age (years)	18–42	26.4 (5.7)
Riding experience (years)	≥ 2	5.6 (3.1)
License status	Valid PTW license	100%
Education level	Graduate / B.Tech	58.8% / 41.2%
Simulator exposure	No / Yes	92.2% / 7.8%
Health status	Medically fit	100%

Fig. 4c: route direction changes and detour prompts, and Fig. 4d: speed change triggers near sensitive areas.

3.1.3. Participants

MoRTH fatality data from 2019 to 2023 consistently shows that male riders constitute the overwhelming share of two-wheeler deaths in India, with yearly proportions of 86.0%, 87.3%, 86.4%, 86.2%, and 85.2%, respectively [23, 24, 25, 26, 28]. The 18–45 age group accounts for the largest share of these deaths. This strong demographic pattern supports our choice to include only male riders in this study. We therefore recruited 51 male riders aged 18 to

42 years, all with a minimum of two years of riding experience. Table 2 summarizes their characteristics, while Fig. 3 shows participants during simulator trials.

The male-only cohort is justified by three converging factors: First, national crash data shows male riders as the dominant high-risk group. According to MoRTH reports, male PTW fatalities accounted for 85.2–87.3% of all PTW deaths from 2019 to 2023 [23, 24, 25, 26, 28], with fatalities concentrated in the 18–45 age group. Statistical tests confirm this pattern: a binomial test shows male fatalities (86%) significantly exceed equal representation ($p < 0.0001$), and a chi-square test confirms consistency across years ($\chi^2 = 262.3$, $df = 5$, $p < 0.0001$).

Second, female exposure to manually geared PTWs is extremely limited. Women hold only about 6% of motor vehicle licenses in India [12, 45, 2], significantly below equal representation ($p < 0.0001$). The number of women riding manually geared motorcycles is even smaller, largely due to socio-cultural factors and traditional norms and practical barriers [45]. Consequently, the pool of female riders with manual transmission experience suitable for simulator-based research is extremely small.

Third, the simulator’s manual transmission requirement creates a practical recruitment constraint. The high-fidelity PTW simulator replicates a manually geared motorcycle with functional clutch and gear controls, requiring prior manual transmission experience. Given the extremely low prevalence of female riders with such experience in India, recruiting a statistically sufficient female sample for robust analysis is infeasible. Thus, male participants were selected to ensure a sufficiently large and statistically robust sample for meaningful spatial risk analysis, consistent with the dominant high-risk demographic. Future studies will include female riders as their PTW participation increases.

3.1.4. Time Pressure (TP) Scenarios

To ethically evoke graduated cognitive load mimicking emergency commuting, participants were tasked with reaching an examination hall under three conditions: (i) NTP: Baseline with ample time; (ii) LTP: Limited to 90% of the baseline time (with a reminder

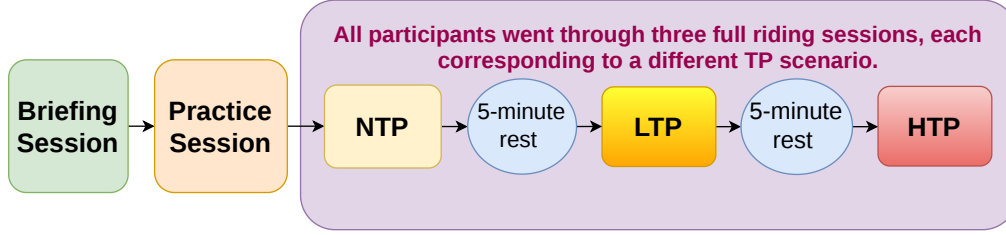


Figure 5: Flow of dataset collection and experimental design.

to avoid being late); and (iii) HTP: Restricted to 80% of the baseline time, accompanied by active urgency cues (e.g., exam hall closing shortly). This standardized exam-deadline scenario simulates high-arousal stress states common in urban transit, providing a controlled environment to study behavioral degradation.

3.1.5. Experimental Design and Protocol

A structured four-phase experimental protocol (Fig. 5) was implemented to ensure data consistency: (i) Briefing: Standardized orientation and informed consent; (ii) Practice: A 5–10 minute familiarization ride (excluded from analysis); (iii) Main Task: Participants completed three riding sessions under NTP, LTP, and HTP conditions, with the session sequence randomized across participants to avoid order effects; and (iv) Rest: 5–8 minute inter-session breaks to stabilize cognitive load and mitigate fatigue.

Ethics and Informed Consent: All participants signed a written consent form before taking part in the research. The research protocol received ethical clearance from the IE Committee at the IIT Indore (Approval No. BSBE/IITI/IHEC-11/2025/11).

Consent to Publish: All participants gave their written consent for the publication of anonymized data and findings derived from this study. To safeguard participant privacy, all data have been fully anonymized.

3.1.6. Behavioral Transition : Why LTP Matters

Adding the LTP condition allows us to capture the transitional phase from safe riding (NTP) to risky riding (HTP), representing the early stages of behavioral decline. The progression $NTP \rightarrow LTP \rightarrow HTP$ illustrates the transition from stable, safe riding to high-stress

operation characterized by speeding and infractions. LTP serves as a critical indicator where subtle warning signs—such as minor speeding or slight lane drifts—first manifest. By incorporating LTP, the model improves its sensitivity to rising stress levels, supporting earlier detection and enabling more timely safety interventions. Moreover, LTP data inform policy-makers and trainers by defining practical stress thresholds, facilitating preventive strategies before riders reach the high-risk HTP state.

3.2. Feature Engineering and Preprocessing

From raw time-series sensor data, we derived 64 domain-informed features that capture vehicle dynamics, rider behavior, and contextual factors as summarized in Table 3. The features include things like average and variation in speed and acceleration, how often the rider brakes, how frequently they change lanes, and the TP level. Min-Max normalization was used to bring all feature values within the $[0, 1]$ range. We segment the time series data into fixed-length overlapping windows to preserve temporal dependencies, enabling the model to capture both gradual risk accumulation and abrupt behavioral deviations. Detailed training configurations are provided in Section 4.1.

3.3. Dataset and Feature Set

The experimental protocol yielded 153 multivariate time-series sessions (51 participants $\times$ 3 conditions). Data were sampled at 100 Hz. The 100 Hz streams were segmented into 960 ms windows (96 time steps) with 50% overlap, yielding 129,209 windows slid continuously. A window is labeled as collision (1) if a collision occurs within it, otherwise 0. The dataset exhibits a class distribution of 21% collision and 79% non-collision events. A single collision spans multiple windows (impact+sliding+aftermath, due to overlap), reflecting the inherent imbalance of safety-critical scenarios. This is window-level, not event-level labeling.

Each input sequence comprises 64 features (summarized in Table 3) categorized into: (i) Vehicle Dynamics: Speed, acceleration, and 3D rotation; (ii) Control Inputs: Throttle, hydraulic brake force, and gear transitions; (iii) Proximity: Headway/tailway distances and

Table 3: Overview of simulator-derived features and their categories

Feature Category	Included Parameters	Count
Vehicle Controls	Ignition status, Engine state, Throttle position, Brake application, Clutch engagement, Handbrake, Steering angle, Gear position, Headlight usage, Horn activation	10
Vehicle Performance	Speed, Engine RPM, Fuel consumption, Distance covered	4
Lighting & Indicators	Indicator signal, Pre-move indication, Junction turning signal, Lane-change signal, Headlight non-compliance	5
Behavioral Violations	Speeding, Incorrect junction/intersection speed, Speed violations on speed breakers, Gap maintenance errors, Hazardous overtaking, Unsigned turns, Lane misuse, Driving on the wrong side, Handbrake use while moving, Riding the clutch, Faulty gear shift order, Improper clutch release, Shifting without clutch engagement, Proper gear selection before start, Gradual clutch engagement	15
Traffic Infractions	White line violations, Yellow line crossings, Stop line breaches, Signal jumping, Entering no-entry zones, Illegal U-turns, Parking infringements	7
Temporal Context	Time stamp, Time Pressure condition (0=HTP, 1=LTP, 2=NTP)	2
Spatial Attributes	3D position coordinates, 3D rotation angles, Lane number, Left/Right lane offsets	9
Motion and Proximity	Lateral/Longitudinal velocity, Headway (distance & time), Tailway (distance & time), Left-/Rightway distance, Steering angle	9
Brake System	Brake test completion, Front and Rear braking force	3

Note: All features associated with collisions (vehicles, objects, obstacles, etc.) were merged into a single binary target variable ‘Target’, where 0 = no collision and 1 = collision. These collision events were excluded from the input feature set to prevent label leakage.

lane offsets; (iv) Time context and scenario: Time Stamp, TP (0=HTP, 1=LTP, 2=NTP); and (v) Behavioral Violations: Overspeeding, improper gap maintenance and traffic rule infractions. All collisions were aggregated into a single binary target variable, with leakage prevented by excluding impact-related features from the input.

3.4. Scope of the Classification and Forecasting Tasks

We define two complementary but distinct tasks, and are explicit about what each does and does not establish.

Classification Task (Same-Window Risk Detection): Given a 960 ms observation window, the model predicts whether a collision event occurs within that window. Because a subset of positive windows overlaps the collision event itself (impact, sliding, or aftermath phases, as noted in Section 3.3), this task is best characterized as *same-window collision risk detection* rather than a fixed-lead-time forecast: it evaluates whether pre-collision and in-progress behavioral signatures are jointly separable from normal riding, not how far in advance a collision can be anticipated.

Forecasting Task (Prospective Kinematic Prediction): Given a lookback window of $L = 96$ timesteps, a separately trained instance of the architecture predicts future

kinematic states (e.g., speed, lean angle, longitudinal force) up to $H \in \{96, 192, 336, 720\}$ timesteps (up to 7.2s) ahead. This task, evaluated in Section 5.2, is a genuine prospective-forecasting result with an explicit lead time, but its target is future sensor state, not future collision probability; it demonstrates that the architecture can anticipate the kinematic precursors of hazardous states, which is complementary to, but distinct from, a direct lead-time collision-probability forecast.

We view a labeled, fixed-lead-time collision-probability forecast (i.e., excluding all windows that overlap a collision event and labeling remaining windows by whether a collision begins within a defined horizon τ afterward) as an important direction for follow-up work, and report the current classification results with this scope explicitly stated rather than implied.

3.5. MotoSafety: Proposed Architecture

3.5.1. Architecture Overview

The proposed MotoSafety architecture, shown in Fig. 6, is a DL architecture based on the Learned Temporal Importance (LTI) principle, developed for the prediction of PTW collisions and forecasting. It employs a multi-stage pipeline: (i) a parallel multi-scale dilated convolutional encoder with three branches (dilation factors 1, 2, 4) extracts local to medium-range temporal patterns; (ii) a 2-layer Bi-LSTM (128 hidden units per direction) captures long-range sequential dependencies; (iii) a novel TIP module (which uses a lightweight Multi-Layer Perceptron (MLP) as a learned scoring function to evaluate the importance of each timestep) is applied independently to the outputs of both the CNN and LSTM paths, learning to adaptively weight the most critical pre-collision timesteps for each feature type and collapsing them into compact vectors; (iv) these vectors are concatenated and processed by a self-gated fusion stage, refined by channel-wise recalibration via a SE block and MHA (4 heads). This representation undergoes normalization, dropout-based regularization, and linear projection to output binary collision risk scores.

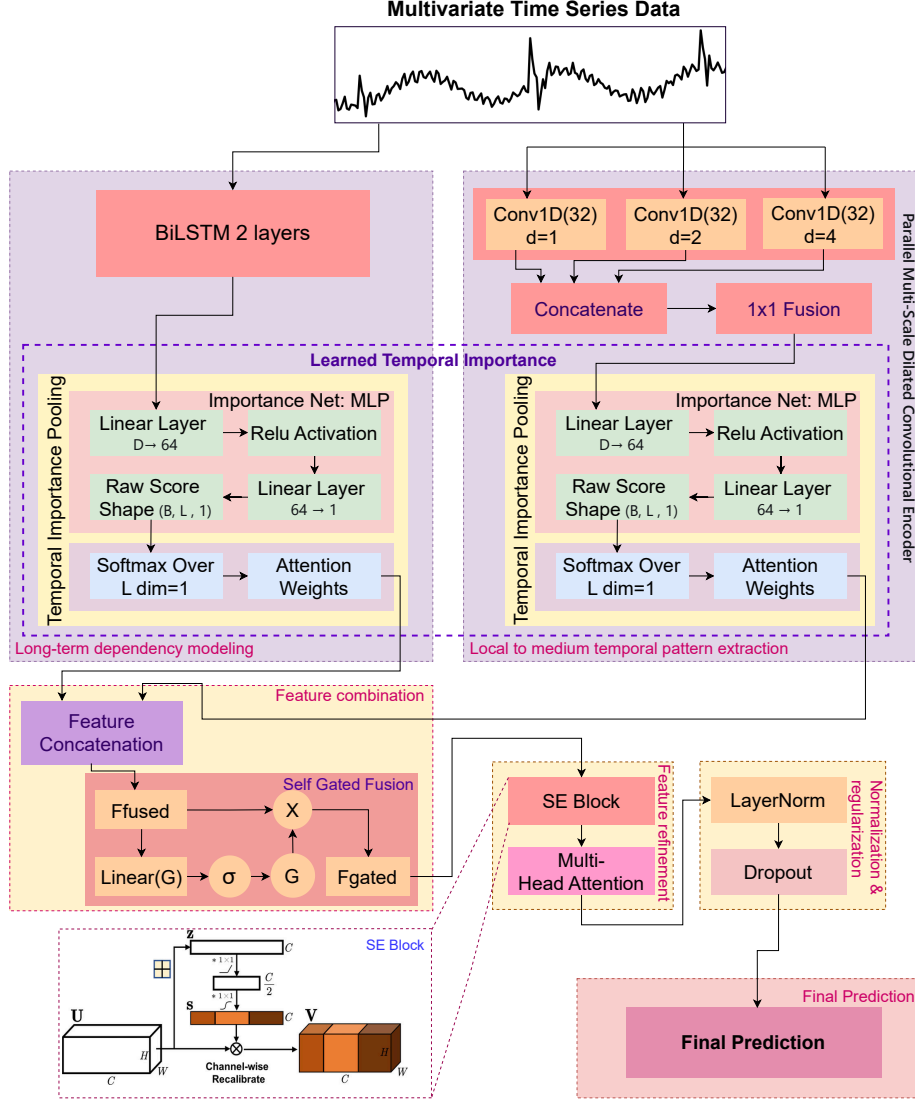


Figure 6: The proposed MotoSafety architecture.

3.5.2. Comment on the Novelty

We distinguish clearly between architectural components adapted from prior work and the components that constitute this paper’s novel contribution.

Adapted components: The parallel dilated-convolution encoder, Bi-LSTM, SE recalibration, and MHA are established building blocks, each individually well-studied in time-series and sequence modeling literature.

Novel contribution: TIP. The core contribution of this work is the TIP module and the LTI principle it instantiates. Rather than collapsing the temporal dimension with a

fixed operator (mean-pool, max-pool, or final-hidden-state extraction, as is standard practice in CNN/RNN-based time-series classifiers), TIP learns a content-aware scoring function that assigns a data-dependent importance weight to every timestep before collapsing the sequence. This is applied independently to both the CNN and BiLSTM representations, allowing the model to learn which pre-collision moments matter most for each feature pathway, rather than treating all timesteps as equally informative. Unlike full self-attention pooling (which computes pairwise timestep interactions at $\mathcal{O}(L^2)$ cost), TIP scores each timestep independently against a shared learned criterion, at $\mathcal{O}(L)$ cost, and produces a fixed-size vector consumed by the downstream fusion stage. Applied independently to both CNN and BiLSTM branches, TIP reduces the downstream fusion, SE-recalibration, and attention stages to $\mathcal{O}(1)$ relative to the sequence length, while the front-end encoder operates at $\mathcal{O}(L)$ complexity, resulting in an overall $\mathcal{O}(L)$ framework (Table 4).

Empirical validation of TIP. The effectiveness of TIP is validated in Section 5.8 (Table 14), where it outperforms standard fixed-pooling strategies (mean-pooling, max-pooling, last-timestep) under an identical architecture, indicating that the performance gain stems from learned, content-aware weighting rather than fixed pooling.

This strategic “content-aware collapse” means the fusion, SE-recalibration, and MHA stages operate on a fixed-size 320-dimensional vector regardless of sequence length, i.e., these downstream stages cost $\mathcal{O}(1)$ per sample once pooling is complete. The upstream CNN and Bi-LSTM encoder remains $\mathcal{O}(L)$, so the model’s end-to-end complexity is $\mathcal{O}(L)$ overall, not $\mathcal{O}(1)$; the practical benefit of TIP is that it removes the $\mathcal{O}(L^2)$ cost a full-sequence self-attention mechanism would otherwise add on top of the encoder. Table 4 details the per-module accounting. This design ensures the low-latency performance required for robust real-time deployment on edge devices.

3.5.3. *MotoSafety: Mathematical Formulation*

We address the problems of binary collision prediction and multi-horizon time-series forecasting from multivariate sensor data under TP. Our input consists of sensor readings

Table 4: Per-module computational complexity. L : sequence length, C : feature channels, H : hidden dimension, k : kernel size.

Module	Complexity
Dilated CNN encoder ($\times 3$ branches)	$\mathcal{O}(L \cdot k \cdot C)$
Bi-LSTM (2 layers)	$\mathcal{O}(L \cdot H^2)$
TIP scoring + weighted sum	$\mathcal{O}(L \cdot H)$
Gated fusion + SE block	$\mathcal{O}(1)$ (fixed-size input)
Multi-head attention (post-pooling)	$\mathcal{O}(1)$ (fixed-size input)
MotoSafety total	$\mathcal{O}(L)$

$X \in \mathbb{R}^{B \times L \times C}$ from a two-wheeler simulator, with B batch size, lookback window L and C feature channels.

Forward Process: The input X is processed by two parallel branches: CNN and BiLSTM. The CNN branch captures local patterns using three dilated convolutional branches with kernel size $k = 3$. To accommodate the channel-first requirement of 1D convolutions, we apply a permutation function $\pi(\cdot)$ to X such that $\hat{X} = \pi(X) \in \mathbb{R}^{B \times C \times L}$:

$$H_d = \text{ReLU}(\text{Conv1D}_d(\hat{X})) \quad (1)$$

where $d = \{1, 2, 4\}$ is the dilation factor. These views are merged into a unified 64-channel representation via a 1×1 convolution: $H_{\text{ms}} = \text{Conv1D}_{1 \times 1}([H_1; H_2; H_4]) \in \mathbb{R}^{B \times 64 \times L}$. In parallel, a bidirectional LSTM captures long-range dynamics:

$$H_{\text{lstm}} = \text{BiLSTM}(X) \in \mathbb{R}^{B \times L \times 256}. \quad (2)$$

To identify safety-critical moments, we utilize TIP to learn importance weights across the temporal dimension. Two distinct networks, MLP_{θ_1} and MLP_{θ_2} , assign scores to the CNN and LSTM features respectively:

$$w_c = \text{Softmax}(\text{MLP}_{\theta_1}(H_{\text{ms}}^T)), \quad f_{\text{cnn}} = \sum_{t=1}^L w_c^{(t)} H_{\text{ms}}^{(t)} \quad (3)$$

$$w_l = \text{Softmax}(\text{MLP}_{\theta_2}(H_{\text{lstm}})), \quad f_{\text{lstm}} = \sum_{t=1}^L w_l^{(t)} H_{\text{lstm}}^{(t)} \quad (4)$$

The resulting vectors, $f_{\text{cnn}} \in \mathbb{R}^{64}$ and $f_{\text{lstn}} \in \mathbb{R}^{256}$, are concatenated to form the joint representation $f = [f_{\text{cnn}}, f_{\text{lstn}}] \in \mathbb{R}^{320}$. The fused representation undergoes refinement through a gating mechanism and a channel-wise recalibration block (SE-Block). Finally, a MHA block performs high-order feature refinement:

$$f_{\text{gate}} = f \odot \sigma(W_g f + b_g), \quad (5)$$

$$f_{\text{se}} = \text{SEBlock}(f_{\text{gate}}), \quad (6)$$

$$f_{\text{attn}} = \text{MHA}(f_{\text{se}}). \quad (7)$$

The final output $\mathbf{y} \in \mathbb{R}^2$ is produced by applying a linear layer to the normalized and regularized context representation:

$$\mathbf{y} = \text{Softmax}\left(\text{Linear}\left(\text{LayerNorm}(\text{Dropout}(f_{\text{attn}}))\right)\right). \quad (8)$$

To address the extreme class imbalance (rare collision events), we utilize Focal Loss with a focusing parameter $\gamma = 2$:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{B} \sum_{i=1}^B (1 - P_{i,y_i})^\gamma \log(P_{i,y_i}), \quad (9)$$

where P_{i,y_i} denotes the model’s estimated probability for the actual binary collision outcome $y_i \in \{0, 1\}$.

Multi-Horizon Forecasting Module. Our forecasting module predicts future target values $\hat{Y} \in \mathbb{R}^{B \times \tau}$ over horizon τ . It uses a *separate but architecturally similar* network with identical CNN and BiLSTM branches, TIP, gating, SE-Block, and MHA. Both the classification and forecasting modules use the same $L = 96$ timestep lookback window; they differ only in their prediction target (same-window collision label vs. future kinematic states) and are trained separately, as described above. This produces a refined context vector $\mathbf{f}_{\text{attn}} \in \mathbb{R}^{B \times 320}$. The

final forecast is obtained via:

$$\hat{Y} = \text{Linear}_{\text{forecast}}(\text{LayerNorm}(\mathbf{f}_{\text{attn}})). \quad (10)$$

The module is trained to minimize the Mean Squared Error (MSE) loss:

$$\mathcal{L}_{\text{forecast}} = \frac{1}{B \cdot \tau} \sum_{i=1}^B \sum_{k=1}^{\tau} \left(y_{i,L+k} - \hat{Y}_{i,k} \right)^2, \quad (11)$$

where $y_{i,L+k}$ is the true target value at future step k . *Both models share identical architectural components but are trained separately for their respective tasks.*

4. Experimental Setup and Baselines

4.0.1. Ground Truth and Human Validation

To ensure the reliability of the simulator-generated labels, a systematic validation protocol was conducted. Two road experts in road safety and human factors (male, post-graduate) reviewed a representative subset of 2,000 randomly selected segments from the total of 129,000-sequence dataset. Experts independently annotated the onset and severity of high-risk episodes (e.g. abrupt swerves, loss of balance, critical near-misses) using a standardized coding scheme based on kinematic thresholds and behavioral cues. We used Cohen’s Kappa (κ) to check how well the two annotators agreed with each other, which yielded a value of 0.87. This indicates near-perfect agreement based on established interpretation standards [19]. This high agreement validates the dataset’s labels as a robust ground truth for model training.

4.0.2. Baselines

We benchmark MotoSafety against ML (RF, CNN, RNN) and modern DL architectures: Informer [53] and iTransformer [22] for Transformer-based long-range modeling. TimesNet [48] for multi-scale frequency analysis; PatchTST [32] for patching-based local semantics.

Table 5: Training configuration for MotoSafety classification and forecasting models.

Component	Setting
Classification Model	
Input sequence length	96 timesteps (960 ms)
Feature dimension	64
CNN hidden channels	32 per branch, 64 fused
BiLSTM hidden dimensions	128 (bidirectional)
Total model parameters	~1.15 million
Dropout rate	0.3
Loss function	Focal Loss ($\gamma = 2$)
Forecasting Model	
Input lookback length	96 timesteps
Prediction horizons	$H \in \{96, 192, 336, 720\}$
Architecture	Identical to classification model
Loss function	MSE
Shared Settings	
Optimizer	AdamW (initial learning rate: 3×10^{-4} , L2 weight decay: 5×10^{-5})
Batch size	64
Training epochs	50
Regularization	EMA, Mixup, MC Dropout (5 samples)
Hardware	NVIDIA T400 GPU (4GB VRAM)

Time-LLM [16] and LLM4TS (GPT-2) [7] as a representative fine-tuned LLM baseline.

4.1. Training and Evaluation

4.1.1. Training Protocol

The MotoSafety model is implemented in PyTorch and optimized using AdamW (initial learning rate: 3×10^{-4} , L2 weight decay: 5×10^{-5}) with Focal Loss ($\gamma = 2$) for classification and MSE loss for forecasting. To enhance generalization, we employed Mixup augmentation, Exponential Moving Average (EMA) of weights, and Monte Carlo (MC) Dropout (5 samples) for uncertainty calibration. For model training, we utilized an NVIDIA T400 GPU (4 GB VRAM) and set the batch size to 64, running the process for 50 epochs. The main training settings are listed in Table 5. For all experiments, we used a participant-wise data split of 80% for training, 10% for validation, and 10% for testing, where all sessions from the same rider are kept together in the same split to prevent data leakage.

4.1.2. Evaluation Protocol

We evaluated classification performance using Accuracy, F1-Score, and ROC-AUC, while for the forecasting task, we used MSE and Mean Absolute Error (MAE) as evaluation metrics.

Table 6: Behavioral differences across TP: NTP/LTP/HTP (Mean $\pm$ Std).

Feature	NTP	LTP	HTP
Speed (km/h)	33.12 $\pm$ 19.42	39.69 $\pm$ 21.67	49.00 $\pm$ 26.48
Sudden braking	1.32 $\pm$ 2.54	1.38 $\pm$ 2.36	1.80 $\pm$ 2.01
Dangerous overtaking	1.07 $\pm$ 1.18	1.21 $\pm$ 1.27	1.36 $\pm$ 1.52
Headway distance (m)	8.99 $\pm$ 14.38	8.20 $\pm$ 14.34	7.75 $\pm$ 13.88

Table 7: Multivariate long-term forecasting results (mean $\pm$ std over 5 runs). Input lookback length $L = 96$. Forecast horizons $H \in \{96, 192, 336, 720\}$. Avg is averaged over all four prediction horizons.

Models Metric	MotoSafety		Time-LLM		iTransformer		PatchTST	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.033 $\pm$ 0.005	0.082 $\pm$ 0.007	0.136	0.148	0.161	0.237	0.188	0.208
192	0.037 $\pm$ 0.006	0.086 $\pm$ 0.008	0.145	0.153	0.170	0.246	0.205	0.223
336	0.041 $\pm$ 0.006	0.097 $\pm$ 0.010	0.192	0.215	0.177	0.258	0.227	0.272
720	0.045 $\pm$ 0.007	0.111 $\pm$ 0.012	0.212	0.293	0.183	0.262	0.267	0.338
Avg	0.039	0.094	0.171	0.210	0.173	0.251	0.222	0.260

Statistical significance was assessed via paired Wilcoxon signed-rank tests over 5 runs, with the Holm–Bonferroni correction ($\alpha = 0.05$).

5. Results

5.1. Impact of Time Pressure on Riding Behavior

Analysis of our PTW dataset reveals that HTP is the primary driver of behavioral volatility, with a clear transition to high-risk strategies as TP increases, as shown in Table 6. Riders under HTP exhibited 48% higher mean speed, 36% increase in sudden braking, and 14% shorter headway distance compared to NTP, along with a 27% increase in dangerous overtaking. These statistically supported trends define a hurry-up strategy of excessive speed, abrupt control, and reduced spacing, highlighting TP as a key behavioral stressor that increases the risk of PTW collisions.

5.2. Multivariate Long-Term Forecasting Performance

To evaluate the proactive safety capabilities of MotoSafety, we conducted multivariate forecasting across four different time horizons ($H \in \{96, 192, 336, 720\}$) on our PTW data. This experiment assesses the models ability to project future kinematic states such as lean angles and longitudinal forces, essential for anticipating hazardous transitions before they

Table 8: Comparison of model performance (mean $\pm$ standard deviation over 5 runs). We used the paired Wilcoxon signed-rank test with Holm–Bonferroni adjustment to determine statistical significance ($\alpha = 0.05$). Significance is denoted by: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. All baseline models were compared against the proposed MotoSafety framework.

Model	Acc. (%)	Prec. Result	Rec. Result	F1-Score (%)	AUC (%)
	Mean $\pm$ Std			Mean $\pm$ Std	Mean $\pm$ Std
RF	84.93 $\pm$ 0.16***	85.24	84.87	85.02 $\pm$ 0.23***	93.68 $\pm$ 0.31***
CNN	90.19 $\pm$ 0.11***	93.82	80.68	86.74 $\pm$ 0.21***	97.62 $\pm$ 0.22***
RNN	90.94 $\pm$ 0.12***	83.41	96.38	89.43 $\pm$ 0.32***	97.27 $\pm$ 0.34***
LLM4TS	90.06 $\pm$ 0.14***	90.58	83.23	86.81 $\pm$ 0.31***	97.31 $\pm$ 0.42***
Informer	93.61 $\pm$ 0.07***	89.72	94.88	92.21 $\pm$ 0.22**	99.12 $\pm$ 0.21**
TST	93.86 $\pm$ 0.09***	93.67	90.74	92.18 $\pm$ 0.29**	99.23 $\pm$ 0.28*
TimesNet	94.06 $\pm$ 0.05***	89.23	96.69	92.77 $\pm$ 0.38**	99.18 $\pm$ 0.24*
iTransformer	94.39 $\pm$ 0.06**	92.41	93.57	93.12 $\pm$ 0.42*	99.22 $\pm$ 0.32*
PatchTST	94.41 $\pm$ 0.05**	90.73	96.18	93.28 $\pm$ 0.24*	99.31 $\pm$ 0.22
Time-LLM	94.46 $\pm$ 0.07*	93.12	93.79	93.41 $\pm$ 0.33*	99.24 $\pm$ 0.31*
MotoSafety	94.97 $\pm$ 0.08	93.80	93.90	93.70 $\pm$ 0.20	99.33 $\pm$ 0.30

culminate in a collision. As shown in Table 7, MotoSafety attains a mean MSE of 0.039 and an MAE of 0.094, representing a $4.4\times$ reduction in error over the Time-LLM (0.171) and iTransformer (0.173) baselines. This high forecasting fidelity under TP is critical for proactive safety; by accurately projecting states up to $H = 720$, the model provides a vital safety buffer for early risk mitigation before a maneuver reaches a non-recoverable limit.

5.3. Comparative Classification Performance

The model performance results on the PTW simulation dataset under TP are presented in Table 8. The proposed MotoSafety achieves 94.97% accuracy, 93.7% F1-score and 99.33% ROC AUC, outperforming ten baselines including TimesNet, PatchTST, iTransformer, Time-LLM and LLM4TS. To assess the robustness of the performance gains, statistical significance is evaluated using paired Wilcoxon signed-rank tests over 5 independent runs (Table 8). For the nine pairwise comparisons against baselines, we applied the Holm–Bonferroni method ($\alpha = 0.05$) to regulate the family-wise error rate (FWER). The high ROC AUC indicates that MotoSafety can maintain a low false-alarm rate, which is a prerequisite for real-world Advanced Rider Assistance Systems (ARAS).

Table 9: Forecasting performance by prediction horizon (mean over 5 runs)

Horizon H	Time (s)	MSE	MAE
96	0.96	0.033 ± 0.005	0.082 ± 0.007
192	1.92	0.037 ± 0.006	0.086 ± 0.008
336	3.36	0.041 ± 0.006	0.097 ± 0.010
720	7.20	0.045 ± 0.007	0.111 ± 0.012

Table 10: Performance comparison of MotoSafety showing improvement from adding TP features. Significance markers indicate improvement over the Without TP baseline: *** $p < 0.001$

Model	Accuracy (%)	AUC (%)
Without TP Feature	94.09 ± 0.04	99.10 ± 0.01
With Predicted TP Feature	$94.82 \pm 0.02^{***}$	$99.24 \pm 0.01^{***}$
With Ground Truth TP Feature	$94.97 \pm 0.01^{***}$	$99.33 \pm 0.01^{***}$

5.4. Forecasting Horizon Analysis

To characterize how forecasting accuracy varies with prediction horizon, Table 9 reports MSE and MAE for the kinematic-state forecasting task (Section 3.5.3) across horizons up to 7.2s. Error increases gradually with horizon (MSE from 0.033 at $H = 96$ to 0.045 at $H = 720$), indicating the model retains meaningful predictive signal even at longer horizons rather than degrading sharply. We note this result characterizes future kinematic-state prediction and is reported separately from the same-window collision classification task described in Section 3.3; the two tasks use different targets and should not be read as a single fused collision-lead-time result.

5.5. Downstream Task: Impact of TP on Model Performance

In practical deployment, direct and accurate measurement of ground truth TP labels, denoted TP_{gt} , may be infeasible due to sensor limitations or real-time constraints. In the proposed data set, the TP characteristic is a categorical variable representing three distinct TP conditions: $HTP = 0$, $LTP = 1$, $NTP = 2$. To address missing or unavailable GT TP labels, we employed a Time Series Transformer (TST) [52] model to predict these categorical TP labels from the multivariate time series input data. The TST achieved 89.26% accuracy for 3-class TP prediction (0=HTP, 1=LTP, 2=NTP). Table 11 shows the confusion matrix (raw counts) per-class performance.

Table 11: Confusion matrix (raw counts) for TST-based TP prediction.

True / Predicted	0: HTP	1: LTP	2: NTP
0: HTP	8878	229	163
1: LTP	76	7326	568
2: NTP	82	1661	6859

Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T) \in \mathbb{R}^{T \times D}$ represent the multivariate time series input from the PTW simulator, which includes all features except the explicit TP indicator. In this notation, T refers to the sequence length, while $D = 63$ specifies the number of telemetry features. Each $\mathbf{x}_t \in \mathbb{R}^D$ captures the sensor states at the t -th time instant. The TST model is then expressed as:

$$TP_{pred} = f_{TST}(\mathbf{X}), \quad (12)$$

where $TP_{pred} \in \{0, 1, 2\}$ is the predicted TP label. The predicted TP feature TP_{pred} , along with all other relevant sensor and behavioral features of the data set, forms the complete input to our collision risk prediction model, MotoSafety. Formally, the model can be expressed as:

$$\hat{y} = g_{MotoSafety}(\mathbf{X}, TP), \quad (13)$$

where TP is the ground truth TP_{gt} or predicted TP_{pred} , and $\hat{y}$ is the predicted safety risk label for the rider. We evaluated the robustness of model by comparing the predictions when using:

$$\hat{y}_{gt} = g_{MotoSafety}(\mathbf{X}, TP_{gt}), \quad (14)$$

$$\hat{y}_{pred} = g_{MotoSafety}(\mathbf{X}, TP_{pred}). \quad (15)$$

The results presented in Table 10 highlight the critical role of TP in predicting collision risk. While the baseline model $f_{MotoSafety}(X)$ achieves an accuracy 94.09% using only PTW simulation data without TP feature. The integration of predicted TP features $f_{MotoSafety}(X \oplus TP_{pred})$ improved this to 94.82%, nearly matching the 94.97% Oracle per-

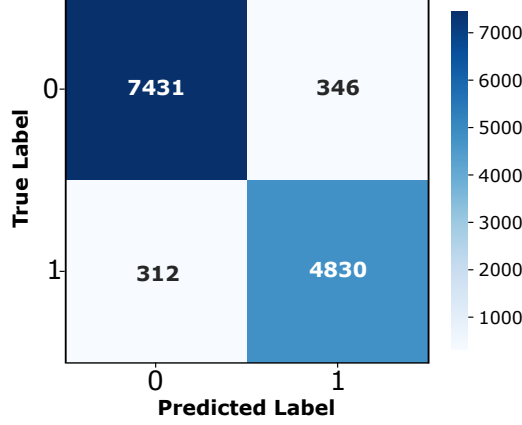


Figure 7: Confusion matrix of MotoSafety predictions on the main PTW simulator dataset under TP.

formance $f_{\text{Motosafety}}(X \oplus TP_{gt})$. This performance gain shows that TP_{pred} captures latent psychological context, such as rider stress, which is not fully represented by telemetry alone. The marginal 0.15% gap between the predicted (94.82%) and Oracle (94.97%) results validates MotoSafety for real-world deployment where ground truth states are not available.

5.6. Model Performance and Reliability

5.6.1. Confusion Matrix Analysis

The confusion matrix provides a detailed summary of MotoSafety model classification outcomes for collision and non-collision cases. Figure 7 shows the distribution of accurate and inaccurate predictions.

5.6.2. Calibration Curve and Risk Scores

The MST calibration curve on the primary PTW simulator dataset under TP (Fig. 8) shows that predicted collision probabilities closely match observed event frequencies, indicating statistically reliable and well-calibrated outputs. In safety-critical contexts, such calibration enables risk-aware adaptive interventions rather than reliance on fixed thresholds. The distribution of predicted risk scores by true outcome (Fig. 9) further illustrates clear separation: collision events consistently receive higher predicted probabilities, while non-collision samples remain concentrated at lower values. This separation reinforces both discriminative ability and calibration quality, supporting the potential for real-world deployment.

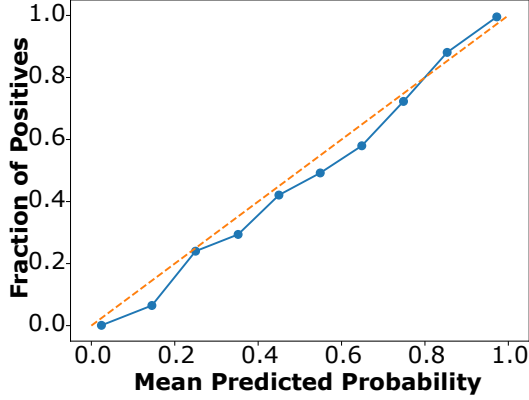


Figure 8: Calibration curve.

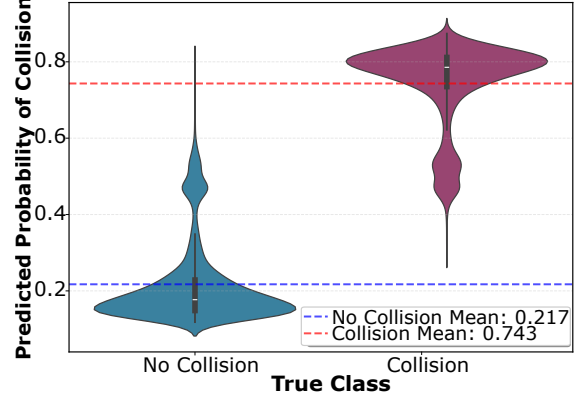


Figure 9: Distribution of predicted collision risk.

Table 12: Cross-Safety-Domain Evaluation of MotoSafety Accuracy (%) and Prior Work Accuracy (%). MST: MotoSafety.

Dataset	Details	Safety main	Do- Application	CountryMST	Prior Work Accuracy
Proposed Dataset	PTW Simulator with GT TP	PTW Safety	Collision Prediction	India	94.97 –
[41]	PTW High-Fidelity Simulator	PTW Safety	Collision Prediction	Germany	99.49 91.0 [RF, GB [42]]
[4]	PTW Real Time Fall Event Data	PTW Safety	Fall Detection	France	96.10 91.59 [DT [8]]
[40]	Human Activity Recognition	Human Safety	Activity Recognition	Italy	97.66 83.35 [K-means + NB [14]]
[50]	Wearable-Sensor Movement	Human Clinical Safety	Exercise Quality Assessment	Turkey	99.65 88.65[MTMM-DTW [49]]

5.7. Architecture Transferability Across Safety Domains

Note that each dataset below is used to train and evaluate a separate MotoSafety instance from scratch (i.e., no weights are transferred from the PTW model); this experiment therefore evaluates the transferability of the architecture to other safety-critical sequence classification tasks, rather than knowledge transfer from the PTW domain itself.

To evaluate the versatility of the MotoSafety framework beyond motorcycle simulators, we performed cross-dataset evaluations across distinct safety-critical domains as shown in Table 12. In the PTW Safety domain, MotoSafety achieved 99.49% accuracy on the high-fidelity German simulator dataset [41] for collision prediction, outperforming traditional ensemble methods (RF/GB) by 8.49%. Similarly, on real-world fall event data [4], achieved improved 96.10% accuracy, a significant improvement over the 91.59% reported for Decision Tree (DT) benchmarks [8]. Beyond vehicle dynamics, the model exhibited exceptional

Table 13: Ablation Study of MotoSafety Components.

Variant	Accuracy (%)	Δ Acc.
Full Model	94.97	–
No BiLSTM	90.42	–4.55
No SE Block	91.24	–3.73
No Attention	90.78	–4.19
No CNN Branch	92.66	–2.31
No Gated Fusion	92.90	–2.07

robustness in Human and Clinical Safety. On the UCI HAR human activity recognition dataset [40], MotoSafety surpassed Naive Bayes (83.35%) by 14.31%, achieving 97.66% accuracy. Most notably, in clinical exercise quality assessment [50], the framework achieved near-perfect classification (99.65%), marking an 11% gain over DTW-based clinical algorithms (88.65%) [49]. These results demonstrate that the MotoSafety architecture is not overfit to PTW-simulator-specific artifacts and performs competitively when retrained on diverse safety-critical sequence classification tasks, supporting the architecture’s applicability beyond its original design domain.

5.8. Ablation Study

We evaluated individual contributions of MotoSafety components through a comprehensive ablation study shown in Table 13. The study was carried out on PTW simulation data under TP. The complete model attains an accuracy of 94.97%. The removal of BiLSTM and MHA resulted in significant performance degradation of 4.55% and 4.19%, respectively, highlighting their role in capturing long-range dependencies. Similarly, the exclusion of the CNN branch and Gated Fusion mechanism led to accuracy drops of 2.31% and 2.07%, underscoring the necessity of multi-scale feature extraction. Furthermore, removing the SE block decreases accuracy by 3.73%, validating its effectiveness in channel-wise feature recalibration. Collectively, these findings justify the necessity of each integrated module to maintain high predictive precision.

To further isolate the effect of TIP, we compare it against three standard temporal-pooling strategies—mean-pooling, max-pooling, and last-timestep extraction—under an otherwise

Table 14: Comparison of MotoSafety with TIP against standard fixed temporal-pooling strategies (mean $\pm$ std over 5 runs).

Pooling Strategy	Accuracy (%)	AUC (%)
Mean-pooling	92.23 $\pm$ 0.15	98.80 $\pm$ 0.22
Max-pooling	93.41 $\pm$ 0.12	99.11 $\pm$ 0.18
Last-timestep	92.87 $\pm$ 0.14	98.97 $\pm$ 0.20
TIP	94.97 $\pm$ 0.08	99.33 $\pm$ 0.30

Table 15: MotoSafety accuracy with real-bike features (IMU+GPS).

Features	Accuracy (%)
3	80.22
5	83.11
7	88.78
9	90.16
21	93.91
All 64 (simulator)	94.97

identical architecture. As shown in Table 14, MotoSafety with TIP achieves 94.97% accuracy, outperforming mean-pooling by 2.74%, max-pooling by 1.56%, and last-timestep by 2.10%. This indicates that the performance gain stems from the learned, content-aware weighting rather than from the surrounding architecture alone.

5.9. Real-World Feature Feasibility

To assess real-world deployment feasibility, we identified that 21 of the 64 features used in this study are available on standard motorcycles via low-cost IMU+GPS (e.g., speed, acceleration, yaw rate, position, brake force, throttle, steering angle). The remaining features (e.g., precise gap metrics, rule violations) are simulator-only. As shown in Table 15, accuracy increases with more features, achieving 93.91% with 21 features (compared to 94.97% with all 64 features), indicating that real-world deployment is feasible with onboard sensors. Testing on actual hardware in real traffic conditions remains essential and is therefore deferred to future work, as discussed in Section 6.1.

5.10. Model Complexity and Edge-AI Deployability

The MotoSafety model is designed for real-time deployment in resource-constrained environments. Latency was measured on an Intel Core i7 CPU (2.8 GHz) with batch size

Table 16: MotoSafety Efficiency Metrics and Comparative Latency.

A. MotoSafety Efficiency & Deployability				
Metric		Value	Significance	
Total Parameters		1,149,048	Low Complexity	
Model Size		4.38 MB	Edge-ready Storage	
Throughput		> 7,400 s/sec	High-speed Inference	
Arch. Complexity		$\mathcal{O}(L)$	Linear Scalability	
Edge-AI Deployability	Yes		Suitable for wearable	
Deployment regions	Re-Global		Operate in low-resource	
B. Latency Comparison (PTW Simulator Dataset Under TP)				
Model (Year)		Latency (ms)	Speed Gap	Parameters
MotoSafety		0.135	$1.0\times$	1,149,048
iTransformer (2024)		0.587	$4.3\times \downarrow$	275,714
PatchTST (2023)		0.720	$5.3\times \downarrow$	187,330
Time-LLM (2024)		1.280	$9.5\times \downarrow$	3,183,618
TimesNet (2023)		2.960	$21.9\times \downarrow$	4,708,226
LLM4TS (2025)		9.710	$71.9\times \downarrow$	124,506,626

32, PyTorch 2.0, CPU-only inference, averaged over 1000 batches. With only 1.15M parameters and a 4.38 MB memory footprint (Table 16A), it maintains a low computational profile without sacrificing accuracy. Benchmarking on our PTW simulation dataset reveals that MotoSafety achieves a per-sample inference latency of 0.135 ms, outperforming all baseline and SOTA architectures (Table 16B). Specifically, MotoSafety is **$5.3\times$, $9.5\times$, $21.9\times$, and $71.9\times$ faster** than PatchTST, Time-LLM, TimesNet, and LLM4TS (GPT-2), respectively. Even when accounting for TST preprocessing (36.88ms overhead) to predict TP labels when ground truth is unavailable, MotoSafety (37.01ms) remains faster than iTransformer (37.47ms), PatchTST (37.60ms), Time-LLM (38.16ms), TimesNet (39.84ms), and LLM4TS (46.59ms). This ultra-low latency ensures near-instantaneous risk assessment, allowing maximum time for emergency interventions. These efficiency metrics satisfy the stringent requirements for real-time deployment on low-cost on-board units (OBUs). Unlike GPU-dependent Transformer baselines, complexity of MotoSafety is $\mathcal{O}(L)$. This balance of efficiency and performance makes it a viable candidate for large-scale ITS deployment in developing regions, providing an accessible AI-driven safety net for vulnerable riders.

6. Conclusion

In this work, we investigated how TP affects PTW riders and contributes to collision risk. To fill this gap, we collected a large-scale dataset comprising over 129,000 labeled multivariate time-series sequences from 153 simulator rides conducted by 51 participants under no, low, and high TP conditions. Each sequence includes 64 features covering vehicle dynamics, control inputs, proximity measures, temporal context, and behavioral violations.

Building on this dataset, we proposed MotoSafety, a novel deep learning architecture grounded in the LTI principle. MotoSafety achieves 94.97% accuracy and 99.33% ROC AUC for collision risk assessment, outperforming ten baselines including TimesNet, PatchTST, iTransformer, Time-LLM, and LLM4TS. For long-term forecasting, it achieves 0.039 MSE and 0.094 MAE ($4.4\times$ lower error than Time-LLM and iTransformer). With only 1.15 million parameters and 0.135 ms inference latency, the model is $21.9\times$ and $71.9\times$ faster than TimesNet and LLM4TS.

Our findings show that explicit TP prediction provides a critical inductive bias, improving collision accuracy from 94.09% to 94.97% (ground truth) and 94.82% with predicted TP. Furthermore, using only 21 real-world features (available via low-cost IMU+GPS), MotoSafety achieves 93.91% accuracy, indicating practical deployment potential; real-world hardware validation remains future work. Beyond PTW safety, the MotoSafety architecture shows improved transferability when retrained on human activity recognition (97.66%) and clinical exercise monitoring (99.65%) datasets.

The model’s small size (1.15M parameters, 4.38 MB) and low latency (0.135 ms) make it well-suited for future edge deployment on handlebar-mounted devices or smart helmets as a potential edge-deployable black box for two-wheelers, enabling real-time collision risk alerts without relying on cloud connectivity.

6.1. Limitations and Future Work

We acknowledge several limitations in the present study. First, our participant pool consisted solely of male riders, which aligns with Indian fatality statistics showing that males account for 85.2–87.3% of PTW deaths. However, this limits the generalizability of our findings to female riders, since perceptions of TP and riding behavior may vary between genders. Future studies should include female riders to systematically examine whether the observed relationships hold across genders. Second, the data were gathered in a controlled simulator setting, which may not completely capture real-world factors like rider fatigue, changing weather, or social pressures. Real-world TP collision data cannot be collected safely, so simulators provide the only ethical controlled setting. To validate in real conditions, we plan phased testing: closed-course trials followed by naturalistic data collection through commercial riding platforms. Third, we did not record physiological stress markers, such as heart rate variability or electrodermal activity. Behavioral markers such as speed variability and control inputs were used as indicators of cognitive load; however, the absence of physiological validation limits direct assessment of the underlying stress response. Future work should incorporate physiological measures to complement behavioral markers. Fourth, the proposed safety system requires prospective validation in real-world operational settings before practical deployment. Hardware validation on handlebar-mounted devices or smart helmets remains future work.

Future research will assess MotoSafety across varied rider populations and road conditions using both simulator-based and naturalistic datasets. We plan to incorporate multimodal signals—including physiological, behavioral, and contextual data—to improve robustness. Adaptive ITS interventions, such as haptic feedback, throttle modulation, and context-sensitive alerts, will be explored to mitigate crash risk and manage kinetic energy transfer. Transfer learning and domain adaptation techniques will also be applied to ensure the framework is scalable across different vehicle types, regions, and cultural contexts.

Data and Code Availability

The data and code supporting this study will be made available by the corresponding author upon reasonable request after publication.

Funding

This research received funding from the IIT Indore Young Faculty Research Catalyzing Grant (YFRCG) Scheme under Project ID: IITI/YFRCG/2023-24/01.

Acknowledgments

The authors gratefully acknowledge all the volunteers who took part in this study. We also extend our sincere thanks to Manvendra T. for his valuable assistance with data collection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

References

- [1] Acı, Ç., Mutlu, G., Ozen, M., Acı, M., 2025. Enhanced multi-class driver injury severity prediction using a hybrid deep learning and random forest approach. *Applied Sciences* 15.
- [2] Atif, M., Sil, G., 2026. Modeling the effects of driver and road geometric characteristics on consecutive horizontal curve perception. *Transportation Research Record* 2680, 286–310.
- [3] Bham, G.H., Leu, M.C., 2018. A driving simulator study to analyze the effects of portable changeable message signs on mean speeds of drivers. *Journal of Transportation Safety & Security* 10, 45–71.

- [4] Boubezoul, A., Dufour, F., Bouaziz, S., Larnaudie, B., Espié, S., 2019. Dataset on powered two wheelers fall and critical events detection. *Data in Brief* 23, 103828.
- [5] Bouhsissin, S., Sael, N., Benabbou, F., Soultana, A., 2023. Enhancing machine learning algorithm performance through feature selection for driver behavior classification. *International Journal of Electrical and Computer Engineering Systems* 35, 354–365.
- [6] Bouhsissin, S., Sael, N., Benabbou, F., Soultana, A., 2024. Enhancing machine learning algorithm performance through feature selection for driver behavior classification. *Indonesian Journal of Electrical Engineering and Computer Science* 35, 354–365.
- [7] Chang, C., Wang, W.Y., Peng, W.C., Chen, T.F., 2025. Llm4ts: Aligning pre-trained llms as data-efficient time-series forecasters. *ACM Trans. Intell. Syst. Technol.* 16.
- [8] Elwy, F., Aburukba, R., Al-Ali, A.R., Nabulsi, A.A., Tarek, A., Ayub, A., Elsayeh, M., 2023. Data-driven safe deliveries: The synergy of iot and machine learning in shared mobility. *Future Internet* 15.
- [9] Gashaw, S., Goatin, P., Härrä, J., 2018. Modeling and analysis of mixed flow of cars and powered two wheelers. *Transportation Research Part C: Emerging Technologies* 89, 148–167.
- [10] Gupta, M., Pawar, N.M., Velaga, N.R., Mishra, S., 2022. Modeling distraction tendency of motorized two-wheeler drivers in time pressure situations. *Safety Science* 154, 105820.
- [11] Gupta, M., Velaga, N.R., 2024. Motorized two-wheeler riders’ rear brake application in sudden hazardous event of animal crossing. *Transportation Letters* 16, 1268–1275.
- [12] Gupta, M., Velaga, N.R., 2026. Dynamic dilemma zone at signalized intersection: attention allocation patterns using cure survival analysis for male riders. *Accident Analysis & Prevention* 228, 108408.

- [13] International Electrotechnical Commission, 2023. Potentiometers for use in electronic equipment. Standard IEC 60393. International Electrotechnical Commission.
- [14] Ismi, D.P., Panchoo, S., Murinto, M., 2016. K-means clustering based filter feature selection on high dimensional data. *International Journal of Advances in Intelligent Informatics* 2, 38–45.
- [15] Jiang, Y., Qu, X., Zhang, W., Guo, W., Xu, J., Yu, W., Chen, Y., 2025. Analyzing crash severity: Human injury severity prediction method based on transformer model. *Vehicles* 7, 5.
- [16] Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.Y., Liang, Y., Li, Y.F., Pan, S., Wen, Q., 2024. Time-llm: Time series forecasting by reprogramming large language models.
- [17] Kong, X., Das, S., Jha, K., Zhang, Y., 2020. Understanding speeding behavior from naturalistic driving data: Applying classification based association rule mining. *Accident Analysis & Prevention* 144, 105620.
- [18] Krug, E., 2022. It’s time to end deaths on our roads.
- [19] Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- [20] Leung, S., Croft, R.J., Jackson, M.L., Howard, M.E., Mckenzie, R.J., 2012. A comparison of the effect of mobile phone use and alcohol consumption on driving simulation performance. *Traffic Injury Prevention* 13, 566–574.
- [21] Li, X., Oviedo-Trespalacios, O., Rakotonirainy, A., Yan, X., 2019. Collision risk management of cognitively distracted drivers in a car-following situation. *Transportation Research Part F: Traffic Psychology and Behaviour* 60, 288–298.

- [22] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M., 2024. itransformer: Inverted transformers are effective for time series forecasting.
- [23] MORTH, 2019. Road Accidents in India – 2019. Technical Report. Transport Research.
- [24] MoRTH, 2020. Road Accidents in India – 2020. Technical Report. Transport Research.
- [25] MoRTH, 2021. Road Accidents in India – 2021. Technical Report. Transport Research.
- [26] MORTH, 2022. Road Accidents in India – 2022. Technical Report. Transport Research.
- [27] MORTH, 2023. Ministry of road transport & highway, 2022-23. Book 2022–2023 .
- [28] MoRTH, 2023. Road Accidents in India – 2023. Technical Report. Transport Research.
- [29] MORTH, 2024. Ministry of road transport & highway, 2023-24. Book 2023–2024 .
- [30] MORTH, 2025. Ministry of road transport & highway, 2024-25. Book 2024–2025 .
- [31] Mostafa, A.M., Aldughayfiq, B., Tarek, M., Alaerjan, A.S., Allahem, H., Elbashir, M.K., Ezz, M., Hamouda, E., 2025. Ai-based prediction of traffic crash severity for improving road safety and transportation efficiency. Scientific Reports 15, 27468.
- [32] Nie, Y., Nguyen, N.H., Sinthong, P., Kalagnanam, J., 2023. A time series is worth 64 words: Long-term forecasting with transformers.
- [33] Pan, G., Wang, G., Wei, H., Chen, Q., Zhang, A., 2024. Development of an automated global crash prediction model with adaptive feature selection of deep neural networks. IEEE Transactions on Industrial Informatics 20, 12010–12020.
- [34] Pavlidis, I., Dcosta, M., Taamneh, S., Manser, M., Ferris, T., Wunderlich, R., Akleman, E., Tsiamyrtzis, P., 2016. Dissecting driver behaviors under cognitive, emotional, sensorimotor, and mixed stressors. Scientific Reports 6, 25651.

- [35] Pawar, N.M., Velaga, N.R., 2022. Analyzing the impact of time pressure on drivers' safety by assessing gap-acceptance behavior at un-signalized intersections. *Safety Science* 147, 105582.
- [36] Pawar, N.M., Velaga, N.R., Mishra, S., 2022a. Impact of time pressure on acceleration behavior and crossing decision at the onset of yellow signal. *Transportation Research Part F: Traffic Psychology and Behaviour* 87, 1–18.
- [37] Pawar, N.M., Velaga, N.R., Sharmila, R., 2022b. Exploring behavioral validity of driving simulator under time pressure driving conditions of professional drivers. *Transportation Research Part F: Traffic Psychology and Behaviour* 89, 29–52.
- [38] Reporter, T.H.S., 2022. Rash driving, fall in income, more pressure: Behind the scenes of 10-minute delivery. *The Hindu* .
- [39] Reuters, 2022. Delivery race brings road safety risks in india. *The Jakarta Post* .
- [40] Reyes-Ortiz, J., Anguita, D., Ghio, A., Oneto, L., Parra, X., 2013. Human Activity Recognition Using Smartphones. *UCI Machine Learning Repository*. DOI: <https://doi.org/10.24432/C54S4K>.
- [41] Rodegast, M., et al., 2024a. Motorcycle collision dataset.
- [42] Rodegast, P., Maier, S., Kneifl, J., Fehr, J., 2024b. On using machine learning algorithms for motorcycle collision detection. *Discover Applied Sciences* 6, 326.
- [43] Sharma, I., Gupta, M., Mishra, S., Velaga, N.R., 2025. Exploring the impact of time pressure on motorized two-wheeler riders' over-speeding behavior. *Transportation Letters* 17, 595–611.
- [44] Staff, F.I., 2025. Racing against time: Quick commerce is pushing delivery riders to the edge, claims study. *Fortune India* .

- [45] Transportation Research and Injury Prevention Centre, 2023. Road Safety in India. Technical Report. Indian Institute of Technology Delhi. New Delhi, India.
- [46] WHO, 2023. Global status report on road safety, 2023. World Health Organization. .
- [47] World Health Organization, 2023. Road traffic injuries.
- [48] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M., 2023. Timesnet: Temporal 2d-variation modeling for general time series analysis.
- [49] Yurtman, A., Barshan, B., 2014a. Automated evaluation of physical therapy exercises using multi-template dynamic time warping on wearable sensor signals. *Computer Methods and Programs in Biomedicine* 117, 189–207.
- [50] Yurtman, A., Barshan, B., 2014b. Physical Therapy Exercises.
- [51] Zeng, A., Chen, M., Zhang, L., Xu, Q., 2023. Are transformers effective for time series forecasting?, in: *Proceedings of the Thirty-Seventh AAAI Conference on AI and Thirty-Fifth Conference on Innovative Applications of AI and Thirteenth Symposium, AAAI Press*.
- [52] Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., Eickhoff, C., 2020. A transformer-based framework for multivariate time series representation learning.
- [53] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W., 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 11106–11115.