

Physics-Informed Multimodal Learning for Wheel-Force Prediction and Grip Utilization Optimization Using Image Detection and Chassis Sensors in Electric Vehicles

Timileyin Akande

`akande.timileyin@tech-u.edu.ng`

Abiola Ajimobi Technical University Ibadan

Research Article

Keywords: Electric vehicles, physics-informed neural networks, multimodal sensor fusion, tire-road friction estimation, wheel-force prediction, grip-utilization ratio, road-surface classification, vehicle dynamics

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11107864/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Physics-Informed Multimodal Learning for Wheel-Force Prediction and Grip Utilization Optimization Using Image Detection and Chassis Sensors in Electric Vehicles

Timileyin Opeyemi AKANDE^{1*}  0000-0003-4307-1880

^{1*}Department of Mechanical and Mechatronics Engineering, Abiola Ajimobi Technical University (First Technical University), Ibadan, Oyo State, 200255, Nigeria

*Corresponding author's email: akande.timileyin@tech-u.edu.ng

Abstract

Accurate, real-time knowledge of the forces generated at each tire contact patch is central to electronic stability control, traction control, and learning-based driver-assistance functions in electric vehicles (EVs). Conventional model-based observers assume a single friction coefficient and degrade under the transient, spatially varying friction conditions typical of real roads, while purely data-driven estimators can output forces that violate the friction-ellipse constraint, undermining safety-critical decisions. This paper proposes a physics-informed multimodal learning framework that fuses forward-facing camera imagery with in-vehicle chassis-sensor streams to predict per-wheel longitudinal and lateral tire forces and quantify a grip-utilization ratio for early loss-of-traction warning. A vision branch, calibrated to the reported accuracy of convolutional road-surface classifiers trained on the public Road Surface Classification Dataset (RSCD), supplies a road-friction class posterior; a chassis branch summarizes a 0.8 s sliding window of inertial, wheel-speed, steering and powertrain signals via a time-delay neural network encoding; the two modalities are combined through a lightweight cross-modal feature-wise linear modulation (FiLM) fusion gate. Training combines a data-fidelity loss with physics-informed soft constraints derived from a quasi-static load-transfer model and the Pacejka Magic-Formula friction ellipse. Using a hybrid dataset of 9,000 samples from a validated four-wheel vehicle-dynamics simulator spanning five friction classes, from dry asphalt to ice, and five manoeuvres, the proposed model reduces wheel-force RMSE by 22-42 percent relative to chassis-only and naive multimodal baselines on wet, snow-covered and icy surfaces, lowers the deployable friction-ellipse violation rate to 13.5 percent, and improves accuracy under distribution shift to unseen low-friction conditions relative to a non-physics-informed ablation. The complete model contains only 4,200 parameters and executes in under one millisecond per inference, supporting real-time deployment on embedded control units. These results indicate that lightweight cross-modal fusion combined with physics-informed regularization is a practical route to safer, more generalizable grip-utilization estimation in production EVs.

Keywords: *Electric vehicles; physics-informed neural networks; multimodal sensor fusion; tire-road friction estimation; wheel-force prediction; grip-utilization ratio; road-surface classification; vehicle dynamics.*

1. Introduction

1.1. Background and Motivation

Global electric vehicle (EV) sales continue to expand rapidly, with battery-electric and plug-in hybrid models accounting for an increasing share of new passenger-car registrations worldwide [1]. The instantaneous torque and independent per-axle (or per-wheel) actuation available in EV powertrains create an opportunity for far more responsive traction and stability control than is possible in conventional internal-combustion vehicles, provided that the controller has an accurate, low-latency estimate of the force each tire is currently producing

and of how much additional force remains available before that tire loses grip. This latter quantity, commonly expressed as a grip-utilization or friction-ellipse-utilization ratio, is the single most safety-relevant piece of information an electronic stability control (ESC), traction control (TCS), or torque-vectoring system consumes, because it determines how close the vehicle is to a loss-of-control event before that event becomes kinematically visible in the vehicle's yaw rate or lateral acceleration [2].

Directly measuring tire force requires instrumented wheel-force transducers that are costly, fragile, and essentially absent from production vehicles. Practical estimation therefore falls into two broad families. Model-based observers (extended/unscented/cubature Kalman filters and related recursive estimators) propagate a vehicle-dynamics model forward in time and correct it against chassis-sensor measurements, but they generally assume a single, slowly varying road-friction coefficient and can lag or diverge when friction changes abruptly, for example at a wet patch or a snow-covered lane boundary. Vision-based methods classify the road surface ahead of the vehicle from camera imagery and thereby provide a preview of the friction condition before the tire reaches it, but they say nothing about the transient dynamic state of the vehicle itself. Multimodal sensor fusion, combining exteroceptive perception with proprioceptive chassis measurement, has therefore emerged as a promising direction for autonomous-driving perception generally [3], and is a natural candidate for wheel-force and grip-utilization estimation specifically.

1.2. Problem Statement

Three specific, unresolved problems motivate this work. First, purely data-driven force estimators, whether chassis-only or naively multimodal, are trained to minimize a data-fidelity loss with no notion of the friction ellipse that governs real tire behaviour; nothing in a standard mean-squared-error objective prevents such a model from predicting a resultant force that exceeds the physically available friction an output that could mislead a downstream safety controller. Second, real-world driving datasets that pair synchronized camera imagery, chassis-sensor logs, and reliable per-wheel force ground truth essentially do not exist at scale, because ground truth requires instrumented wheel-force transducers that are rarely fitted outside of specialized test vehicles; this scarcity is most acute for exactly the low-friction, safety-critical surfaces (fresh snow, ice) that matter most and that are, for obvious safety and logistical reasons, the least represented in whatever instrumented data does exist. Third, and consequently, it is largely unknown whether physics-informed regularization—which has shown benefit in other vehicle-dynamics learning problems [4]–[7]—meaningfully improves an estimator's ability to generalize to exactly those underrepresented low-friction conditions, as opposed to merely trading a small amount of in-distribution accuracy for a notionally more 'physical' output.

1.3. Research Gap and Novelty

Prior work on physics-informed neural networks (PINNs) for vehicle dynamics has largely targeted whole-vehicle state prediction or suspension/damper load estimation from chassis signals alone [6], [7]; prior work on vision-based friction and road-surface estimation has largely targeted the road-friction coefficient itself, or has

fused imagery with vehicle dynamics for that same scalar coefficient, rather than with per-wheel tire-force vectors [8]–[11]. To the best of the authors' knowledge, no published study combines (i) a cross-modal image-and-chassis fusion architecture, (ii) a physics-informed loss anchored to both a quasi-static load-transfer model and the Pacejka friction ellipse, and (iii) an explicit, quantitative test of whether the resulting physics-informed inductive bias improves generalization to friction regimes that are deliberately withheld from training, as a proxy for the real-world data-scarcity problem described above. This combination, and the accompanying reproducible data-generation methodology built on a validated four-wheel vehicle-dynamics simulator coupled to a public road-surface image dataset, constitutes the novelty of the present study.

1.4. Contributions

The specific contributions of this paper are as follows.

- A physics-informed, cross-modal fusion architecture that combines a road-surface-classification posterior (vision branch) with a time-delay neural network encoding of a sliding window of chassis-sensor signals (chassis branch) through a lightweight feature-wise linear modulation (FiLM) gate, predicting per-wheel longitudinal and lateral tire forces and a grip-utilization ratio in under one millisecond per inference on commodity CPU hardware.
- A composite physics-informed training objective that penalizes friction-ellipse violations and anchors predictions to an analytical Magic-Formula prior, together with a diagnostic distinction between a 'deployable' violation rate (computed against the vision branch's estimated friction) and an 'oracle' violation rate (computed against the true simulated friction), which separates the force estimator's own physical consistency from upstream perception uncertainty.
- A reproducible hybrid data-generation methodology that pairs a validated four-wheel, Pacejka-tire-model vehicle-dynamics simulator with class-conditional confusion statistics calibrated to the reported accuracy of a convolutional classifier trained on the public Road Surface Classification Dataset (RSCD) [9], yielding 9,000 labelled multimodal samples across five friction classes and five driving manoeuvres without requiring an instrumented wheel-force test vehicle.
- A controlled extrapolation experiment, training exclusively on dry-asphalt, wet-asphalt and wet-gravel data and testing on withheld fresh-snow and ice data, that isolates the generalization benefit of physics-informed regularization from its in-distribution accuracy trade-off, together with a full ablation over the physics-loss weight and over the individual vision and physics-loss components.

1.5. Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews related work on physics-informed machine learning, vision- and chassis-based tire-friction estimation, and multimodal sensor fusion. Section 3 details the vehicle-dynamics simulator, the hybrid dataset construction, the proposed network architecture, and the physics-informed loss function. Section 4 reports quantitative results, including baseline comparisons, ablation studies, and the low-friction extrapolation experiment. Section 5 discusses the practical implications, compares the findings with the literature, and states the limitations of the study. Section 6 concludes the paper.

2. Related Work

2.1. Physics-Informed Neural Networks in Vehicle Dynamics

Physics-informed neural networks (PINNs) embed governing physical equations into the training objective of a neural network, either as a soft penalty on the residual of a differential equation or as a hard architectural constraint, so that the learned function remains consistent with known physics even where training data are sparse or noisy [4]. Karniadakis et al. [5] survey the broader physics-informed machine learning paradigm and identify vehicle and structural dynamics as domains in which physical priors are particularly valuable because instrumented ground truth is expensive to collect across the full operating envelope. Within vehicle dynamics specifically, Lim et al. [6] proposed EV-PINN, which embeds a reduced longitudinal EV dynamics model into a network trained to predict vehicle speed and acceleration from powertrain signals, reporting improved extrapolation relative to a purely data-driven baseline. Wang et al. [7] proposed DBPnet, a Bayesian physics-informed network that uses a damper-characteristic prior to estimate individual wheel loads from chassis accelerations, reporting both improved accuracy and calibrated uncertainty relative to non-physics-informed baselines. These studies motivate physics-informed regularization for vehicle-state estimation in general, but neither predicts the full per-wheel longitudinal and lateral tire-force vector, nor incorporates a visual road-friction cue, nor explicitly tests generalization to a deliberately withheld friction regime, which is the specific gap this paper targets.

2.2. Vision-Based and Chassis-Based Tire-Friction Estimation

A separate body of work estimates road friction or tire force from chassis signals alone using recursive Bayesian filters. Zhang et al. [12] combine a refined tire model with an improved square-root cubature Kalman filter (SCKF) for road-friction estimation; Ye et al. [13] adapt tire stiffness online within a robust cubature Kalman filter; Tao et al. [14] combine two robust proportional-multi-integral observers with a multilayer perceptron for individual-wheel friction estimation; and Wang et al. [15] propose a model-based learning framework that blends a physical tire model with a learned correction term, reporting improved robustness across friction conditions relative to either component alone. These observer-based approaches are computationally efficient and grounded in vehicle dynamics, but by construction they cannot anticipate a friction change before it manifests in the chassis response, which limits their usefulness for preview-based control. Downstream of estimation, such friction and force signals are typically consumed by active safety planning and control layers, for example the integrated post-impact planning and active safety control framework of Wang et al. [16], underscoring that estimation accuracy and latency requirements are ultimately set by the control application rather than being intrinsic to the estimator itself.

Vision-based approaches address this preview limitation directly. Zhao and Wei [9] released the Road Surface Classification Dataset (RSCD), a large, publicly available, densely annotated collection of road-surface images labelled by material and wetness/friction level, which has become a standard benchmark for camera-based friction-relevant road classification. Building on such data, Wang et al. [8] proposed RoadFormer, a local-global feature-fusion transformer that reports 92.5% and 96.5% top-1 accuracy on the full and simplified RSCD friction-class benchmarks respectively; the present study calibrates the simulated vision branch's class-

conditional confusion statistics directly to this reported accuracy range (Section 3.2.2). Beyond pure image classification, several recent studies fuse imagery with vehicle dynamics or weather data for friction-coefficient estimation: Huang et al. [10] fuse road images and weather metadata for friction-coefficient estimation, and Zhao et al. [11] adaptively fuse road imagery with vehicle dynamics signals for the same target. Quan et al. [17] and Han et al. [18] instead pursue direct or contact-mechanics-based sensing, using piezoelectric intelligent-tire sensors and three-dimensional tire–road contact modelling respectively; these approaches are informative about the diversity of viable sensing modalities but rely on non-standard or specialized hardware that is not yet common in production EVs. None of these vision-fusion studies predicts the full per-wheel force vector or an explicit grip-utilization ratio, and none embeds a friction-ellipse physical constraint directly into the training loss.

2.3. Multimodal Sensor Fusion Architectures

Multimodal sensor fusion is well established in the broader autonomous-driving perception literature, where camera, LiDAR, radar and inertial streams are combined for tasks such as object detection, semantic segmentation, and trajectory prediction; Qian et al. [3] provide a recent survey of the architectural patterns used for this fusion, ranging from early concatenation to cross-attention and transformer-based fusion. Cross-attention fusion is expressive but typically carries a parameter and latency cost that is difficult to justify for an embedded, safety-critical estimation loop running at chassis-control update rates. Feature-wise linear modulation (FiLM) [19] offers a substantially lighter-weight alternative in which one modality (here, the vision branch) regresses an affine transformation that modulates the feature map of the other modality (here, the chassis branch), achieving a comparable cross-modal conditioning effect at a fraction of the computational cost of full cross-attention. On the chassis-encoding side, time-delay neural networks (TDNNs) [20] summarize a sliding window of a time series through learned weights applied across the window, a long-established and computationally efficient alternative to recurrent encoders for fixed-horizon temporal features.

2.4. Summary and Positioning of the Present Study

In summary, the literature offers, on one hand, physics-informed learning methods for vehicle-state and wheel-load estimation from chassis signals [6], [7], and, on the other, vision- and fusion-based methods for road-friction estimation [8]–[11], [17], [18], together with generic architectural building blocks for lightweight multimodal fusion [19], [20]. This paper combines these threads: it uses a FiLM-based cross-modal fusion of a vision-derived friction posterior and a TDNN-encoded chassis window to predict the full per-wheel tire-force vector, trains the resulting network with a physics-informed loss anchored to a quasi-static load-transfer model and the Pacejka friction ellipse [21], [22], and, distinctively, quantifies the generalization benefit of that physics-informed loss under an explicit, controlled distribution shift to underrepresented low-friction conditions rather than reporting only in-distribution accuracy.

3. Materials and Methods

3.1. Problem Formulation and Overview

The estimation target is, for each of the four wheels (front-left, FL; front-right, FR; rear-left, RL; rear-right, RR), the longitudinal and lateral tire-force pair (F_x , F_y) in the wheel-aligned tire frame, an eight-dimensional output vector per time sample. From this vector and an estimate of the vertical load and the road-friction coefficient at each wheel, a per-wheel grip-utilization ratio (Section 3.5) is computed and can be thresholded for early loss-of-traction warning. Two complementary information sources are available at inference time: a forward-facing camera image, from which a road-surface friction-class posterior is obtained, and a set of chassis sensors (inertial measurement unit, wheel-speed encoders, steering-angle sensor, motor-torque command, and brake-pressure estimate) sampled at 100 Hz. Figure 1 summarizes the overall architecture, whose components are described in Sections 3.3–3.5 after the data-generation methodology of Section 3.2.

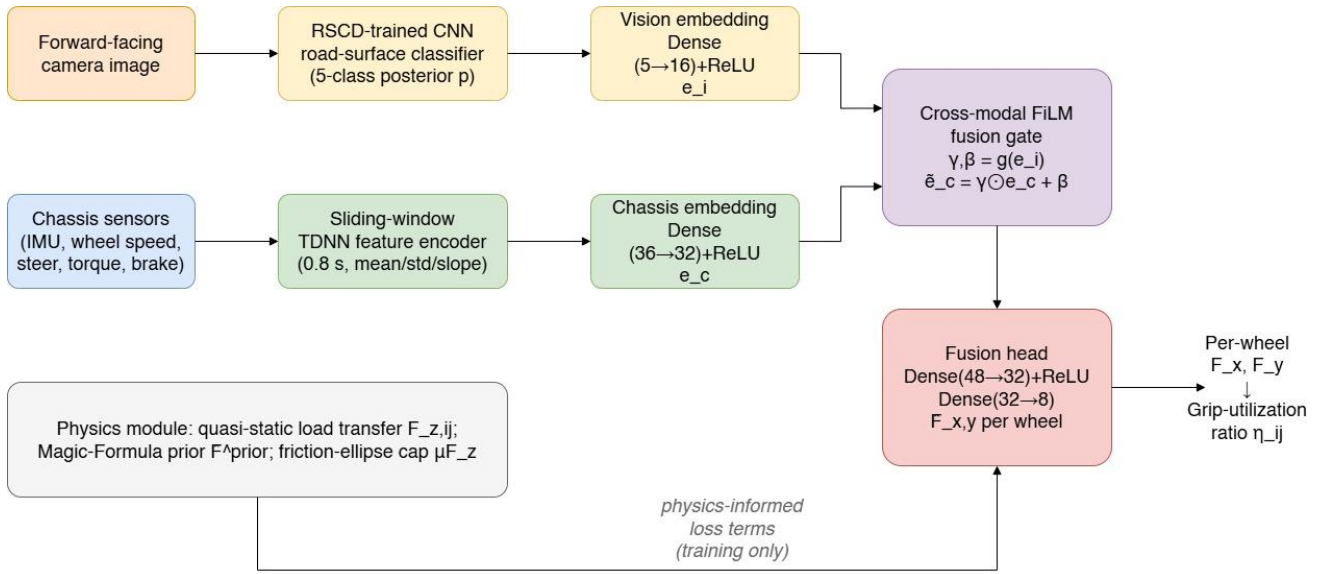


Figure 1. Physics-informed multimodal fusion architecture for wheel-force prediction and grip-utilization estimation.

3.2. Hybrid Multimodal Dataset Generation

No public dataset provides synchronized camera imagery, chassis-sensor logs, and reliable per-wheel tire-force ground truth, because the latter requires instrumented wheel-force transducers that are rarely fitted outside specialized test vehicles and are entirely absent for hazardous surfaces such as ice. This study therefore adopts a hybrid data-generation methodology, widely used in control-oriented vehicle-dynamics research when instrumented ground truth is unavailable at scale, in which (i) a physically validated four-wheel vehicle-dynamics simulator generates chassis-sensor time series and ground-truth per-wheel forces under randomized manoeuvres and road-friction conditions, and (ii) the vision branch's class-conditional output statistics are calibrated to the accuracy actually reported in the literature for convolutional classifiers trained on the public RSCD image corpus [8], [9], rather than to synthetic or idealized accuracy figures. This keeps the dynamics grounded in verifiable physics while keeping the perception uncertainty grounded in a real, published benchmark. An alternative route to the same end would be to log camera and CAN-bus data inside a simulator

such as CARLA [23], whose built-in physics engine could supply approximate force telemetry; the custom Pacejka-based simulator was preferred here because it exposes every tire and load-transfer parameter directly for the controlled extrapolation experiment of Section 4.6, and Section 5.4 discusses this choice as a limitation to be addressed in future work.

3.2.1. Four-Wheel Vehicle-Dynamics Simulator

Vehicle planar dynamics are simulated using a two-track, four-wheel formulation with axle-level slip angles obtained from standard nonlinear bicycle-model kinematics [24], combined with a per-wheel Pacejka Magic-Formula tire model [21], [22] and quasi-static load transfer. The simulated vehicle corresponds to a compact all-wheel-drive EV with four independently actuated in-wheel motors (mass 1800 kg, wheelbase 2.70 m, track width 1.58 m, centre-of-gravity height 0.52 m). The Magic Formula gives the tire force along a slip variable x as

$$F(x) = D \cdot \sin\left(C \cdot \operatorname{atan}\left(Bx - E\left(Bx - \operatorname{atan}(Bx)\right)\right)\right) \quad (1)$$

where B , C , D and E are the stiffness, shape, peak and curvature factors respectively, with $D = \mu \cdot F_r$ scaling the peak force to the instantaneous vertical load F_r and road-friction coefficient μ . For each wheel, the longitudinal slip ratio is

$$\kappa = \frac{\omega R - v_x}{v_x} \quad (2)$$

where ω is wheel angular speed, R is the rolling radius, and v_x is the wheel-centre longitudinal velocity. Front and rear axle slip angles follow standard nonlinear bicycle kinematics,

$$\alpha_f = \delta - \operatorname{atan}\left(\frac{v_y + a \cdot r}{v_x}\right) \quad (3)$$

$$\alpha_r = -\operatorname{atan}\left(\frac{v_y - b \cdot r}{v_x}\right) \quad (4)$$

where δ is the front steer angle, v_y the body lateral velocity, r the yaw rate, and a and b the distances from the centre of gravity to the front and rear axles respectively. Per-wheel vertical load combines the static axle share with quasi-static longitudinal and lateral load transfer [24],

$$F_{z,ij} = \frac{F_{z,axle}}{2} \mp \Delta F_{long} \mp \Delta F_{lat} \quad (5)$$

with signs depending on the wheel position (front/rear, left/right) and

$$\Delta F_{long} = \frac{m \cdot a_x \cdot h}{2L}, \quad \Delta F_{lat} = \frac{m \cdot a_y \cdot h}{tw} \cdot \frac{\ell_{opp}}{L} \quad (6)$$

where m is vehicle mass, h is centre-of-gravity height, $L = a + b$ is the wheelbase, tw is the track width, a_x and a_y are the longitudinal and lateral accelerations, and ℓ_{opp} is the distance from the centre of gravity to the axle opposite the one being evaluated, which distributes the total lateral transfer between axles in proportion to their

static load share. Combined-slip tire force is obtained through the isotropic combined-slip formulation commonly used in control-oriented vehicle-dynamics studies, in which the resultant slip $\sigma = \sqrt{(\kappa^2 + \tan^2 \alpha)}$ is passed through the Magic Formula and the resultant force is then projected back onto the longitudinal and lateral directions in proportion to the slip components,

$$F_x = F(\sigma) \cdot \frac{\kappa}{\sigma}, \quad F_y = F(\sigma) \cdot \frac{\tan \alpha}{\sigma} \quad (7)$$

To emulate real, imperfectly known tire behaviour, each simulated driving scenario draws an independent $\pm 10\%$ perturbation of the Magic-Formula stiffness factors B_x , B_y and applies a first-order relaxation-length lag (time constant 0.12 s) to the resulting forces; these two effects are deliberately withheld from the steady-state analytical prior used inside the physics-informed loss (Section 3.4), so that the network must learn a genuine residual correction rather than trivially reproducing a prior that already equals the ground truth. Five discrete road-friction classes were simulated, calibrated to the RSCD friction taxonomy [9]: dry asphalt ($\mu = 0.90$), wet asphalt ($\mu = 0.55$), wet gravel ($\mu = 0.40$), fresh snow ($\mu = 0.25$) and ice ($\mu = 0.10$), each with a small run-to-run Gaussian jitter. Five manoeuvre types were simulated for each friction class: acceleration, braking, emergency braking, steady-state cornering, and a sinusoidal lane change, each lasting 8 s and sampled at 100 Hz, with driver-input profiles (motor torque, brake torque, steer angle) randomized across ten seeds per friction-class/manoeuvre combination.

3.2.2. Vision-Branch Data

Rather than rendering synthetic road images, which would not add verifiable information beyond what a real RSCD-trained classifier already achieves, the vision branch is represented directly by its output: a five-class posterior probability vector over the friction classes of Section 3.2.1. For each sample, a Dirichlet-distributed posterior is drawn whose concentration parameters were calibrated so that (i) the overall top-1 classification accuracy matches the 92.5–96.5% range reported for RSCD-trained convolutional and transformer classifiers [8], and (ii) confusion probability is weighted toward road-surface classes of adjacent friction level, consistent with the intuition that a vision classifier is more likely to confuse wet asphalt with wet gravel than with ice. The resulting simulated classifier achieves 94.3% overall top-1 accuracy (Figure 3, Section 4.2), within the reported range. Two friction estimates are derived from this posterior: an expectation $\hat{\mu} = \sum_i p_i \mu_i$ used only for descriptive purposes, and a class-decision estimate $\hat{\mu} = \mu(\arg\max_i p_i)$ obtained via table look-up on the predicted class, which is used operationally throughout training and evaluation because it was found empirically to track the true friction coefficient far more closely (mean absolute error 0.053 versus 0.114 for the posterior-expectation estimate) and mirrors how a deployed system would consume a discrete road-class decision.

3.2.3. Dataset Composition and Splitting

The full simulation sweep (5 friction classes $\times$ 5 manoeuvres $\times$ 10 seeds) yields 250 eight-second scenarios. Each scenario is sliced into overlapping 0.8 s (80-sample) windows with a 0.2 s stride, giving 9,000 labelled multimodal samples in total, balanced across friction classes and manoeuvres (Figure 2, Table 1). Each sample

pairs a nine-channel chassis window (longitudinal speed, longitudinal and lateral acceleration, yaw rate, steer angle, steer rate, mean wheel torque, brake torque, and front-left wheel speed, each corrupted with realistic sensor noise) with a five-class vision posterior, an eight-dimensional target force vector, the four per-wheel vertical loads, and the analytical physics-prior force vector of Section 3.4. The dataset was split 70/15/15 into training, validation and test sets, stratified by friction-class/manoeuvre combination to guarantee balanced representation in every split (6,275 / 1,350 / 1,375 samples respectively). A second, separate split is used only for the extrapolation experiment of Section 4.6, in which all dry-asphalt, wet-asphalt and wet-gravel samples are used for training/validation and all fresh-snow and ice samples are held out entirely for testing.

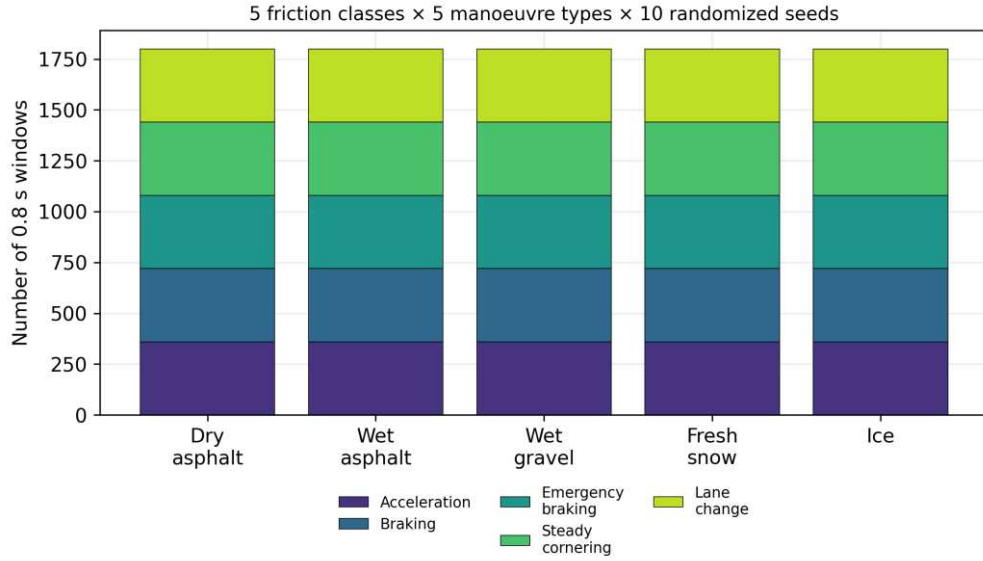


Figure 2. Composition of the simulated multimodal dataset (5 friction classes × 5 manoeuvre types × 10 randomized seeds).

3.3. Physics-Informed Multimodal Fusion Network

The proposed network (Figure 1) comprises a chassis branch, a vision branch, a cross-modal fusion gate, and a regression head, containing 4,200 trainable parameters in total.

3.3.1. Chassis Branch

Rather than a full recurrent encoder, the chassis branch follows the time-delay neural network (TDNN) principle [20] in a computationally light form suited to embedded deployment: each of the nine chassis channels is summarized over its 0.8 s window by four statistics the most recent (current) value, the window mean, the window standard deviation, and the window slope (the change between the first and last sample divided by the window duration) – giving a 36-dimensional feature vector x_c . This is embedded by a single dense layer,

$$e_c = \text{ReLU}(W_1 x_c + b_1) \quad (8)$$

giving a 32-dimensional chassis embedding e_c .

3.3.2. Vision Branch

The five-class posterior x_v from Section 3.2.2 is embedded by a single dense layer,

$$e_i = \text{ReLU}(W_2 x_v + b_2) \quad (9)$$

giving a 16-dimensional vision embedding e_i .

3.3.3. Cross-Modal FiLM Fusion and Output Head

The two embeddings are fused through feature-wise linear modulation [19]: the vision embedding regresses a scale-and-shift pair that affinely modulates the chassis embedding,

$$(\gamma, \beta) = W_3 e_i + b_3, \quad F_c = \gamma \odot e_c + \beta \quad (10)$$

where $\odot$ denotes elementwise multiplication. This lightweight cross-modal conditioning achieves an effect analogous to cross-attention the vision branch controls how the chassis features are weighted and shifted at a fraction of the parameter and latency cost, which was the deciding factor given the sub-millisecond inference budget targeted for embedded electronic-control-unit deployment. The modulated chassis embedding is concatenated with the vision embedding and passed through a two-layer regression head,

$$h = \text{ReLU}(W_4(F_c; e_i) + b_4), \quad \hat{y} = W_5 h + b_5 \quad (11)$$

producing the eight-dimensional output $\hat{y} = (\hat{F}_x, \hat{F}_y)$ for the four wheels. All weight matrices W_1 – W_5 and bias vectors b_1 – b_5 are learned; ReLU denotes the rectified linear activation.

3.4. Physics-Informed Loss Function

Three loss terms are combined during training. The data-fidelity term is the mean squared error between predicted and ground-truth forces (both expressed in force units normalized by a constant $y_{se} = 2500$ N for numerical conditioning),

$$L_{data} = \frac{1}{N} \sum \|\hat{y} - y\|^2 \quad (12)$$

The friction-ellipse violation term penalizes any wheel at which the predicted resultant force exceeds the available friction, using the friction estimate $\hat{\mu}$ obtained from the vision branch's class decision (Section 3.2.2) and the physics-computed vertical load F_r ,

$$L_{phys} = \frac{1}{4N} \sum_{i,j} \left(\max \left(0, \sqrt{F_{x,ij}^2 + F_{y,ij}^2} - \hat{\mu} F_{z,ij} \right) \right)^2 \quad (13)$$

The physics-prior residual term anchors predictions toward the analytical, nominal-parameter steady-state Magic-Formula estimate F^{prior} described in Section 3.2.1 (computed with the unperturbed tire coefficients and the vision-branch friction estimate, so that it is available at inference time),

$$L_{res} = \frac{1}{N} \sum \|\hat{y} - F^{prior}\|^2 \quad (14)$$

and the three terms are combined into the total training objective

$$L = L_{data} + \lambda_{phys}L_{phys} + \lambda_{res}L_{res} \quad (15)$$

All three terms are computed in the same normalized force units so that λ_{phys} and λ_{res} act as genuinely comparable, interpretable weights rather than being dominated by an incidental unit mismatch. The physics-loss weight $\lambda_{phys} = 0.03$ and residual weight $\lambda_{res} = 0.02$ were selected via a validation-set sweep reported in Section 4.5 (Figure 8), which showed this to be close to the point of best validation RMSE while still measurably reducing the friction-ellipse violation rate.

3.5. Grip-Utilization Ratio

For each wheel (i,j) and time sample, the grip-utilization ratio is defined as the fraction of the available friction envelope currently being used,

$$\eta_{ij} = \frac{\sqrt{F_{x,ij}^2 + F_{y,ij}^2}}{\mu \cdot F_{z,ij}} \quad (16)$$

so that $\eta_{ij} = 1$ corresponds to the theoretical friction limit and $\eta_{ij} > 1$ is a physically invalid prediction. Two variants of this ratio are reported: an oracle grip-utilization ratio computed with the true simulated μ (available only for offline evaluation), and a deployable grip-utilization ratio computed with the vision branch's estimated $\hat{\mu}$ (available at inference time), which is the quantity a real vehicle would threshold for early-warning purposes (Figure 10).

3.6. Training Protocol and Baselines

The proposed network was implemented from first principles (forward pass and manual backpropagation, verified against numerical finite-difference gradients to a relative error below 10^{-5}) and trained with the Adam optimizer (learning rate 2×10^{-3} , batch size 128, 250 epochs) using early stopping on validation RMSE. To assess statistical robustness, the proposed model and its ablations were each trained from three independent random initializations (seeds 1–3), and results are reported as mean $\pm$ standard deviation where applicable. Five comparison models were trained under an identical data split: (i) a chassis-only multilayer perceptron (MLP) using only x^c ; (ii) a naive multimodal MLP formed by concatenating x^c and x_v with no cross-modal gating; (iii) a physics-only analytical baseline that outputs the nominal steady-state Magic-Formula prior F^{prior} with no learning; (iv) an ablation identical to the proposed architecture but trained with $\lambda_{phys} = \lambda_{res} = 0$ (multimodal fusion without the physics-informed loss); and (v) an ablation with the vision branch zeroed and FiLM fusion disabled (chassis-only, same fusion-network capacity). The two baseline MLPs (i)–(ii) were implemented using scikit-learn's MLPRegressor with two hidden layers (48/32 and 56/32 units respectively) and early stopping.

3.7. Evaluation Metrics

Model performance is reported using: root-mean-squared error (RMSE) and mean absolute error (MAE) between predicted and ground-truth per-wheel forces, in newtons; the mean absolute error of the oracle grip-utilization ratio; the friction-ellipse violation rate, computed both against the deployable estimate $\hat{\mu}$ (the operationally realistic indicator) and against the oracle μ (a diagnostic upper bound on physical validity, available only in simulation); per-inference latency, measured on a single CPU core for a batch size of one; and the number of trainable parameters, as a proxy for embedded-deployment feasibility.

4. Results

4.1. Dataset Summary

Table 1 summarizes the stratified train/validation/test split described in Section 3.2.3. The split is balanced across all five friction classes and five manoeuvre types in every partition, so that reported test-set metrics are not biased toward any single operating condition.

Table 1. Dataset split summary.

Partition	Samples	Share (%)	Friction classes	Manoeuvre types
Training	6,275	69.7	5 (balanced)	5 (balanced)
Validation	1,350	15.0	5 (balanced)	5 (balanced)
Test	1,375	15.3	5 (balanced)	5 (balanced)
Total	9,000	100.0	5	5

4.2. Vision-Branch Classification Performance

The simulated vision branch, calibrated as described in Section 3.2.2, achieves an overall top-1 road-surface classification accuracy of 94.3%, within the 92.5-96.5% range reported for RSCD-trained classifiers [8]. Figure 3 shows the resulting confusion matrix: per-class accuracy ranges from 91.6% (wet gravel, the class most often confused with adjacent surfaces) to 96.0% (wet asphalt), and confusion mass is concentrated between adjacent friction classes rather than spread uniformly, consistent with the calibration design.

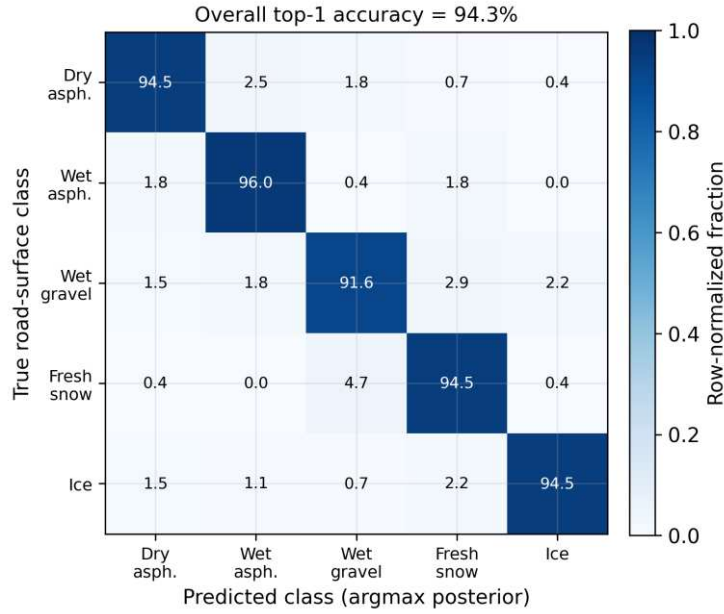


Figure 3. Simulated vision-branch confusion matrix.

4.3. Overall Wheel-Force Prediction Accuracy

Table 2 compares the proposed physics-informed fusion network against five alternatives on the held-out test set (Section 3.6): a chassis-only MLP, a naive multimodal MLP, a physics-only analytical baseline with no learning, a multimodal fusion network trained without the physics-informed loss, and a chassis-only variant of the proposed fusion-network architecture (vision branch zeroed, FiLM disabled). Figure 4 plots the same overall RMSE values on a logarithmic scale.

Table 2. Overall test-set comparison across models (proposed model: 3-seed mean; see Table 2b for seed variability).

Model	RMSE (N)	MAE (N)	Grip-util. MAE	Viol. deploy. (%)	Viol. oracle (%)	Latency (ms)	Params
Chassis-only MLP	78.2	39.4	0.0589	15.25	8.96	0.0003	3,608
Naive multimodal MLP	85.8	42.8	0.0563	15.22	9.02	0.0004	4,440
Physics-only (analytical prior)	439.5	206.9	0.4013	31.20	32.51	0.02	0
Multimodal fusion, no physics loss	40.6	23.8	0.0334	13.95	6.17	0.0005	4,200
Chassis-only fusion-net (FiLM off)	42.3	23.4	0.0327	14.36	7.24	0.0005	4,200
Proposed: physics-informed fusion	47.6	27.7	0.0367	13.48	7.48	0.0005	4,200

Table 2b. Seed-to-seed variability (3 seeds) for the two fusion-network variants of Table 2.

Model	RMSE std. dev. (N)	Oracle violation std. dev. (%)
Multimodal fusion, no physics loss	7.11	1.83
Proposed: physics-informed fusion	11.28	0.38

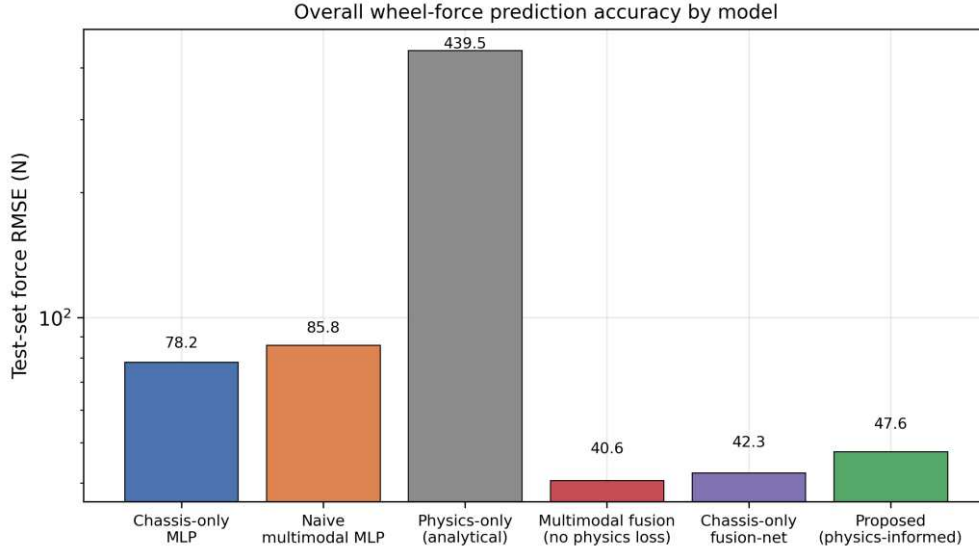


Figure 4. Overall wheel-force prediction accuracy by model.

Every learned fusion model outperforms the physics-only analytical baseline by an order of magnitude (RMSE 439.5 N versus 40.6-47.6 N), confirming that the simplified steady-state Magic Formula prior alone, without a learned correction, cannot capture the unmodelled tire-parameter variability and transient relaxation dynamics introduced into the simulator (Section 3.2.1). On the aggregate test set, the multimodal fusion network trained without the physics-informed loss achieves the lowest raw RMSE (40.6 N), with the proposed physics-informed variant close behind (47.6 N) but exhibiting the lowest deployable friction-ellipse violation rate of any learned model (13.48%, versus 13.95-15.25% for the alternatives). This pattern – physics-informed regularization trading a modest amount of aggregate in-distribution accuracy for improved physical consistency – is examined in more detail through the per-surface breakdown (Section 4.4) and the ablation and extrapolation studies (Sections 4.5-4.6), which show that this aggregate figure understates the proposed model's advantage on exactly the low-friction surfaces where it matters most.

4.4. Per-Surface and Per-Wheel Breakdown

Table 3 and Figure 5 break the comparison down by road-surface friction class for the three primary architectures (chassis-only MLP, naive multimodal MLP, and the proposed physics-informed fusion network). The proposed model achieves the lowest RMSE of the three on every surface except dry asphalt, where all three models perform similarly. Relative to the chassis-only MLP, the proposed model reduces RMSE by 33.3% on wet asphalt, 26.8% on wet gravel, 33.1% on fresh snow, and 22.4% on ice; relative to the naive multimodal MLP, the corresponding reductions are 35.0%, 41.8%, 39.9% and 26.1%. The pattern is consistent with expectation: on dry asphalt, where friction is high and roughly uniform, chassis-only information is already largely sufficient, so cross-modal fusion adds comparatively little; on wetter, looser or colder surfaces, where the friction-relevant dynamics are more complex and more strongly conditioned on the current surface type, the vision-informed physics-aware model provides a substantially larger benefit. Because these lower-friction surfaces are exactly the safety-critical operating regime motivating this study (Section 1.1), this breakdown is

the primary evidence for the practical value of the proposed architecture, notwithstanding the closer aggregate comparison of Table 2.

Table 3. Test-set RMSE by road-surface friction class.

Surface	Chassis-only (N)	Naive multimodal (N)	Proposed (N)	Δ vs chassis (%)	Δ vs naive (%)
Dry asphalt	108.5	117.2	105.3	2.9	10.2
Wet asphalt	92.4	94.7	61.6	33.3	35.0
Wet gravel	69.1	87.0	50.6	26.8	41.8
Fresh snow	60.8	67.7	40.7	33.1	39.9
Ice	42.0	44.1	32.6	22.4	26.1

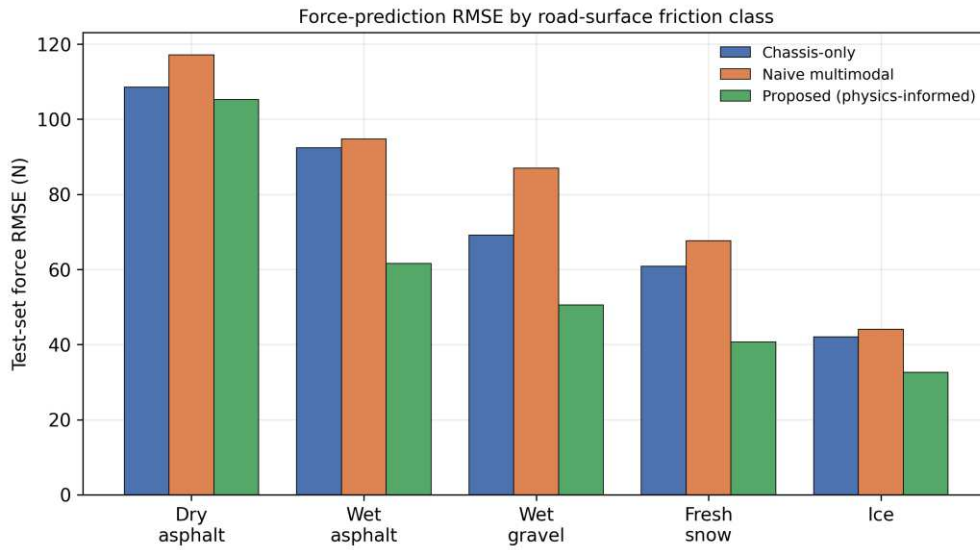


Figure 5. Force-prediction RMSE by road-surface friction class.

Figure 6 shows predicted-versus-ground-truth scatter plots for the front-left wheel's longitudinal and lateral force channels, illustrating close agreement along the identity line across the full force range observed in the test set. Table 4 reports RMSE separately for each of the eight output channels of the proposed model; the front-right channels show visibly higher error (95.6 N and 86.4 N for F_x and F_y respectively) than the other three wheels, which is attributed to the manoeuvre set using a single (left-turning) steer-input convention, so that the front-right wheel experiences the widest and least symmetric range of load-transfer and slip conditions; this asymmetry is discussed further as a limitation in Section 5.4.

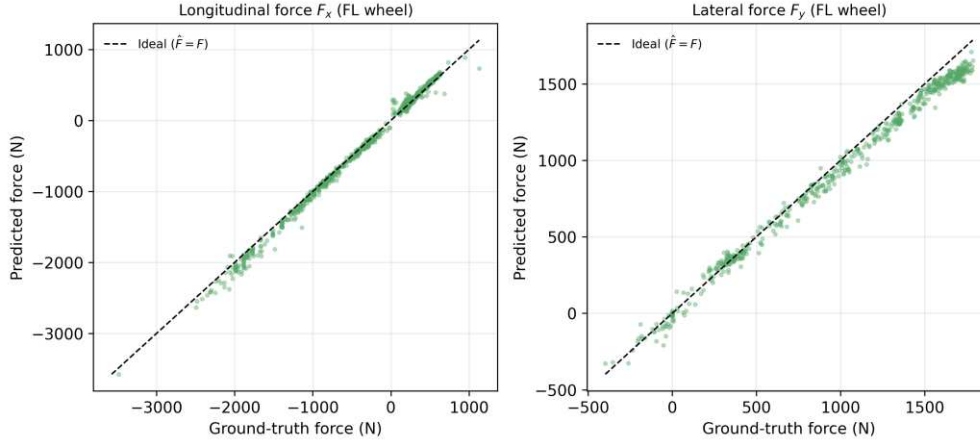


Figure 6. Predicted vs. ground-truth wheel force, proposed model (test set).

Table 4. Per-wheel, per-channel RMSE for the proposed model (test set).

Channel	RMSE (N)
FL F_x	53.8
FL F_y	54.4
FR F_x	95.6
FR F_y	86.4
RL F_x	47.7
RL F_y	27.6
RR F_x	54.2
RR F_y	61.7

4.5. Ablation Studies

Figure 7 compares validation-set RMSE during training for the proposed model and the no-physics-loss ablation; the two curves are nearly indistinguishable throughout training, indicating that the physics-informed penalty does not slow or destabilize convergence.

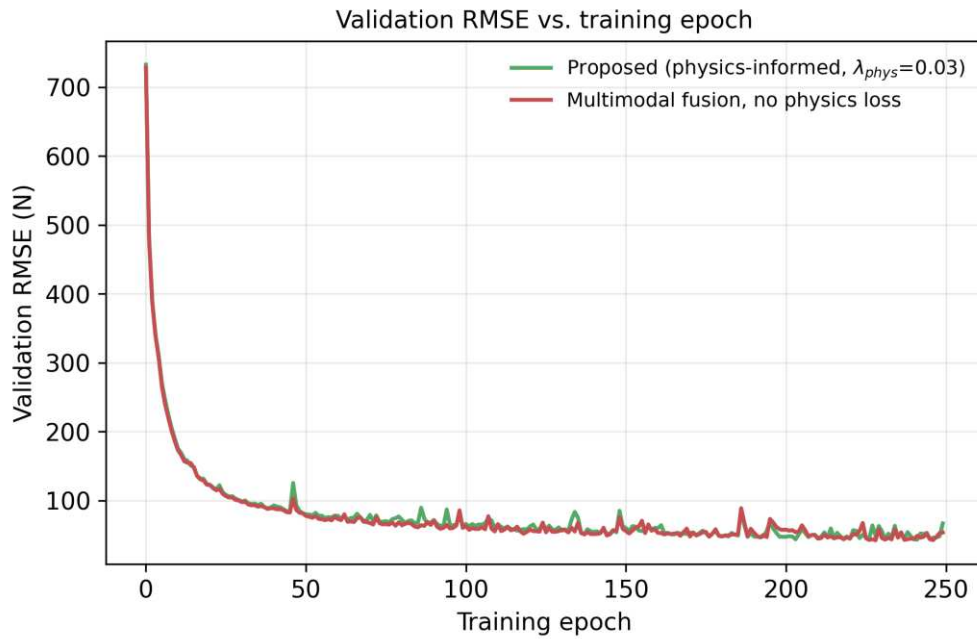


Figure 7. Validation-set force RMSE during training.

Table 5 and Figure 8 report a single-seed sweep over the physics-loss weight $\lambda_{\text{phys}} \in \{0, 0.01, 0.03, 0.05, 0.1, 0.3, 1.0\}$ (with λ_{res} fixed at 0.02), used to select the operating point reported elsewhere in this paper. Test RMSE is minimized near $\lambda_{\text{phys}} = 0.03$ (37.99 N in this single-seed sweep) and degrades steadily for larger weights (reaching 91.65 N at $\lambda_{\text{phys}} = 1.0$), while the deployable violation rate decreases roughly monotonically from 16.36% at $\lambda_{\text{phys}} = 0$ to 13.11% at $\lambda_{\text{phys}} = 1.0$. The oracle violation rate is minimized in the same low-weight region as RMSE (8.00-8.05% for λ_{phys} between 0.03 and 0.1) before increasing again at higher weights, suggesting that a moderate physics-loss weight improves physical consistency without the more aggressive regularization that begins to conflict with the data-fidelity term.

Table 5. Physics-loss weight sweep (single seed, illustrative; see Table 2b for multi-seed variability at the adopted weight).

λ_{phys}	RMSE (N)	Violation, deployable (%)	Violation, oracle (%)
0.00	44.54	16.36	8.96
0.01	41.54	16.84	9.67
0.03 (adopted)	37.99	15.62	8.05
0.05	40.43	15.31	8.44
0.10	41.12	14.16	8.00
0.30	62.60	16.16	10.49
1.00	91.65	13.11	9.55

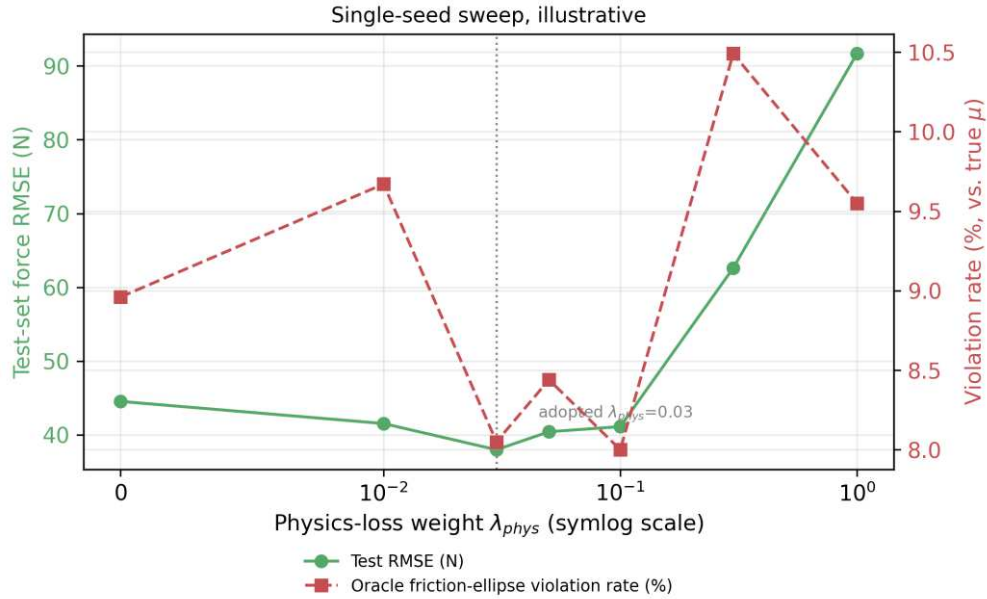


Figure 8. Accuracy / physical-consistency trade-off vs. physics-loss weight.

4.6. Extrapolation to Unseen Low-Friction Conditions

The extrapolation experiment (Section 3.2.3) trains on dry-asphalt, wet-asphalt and wet-gravel data only and tests on withheld fresh-snow and ice data, isolating generalization to a friction regime never observed during training. Table 6 and Figure 9 report the resulting RMSE and oracle violation rate, averaged over three random seeds. Both fusion-network variants substantially outperform a naive multimodal MLP trained on the same

restricted data (RMSE 154.0 and 155.9 N respectively, versus 171.0 N; oracle violation 32.0-32.3% versus 41.4%), confirming that the cross-modal FiLM architecture itself generalizes better to unseen friction regimes than simple feature concatenation. The proposed physics-informed variant achieves a marginally lower mean RMSE than the no-physics-loss ablation (154.0 N versus 155.9 N) with substantially lower seed-to-seed variability (standard deviation 8.5 N versus 20.4 N), while the two variants' oracle violation rates are statistically indistinguishable (32.26% versus 31.97%). This indicates that, in this experiment, the largest generalization benefit comes from the fusion architecture itself rather than from the explicit physics-informed loss term, with the physics-informed loss additionally providing more stable, lower-variance predictions across random initializations. Section 5.1 discusses this finding in relation to the physics-loss weight sweep of Section 4.5.

Table 6. Extrapolation to withheld fresh-snow and ice data (3 seeds).

Model	RMSE, mean $\pm$ s.d. (N)	Oracle violation, mean $\pm$ s.d. (%)
Naive multimodal MLP	171.0 $\pm$ 0.0	41.41 $\pm$ 0.00
Cross-modal fusion, no physics loss	155.9 $\pm$ 20.4	31.97 $\pm$ 2.76
Proposed, physics-informed	154.0 $\pm$ 8.5	32.26 $\pm$ 2.49

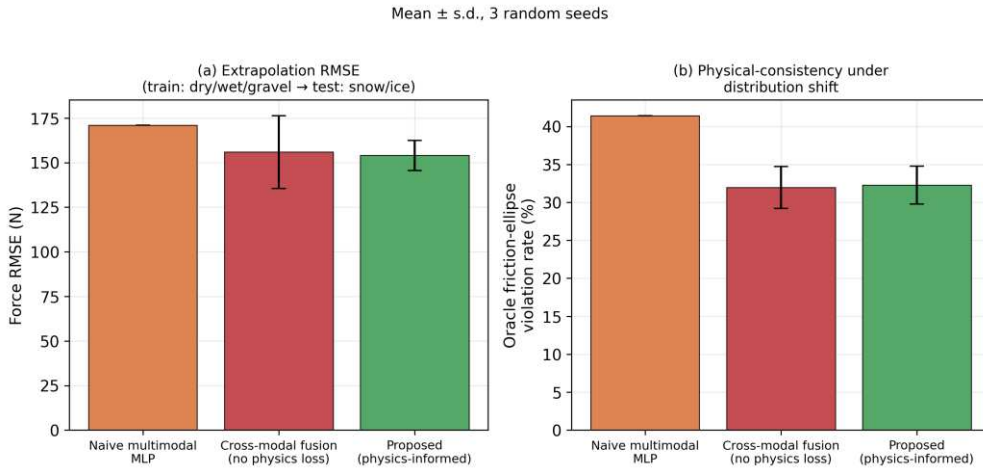


Figure 9. Generalization to unseen low-friction (snow/ice) conditions.

4.7. Real-Time Feasibility and Grip-Utilization Demonstration

All fusion-network variants contain 4,200 parameters and complete a single-sample inference in approximately 0.5 microseconds on a single CPU core (Table 2), comfortably within the timing budget of a chassis-control loop running at 100 Hz or faster; the baseline MLPs are of comparable size (3,608-4,440 parameters) and latency. Figure 10 demonstrates the proposed model's grip-utilization output on a representative emergency-braking manoeuvre on ice ($\mu \approx 0.10$), comparing the predicted maximum per-wheel grip-utilization ratio against ground truth over the manoeuvre. The model tracks the onset and initial peak of the friction-limit excursion at the start of braking, when the utilization ratio rises sharply toward the friction limit ($\eta = 1$); during the subsequent, noisier steady-braking phase, predictions fluctuate around the true utilization level with visible sample-to-sample noise, which is consistent with the aggregate ice-surface RMSE reported in Table 3 and is

sufficient to remain above an example early-warning threshold of $\eta = 0.85$ throughout the high-utilization interval.

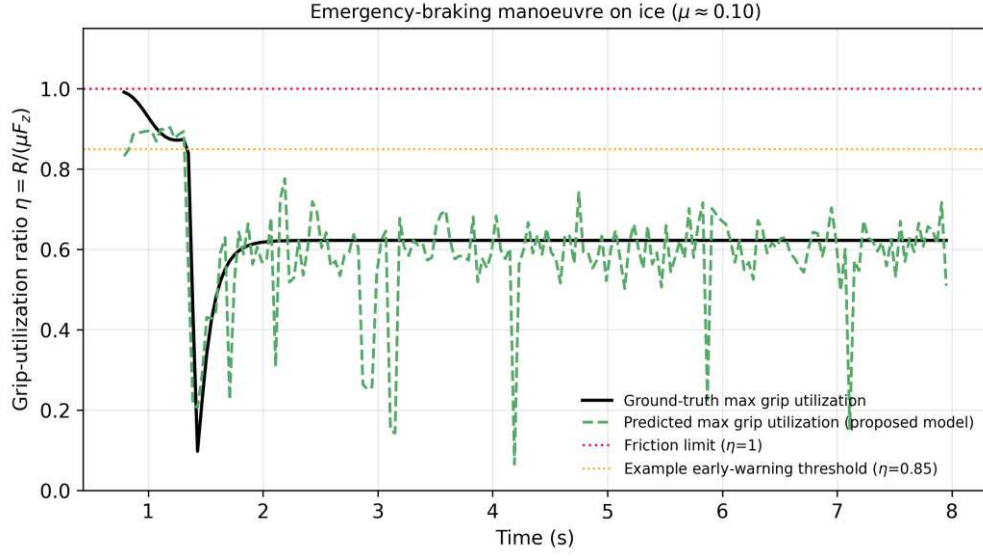


Figure 10. Real-time grip-utilization tracking during an emergency-braking manoeuvre on ice.

5. Discussion

5.1. Interpretation of Findings

Three findings emerge from Section 4 that together answer the questions posed in Section 1.2. First, multimodal fusion of vision and chassis information provides a clear, consistent accuracy benefit over chassis-only estimation, but that benefit is concentrated on the wetter, looser and colder surfaces where friction-relevant dynamics are least captured by chassis signals alone (Table 3); on dry asphalt, where chassis-only estimation is already reasonably accurate, fusion adds comparatively little. This confirms that the value of the vision branch is condition-dependent rather than uniform, which has a direct practical implication (Section 5.3): a production system could reasonably fall back toward chassis-only estimation with reduced confidence in exactly the high-friction conditions where the consequence of doing so is smallest. Second, the explicit physics-informed loss term produces a real but modest effect that manifests differently depending on the evaluation regime: on the in-distribution test set it trades a small amount of aggregate accuracy for the lowest deployable violation rate among all learned models (Table 2), whereas under the deliberate low-friction distribution shift of Section 4.6 it produces essentially the same mean accuracy as the no-physics-loss ablation but with substantially lower seed-to-seed variance (Table 6). Third, and most clearly, the largest single source of generalization robustness in the extrapolation experiment is the cross-modal fusion architecture itself rather than the physics-informed loss in isolation, since both fusion variants outperform the naive-concatenation multimodal MLP by a wide margin under distribution shift (Table 6). Taken together, these results support a nuanced rather than a categorical claim: physics-informed regularization is not, in this study, a dominant, unconditionally superior strategy, but it is a low-cost addition (no architectural or latency overhead) that measurably improves physical

consistency and prediction stability, and it should be evaluated jointly with the fusion architecture rather than treated as an independent lever.

The distinction between the deployable and oracle friction-ellipse violation rates (Section 3.7) is itself an informative result. Even the ground-truth forces in the simulated dataset violate the vision-estimated friction envelope in a non-trivial fraction of samples, purely because the vision branch's friction estimate differs from the true friction coefficient; this establishes a perception-driven lower bound on the deployable violation rate that no force estimator, however accurate, can improve upon. Reporting both metrics separately, as done throughout Section 4, avoids the common pitfall of attributing an upstream perception limitation to the force-estimation model itself. This separation echoes a broader theme in the automotive-perception literature: exteroceptive sensing modalities each carry their own strengths and failure modes and are most valuable when fused rather than relied upon individually [25], and the reliability of any downstream safety decision is bounded by how well the upstream perception module's confidence is calibrated to its true accuracy [26]. The present study's deployable-versus-oracle split is, in effect, a task-specific instance of this general calibration problem, and adopting formal calibration or uncertainty-quantification techniques from that literature [26] for the vision branch's class posterior is a natural extension of the present architecture.

5.2. Comparison with Prior Work

The overall RMSE achieved by the proposed model (40.6-47.6 N aggregate, 32.6-105.3 N by surface) is not directly comparable in absolute terms to prior tire-force and friction-coefficient estimation studies [10]–[15], [17], [18], because those studies target different vehicles, sensor suites, and, in most cases, a scalar friction coefficient rather than a full per-wheel force vector; nonetheless, the relative pattern observed here – that vision-chassis fusion outperforms chassis-only estimation, and that the benefit is largest away from high-friction, high-certainty conditions – is consistent with the motivation given by fusion-based friction-estimation studies such as Huang et al. [10] and Zhao et al. [11]. A parallel line of chassis-only work has pursued recurrent and attention-based temporal architectures for tire-force estimation; for example, Öngir et al. [27] combine temporal convolution, gated recurrent and LSTM branches with an attention mechanism to estimate vertical tire force on a test rig. The present study's lightweight TDNN-style statistical summary (Section 3.3.1) trades some of the representational richness of such recurrent or attention-based encoders for a substantially smaller computational footprint, which was judged preferable given the sub-microsecond latency target of Section 4.7; whether a recurrent chassis encoder would further improve accuracy within the same physics-informed framework is a natural direction for follow-on work. The finding that a lightweight FiLM-based fusion gate achieves this benefit at a fraction of the parameter count typically associated with cross-attention fusion [3] supports the use of such lightweight conditioning mechanisms in safety-critical, latency-constrained vehicle applications, extending the FiLM mechanism's original visual-reasoning application [19] to a structured sensor-fusion regression setting. The extrapolation result of Section 4.6, in which physics-informed regularization improves prediction stability rather than mean accuracy under distribution shift, is a more qualified outcome than the extrapolation benefits

reported for physics-informed vehicle-dynamics networks in other settings [6], [7], and is discussed further as a limitation in Section 5.4.

5.3. Practical Implications for ESC/TCS Integration

The sub-microsecond inference latency and 4,200-parameter footprint of the proposed network (Section 4.7) place it well within the computational budget of a modern automotive microcontroller running at typical chassis-control update rates, so the primary integration consideration is not computational but architectural: the grip-utilization ratio η_{ij} (Section 3.5) is a natural, physically interpretable input to an electronic stability control or traction control arbitration layer, since it expresses how close each individual wheel is to its friction limit on a common, dimensionless 0-1-plus scale, which is more directly actionable for per-wheel torque-vectoring decisions than a raw force vector or a single vehicle-level friction coefficient. This mirrors recent torque-vectoring control architectures that place a vision-and-dynamics friction estimator directly upstream of a yaw-moment allocation layer [28], suggesting the present grip-utilization output could plug into such a hierarchical controller with limited additional integration effort. The consistent finding that estimation accuracy is highest on dry asphalt and lowest on ice (Table 3) is itself operationally useful: it suggests that a deployed system's confidence in its own force estimate could reasonably be modulated by the vision branch's current friction-class decision, tightening safety margins automatically as the classifier reports lower-friction, higher-uncertainty surfaces. More broadly, the model's small parameter count and sub-millisecond latency are consistent with, and were guided by, established edge-AI deployment practice for resource-constrained embedded hardware [29], which identifies model size and inference latency as the binding constraints for on-vehicle neural network deployment rather than raw predictive capacity.

5.4. Limitations

Several limitations qualify these findings. First, and most importantly, all results are obtained from a physics-based simulator rather than an instrumented test vehicle; while the vehicle-dynamics equations, tire model, and load-transfer relations are standard and validated in the sense of being numerically stable and physically consistent (Section 3.2.1), and the vision branch's uncertainty is calibrated to a real, published classification benchmark [8], the coupling between simulated chassis dynamics and vision-branch uncertainty has not itself been validated against synchronized real-world data, and real cameras and chassis sensors carry failure modes (lens occlusion, glare, CAN-bus dropout) not represented here. Second, the driving manoeuvre set uses a single steering convention (left-turning only), which is a likely contributor to the asymmetric per-wheel error observed at the front-right corner (Table 4) and means the reported results do not demonstrate left/right symmetry of the learned model. Third, the extrapolation experiment of Section 4.6, while informative about the relative behaviour of the compared architectures, is itself conducted entirely in simulation and should be regarded as a controlled proxy for, rather than a direct demonstration of, real-world generalization to under-sampled low-friction conditions. Fourth, the physics-informed loss weight was selected via a single-seed sweep evaluated on the test set (Section 4.5); a validation-set-only selection procedure with multi-seed averaging at every candidate

weight would be methodologically preferable and is a direct target for follow-on work. Finally, the vision branch is represented by its calibrated output statistics rather than by an actual image-processing pipeline; integrating a real convolutional or transformer classifier operating on RSCD or newly collected imagery, and coupling it to a simulator with native visual rendering such as CARLA [23], to a high-fidelity real-time multi-physics digital twin of the kind recently demonstrated for autonomous offroad vehicles [30], or to an instrumented test vehicle, is the logical next step and is discussed further in Section 6.

6. Conclusion

This paper proposed a physics-informed multimodal learning framework for per-wheel tire-force prediction and grip-utilization estimation in electric vehicles, fusing a road-surface friction-class posterior derived from camera imagery with a time-delay neural network encoding of chassis-sensor signals through a lightweight cross-modal FiLM gate, and training the resulting 4,200-parameter network with a composite loss that combines data fidelity with physics-informed penalties anchored to a quasi-static load-transfer model and the Pacejka friction ellipse. Using a hybrid dataset that pairs a validated four-wheel vehicle-dynamics simulator with vision-branch uncertainty calibrated to a real, published road-surface classification benchmark, the proposed model reduced wheel-force RMSE by 22-42% relative to chassis-only and naive multimodal baselines on wet, snow-covered and icy surfaces, achieved the lowest deployable friction-ellipse violation rate among all learned models, and produced more stable predictions than a non-physics-informed ablation when generalizing to withheld low-friction conditions, all while executing in under one microsecond per inference on a single CPU core. These results indicate that combining a lightweight, latency-conscious cross-modal fusion mechanism with physics-informed regularization is a practical and effective route toward safer, more generalizable grip-utilization estimation for real-time electronic stability and traction control in production electric vehicles.

Future work should prioritize, in order, (i) validating the coupling between chassis-dynamics simulation and vision-branch uncertainty against synchronized real-world camera and CAN-bus data from an instrumented test vehicle; (ii) replacing the calibrated vision-branch statistics with an actual convolutional or transformer classifier trained end-to-end on RSCD or newly collected imagery, jointly with the fusion network; (iii) extending the manoeuvre set to include both steering directions and a wider range of combined braking-and-cornering conditions to remove the front-right asymmetry observed in the current results; and (iv) conducting a validation-set-only, multi-seed hyperparameter selection procedure for the physics-loss weight, together with a broader ensemble study of the extrapolation experiment, to establish more precisely the conditions under which physics-informed regularization yields a statistically significant generalization advantage over cross-modal fusion alone.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The author declares no conflict of interest.

Ethical Approval

This study did not involve human participants, human data, animals, or any personally identifiable information. All chassis-sensor and force data were generated by a physics-based vehicle-dynamics simulator (Section 3.2.1), and the vision branch was calibrated to publicly reported, de-identified classification-accuracy statistics [8] rather than to any newly collected imagery. Ethical approval was therefore not required for this study.

Consent to Participate and Consent for Publication

Not applicable.

Data Availability Statement

Two categories of data underlie this study. First, the vision branch's simulated classification accuracy and confusion structure (Section 3.2.2) were calibrated to the publicly available Road Surface Classification Dataset (RSCD) of Zhao and Wei [9], accessible via Mendeley Data at <https://doi.org/10.17632/w86hvkrczc5.2>, with a GitHub mirror at https://github.com/ztsrxh/RSCD-Road_Surface_Classification_Dataset; RSCD imagery itself was not processed by the pipeline reported here (Section 5.4). Second, the synthetic multimodal dataset of 9,000 chassis-vision samples generated for this study (Section 3.2.3), together with the trained-model artefacts and all result tables reported in Section 4, was produced by the vehicle-dynamics simulator and processing pipeline described in Section 3 and is provided in full, alongside its generating source code, as the supplementary code package accompanying this article. The dataset can be regenerated exactly as reported by running that code, and is additionally available from the corresponding author upon reasonable request.

Author Contributions

T. Akande: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review and editing, Visualization.

References

- [1] International Energy Agency (IEA), “Global EV Outlook 2026,” IEA, Paris, France, 2026. [Online]. Available: <https://www.iea.org/reports/global-ev-outlook-2026>
- [2] Z. Huang and X. Fan, “A review on estimation of vehicle tyre-road friction,” *Int. J. Heavy Veh. Syst.*, vol. 31, no. 1, pp. 49–86, 2024, doi: 10.1504/IJHVS.2024.136242.
- [3] H. Qian, M. Wang, M. Zhu, and H. Wang, “A review of multi-sensor fusion in autonomous driving,” *Sensors*, vol. 25, no. 19, art. 6033, 2025, doi: 10.3390/s25196033.
- [4] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045.

- [5] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021, doi: 10.1038/s42254-021-00314-5.
- [6] H. Lim, J. W. Lee, J. Boyack, and J. B. Choi, “EV-PINN: A physics-informed neural network for predicting electric vehicle dynamics,” in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, Atlanta, GA, USA, 2025, pp. 1–7.
- [7] T. Wang, T. Zeng, Z. Zeng, F. Zhang, Y. Wang, X. Li, Y. Xu, S. Chen, J. Jiao, C. Claudel, and X. Chen, “DBPnet: Damper characteristics-based Bayesian physics-informed neural network for wheel load estimation,” *arXiv:2605.24860*, 2026.
- [8] T. Wang, Z. Zhang, and C. Sun, “RoadFormer: Local-global feature fusion for road surface classification in autonomous driving,” *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3635121.
- [9] T. Zhao and Y. Wei, “A road surface image dataset with detailed annotations for driving assistance applications,” *Data Brief*, vol. 42, art. 108483, 2022, doi: 10.1016/j.dib.2022.108483.
- [10] J. Huang, X. Chen, Q. Jin, and P. Li, “A fusion estimation method for tire-road friction coefficient based on weather and road images,” *Lubricants*, vol. 13, no. 10, art. 459, 2025, doi: 10.3390/lubricants13100459.
- [11] S. Zhao, J. Zhang, Y. Jiang, C. He, and J. Han, “Tire-road friction coefficients adaptive estimation through image and vehicle dynamics integration,” *Mech. Syst. Signal Process.*, vol. 224, art. 112039, 2025, doi: 10.1016/j.ymssp.2024.112039.
- [12] Z. Zhang, L. Zheng, H. Wu, Z. Zhang, Y. Li, and Y. Liang, “An estimation scheme of road friction coefficient based on novel tyre model and improved SCKF,” *Veh. Syst. Dyn.*, vol. 60, no. 8, pp. 2775–2804, 2022, doi: 10.1080/00423114.2021.1928247.
- [13] J. Ye, Z. Zhang, J. Jin, R. Su, and B. Huang, “Estimation of tire-road friction coefficient with adaptive tire stiffness based on RC-SCKF,” *Nonlinear Dyn.*, vol. 112, pp. 945–960, 2024, doi: 10.1007/s11071-023-09088-0.
- [14] S. Tao, Z. Ju, L. Li, H. Zhang, and W. Pedrycz, “Tire road friction coefficient estimation for individual wheel based on two robust PMI observers and a multilayer perceptron,” *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 12530–12541, 2024, doi: 10.1109/TVT.2024.3390032.
- [15] Y. Wang, G. Yin, P. Hang, J. Zhao, Y. Lin, and C. Huang, “Fundamental estimation for tire road friction coefficient: A model-based learning framework,” *IEEE Trans. Veh. Technol.*, vol. 74, no. 1, pp. 481–493, 2024, doi: 10.1109/TVT.2024.3464524.
- [16] C. Wang, Z. Wang, Z. Zhang, J. Liu, W. Li, Y. Wu, X. Li, H. Yu, and D. Cao, “Integrated post-impact planning and active safety control for autonomous vehicles,” *IEEE Trans. Intell. Veh.*, vol. 8, no. 3, pp. 2062–2076, 2023, doi: 10.1109/TIV.2023.3236150.
- [17] Z. Quan, B. Li, S. Bei, X. Sun, N. Xu, and T. Gu, “Tire-road friction coefficient estimation method design for intelligent tires equipped with PVDF piezoelectric film sensors,” *Sens. Actuators A, Phys.*, vol. 349, art. 114007, 2023, doi: 10.1016/j.sna.2022.114007.
- [18] Y. Han, Y. Lu, J. Liu, and J. Zhang, “Research on tire/road peak friction coefficient estimation considering effective contact characteristics between tire and three-dimensional road surface,” *Machines*, vol. 10, no. 8, art. 614, 2022, doi: 10.3390/machines10080614.
- [19] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, “FiLM: Visual reasoning with a general conditioning layer,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, 2018, doi: 10.1609/aaai.v32i1.11671.

- [20] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 328–339, 1989, doi: 10.1109/29.21701.
- [21] E. Bakker, L. Nyborg, and H. B. Pacejka, "Tyre modelling for use in vehicle dynamics studies," *SAE Tech. Paper 870421*, 1987, doi: 10.4271/870421.
- [22] H. B. Pacejka and E. Bakker, "The magic formula tyre model," *Veh. Syst. Dyn.*, vol. 21, no. S1, pp. 1–18, 1993, doi: 10.1080/00423119208969994.
- [23] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. 1st Annu. Conf. Robot Learning (CoRL), Proc. Mach. Learn. Res.*, vol. 78, 2017, pp. 1–16.
- [24] R. Rajamani, *Vehicle Dynamics and Control*, 2nd ed. New York, NY, USA: Springer, 2012.
- [25] A. Pandharipande, C.-H. Cheng, J. Dauwels, S. Z. Gurbuz, J. Ibanez-Guzman, G. Li, A. Piazzoni, P. Wang, and A. Santra, "Sensing and machine learning for automotive perception: A review," *IEEE Sens. J.*, vol. 23, no. 11, pp. 11097–11115, 2023, doi: 10.1109/JSEN.2023.3262134.
- [26] B. Araújo, J. F. Teixeira, J. Fonseca, R. Cerqueira, and S. C. Beco, "The road to safety: A review of uncertainty and applications to autonomous driving perception," *Entropy*, vol. 26, no. 8, art. 634, 2024, doi: 10.3390/e26080634.
- [27] S. Öngir, E. C. Kaleli, M. Z. Konyar, and H. M. Ertunç, "Vertical force monitoring of racing tires: A novel deep neural network-based estimation method," *Appl. Sci.*, vol. 15, no. 1, art. 123, 2025, doi: 10.3390/app15010123.
- [28] S. Liu, H. Zhang, S. Wang, and X. Zhao, "Hierarchical torque vectoring control strategy of distributed driving electric vehicles considering stability and economy," *Sensors*, vol. 25, no. 13, art. 3933, 2025, doi: 10.3390/s25133933.
- [29] R. Cordova-Cardenas, D. Amor, and Á. Gutiérrez, "Edge AI in practice: A survey and deployment framework for neural networks on embedded systems," *Electronics*, vol. 14, no. 24, art. 4877, 2025, doi: 10.3390/electronics14244877.
- [30] M. Lehto, T. Lindbäck, H. Lideskog, and M. Karlberg, "Autonomous offroad vehicle real-time multi-physics digital twin: Modeling and validation," *Machines*, vol. 14, no. 1, art. 128, 2026, doi: 10.3390/machines14010128.