Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.079056

ARTICLE

Road Surface Classification Using IMU Data Based on the CGB-Net Deep Learning Architecture

Duong Do The[1,2], Duc-Nghia Tran[3], Hoang-Dieu Vu[4], Manh-Tuyen Vi[4,*] and Duc-Tan Tran[4,*]

[1]Academy of Policy and Development, Hanoi City, Vietnam
[2]Graduate University of Sciences and Technology, Vietnam Academy of Science and Technology, Hanoi City, Vietnam
[3]Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi City, Vietnam
[4]Faculty of Electrical and Electronic Engineering, Phenikaa University, Hanoi City, Vietnam

*Corresponding Authors: Manh-Tuyen Vi. Email: tuyen.vimanh@phenikaa-uni.edu.vn; Duc-Tan Tran. Email: tan.tranduc@phenikaa-uni.edu.vn

Received: 13 January 2026; Accepted: 13 April 2026; Published: 08 May 2026

ABSTRACT: Road-surface identification is important for transportation monitoring and maintenance. However, this task is challenging due to the complexity of vibration signals, feature overlap among different surface types, and variations in real-world operating conditions. These challenges become more significant in time-series classification, where models must achieve high accuracy while remaining computationally efficient and suitable for low-cost hardware. This study investigates the design and evaluation of an automatic road-surface classification system using motion data collected from inertial sensors mounted on a vehicle, including accelerometers and gyroscopes. The system segments synchronized IMU signals into fixed-length windows and assigns each window to a predefined road-surface category. To address this problem, this study proposes CGB-Net, a lightweight and efficient deep learning architecture for road-surface classification. In this architecture, the integration of modules is designed to capture hierarchical features: one-dimensional convolutional neural networks (1D-CNN) are used to extract local temporal features from accelerometer and gyroscope signals, Gated Recurrent Units (GRU) model short-term temporal dependencies, and Bidirectional Long Short-Term Memory (Bi-LSTM) networks capture global long-term temporal context in both directions. The model is trained to distinguish among three types of road surfaces representing different material properties and degradation states: Asphalt_10 (new asphalt, less than 10 years old), Asphalt_15 (aged asphalt, more than 15 years old), and Concrete. Experimental results show that the proposed system achieves high classification performance, with both accuracy and F1-score exceeding 95%. These results indicate strong potential for practical applications in automated road monitoring and maintenance systems.

KEYWORDS: Road surface classification; deep learning; inertial sensors; accelerometer and gyroscope; CGB-Net; one-dimensional convolutional neural networks; recurrent neural networks; time-series analysis

1 Introduction

Road surface quality is one of the key determinants of traffic safety, ride comfort, and vehicle operating costs [1]. Over time, under the combined effects of repeated traffic loading and environmental conditions, pavement structures tend to deteriorate and exhibit various forms of damage such as cracking, potholes, rutting, and structural degradation. These changes not only increase vehicle-induced vibrations but also degrade service quality and driving experience [2–4]. At the network level, road construction and maintenance require substantial financial resources—particularly in low- and middle-income countries—while

empirical evidence indicates that rough or severely rutted pavements are associated with higher crash risk and increased accident severity [5–7].

Given the significant impact of road surface conditions, frequent and up-to-date monitoring has become an essential requirement for supporting proactive maintenance planning and rational resource allocation. However, achieving continuous and large-scale monitoring requires an assessment approach that can be reliably deployed in real-world environments, remains cost-effective, and enables standardized comparisons across road segments and survey campaigns.

Traditionally, pavement conditions have been assessed through field surveys and visual inspections, with results summarized using standardized indices such as the Pavement Condition Index (PCI), which reflects the overall condition and damage level of pavement segments. Despite their widespread use, manual survey methods are labor-intensive and time-consuming, difficult to conduct frequently with dense spatial coverage over large road networks, and susceptible to organizational constraints and subjective judgment by inspectors [8–10]. Consequently, automated monitoring approaches based on vehicle-mounted sensors and data-driven analysis have attracted increasing attention as scalable and efficient alternatives.

From the perspective of sensing modalities and data acquisition technologies, automated road surface monitoring solutions can generally be classified into three categories.

(i) Vision-based approaches, which utilize images or videos combined with machine learning or deep learning techniques to detect pavement distress or recognize surface types [11–13]. Beyond road-level imagery, some studies have also leveraged high-resolution remote sensing data and deep learning models to extract infrastructure-related information for large-scale road monitoring and analysis [14].

(ii) LiDAR- or laser-scanning-based approaches, which reconstruct 2D or 3D road surface geometry for assessment, mapping, and feature analysis using point clouds [15,16].

Although these two categories provide rich spatial information, they typically require expensive equipment, complex calibration procedures, and substantial computational resources, limiting their applicability in cost-sensitive or large-scale deployments.

(iii) Vibration-based approaches, which infer road surface type and degradation level from vehicle motion data collected by onboard sensors or mobile devices [17,18].

Among these categories, vibration-based methods using inertial measurement units (IMUs), comprising accelerometers and gyroscopes, are particularly attractive due to their low cost, ease of installation, and suitability for using ordinary vehicles as probe vehicles for large-scale data collection. Numerous studies have demonstrated the feasibility of low-cost or embedded sensors for pavement monitoring [9], low-cost IMU-based systems for urban roads [19,20], and even two-wheeled vehicles or scooters as data collection platforms [21]. Nevertheless, vibration signals are highly sensitive to driving conditions: speed, driving maneuvers, suspension systems, and sensor placement can significantly affect signal characteristics, posing challenges to model robustness in real-world deployment [22–24].

From an algorithmic perspective, early studies primarily employed rule-based or threshold-based methods to detect anomalies such as potholes or speed bumps. While computationally lightweight, these approaches are highly sensitive to operating conditions and often require manual threshold tuning for each deployment scenario; moreover, they typically address binary anomaly detection rather than multi-class road surface classification [25–27]. Subsequently, traditional machine learning methods were introduced, relying on handcrafted features extracted from time-domain, frequency-domain, or time–frequency representations and classified using Support Vector Machines, Decision Trees, and related models. Several studies reported promising results when carefully designed features were employed [28–30]. However, these approaches

remain heavily dependent on feature selection, window configuration, and generalization capability across different vehicles, road conditions, and sensing setups [31].

More recently, deep learning has gained increasing attention due to its ability to learn features directly from raw time-series signals, reducing reliance on handcrafted features and providing improved temporal modeling capabilities. CNN- and RNN-based architectures, including LSTM and GRU variants, have been applied to road surface classification and demonstrated competitive performance, particularly in multi-class scenarios [32,33]. However, selecting an appropriate model architecture for road-induced vibration signals remains nontrivial: such signals exhibit both short impulsive events and longer oscillatory trends within time windows. Moreover, deep learning models may impose significant computational demands, which must be carefully considered when targeting future low-cost deployment [24,34].

Based on the above analysis, IMU-based road surface classification requires a modeling framework capable of: (i) extracting local and discriminative patterns from vibration signals, such as short impulses, abrupt amplitude changes, or repetitive patterns related to surface roughness and material properties; (ii) learning and modeling temporal relationships at multiple scales to capture both instantaneous variations and longer-term contextual trends within each signal window, thereby improving discrimination between overlapping classes; and (iii) achieving high classification performance while maintaining potential deployability under computational and memory constraints, as the long-term goal of road monitoring systems is large-scale, low-cost, and easily deployable operation.

In this work, road-surface classification is formulated through a supervised three-class time-series recognition framework, which serves as the basis for designing and evaluating an automated monitoring system using the CGB-Net architecture. Given a fixed-length window of synchronized multi-axis IMU signals—comprising tri-axial acceleration and tri-axial angular velocity—the objective is to systematically identify pavement types and degradation states rather than merely detecting isolated anomalies. Within the proposed CGB-Net, one-dimensional convolutional neural networks (1D-CNN) extract local temporal patterns, Gated Recurrent Units (GRU) model short-term temporal dependencies, and Bidirectional Long Short-Term Memory (Bi-LSTM) networks capture global long-term temporal context in both forward and backward directions. The model is trained to distinguish among three categories: Asphalt_10 (new asphalt, less than 10 years old), Asphalt_15 (aged asphalt, more than 15 years old), and Concrete. These classes are selected to represent distinct material compositions and structural integrity, utilizing pavement service age as a practical proxy for variations in surface roughness and macro-texture—factors that significantly influence the vibration characteristics captured by the vehicle-mounted sensors.

In addition, asphalt surfaces are differentiated by service age because this material is more susceptible to climate-induced oxidation and surface wear, which tend to evolve relatively rapidly over time. In contrast, concrete pavements possess substantially higher structural durability; therefore, their vibration characteristics change much more slowly over time compared with asphalt surfaces. Moreover, the concrete road segments examined in this study exhibited high uniformity in the IMU signals recorded at the time the experiments were conducted.

The main challenges encountered in the development of the road surface monitoring system are summarized as follows.

- Challenge 1: Signal complexity and class overlap in road-induced vibrations. IMU signals generated by road–vehicle interaction are highly complex and may exhibit substantial overlap among surface types, particularly when distinguishing asphalt pavements with different aging levels (Asphalt_10 vs. Asphalt_15). This overlap is further exacerbated under real driving conditions, where speed, vehicle dynamics, and road texture variability influence vibration signatures.

- Challenge 2: Effective exploitation of multi-axis IMU data (accelerometer and gyroscope). Accelerometer and gyroscope channels provide complementary motion information; however, effectively leveraging both modalities requires a model capable of learning discriminative representations from multi-axis time-series data while remaining robust to variations in operating conditions, sensor placement, and driving behavior.
- Challenge 3: Multi-level temporal dependency modeling. Road surface characteristics are reflected not only in instantaneous vibration peaks but also in their temporal evolution within a signal window. Accurate classification therefore requires modeling both short-term transients and long-term contextual patterns—an aspect insufficiently addressed by purely feature-based or single-stage models.
- Challenge 4: Balancing performance and complexity for practical deployment. Although deep learning can deliver high classification accuracy, computational and memory costs may become limiting factors for deployment on low-cost or resource-constrained platforms. Beyond performance evaluation, it is therefore necessary to assess whether a deep sequential architecture can achieve strong classification quality while maintaining reasonable complexity for deployment-oriented development.

From these challenges, it is evident that an effective road surface classification system must be sufficiently powerful to learn complex vibration patterns and separate overlapping classes, while also exploiting the complementary information from accelerometer and gyroscope signals and maintaining practical computational complexity. The main contributions of this paper are summarized as follows:

- Contribution 1: This study establishes a vehicle-mounted IMU-based road-surface classification framework using synchronized accelerometer and gyroscope signals collected under realistic driving conditions. The framework addresses a three-class recognition task involving Asphalt_10, Asphalt_15, and Concrete, thereby enabling the identification of pavement material and degradation state from road-induced vibration data.
- Contribution 2: This study proposes CGB-Net, a hierarchical deep learning architecture specifically designed for vibration-based road-surface classification. By integrating 1D-CNN, GRU, and Bi-LSTM in a unified framework, the proposed model is able to capture local temporal patterns, short-term dependencies, and long-range bidirectional contextual information, thereby improving discrimination between surface classes with highly overlapping vibration characteristics.
- Contribution 3: A comprehensive experimental evaluation is conducted through comparisons with multiple baseline and ablation models, repeated trials using different random seeds, and testing on an independent dataset. The results demonstrate that CGB-Net achieves high and stable classification performance, with an average accuracy of 95.07% and an F1-score of 95.03%, while maintaining a compact architecture with 9600 parameters and a Memory footprint of 90 kB, highlighting its potential for deployment in low-cost road-monitoring systems.

The remainder of this paper is organized as follows. Section 1 introduces the research background, challenges, and main contributions. Section 2 presents the IMU data (accelerometer and gyroscope), signal processing pipeline, the CGB-Net model for three-class road surface classification, describes experimental setup and evaluation metrics. Section 3 reports and analyzes the results. Section 4 discusses the findings and compares them with related studies. Section 5 concludes the paper and outlines directions for future research.

2 Materials and Methods

2.1 IoT-Based Vehicle-Mounted IMU Road-Surface Monitoring System

The system proposed in this study targets the problem of classifying three types of road surfaces based on motion data collected from a vehicle-mounted inertial measurement unit (IMU), including an accelerometer

and a gyroscope. Fig. 1 illustrates the overall system architecture, in which the workflow is implemented through two main modules: an on-vehicle IMU data acquisition module and a road-surface processing and classification module that operates on the acquired signal sequences.

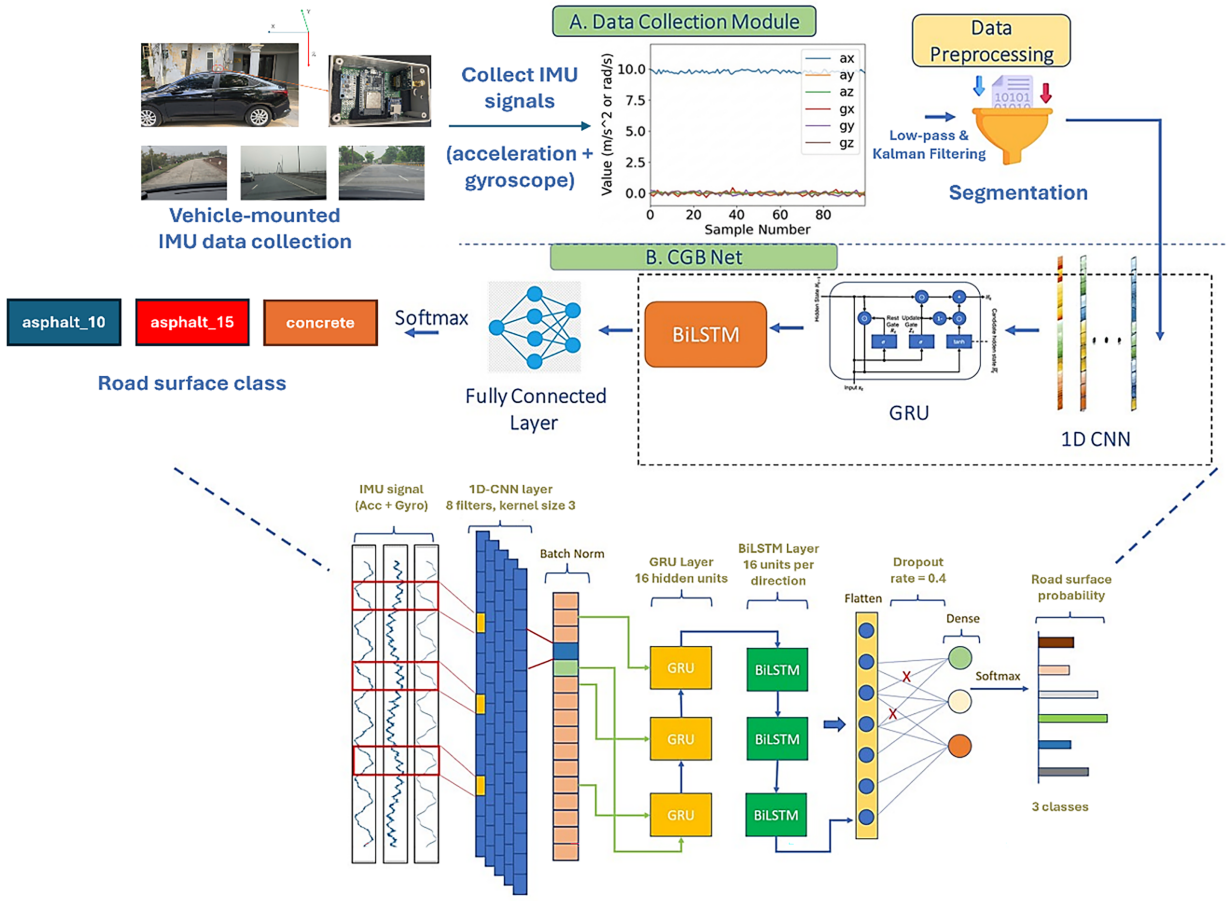

Figure 1: Overview of the proposed IMU-based road-surface classification framework: (**A**) Vehicle-mounted IMU sensing and preprocessing pipeline. (**B**) CGB-Net classifier (1D-CNN–GRU–Bi-LSTM) for three-class road-surface identification.

During the data collection stage, the IMU is rigidly mounted on the vehicle to continuously record IMU time-series signals consisting of tri-axial acceleration and tri-axial angular velocity while the vehicle travels over road segments belonging to three target surface categories: Asphalt_10, Asphalt_15, and Concrete.

Road-surface types can be differentiated using IMU data because variations in pavement structural properties and surface texture lead to distinct vehicle–suspension dynamic responses. Variations in acceleration reflect the impact forces transmitted from the road surface to the vehicle body [35]. Asphalt_10 typically maintains a relatively smooth macro-texture with minimal structural degradation, resulting in stable vibrations with low amplitude. In contrast, Asphalt_15 exhibits increased micro-roughness and bitumen degradation, which act as high-frequency excitation sources and significantly increase the variance of vertical acceleration.

Furthermore, angular velocity signals reflect the rotational response of the vehicle to surface irregularities. For example, concrete pavements often induce characteristic periodic oscillations in pitch and roll due to the presence of joints between concrete slabs [36]. In contrast, aged asphalt surfaces tend to produce more

stochastic fluctuations in gyroscope signals, caused by non-uniform material degradation and longitudinal rutting along the direction of travel.

The collected data is subsequently preprocessed and segmented into fixed-length time windows before being fed into the classification module for model training and performance evaluation.

The IoT-based data acquisition device is designed to continuously record vehicle motion data for the analysis and assessment of road-surface conditions. Fig. 2 illustrates the overall architecture of the device. The system is powered directly by the vehicle battery, with an ESP32 microcontroller serving as the central processing unit and communicating with a backend server via wireless Bluetooth or Wi-Fi connectivity. Peripheral hardware components, including the GPS module, inertial sensor, and local storage, are directly interfaced with the ESP32 microcontroller.

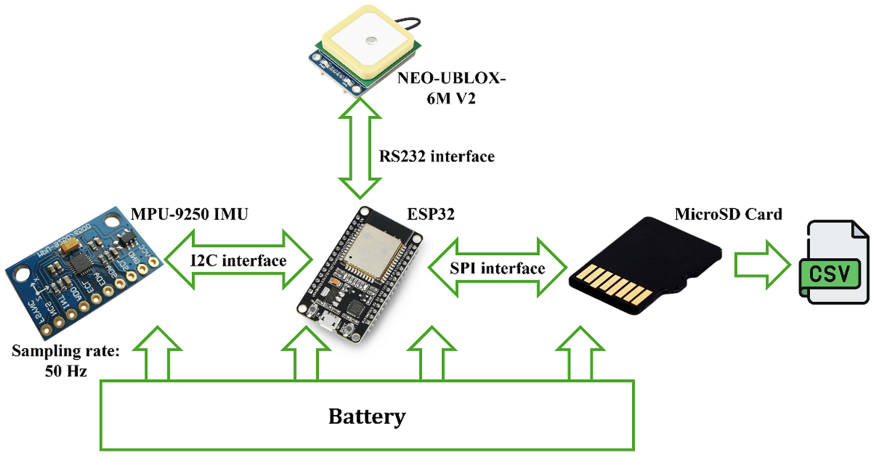

Figure 2: IoT-based road-surface data acquisition device.

When network connectivity is available, the ESP32 can transmit data to the server via Wi-Fi or Bluetooth Low Energy (BLE) for storage, remote monitoring, and analysis, while also supporting updates to digital maps. To ensure accurate labeling of the motion data, the research team integrated and synchronized dash-cam video during data collection. This synchronization provides visual ground truth for the corresponding road segments, thereby improving the reliability of the dataset used for training and testing the road-surface classification model.

The peripheral modules include a NEO-UBLOX-6M V2 GPS receiver, an MPU-9250 inertial sensor, and a microSD memory card, all connected to the ESP32 through standard serial interfaces such as UART and I^2C/SPI. The GPS module operates at 3.3 V and integrates a ceramic antenna, EEPROM for configuration storage, and a backup battery; it provides positioning, speed, and timestamp information. The MPU-9250 is a 9-axis IMU (3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer) with typical acceleration ranges of ±2 g/±4 g/±8 g/±16 g and angular velocity ranges of ±250/±500/±1000/±2000°/s, making it suitable for capturing vibrations, short impacts, and vehicle body oscillations.

During data acquisition, the MPU-9250 measures raw acceleration and angular velocity along the X, Y, and Z axes at a sampling frequency of 50 Hz per axis. The IMU data are transmitted to the ESP32 via I^2C/SPI, after which the 32-bit microcontroller performs real-time preprocessing, including low-pass filtering to remove high-frequency noise, gravity compensation on acceleration channels when necessary, and coordinate-frame normalization. The ESP32 then packages the data into time frames and continuously writes them to the microSD card to minimize data loss in the event of network interruptions.

Field experiments were conducted on urban and suburban roads in Hanoi, Vietnam, where the tropical monsoon climate induces substantial variations in temperature and humidity, contributing to pavement deterioration. To ensure that the collected inertial signals reflect a representative spectrum of pavement conditions, the experimental route was designed to traverse road segments with clearly different service ages and structural configurations. Specifically, the route began on Hoang Quoc Viet Street (most recently renovated in 2024), continued along the elevated Ring Road 3 section from Mai Dich to Nam Thang Long (opened to traffic in 2020), then proceeded to Thang Long Boulevard (opened in 2010), followed by a concrete roadway segment on the Red River dike, and finally crossed Thanh Tri Bridge (opened in 2007). This combination provides pavement segments younger than 10 years, older than 15 years, and concrete pavement within a single continuous trajectory, thereby enabling controlled comparisons between different road-surface types under consistent vehicle operating conditions.

The test vehicle was a compact sedan operating under normal traffic conditions. The device, with dimensions of 85 mm × 35 mm × 35 mm, was mounted near the center of the vehicle roof, as shown in Fig. 3. This mounting position, close to the midpoint of the wheelbase, was selected to reduce errors caused by mechanical looseness or localized vibrations. Because raw IMU data are often noisy, the processing pipeline also incorporates a Kalman filter to smooth the signals prior to storage and subsequent analysis. Since driving maneuvers such as speed changes, acceleration, braking, or turning can introduce noise into IMU signals, the data collection was conducted following a consistent driving protocol in which the driver maintained a relatively steady speed and avoided abrupt maneuvers whenever possible.

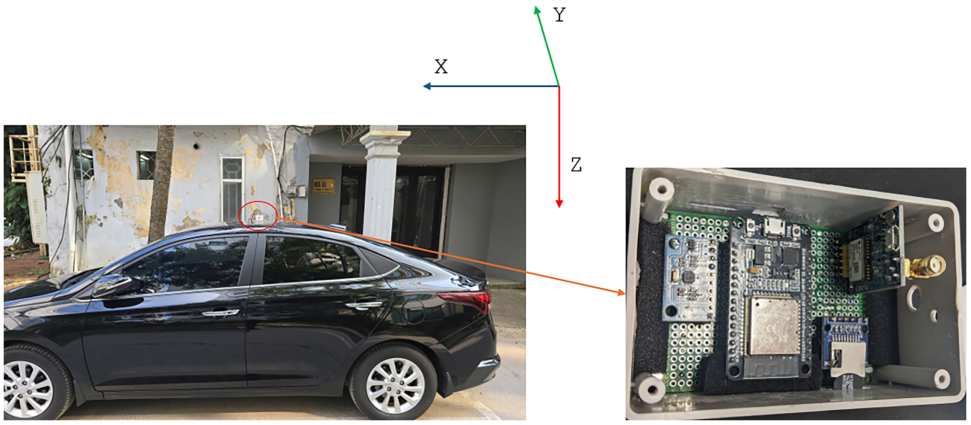

Figure 3: Mounting location of the data acquisition device on the vehicle.

2.2 Data Collection

Table 1 presents detailed information on the label classes used in this study. The labeling process was conducted following a consistent cross-verification procedure to ensure reliability. In addition, data windows corresponding to intersections, lane changes, or U-turn maneuvers were removed through video inspection to ensure that the IMU signals primarily reflect the interaction between the vehicle and the road surface. This procedure helps ensure high-quality training and testing datasets while minimizing labeling bias.

Table 1: Description of road-surface classes and label definitions.

Road-Surface Label	Explain
Asphalt_10	Asphalt pavement with a service life of less than 10 years.
Asphalt_15	Asphalt pavement with a service life of more than 15 years.
Concrete	Cement concrete pavement (structurally and materially different from asphalt).

Each survey route corresponding to a specific road-surface type was assigned a unique identifier to facilitate data management and traceability. During data acquisition, the inertial sensor simultaneously recorded tri-axial acceleration (*Ax, Ay, Az*) and tri-axial angular velocity (*Gx, Gy, Gz*) at a sampling frequency of 50 Hz for each axis. In addition to IMU data, the dataset includes auxiliary information fields such as GPS coordinates (latitude and longitude), altitude (Alt), vehicle speed (Speed), number of satellites (Sats), and the road-surface label (ROADSURFACE) assigned during the survey process.

The sampling frequency of 50 Hz was selected to ensure adequate capture of rapid oscillations and short-term vibrations of the vehicle body as it traverses surface irregularities characteristic of each road-surface type. Table 2 illustrates the accelerometer and gyroscope data fields collected in this study.

Table 2: Example IMU signals (Accelerometer and Gyroscope) on Asphalt_15.

Asphalt_15					
Ax	***Ay***	***Az***	***Gx***	***Gy***	***Gz***
−0.08	−0.1	−0.98	2.87	−6.04	0.12
−0.1	−0.01	−0.97	4.03	−7.26	−0.24
−0.11	−0.03	−0.97	3.48	−8.18	0.06
0	0.08	−1.09	2.44	−5.55	−0.73
−0.01	0.11	−0.91	0.61	−5.43	−0.61
0.03	0.11	−1.01	1.16	−3.3	−0.49
0.03	0.11	−0.97	0.92	−4.64	−0.85
0.13	0.06	−1.09	1.22	0.73	−0.24
−0.08	0.02	−0.9	1.34	0.43	−0.79
−0.04	0.03	−0.87	1.04	5	−0.31

The accelerometer and gyroscope data collected from the vehicle-mounted device were temporally synchronized to construct the dataset for classification. A time-based sliding window approach was applied. With a sampling frequency of 50 Hz, window lengths ranging from 1.0 to 5.0 s were investigated. In addition, the overlap ratio between consecutive windows was examined with values ranging from 20% to 80%. For example, with a 5-s window and a 40% overlap, each window contains 250 samples per axis, and the sliding step is 3 s (i.e., 150 samples per axis). Other window configurations were implemented in a similar manner to evaluate the effects of window length and overlap ratio on classification performance. After window generation, the dataset was split into training and testing sets using an 80%/20% ratio. Tables 3 and 4 report the number of windows for each road-surface class in the training and testing sets, respectively, for the configuration with a 5-s window and 40% overlap. Furthermore, to comprehensively validate the accuracy and robustness of the proposed model, additional data were collected along a different driving route. This supplementary dataset was acquired along road corridors connecting Hanoi and Thai Nguyen

City, Vietnam at different time periods, incorporating natural variations in environmental conditions in order to better reflect real-world driving scenarios. Notably, the road surface characteristics in this route share a certain degree of similarity with the previous dataset. Table 5 summarizes the number of windows for each road-surface class in this Independent Test Dataset using the same 5-s window and 40% overlap configuration.

Table 3: Distribution of samples by class in the training dataset (5 s window, 40% overlap).

Road Surface	Total Observations
Asphalt_10	14,113
Asphalt_15	15,347
Concrete	10,423
Total	39,883

Table 4: Distribution of samples by class in the testing dataset (5 s window, 40% overlap).

Road Surface	Total Observations
Asphalt_10	3386
Asphalt_15	3885
Concrete	2576
Total	9847

Table 5: Distribution of samples by class in the independent test dataset (5-s window, 40% overlap).

Road Surface	Total Observations
Asphalt_10	2105
Asphalt_15	2253
Concrete	1674
Total	6032

2.3 IMU Signal Processing and Deep Learning–Based Classification Model

The IMU signal sequences collected from the vehicle-mounted device—including tri-axial acceleration (*Ax, Ay, Az*) and tri-axial angular velocity (*Gx, Gy, Gz*)—were subjected to a structured preprocessing pipeline to enhance data reliability and consistency prior to training the classification model. First, all data were organized according to road-surface labels (Asphalt_10, Asphalt_15, and Concrete) and split into training and testing sets, with the testing set kept independent to ensure objective evaluation. Through empirical analysis, a Train/Test split of 80%/20% was found to yield the best overall performance.

Next, a time-based sliding window approach was employed to transform continuous IMU time-series data into fixed-size samples. Specifically, with a sampling frequency of approximately 50 Hz, the signals were segmented into windows of length T seconds (examined in the range of 1–5 s), while the overlap ratio between consecutive windows was varied from 20% to 80%. This segmentation strategy enables the model to learn characteristic vibration patterns within each temporal interval, increases the number of training samples, and improves the ability to model signal variability induced by real-world operating conditions (speed changes, vehicle body vibrations, and surface transitions).

To reduce the influence of transient driving maneuvers on IMU signals, a window-level filtering step was applied prior to model training and evaluation, based on GPS speed information and gyroscope signals. First, data windows corresponding to near-stationary vehicle conditions were removed using a minimum speed threshold in order to avoid noise caused by engine vibrations during vehicle start-up or stopping. Next, windows containing strong acceleration or braking phases were detected using longitudinal acceleration estimated from the derivative of the GPS speed and were discarded if the value exceeded a predefined threshold. Finally, windows with significant rotational motion during cornering were identified using the angular velocity around the *Gz* axis and were also removed. After this filtering step, the remaining data windows were processed according to the previously described preprocessing pipeline, including low-pass filtering, Kalman filter–based smoothing, and coordinate system normalization.

From a signal processing perspective, the CGB-Net model is designed as a sequence of transformations that progressively convert raw IMU data into discriminative representations for road-surface classification. Each IMU signal window can be represented as a multi-channel sequence $X \in \mathrm{R}^{T \times C}$, where T denotes the number of temporal samples within the window and C denotes the number of sensor channels (including the accelerometer and gyroscope axes).

The vibration signals measured by the IMU typically contain multiple components, including short-duration oscillations caused by localized road irregularities, mid-term vibrations associated with vehicle dynamics, and longer-term trends reflecting the overall roughness characteristics of the pavement surface. Therefore, the signal processing pipeline must progressively emphasize informative vibration patterns while reducing the influence of noise and random fluctuations in the data.

In CGB-Net, the preprocessed IMU signals are transformed through several consecutive processing layers, each designed to capture different aspects of the vibration sequence. This hierarchical transformation gradually converts the raw sensor data into more stable feature representations, enabling the model to highlight subtle differences between road-surface classes that exhibit relatively similar vibration patterns.

The preprocessed IMU windows (in multi-channel sequence form) are then fed into the CGB-Net architecture, as illustrated in Fig. 1. The architecture consists of three main components—1D-CNN, GRU, and Bi-LSTM—designed to jointly exploit local vibration patterns and temporal dependencies within the signals. The 1D-CNN block learns local temporal patterns, such as short vibration impulses, abrupt amplitude variations, or short-term repetitive oscillations, which often manifest differently across road-surface types. To capture temporal dependencies beyond the receptive field of the CNN, a GRU block is subsequently employed to model short- to mid-term temporal relationships with relatively low computational complexity. Finally, the Bi-LSTM block enhances contextual modeling by processing the sequence in both forward and backward directions, allowing the model to better exploit long-term temporal structures and subtle transitions within each signal window—an aspect particularly important for distinguishing overlapping classes such as Asphalt_10 and Asphalt_15.

To improve training stability, Batch Normalization layers were integrated after the CNN blocks to reduce internal covariate shift, accelerate convergence, and mitigate unstable gradients. The output of the sequential modeling blocks is then passed through classification layers to predict the probability of each of the three road-surface classes.

2.4 Experimental Design

To objectively evaluate the performance of CGB-Net in the three-class road-surface classification task (Asphalt_10, Asphalt_15, Concrete), this study implements a set of comparative models consisting of

commonly used time-series classification architectures and several controlled ablation variants to clarify the contribution of each component in CGB-Net.

For the baseline time-series models, LSTM is selected as a representative recurrent architecture and used as a reference benchmark to assess the benefits of incorporating 1D-CNN layers at the input stage for local feature extraction prior to temporal modeling. In addition, a CNN–LSTM architecture is employed, which combines 1D-CNN for local pattern extraction with LSTM for learning long-term dependencies; Batch Normalization is placed after the convolutional layers to stabilize training.

For the standard CGB-Net architecture and its ablation variants, all models are constructed to maintain consistency in input/output formats and training procedures to ensure fair comparison. The variants differ only in whether the GRU and/or Bi-LSTM components are retained or removed:

- 1D-CNN: A convolution-only variant that removes both GRU and Bi-LSTM. The architecture consists of two Conv1D layers with 8 and 16 filters, respectively, each followed by normalization (when enabled) and Dropout = 0.4. This is followed by a compact dense layer (16 units) and a 3-class softmax layer. This model learns only local patterns and does not explicitly model long-term temporal dependencies.
- 1D-CNN + GRU: This variant retains the GRU while removing the Bi-LSTM to isolate the effect of unidirectional recurrent modeling. The model begins with a Conv1D layer (8 filters) and normalization, followed by pooling/aggregation before feeding into a GRU layer, and finally a 3-class classification layer.
- 1D-CNN + Bi-LSTM: This variant retains the Bi-LSTM while removing the GRU. After the Conv1D block (8 filters), the sequence is passed through one or more Bi-LSTM layers (stacked according to the design), followed by the 3-class classification head.

The simulation and implementation environment was standardized to ensure reproducibility and fairness across all experiments. All models were implemented in Python using the TensorFlow/Keras framework. Model training and evaluation were performed on a local workstation equipped with an Intel Core i5-12400F processor, an NVIDIA GeForce RTX 2060 GPU with 6 GB VRAM, and 16 GB of RAM. In addition, the same software environment, preprocessing procedure, data partition strategy, and training platform were consistently applied to CGB-Net and all baseline models, so that the comparative results primarily reflect differences in model architecture rather than differences in experimental setup.

2.5 Training Configuration

To train the CGB-Net and its comparative models described in Section 2.4, the hyperparameters were determined through a preliminary empirical tuning process on the training dataset. The primary objective was to balance classification accuracy, convergence stability, and the computational efficiency required for low-cost IoT deployment.

Through a series of initial experiments, the optimal structural hyperparameters for the complete CGB-Net were finalized as follows: a 1D-CNN block with 8 filters (kernel size 3), a GRU layer with 16 hidden units, and a Bi-LSTM layer with 16 units in each direction. Following feature extraction, the network employs Global Average Pooling, a Dropout rate of 0.4, and concludes with a 3-class softmax layer. As outlined in Section 2.4, these core structural parameters were applied consistently across all baseline and ablation models to guarantee a fair comparative analysis.

For the optimization and learning configuration, all models were trained consistently using the Adam optimizer with an initial learning rate of 0.001, a batch size of 128, and a training duration of 100 epochs. To further mitigate overfitting, regularization strategies (such as L1/L2) were applied alongside the dropout mechanism. This selected setting was confirmed during the preliminary stage to provide stable convergence on the road-induced vibration dataset.

To assess robustness with respect to random initialization, each model is trained independently five times using different random seeds. Final results are reported as the mean and standard deviation over the five runs; no seed selection is performed based on performance.

For temporal sample generation, a time-based sliding window approach is adopted rather than a fixed number of raw samples. With a sampling frequency of 50 Hz, window lengths in the range of 1.0–5.0 s are investigated. The overlap ratio between consecutive windows is evaluated at 20%, 40%, 60%, and 80%. After window generation, the dataset is split into training and testing sets using a 80%/20% ratio, as described in the Data Collection section. The distribution of windows by class for the illustrative configuration (5 s window, 40% overlap) is reported in Tables 3 and 4.

2.6 Evaluation Metrics

To evaluate the performance of the road-surface classification system (Asphalt_10, Asphalt_15, and Concrete), this study employs two primary metrics: Accuracy and F1-score. Accuracy represents the proportion of correctly predicted samples over the entire dataset, whereas the F1-score jointly summarizes Precision (class-wise exactness) and Recall (class-wise coverage). The F1-score is particularly suitable for multi-class classification problems in which classes may exhibit overlap and non-uniform distributions.

The metrics are computed using the following formulations:

$$Accuracy = \frac{Number\ of\ Correctly\ predicted\ samples}{Total\ number\ of\ samples} \times 100 \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100 \tag{4}$$

where TP denotes the number of true positives for the class under consideration, FP is the number of samples incorrectly predicted as that class, and FN is the number of samples belonging to that class but misclassified as other classes.

In addition to these aggregate metrics, a confusion matrix is utilized to provide a detailed analysis of misclassification patterns between classes—particularly the confusion between Asphalt_10 and Asphalt_15, which tend to exhibit overlapping vibration characteristics.

Beyond accuracy-oriented evaluation, practical deployment on resource-constrained IoT microcontrollers necessitates a thorough assessment of computational efficiency. To comprehensively benchmark this, the study evaluates specific deployment-related indicators, notably Floating Point Operations (FLOPs) and Memory footprint, alongside processing time and latency. FLOPs represent the total number of arithmetic operations required for a single forward pass, serving as a theoretical, hardware-independent measure of computational complexity. Memory, in this context, refers to the static storage required to save the trained model's learnable parameters (weights and biases). Together, these metrics complement the accuracy evaluation by ensuring that the selected architecture not only achieves high predictive performance but also meets the strict resource constraints required for low-cost edge deployment. Finally, to verify reliability, model stability is further characterized and reported across repeated experiments.

3 Results and Discussions

To evaluate the classification performance of the CGB-Net deep learning architecture for road-surface monitoring, we conducted a comparative analysis across six models: LSTM, 1D-CNN, and hybrid variants including 1D-CNN + LSTM, 1D-CNN + GRU, 1D-CNN + Bi-LSTM, along with the proposed CGB-Net. The evaluation considers both classification metrics (Accuracy and F1-score) and computational efficiency indicators (number of parameters, FLOPs, and Memory footprint).

3.1 Model Performance and Computational Efficiency

Fig. 4 illustrates the Accuracy and F1-score of the models under four different experimental settings. CGB-Net demonstrates consistently superior performance across all four settings, maintaining a leading position in both Accuracy and F1-score compared with the baseline models. In particular, under Setting 2, CGB-Net achieves peak performance with an accuracy of approximately 95%, while also exhibiting high robustness, as evidenced by a less fluctuating performance curve relative to other models when input configurations change.

Figure 4: Accuracy and F1-score of the models under four different experimental settings.

The hybrid models, including 1D-CNN + Bi-LSTM and 1D-CNN + GRU, form a second performance tier with strong results (exceeding 90% under Setting 2), confirming the effectiveness of combining spatial (local) feature extraction with temporal modeling. However, these models show a slight performance degradation under Settings 3 and 4, indicating certain limitations in generalization when a comprehensive multi-stage temporal modeling mechanism—such as that employed by CGB-Net—is absent.

In contrast, the standalone 1D-CNN model records the lowest performance with substantial variability, dropping sharply to approximately 85% under Setting 3. This pronounced gap highlights that relying solely on convolutional layers for local feature extraction is insufficient to capture the complex temporal dependencies inherent in road-induced vibration signals, thereby underscoring the necessity of integrating dedicated sequential modeling components, as realized in the CGB-Net architecture.

To provide deeper insight into the trade-off between predictive performance and operational cost, Table 6 reports detailed quantitative metrics, including the number of parameters, FLOPs, and Memory

footprint of the evaluated models on the test dataset. To ensure fairness in the comparison, all baseline models were implemented and trained under the same experimental settings as the CGB-Net model. Specifically, all models used the same data preprocessing pipeline, the same training–testing data split, and the same training configurations described in Section 2.5. As a result, the performance differences among the models primarily reflect differences in model architecture rather than factors related to the training procedure.

Table 6: Evaluation of model performance and computational resources (parameters, FLOPs, and memory).

Model	Accuracy (%)	F1-Score (%)	Params	FLOPS (M)	Memory (kB)
LSTM	89.50 ± 1.20	88.10 ± 1.50	3212	0.2806	31
1D-CNN + LSTM	91.20 ± 0.85	90.50 ± 1.10	1916	0.1603	34
1D-CNN	88.40 ± 1.10	87.20 ± 1.30	25,560	0.145	136
1D-CNN + GRU	92.10 ± 0.60	91.50 ± 0.80	16,820	1.48	135
1D-CNN + Bi-LSTM	93.50 ± 0.50	93.10 ± 0.60	21,490	1.98	154
CGB-Net	95.07 ± 0.45	95.03 ± 0.52	9600	1.75	90

The experimental results confirm the overall superiority of CGB-Net, which achieves an average accuracy of 95.07% (±0.45) and an F1-score of 95.03% (±0.52). Notably, despite delivering the highest performance, CGB-Net maintains a compact architecture with only 9600 parameters and a Memory footprint of 90 kB. Compared with its closest competitor, 1D-CNN + Bi-LSTM (accuracy of 93.50%), CGB-Net not only improves performance by approximately 1.5% but also reduces the number of parameters by more than half (from 21,490 parameters) and significantly lowers Memory footprint (from 154 kB). These results demonstrate that the careful integration of GRU and Bi-LSTM blocks enables the model to learn richer representations without unnecessarily increasing network size.

Besides the average improvements in Accuracy and F1-score, the stability of CGB-Net across independent training runs serves as a crucial indicator of the model's advantage. The performance of CGB-Net consistently outpaces the strongest competing baseline (1D-CNN + Bi-LSTM) in both evaluation metrics. When comparing the paired results from each run using a paired statistical test, the observed differences achieve statistical significance with $p < 0.05$. This reinforces the conclusion that the hybrid design (combining GRU and Bi-LSTM) enables the model to more effectively exploit temporal features and robustly reconstruct multi-scale vibration patterns in the IMU signals.

In contrast, the standalone 1D-CNN model exhibits inefficient design for this task. Although it has the largest number of parameters (25,560), it yields the lowest classification performance (88.40%), indicating that much of the model capacity is wasted due to its inability to capture essential temporal dependencies. Meanwhile, lightweight models such as 1D-CNN + LSTM, although computationally efficient (only 1916 parameters and 34 kB of Memory footprint), suffer from a substantial drop in accuracy (approximately 3.8% lower than CGB-Net), making them less suitable for applications that require high reliability.

3.2 Evaluation of Stability across Independent Experimental Trials

The data were collected from a single test vehicle; therefore, it is critically important to ensure that the model's performance does not depend on a particular random data split. Accordingly, this study conducted an evaluation based on five independent experimental runs (Run 1–Run 5). In each run, a different random seed was used to shuffle and re-split the dataset into 80% for training and 20% for testing, enabling an assessment of CGB-Net robustness to variations in the input data. In addition, a sixth evaluation (Run 6) was performed using the independent test dataset. In this setting, the model was trained on the original dataset

and tested directly on the independent test dataset to examine cross-route generalization under naturally varying environmental conditions.

The detailed statistical results are visualized in Fig. 5, where the left chart presents the F1-score and the right chart shows the Accuracy. The yellow bars represent the mean performance of each individual run, while the blue dashed line indicates the overall average performance across all five runs.

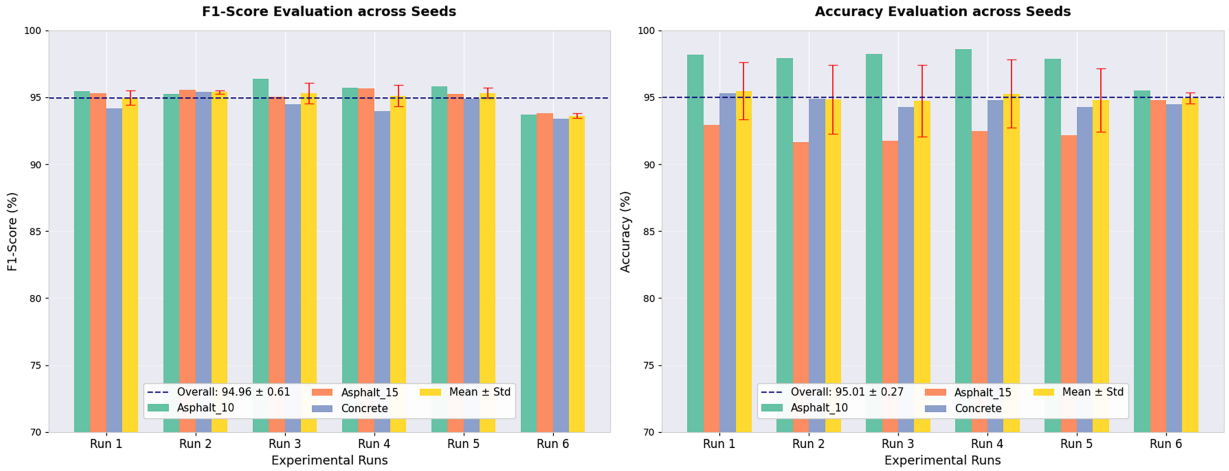

Figure 5: Evaluation of accuracy and F1-score for the three road-surface classes after independent experimental runs.

The results demonstrate that CGB-Net maintains highly stable performance despite changes in the training/testing data partitions (Run 1–5). Specifically, the global average Accuracy and F1-score reach 95.07% (±0.45) and 95.03% (±0.52), respectively. A class-wise analysis further shows that the Asphalt_10 class exhibits the highest stability, while the Asphalt_15 and Concrete classes show only minor and negligible variations across runs. The very low standard deviations provide strong evidence that CGB-Net generalizes well, learning the fundamental characteristics of road surfaces rather than memorizing localized data patterns. In addition, evaluation on the independent test dataset (Run 6) yields an average Accuracy of 94.71% and an average F1-score of 93.63%. While Run 1–5 are evaluated under the original-route distribution, Run 6 introduces a domain shift through new road corridors and varying environmental conditions. The performance remains high in Run 6 as the road surface characteristics in this independent dataset share a certain degree of similarity with the original route. The slightly increased standard deviation reflects the inherent heterogeneity of these real-world driving scenarios.

3.3 Confusion Matrix Evaluation

To comprehensively evaluate the classification capability of the CGB-Net model, this study conducts a confusion matrix analysis on the test dataset. Fig. 6 illustrates the detailed classification results of the three best-performing models: CGB-Net, 1D-CNN + GRU, and 1D-CNN + LSTM.

By examining the elements along the main diagonal (true positives) of the confusion matrices, CGB-Net demonstrates clear superiority in correctly identifying all three classes. Specifically, CGB-Net correctly classifies 1899 samples of Asphalt_10, 1604 samples of Asphalt_15, and 1487 samples of Concrete. Compared with the 1D-CNN + LSTM model, CGB-Net substantially improves recall for the Asphalt_15 class (from 1500 to 1604 samples) and for Asphalt_10 (from 1830 to 1899 samples). These gains indicate that the proposed architecture not only learns general patterns but also captures subtle variations within each road-surface type.

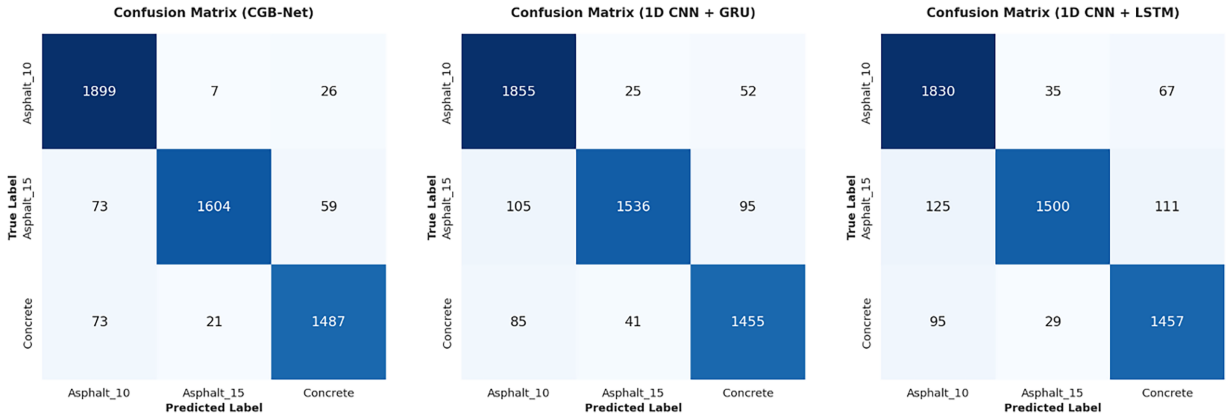

Figure 6: Confusion matrices of the CGB-Net, 1D-CNN + GRU, and 1D-CNN + LSTM models.

The 1D-CNN + LSTM model exhibits the greatest difficulty in separating Asphalt_10 and Asphalt_15. From a physical perspective, vibration signals from these two surfaces exhibit significant overlaps in both frequency content and amplitude, making them prone to confusion by simpler models. In particular, 125 samples that are actually Asphalt_15 are misclassified as Asphalt_10 (false negatives for Asphalt_15). This suggests that the unidirectional gating mechanism of LSTM is insufficient to distinguish pavement aging characteristics, which are often manifested through complex long-term vibration patterns, from newer asphalt surfaces. The 1D-CNN + GRU model partially improves accuracy due to the more efficient gating structure of GRU, reducing the number of misclassified samples to 105; however, the error rate remains non-negligible.

In contrast, CGB-Net achieves markedly superior separation by reducing the number of Asphalt_15 samples misclassified as Asphalt_10 to only 73, representing an error reduction of 41.6% compared with 1D-CNN + LSTM. This improvement is primarily attributed to the Bi-LSTM block in CGB-Net. While the CNN extracts local features (e.g., isolated shock events) and the GRU captures short-term dependencies, the Bi-LSTM enables the model to leverage global temporal context in both directions. As a result, CGB-Net can distinguish whether a vibration sequence arises from the inherently rough and continuously degraded characteristics of aged asphalt (Asphalt_15) or from localized disturbances on newer asphalt (Asphalt_10), thereby significantly reducing false positives between these closely related classes.

For the Concrete class, which is characterized by expansion joints and a rigid surface structure, CGB-Net also demonstrates higher specificity. The number of Concrete samples misclassified as Asphalt_15 is reduced substantially—from 41 samples in the GRU-based model to only 21 samples with CGB-Net—further confirming the robustness of the proposed architecture across all road-surface types.

3.4 Training Dynamics and the Impact of Sequence Length and Overlap Ratio

To evaluate the learning dynamics and generalization capability of the proposed model, Fig. 7 illustrates the evolution of Loss and Accuracy on both the training and validation sets over 100 epochs. During the initial convergence phase (epochs 0–20), CGB-Net exhibits a notably fast convergence rate: the training loss drops sharply from approximately 0.8 to below 0.2, while accuracy increases rapidly from around 70% to above 90%. This rapid improvement indicates that the hybrid architecture—supported by Batch Normalization layers following the CNN block—facilitates effective gradient propagation, enabling the model to quickly capture salient road-induced vibration patterns in the early training stages.

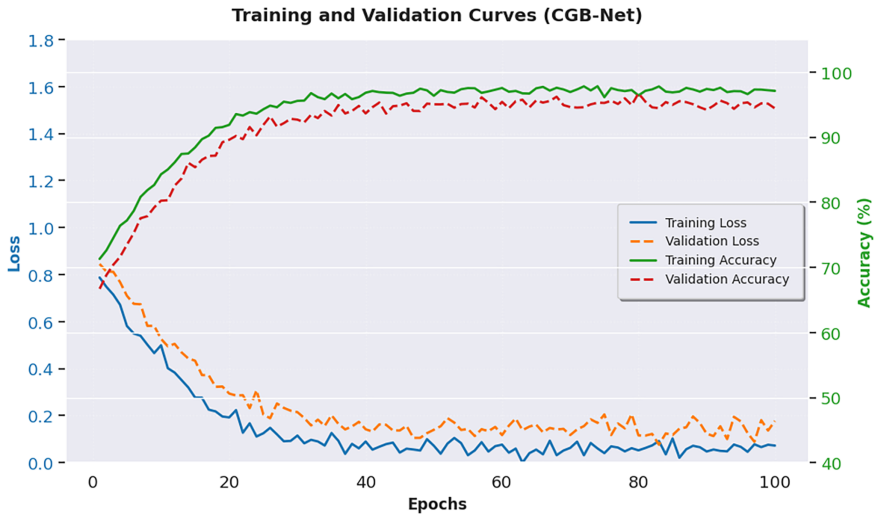

Figure 7: Training–validation curves of CGB-Net (loss and accuracy).

Another important indicator of model reliability is the close alignment between the training and validation curves. As shown in the figure, the validation loss closely follows the training loss without any sign of divergence. Similarly, the validation accuracy remains consistently close to the training accuracy, maintaining a narrow and stable gap. Owing to the regularization mechanisms incorporated in CGB-Net, including Dropout and normalization layers, validation accuracy is sustained at approximately 95% without noticeable degradation. After roughly epoch 40, the model enters a stable saturation regime, where the metrics fluctuate only slightly around a local optimal minimum without abrupt changes, further confirming the stability of the optimization process.

Optimizing the input data segmentation parameters plays a decisive role in effectively exploiting the feature extraction capability of CGB-Net. As shown in the left panel of Fig. 8, model performance exhibits a clear dependency on the temporal window size: as the window length increases from 1 to 4 s, both Accuracy and F1-score improve steadily. Short windows (1–2 s) yield inferior results because they do not provide sufficient temporal context, making it difficult for the recurrent blocks (GRU/Bi-LSTM) to fully capture oscillatory cycles and the characteristic vehicle dynamics associated with each road-surface type. Performance reaches its optimum at a window length of 4 s, achieving a peak accuracy of 95.07%, indicating a suitable balance between information richness and generalization capability. When the window is further extended to 5 s, performance slightly degrades, as the signal becomes diluted by background noise and redundant information, reducing feature discriminability and adversely affecting convergence.

In addition, the right panel of Fig. 8 shows that an overlap ratio of 0.4 (40%) provides the most effective configuration and coincides with the high-performance region of the 4-s window. This overlap increases the number of training samples in a balanced manner while preserving temporal continuity without introducing excessive redundancy between adjacent windows. In contrast, when the overlap ratio is increased to 0.6 and 0.8, classification accuracy drops markedly (to below 88%), reflecting the negative impact of high correlation among neighboring samples—an effect that can bias the model toward nearly identical patterns and reduce its generalization ability. Based on these experimental findings, the study selects a 4-s window combined with a 40% overlap ratio as the standard configuration to ensure optimal performance and stability for the road-surface classification task.

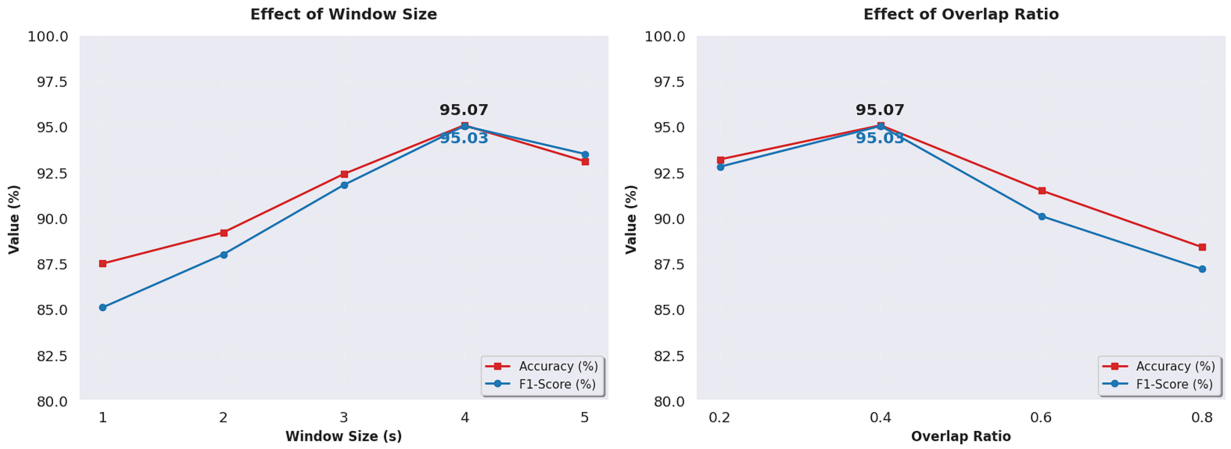

Figure 8: Impact of window size and overlap ratio on accuracy and F1-score.

4 Discussions

4.1 Comparative Analysis

The choice of model architecture and feature extraction strategy plays a critical role in the performance of road-surface classification systems based on inertial sensors. In traditional approaches, vibration signals are usually preprocessed and transformed into handcrafted features in the time domain or frequency domain, which are then fed into conventional machine learning classifiers such as Support Vector Machines (SVM) or Random Forests. While these methods are relatively simple to implement, they strongly depend on expert-driven feature selection and often suffer from limited generalization when operating conditions change, such as variations in vehicle speed, vehicle type, or driving behavior [26].

To reduce reliance on handcrafted features, deep learning methods have been increasingly adopted to learn representations directly from raw sensor data. Menegazzo and von Wangenheim conducted a benchmark study across multiple models, including both classical and deep learning approaches, and reported that CNN-based configurations achieved the best performance at around 93% accuracy [32]. Other studies have explored recurrent neural networks (RNNs/LSTMs) to model the temporal continuity of vibration signals. For example, Park et al. combined LSTM with ensemble strategies and sensor feature selection, achieving an accuracy of 94.6% on real-world test data [37]. More recent trends indicate that hybrid architectures (CNN + RNN/LSTM) often yield further improvements by jointly capturing local patterns and long-term temporal dependencies. Arce-Saenz et al. showed that a CNN/LSTM configuration achieved the highest macro-F1 score (93.38%) when exploiting multi-IMU data [38].

The experimental results of this study also confirm the limitations of single-architecture models. A standalone 1D-CNN, despite having a relatively large number of parameters, focuses mainly on local patterns and is insufficient to fully model temporal dynamics. Conversely, a pure LSTM is effective at sequence modeling but lacks strong local feature extraction capability. Hybrid architectures such as 1D-CNN + LSTM significantly improve performance; however, CGB-Net further advances this direction by adopting a lightweight hybrid design. Specifically, it combines a GRU block for efficient short-term dependency modeling with a Bi-LSTM block to capture bidirectional long-term context. As a result, CGB-Net achieves the highest performance (95.07%) while substantially reducing the number of parameters and Memory footprint usage compared with heavier hybrid variants under the same experimental setting.

Compared with more complex heavy hybrid models—such as ResBiGRU-SE, which integrates residual connections and squeeze-and-excitation mechanisms and can reach very high accuracy (98.41%) on public

datasets—such approaches come at the cost of increased architectural complexity and computational demand [39]. In addition, some studies focus on anomaly detection tasks (e.g., potholes or speed bumps) or vision-based approaches. For instance, CNN models using smartphone data have reported accuracies of up to 98% for distinguishing potholes and speed bumps [40], while image- or camera-based methods typically require computationally expensive vision pipelines. In contrast, CGB-Net operates directly on one-dimensional IMU time-series data, making it more suitable for real-time road-surface classification and deployment on resource-constrained edge platforms.

An important advantage of the proposed system is its ability to be deployed at a low cost. The data acquisition system used in this study employs a low-cost IMU sensor combined with an embedded processor mounted on the vehicle. In practice, commonly used IMU modules such as the MPU-9250 typically cost around 8–10 USD, while capable microcontrollers such as the ESP32 are priced at approximately 5–8 USD. Therefore, the total hardware cost for a basic data collection node is only about 10–20 USD. With this reasonable cost, the system can be deployed on a wide range of vehicles in normal operation, such as private cars, service vehicles, or public transportation, to collect vibration data during regular driving. This approach enables the development of large-scale road monitoring systems based on distributed data collected from multiple vehicles, allowing road conditions to be updated continuously while maintaining low deployment costs.

4.2 Limitations and Future Work

Although CGB-Net demonstrates high classification performance and stability, several limitations must be acknowledged for broader real-world deployment. While the current study establishes a strong baseline, the generalizability of the model is constrained by the experimental setting, including the use of a single test vehicle, specific road corridors, and relatively stable operating conditions.

First, the influence of vehicle dynamics and sensor configuration. The dataset was collected using a single compact sedan. Since different vehicle types (e.g., SUVs, buses, or trucks) possess distinct suspension systems and dynamic responses, the resulting IMU vibration signals may vary significantly. While CGB-Net is not inherently tied to one vehicle type, its parameters—such as normalization factors and filtering cut-off frequencies—may require re-tuning to match the response of different platforms. Furthermore, the quality and mounting position of the IMU sensor can influence the captured vibration characteristics. Future research should incorporate diverse vehicle types and investigate transfer learning or domain adaptation to improve cross-platform robustness.

Second, the impact of environmental and operational variability. This study primarily considered stable driving conditions. Practical factors such as wet road surfaces from heavy rain, sudden maneuvers, or differences in tire pressure were not explicitly modeled, though they can introduce significant noise into IMU signals. Additionally, while the sliding-window strategy effectively increased the training sample size, it introduced inherent dependencies between adjacent windows. Future studies should integrate these transient variables to evaluate model performance under more diverse and unpredictable real-world scenarios.

Third, the complexity of pavement degradation modeling. Pavement service age (10 vs. 15 years) was employed as a practical proxy for surface degradation. However, this metric does not fully capture the physical complexity of pavement aging, which is driven by traffic volume, axle loading, and maintenance history. Although the current label design is suitable for initial validation, incorporating longitudinal survey data, traffic statistics, and official maintenance records would enhance both label interpretability and the model's ability to estimate continuous condition indices, such as the International Roughness Index (IRI).

Finally, challenges in edge deployment and real-time processing. Transitioning from the current offline evaluation to real-time deployment on resource-constrained embedded platforms like the ESP32 remains a practical challenge. Despite its relatively compact architecture, the computational cost of recurrent layers (GRU and Bi-LSTM) and the model's Memory footprint may introduce latency on low-power hardware. To improve feasibility for large-scale IoT deployment, future work will explore optimization techniques such as pruning, quantization, and lightweight network redesign to reduce inference latency and power consumption.

5 Conclusions

This study employs CGB-Net, a deep learning architecture designed for automatic road-surface condition classification using motion data collected from vehicle-mounted inertial sensors (accelerometers and gyroscopes). Unlike traditional approaches that rely on handcrafted features or computationally heavy computer-vision pipelines, this work adopts a lightweight yet powerful hybrid design optimized for time-series vibration signal processing.

CGB-Net integrates three complementary components: a 1D-CNN layer to extract local morphological features from raw sensor data, a GRU layer to efficiently model short-term dependencies, and a Bi-LSTM network to capture global contextual information in both temporal directions. This architecture effectively addresses the challenge of discriminating road-surface types with highly similar vibration characteristics, particularly between new asphalt (Asphalt_10) and aged asphalt (Asphalt_15).

Experimental evaluations demonstrate that CGB-Net achieves an average classification accuracy of 95.07% and an F1-score of 95.03%, outperforming baseline models including standalone 1D-CNN, LSTM, and other hybrid variants. Notably, the model maintains computational efficiency with only 9600 parameters, making it well suited for deployment on resource-constrained IoT devices. An analysis of window size and overlap ratio further identifies that a 4-s window with 40% overlap provides an optimal balance between performance and data density.

Overall, this study confirms the feasibility of leveraging low-cost vehicle-mounted IMU sensors combined with deep sequential modeling for large-scale transportation infrastructure monitoring. The proposed system offers a reliable and cost-effective solution for identifying stages of pavement deterioration, thereby supporting transportation agencies in proactive maintenance planning and improving traffic safety. Future work will focus on extending the system to detect specific anomalies such as potholes and cracks in real time, as well as investigating the model's adaptability across different vehicle types.

Acknowledgement: Thanks to Institute of Information Technology (VAST) for allowing us to use the "IoT and Robot intensive laboratory" equipment and supporting this research, code NVKN02.01/26-27.

Funding Statement: This research was funded by Institute of Information Technology (VAST) grant number NVKN02.01/26-27 and The APC was funded by Academy of Policy and Development (Duong Do The).

Author Contributions: Manh-Tuyen Vi conducted data analysis, applied techniques to the research model, and finalized the manuscript. Duc-Tan Tran and Duong Do The provided expert guidance on applying deep learning methods. Hoang-Dieu Vu and Duc-Nghia Tran collected data and proposed the modeling process, and Duong Do The provided technical advice and wrote the first draft of the study. Duc-Tan Tran and Manh-Tuyen Vi, as lead authors, conceived the research idea and contributed to the investigation, development, and coordination of the research. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The processed IMU dataset supporting the findings of this study is openly available in the GitHub repository at: https://github.com/tuyenvimanh-Phenikaa/Data_Car. The data was last accessed

and verified in February 2026. Additional inquiries regarding the data can be directed to the corresponding author, [Duc-Tan Tran].

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Celaya-Padilla JM, Galván-Tejada CE, López-Monteagudo FE, Alonso-González O, Moreno-Báez A, Martínez-Torteya A, et al. Speed bump detection using accelerometric features: a genetic algorithm approach. Sensors. 2018;18(2):443. doi:10.3390/s18020443.
2. Varona B, Monteserin A, Teyseyre A. A deep learning approach to automatic road surface monitoring and pothole detection. Pers Ubiquitous Comput. 2020;24(4):519–34. doi:10.1007/s00779-019-01234-z.
3. Martinez-Ríos EA, Bustamante-Bello R, Navarro-Tuch S, Perez-Meana H. Applications of the generalized morse wavelets: a review. IEEE Access. 2023;11:667–88. doi:10.1109/ACCESS.2022.3232729.
4. Lekshmipathy J, Velayudhan S, Mathew S. Effect of combining algorithms in smartphone based pothole detection. Int J Pavement Res Technol. 2021;14(1):63–72. doi:10.1007/s42947-020-0033-0.
5. Collier P, Kirchberger M, Söderbom M. The cost of road infrastructure in low- and middle-income countries. World Bank Econ Rev. 2016;30(3):522–48. doi:10.1093/wber/lhv037.
6. Alhasan A, Nlenanya I, Smadi O, MacKenzie CA. Impact of pavement surface condition on roadway departure crash risk in Iowa. Infrastructures. 2018;3(2):14. doi:10.3390/infrastructures3020014.
7. Huynh VN, Truong LT, De Gruyter C. Examining the impacts of road pavement roughness and rutting on traffic safety: a macrolevel analysis. Traffic Inj Prev. 2025;26(6):720–6. doi:10.1080/15389588.2024.2448838.
8. Shaghlil N, Khalafallah A. Automating highway infrastructure maintenance using unmanned aerial vehicles. In: Proceedings of the Construction Research Congress 2018; 2018 Apr 2–4; New Orleans, LA, USA. p. 486–95.
9. Abbondati F, Biancardo SA, Veropalumbo R, Dell'Acqua G. Surface monitoring of road pavements using mobile crowdsensing technology. Measurement. 2021;171(11):108763. doi:10.1016/j.measurement.2020.108763.
10. Loprencipe G, Pantuso A. A specified procedure for distress identification and assessment for urban road surfaces based on PCI. Coatings. 2017;7(5):65. doi:10.3390/coatings7050065.
11. Maeda H, Sekimoto Y, Seto T, Kashiyama T, Omata H. Road damage detection and classification using deep neural networks with smartphone images. Comput Aided Civ Infrastruct Eng. 2018;33(12):1127–41. doi:10.1111/mice.12387.
12. Sun T, Pan W, Wang Y, Liu Y. Region of interest constrained negative obstacle detection and tracking with a stereo camera. IEEE Sens J. 2022;22(4):3616–25. doi:10.1109/JSEN.2022.3142024.
13. Kulambayev B. A deep learning-based approach for road surface damage detection. Comput Mater Contin. 2022;73(2):3403–18. doi:10.32604/cmc.2022.029544.
14. Wang J, Zhang C, Lin T. ConvNeXt-UperNet-based deep learning model for road extraction from high-resolution remote sensing images. Comput Mater Contin. 2024;80(2):1907–25. doi:10.32604/cmc.2024.052597.
15. Guan H, Li J, Yu Y, Chapman M, Wang H, Wang C, et al. Iterative tensor voting for pavement crack extraction using mobile laser scanning data. IEEE Trans Geosci Remote Sens. 2015;53(3):1527–37. doi:10.1109/TGRS.2014.2344714.
16. del Río-Barral P, Soilán M, González-Collazo SM, Arias P. Pavement crack detection and clustering via region-growing algorithm from 3D MLS point clouds. Remote Sens. 2022;14(22):5866. doi:10.3390/rs14225866.
17. Dehnad MH, Yazdi A. A review of numerical and experimental studies on hydroplaning of vehicles in motion on road surfaces. Results Eng. 2024;23(2):102438. doi:10.1016/j.rineng.2024.102438.
18. Nyirandayisabye R, Li H, Dong Q, Hakuzweyezu T, Nkinahamira F. Automatic pavement damage predictions using various machine learning algorithms: evaluation and comparison. Results Eng. 2022;16:100657. doi:10.1016/j.rineng.2022.100657.
19. Al-Sabaeei AM, Souliman MI, Jagadeesh A. Smartphone applications for pavement condition monitoring: a review. Constr Build Mater. 2024;410(2):134207. doi:10.1016/j.conbuildmat.2023.134207.

20. Bruno S, Loprencipe G, Marchetti V. Proposal for a low-cost monitoring system to assess the pavement deterioration in urban roads. Eur Transp. 2023;91:1–10. doi:10.48295/ET.2023.91.10.
21. Cafiso S, Di Graziano A, Marchetta V, Pappalardo G. Urban road pavements monitoring and assessment using bike and e-scooter as probe vehicles. Case Stud Constr Mater. 2022;16(6):e00889. doi:10.1016/j.cscm.2022.e00889.
22. Martinez-Ríos EA, Bustamante-Bello MR, Arce-Sáenz LA. A review of road surface anomaly detection and classification systems based on vibration-based techniques. Appl Sci. 2022;12(19):9413. doi:10.3390/app12199413.
23. Alqaydi S, Zeiada W, El Wakil A, Alnaqbi AJ, Azam A. A comprehensive review of smartphone and other device-based techniques for road surface monitoring. Eng. 2024;5(4):3397–426. doi:10.3390/eng5040177.
24. Alatoom YI, Smadi O. Embedded framework for low-cost pavement condition evaluation using microcontroller and single-board computer platforms. Autom Constr. 2025;178:106442. doi:10.1016/j.autcon.2025.106442.
25. Sebestyen G, Muresan D, Hangan A. Road quality evaluation with mobile devices. In: Proceedings of the 2015 16th International Carpathian Control Conference (ICCC); 2015 May 27–30; Szilvasvarad, Hungary. p. 458–64.
26. Egaji OA, Evans G, Griffiths MG, Islas G. Real-time machine learning-based approach for pothole detection. Expert Syst Appl. 2021;184(2):115562. doi:10.1016/j.eswa.2021.115562.
27. Xin H, Ye Y, Na X, Hu H, Wang G, Wu C, et al. Sustainable road pothole detection: a crowdsourcing based multi-sensors fusion approach. Sustainability. 2023;15(8):6610. doi:10.3390/su15086610.
28. Wang S, Kodagoda S, Shi L, Dai X. Two-stage road terrain identification approach for land vehicles using feature-based and Markov random field algorithm. IEEE Intell Syst. 2018;33(1):29–39. doi:10.1109/MIS.2017.2581327.
29. Li X, Huo D, Goldberg DW, Chu T, Yin Z, Hammond T. Embracing crowdsensing: an enhanced mobile sensing solution for road anomaly detection. ISPRS Int J Geo Inf. 2019;8(9):412. doi:10.3390/ijgi8090412.
30. Singh G, Bansal D, Sofat S, Aggarwal N. Smart patrolling: an efficient road surface monitoring using smartphone sensors and crowdsourcing. Pervasive Mob Comput. 2017;40(6):71–88. doi:10.1016/j.pmcj.2017.06.002.
31. Maeda H, Sekimoto Y, Seto T, Kashiyama T, Omata H. Road damage detection using deep neural networks with images captured through a smartphone. arXiv:1801.09454. 2018.
32. Menegazzo J, von Wangenheim A. Road surface type classification based on inertial sensors and machine learning. Computing. 2021;103(10):2143–70. doi:10.1007/s00607-021-00914-0.
33. Hnoohom N, Mekruksavanich S, Jitpattanakul A. A comprehensive evaluation of state-of-the-art deep learning models for road surface type classification. Intell Autom Soft Comput. 2023;37(2):1275–91. doi:10.32604/iasc.2023.038584.
34. Sattar S, Li S, Chapman M. Developing a near real-time road surface anomaly detection approach for road surface monitoring. Measurement. 2021;185(4):109990. doi:10.1016/j.measurement.2021.109990.
35. Arbabpour Bidgoli M, Golroo A, Sheikhzadeh Nadjar H, Ghelmani Rashidabad A, Ganji MR. Road roughness measurement using a cost-effective sensor-based monitoring system. Autom Constr. 2019;104(8):140–52. doi:10.1016/j.autcon.2019.04.007.
36. Basavaraju A, Du J, Zhou F, Ji J. A machine learning approach to road surface anomaly assessment using smartphone sensors. IEEE Sens J. 2020;20(5):2635–47. doi:10.1109/JSEN.2019.2952857.
37. Park J, Min K, Kim H, Lee W, Cho G, Huh K. Road surface classification using a deep ensemble network with sensor feature selection. Sensors. 2018;18(12):4342. doi:10.3390/s18124342.
38. Arce-Saenz LA, Izquierdo-Reyes J, Bustamante-Bello R. Exploring single-head and multi-head CNN and LSTM-based models for road surface classification using on-board vehicle multi-IMU data. Sci Rep. 2025;15(1):24595. doi:10.1038/s41598-025-10573-2.
39. Mekruksavanich S, Rojanavasu P, Srisungsittisunti B, Plengvittaya C, Phaphan W, Jitpattanakul A. Enhancing intelligent transportation systems: a deep learning approach for terrain recognition using vehicular inertial sensors. Lobachevskii J Math. 2024;45(12):6324–42. doi:10.1134/s1995080224607628.
40. Zareei M, Castañeda CAL, Alanazi F, Granda F, Pérez-Díaz JA. Machine learning model for road anomaly detection using smartphone accelerometer data. IEEE Access. 2025;13:122841–51. doi:10.1109/ACCESS.2025.3586812.