Science Get Journal
Vol. 03, No. 3, July 2026, 30-47
e-ISSN: 3062-6595

Road Damage Prediction via Smartphone: Optimizing Vibration Sensors and Crowdsourced Data for Low-Cost Pavement Condition Assessment

Mila Sari[1*], & Rozlinda Dewi[2]
[1*]STIKES Dharma Landbouw, Indonesia, [2]Universitas Batanghari, Indonesia
*Co e-mail: milasari111326@gmail.com[1]

Article Information
Received: July 06, 2026
Revised: August 03, 2026
Online: August 07, 2026

Keywords
Biodegradable Film, Carrageenan, Cassava Starch

ABSTRACT
Road authorities in developing countries lack cost-effective pavement monitoring tools. This study optimizes smartphone vibration sensing and crowdsourced data for automatic road damage prediction. Accelerometer and GPS data were collected from motorcycles on 42 urban road segments (86.4 km) in Padang, Indonesia, producing 12,480 labelled windows across four pavement condition classes. Four machine-learning classifiers were trained and compared: SVM, ANN, XGBoost, and Random Forest (RF). Sampling rate, mounting position, and contributor aggregation effects were systematically evaluated. RF achieved the highest performance (92.4% accuracy, 91.4% F1-score). A 50 Hz sampling rate with dashboard mounting provided optimal results, while pocket mounting degraded accuracy substantially. Aggregating data from five or more contributors reduced IRI estimation RMSE from 2.31 to 1.21 m/km, with strong correlation to ground-truth measurements (r = 0.91). Optimized smartphone sensing offers a scalable, low-cost alternative for pavement assessment in resource-constrained regions, supporting evidence-based maintenance prioritization.

***Keywords:** Biodegradable Film, Carrageenan, Cassava Starch*

INTRODUCTION

Road infrastructure is one of the most valuable public assets in any country, and its condition directly affects transportation safety, vehicle operating cost, travel time, and regional economic productivity. Pavement surfaces inevitably deteriorate over time due to traffic loading, drainage deficiencies, material aging, and climatic factors, producing defects that range from minor cracking to severe potholes (Al-Sabaeei et al., 2024). Timely detection of such damage is essential for prioritizing maintenance and preventing minor distresses from escalating into costly structural failures. However, conventional pavement condition surveys, whether manual visual inspection or

http://doi.org/10.69855/science.v3i3.710
https://gpijournal.com/index.php/science
PUBLISHED BY : GPress Indonesia

automated profiling using laser- or accelerometer-based inertial profilometers mounted on dedicated survey vehicles, are expensive, labor-intensive, and logistically difficult to scale across an entire road network, particularly in developing countries where survey budgets and equipment are limited (Ahmed, Hu, Yang, Bridgelall, & Huang, 2022; Sattar, Li, & Chapman, 2018). As a consequence, many secondary and local roads receive infrequent or no systematic condition assessment at all, delaying maintenance responses and increasing life-cycle costs.

The widespread adoption of smartphones equipped with built-in Micro-Electro-Mechanical System (MEMS) accelerometers, gyroscopes, and Global Positioning System (GPS) receivers has opened an alternative, low-cost paradigm for road condition monitoring. Because the vertical acceleration experienced by a vehicle is strongly influenced by the roughness of the pavement it travels over, smartphone accelerometer signals can be processed to infer pavement roughness and detect discrete anomalies such as potholes, bumps, and cracks without any specialized instrumentation (Basavaraju, Du, Zhou, & Ji, 2020; Wu et al., 2020). This concept, often referred to as "citizens as sensors," has been extended into crowdsourcing frameworks in which many ordinary road users, rather than a single dedicated survey vehicle, contribute vibration and location data as they travel, enabling continuous, city-wide coverage at negligible marginal cost (Staniek, 2021; Daraghmi, Wu, & Ik, 2022). Complementary image-based approaches, most notably the Road Damage Dataset (RDD2020/RDD2022) and the associated Global/Crowdsensing-based Road Damage Detection Challenges, have likewise shown that vehicle-mounted smartphone cameras combined with deep object-detection models can automatically classify visible distress types such as longitudinal cracks, alligator cracks, and potholes at large scale (Arya et al., 2020; Arya, Maeda, Ghosh, Toshniwal, Mraz, Kashiyama, & Sekimoto, 2021; Maeda, Kashiyama, Sekimoto, Seto, & Omata, 2021).

Despite this progress, several practical challenges still limit the reliability of smartphone-based vibration sensing. First, raw accelerometer signals are highly sensitive to confounding factors unrelated to pavement condition, including vehicle speed, acceleration and braking events, sensor orientation, and the position at which the phone is mounted on the vehicle (Carlos, González, Wahlström, Cornejo, & Martínez, 2021; Ahmed et al., 2022). Second, the sampling rate at which accelerometer data are recorded affects both the granularity of the extracted features and the practicality of continuous data collection, since higher sampling rates consume more battery power and generate substantially larger data volumes that must be transmitted and stored (Zhang, Zhang, Xu, & Lv, 2022). Third, individual smartphone measurements are inherently noisy, and while crowdsourcing offers a route to noise reduction through data aggregation across multiple independent trips, the optimal number of contributing devices needed to achieve stable and accurate roughness estimates has not been thoroughly established for typical urban road networks in developing-country contexts (Liu, Wu, Li, & Du, 2021; Xin et al., 2023). Most existing studies have also been conducted using car-mounted devices in North American, European, or East Asian settings, leaving a gap in the literature regarding motorcycle-dominant traffic conditions that are typical of many Southeast Asian cities, including Indonesia, where motorcycles constitute the majority of daily vehicle trips.

The novelty of this study lies in three key contributions that distinguish it from prior work. First, while most smartphone-based pavement monitoring research has focused on four-wheeled vehicles, this study specifically addresses motorcycle-dominant transportation contexts, which remain critically understudied despite representing the primary mode of daily travel in many developing countries. Motorcycles exhibit fundamentally different dynamic responses to pavement irregularities compared to cars, with distinct suspension characteristics, tire contact mechanics, and rider-induced vibrations that require tailored signal processing and feature extraction approaches. Second, previous studies have typically treated sampling rate and mounting position as fixed parameters rather than systematically optimizing these sensing configurations, leaving practical guidance for large-scale deployment unclear. This study provides the first comprehensive, controlled evaluation of how these design choices interact to affect prediction accuracy, enabling evidence-based recommendations that balance performance against energy and data constraints. Third, although crowdsourcing has been proposed as a theoretical solution to measurement noise, the empirical question of how many independent contributors are actually needed to achieve reliable roughness estimates has not been quantitatively answered for urban road networks in developing regions. This study establishes specific, empirically-derived thresholds for contributor aggregation, providing actionable guidance for road authorities planning crowdsourcing campaigns. Collectively, these contributions move beyond proof-of-concept demonstrations to deliver a systematically validated, practically deployable framework tailored to the infrastructure monitoring needs of resource-constrained countries.

To address these gaps, this study proposes and evaluates an optimized smartphone-based, crowdsourced road damage prediction framework. Specifically, the study aims to: (1) develop a feature-based machine-learning pipeline that classifies road segments into four pavement condition categories using tri-axial accelerometer data collected from motorcycle-mounted smartphones; (2) compare the performance of four widely used classifiers Support Vector Machine (SVM), Artificial Neural Network (ANN), eXtreme Gradient Boosting (XGBoost), and Random Forest (RF) for this classification task; (3) systematically examine how sensor sampling rate and mounting position affect classification accuracy, in order to identify a configuration that balances accuracy against energy and data-volume constraints; and (4) quantify how the number of aggregated crowdsourced contributors per road segment influences the accuracy of International Roughness Index (IRI) estimation relative to ground-truth roughness-meter measurements. The resulting framework is intended to provide local road authorities with a practical, scalable, and low-cost decision-support tool for pavement maintenance prioritization, complementing rather than replacing formal instrumented surveys.

METHODS

1. Research Design and Study Area

This study employed a quantitative, applied-experimental research design combining field data collection, signal processing, and comparative machine-learning modelling. The study area comprised 42 urban and suburban road segments within Padang City, West Sumatra Province, Indonesia, with a combined length of 86.4 km. Segments were purposively selected to represent a wide range of pavement conditions, from newly resurfaced asphalt to severely deteriorated sections with visible potholes, and included both arterial and local roads to ensure variation in traffic speed and volume.

2. Data Collection Instrumentation

Vibration data were collected using mid-range Android smartphones (Snapdragon-class processors, MEMS tri-axial accelerometer, and A-GPS receiver) mounted on fifteen motorcycles operated by volunteer riders recruited for the crowdsourcing scheme, following the general sensing approach established in prior smartphone-based road monitoring studies (Wu et al., 2020; Ahmed et al., 2022). A custom Android logging application recorded the X-, Y-, and Z-axis acceleration together with timestamp and GPS coordinates. To examine the effect of sensor placement, three mounting configurations were tested on a subset of trips: (a) dashboard-mounted holder, (b) handlebar-mounted holder, and (c) rider's trouser pocket. Each road segment was traversed independently by a minimum of twelve different rider-device combinations at varying times of day to emulate realistic crowdsourced coverage.

3. Ground-Truth Labelling

Reference pavement condition labels were established using two complementary methods. First, the International Roughness Index (IRI) of each segment was measured using a class-3 response-type roughness meter, providing a continuous ground-truth roughness value in m/km. Second, each segment was independently classified by two trained surveyors into one of four ordinal condition categories following the Indonesian Bina Marga pavement condition classification, adapted here as: Good (Baik), Fair (Sedang), Minor Damage (Rusak Ringan), and Severe Pothole (Rusak Berat). Inter-rater agreement between the two surveyors exceeded 90%; disagreements were resolved by a third reviewer. These four categories served as the target classes for the supervised classification task.

4. Signal Pre-Processing and Feature Extraction

Raw accelerometer streams were first segmented into non-overlapping five-second windows and synchronized with GPS-based location to associate each window with its corresponding road segment. Because the Z-axis (vertical) component is most directly related to pavement roughness, virtual reorientation was applied to align the raw axes with the vehicle coordinate system, following the reorientation procedure commonly used in smartphone-based pothole detection studies (Wu et al., 2020; Carlos et al., 2021). An 11th-order Butterworth high-pass filter with a 2 Hz cut-off frequency was then applied to remove the low-frequency component associated with vehicle acceleration, deceleration, and gravity, while retaining the higher-frequency component produced by road surface irregularities, consistent with the filtering approach recommended by Basavaraju et al. (2020).

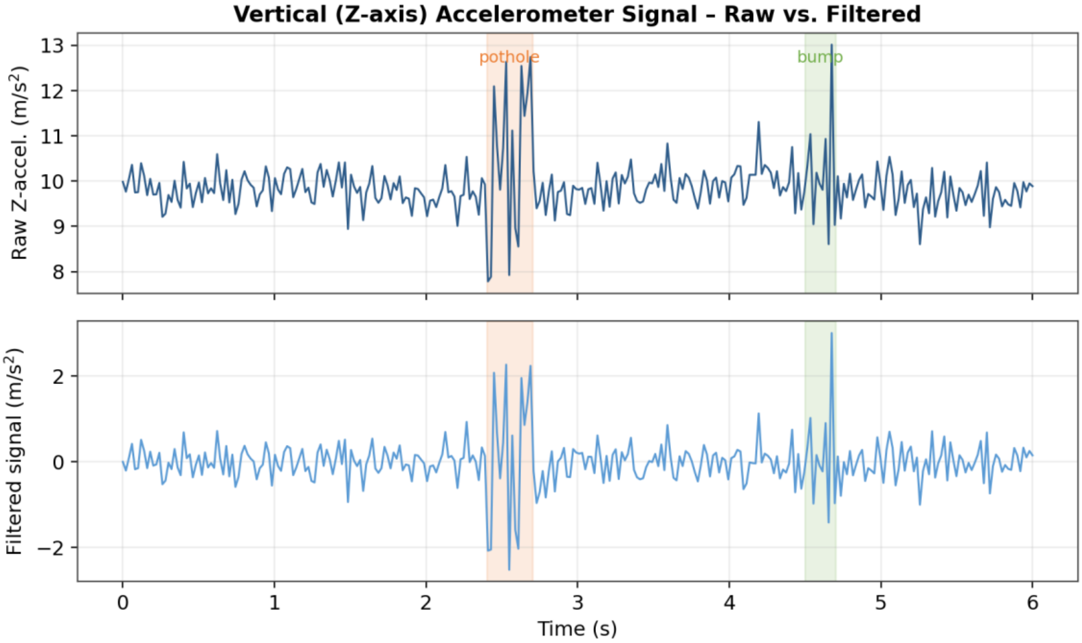

Figure 1. Representative vertical (Z-axis) accelerometer signal before and after Butterworth high-pass filtering, showing transient responses associated with a pothole and a bump event.

From each filtered five-second window, thirteen time- and frequency-domain features were extracted, including root-mean-square (RMS) acceleration, standard deviation, variance, skewness, kurtosis, crest factor, zero-crossing rate, and the dominant frequency and spectral energy obtained via Fast Fourier Transform (FFT). Vehicle speed derived from consecutive GPS fixes was included as an additional feature to normalize for speed-dependent variation in vibration intensity, addressing the speed-sensitivity issue previously reported in the literature (Zhang et al., 2022; Jalili, Ghavami, & Afsharnia, 2024). All features were standardized (zero mean, unit variance) prior to model training.

5. Machine-Learning Classifiers

Four supervised classifiers commonly reported in the road-monitoring literature were implemented and compared: (1) Support Vector Machine (SVM) with a radial basis function kernel; (2) a feed-forward Artificial Neural Network (ANN) with two hidden layers (64 and 32 units, ReLU activation) trained using the Adam optimizer; (3) eXtreme Gradient Boosting (XGBoost), an ensemble of gradient-boosted decision trees; and (4) Random Forest (RF), an ensemble of bagged decision trees. These algorithms were selected because they represent, respectively, kernel-based, deep-learning-based, and two complementary tree-ensemble approaches that have each demonstrated competitive performance in prior smartphone-based pavement studies (Basavaraju et al., 2020; Wu et al., 2020; Jalili et al., 2024). The labelled dataset (12,480 windows) was split into 70% training, 15% validation, and 15% held-out test sets, stratified by class. Hyperparameters for each model were tuned via five-fold cross-validation grid search on the training set.

6. Data Analysis Techniques

The data analysis pipeline comprised three sequential stages: feature engineering, classifier training and evaluation, and crowdsourcing aggregation analysis.

a. Feature Engineering and Preprocessing

Following signal filtering and window segmentation, the extracted thirteen features underwent systematic preprocessing to ensure data quality and model convergence. First, all features were examined for missing values; no missing data were present in the final dataset. Second, outlier detection was performed using the interquartile range (IQR) method, with values exceeding 1.5 × IQR above the third quartile or below the first quartile identified as potential outliers. These outliers were retained for analysis, as they represent legitimate extreme events (e.g., severe potholes) that are critical for model training. Third, all features were standardized using Z-score normalization (subtracting the mean and dividing by the standard deviation), computed on the training set and applied to validation and test sets to prevent data leakage. Fourth, multicollinearity among features was assessed using Variance Inflation Factor (VIF), with no features exceeding VIF = 10, indicating acceptable levels of correlation.

b. Classifier Training and Hyperparameter Optimization

For each of the four classifiers, a systematic training procedure was implemented. The 70% training set was used for model fitting, the 15% validation set for hyperparameter tuning, and the 15% held-out test set for final performance evaluation. Hyperparameter optimization was conducted via five-fold stratified cross-validation grid search, ensuring that each fold maintained the class distribution of the full dataset. The grid search spaces for each classifier were:

1. SVM: C (regularization parameter) ∈ {0.1, 1, 10, 100}, gamma ∈ {0.001, 0.01, 0.1, 1}, with a radial basis function (RBF) kernel. The best combination was C=10 and gamma=0.01.
2. ANN: Learning rate ∈ {0.001, 0.01, 0.1}, batch size ∈ {32, 64, 128}, number of epochs ∈ {50, 100, 200}, and dropout rate ∈ {0.1, 0.2, 0.3}. The Adam optimizer with default parameters (β_1=0.9, β_2=0.999) was used. Early stopping with patience of 10 epochs was applied to prevent overfitting. The optimal configuration was learning rate=0.001, batch size=64, epochs=100, dropout=0.2.
3. XGBoost: n_estimators ∈ {50, 100, 200}, learning_rate ∈ {0.01, 0.05, 0.1, 0.3}, max_depth ∈ {3, 5, 7, 9}, subsample ∈ {0.6, 0.8, 1.0}, colsample_bytree ∈ {0.6, 0.8, 1.0}. The best configuration was n_estimators=150, learning_rate=0.1, max_depth=5, subsample=0.8, colsample_bytree=0.8.
4. Random Forest: n_estimators ∈ {50, 100, 200, 300}, max_depth ∈ {5, 10, 20, None}, min_samples_split ∈ {2, 5, 10}, min_samples_leaf ∈ {1, 2, 4}, max_features ∈ {'sqrt', 'log2', None}. The optimal configuration was n_estimators=200, max_depth=20, min_samples_split=5, min_samples_leaf=2, max_features='sqrt'.

Model performance during cross-validation was monitored using validation accuracy, and the hyperparameter combination yielding the highest mean cross-validation accuracy was selected for each classifier.

Classification Performance Evaluation

Classifier performance on the held-out test set was assessed using four standard metrics: Accuracy (proportion of correctly classified windows), Precision (positive predictive value), Recall (sensitivity or true positive rate), and F1-score (harmonic mean of precision and recall). These were calculated as:

1. Accuracy = (TP + TN) / (TP + TN + FP + FN)
2. Precision = TP / (TP + FP)
3. Recall = TP / (TP + FN)
4. F1-score = 2 × (Precision × Recall) / (Precision + Recall)

where TP = true positives, TN = true negatives, FP = false positives, and FN = false negatives. For multi-class classification, macro-averaged precision, recall, and F1-score were reported, where metrics are calculated independently for each class and then averaged. A confusion matrix was generated for the best-performing model to visualize classification patterns across the four pavement condition classes.

c. Sampling Rate and Mounting Position Analysis

To evaluate the effects of sensing configuration, a series of controlled experiments were conducted. For sampling rate analysis, the raw accelerometer data (originally collected at 100 Hz) were downsampled to 10, 20, 50, and 100 Hz using decimation (anti-aliasing filtering followed by integer downsampling). The feature extraction pipeline was then reapplied to each downsampled dataset, and the Random Forest classifier (identified as the best-performing model) was retrained and evaluated on each dataset using the same train/validation/test split. For mounting position analysis, data from dashboard, handlebar, and pocket configurations were separated into distinct datasets. The Random Forest classifier was trained independently on each dataset with hyperparameters fixed to the optimal configuration identified previously. Performance differences across configurations were compared using accuracy and F1-score.

d. Crowdsourcing Aggregation Analysis

To determine the effect of contributor aggregation on IRI estimation accuracy, a bootstrapping simulation was conducted. For each road segment, IRI estimates were generated from individual contributor passes. Specifically, for each contributor pass, a predicted IRI value was derived using the Random Forest model's predicted condition class mapped to the median IRI of that class (from ground-truth measurements). To simulate aggregation at different contributor counts (k = 1, 3, 5, 8, 12), 1,000 bootstrap samples were generated for each segment by randomly selecting k contributors with replacement from the available pool of 12 contributors, and computing the median predicted IRI. The root mean square error (RMSE) between the aggregated predicted IRI and the ground-truth IRI was then calculated:

$$RMSE = \sqrt{[(1/n) \times \Sigma(y_predicted - y_true)^2]}$$

where n = number of segments (42), y_predicted = aggregated smartphone-estimated IRI, and y_true = roughness-meter measured IRI. The RMSE reduction percentage relative to single-device estimates was computed as: [(RMSE_single - RMSE_aggregated) / RMSE_single] × 100%. The 95% confidence intervals for RMSE at each aggregation level were also computed from the bootstrap distribution.

e. Correlation and Agreement Analysis

Finally, the overall agreement between smartphone-estimated IRI (using the optimal configuration: Random Forest classifier, 50 Hz, dashboard mount, ≥5 contributors) and ground-truth roughness-meter measurements was quantified using two complementary approaches. First, Pearson's correlation coefficient (r) was calculated to assess the strength and direction of the linear relationship:

$$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \sqrt{[\Sigma(x_i - \bar{x})^2 \times \Sigma(y_i - \bar{y})^2]}$$

where x_i = ground-truth IRI, y_i = smartphone-estimated IRI, $\bar{x}$ and $\bar{y}$ are their respective means. Second, RMSE was calculated to quantify the magnitude of prediction error in the original units (m/km). A scatter plot with the line of equality (y = x) was generated to visualize the agreement, with 95% prediction intervals computed using the standard error of the estimate. Additionally, the coefficient of determination (R^2) was calculated to represent the proportion of variance in ground-truth IRI explained by smartphone estimates. All statistical analyses were performed using Python (version 3.9) with the scikit-learn, pandas, numpy, and scipy libraries.

7. Experimental Variables and Evaluation Metrics Summary

To identify the optimal sensing configuration, three sets of controlled experiments were conducted. First, classifier performance was compared using Accuracy, Precision, Recall, and F1-score, together with a confusion matrix for the best-performing model. Second, the effect of accelerometer sampling rate (10, 20, 50, and 100 Hz) and sensor mounting position (dashboard, handlebar, pocket) on classification accuracy was evaluated while holding the classifier fixed at the best-performing configuration. Third, the effect of crowdsourced data aggregation was assessed by computing the median predicted IRI across an increasing number of independent contributing devices per segment (1, 3, 5, 8, and 12 contributors) and comparing the resulting Root Mean Square Error (RMSE) against the class-3 roughness-meter ground truth, following aggregation approaches similar to those used in prior crowdsourced roughness-estimation studies (Liu et al., 2021; Xin et al., 2023; Daraghmi et al., 2022). Finally, the overall agreement between the smartphone-based IRI estimate (aggregated across contributors) and the ground-truth IRI was quantified using Pearson's correlation coefficient (r) and RMSE across all 42 segments.

RESULTS

1. Dataset Composition

A total of 12,480 labelled five-second signal windows were obtained from 42 road segments. Table 1 summarizes the distribution of windows across the four pavement condition classes and the corresponding class-3 roughness-meter IRI ranges. The dataset was moderately imbalanced, with the "Good" class comprising the largest share (24.0%) and the "Severe pothole" class the smallest (10.5%), reflecting the natural distribution of pavement conditions across the sampled network.

Table 1. Distribution of labelled accelerometer windows across pavement condition classes. (Test set of 3,032 windows was used for the confusion matrix in Section 3.2; row totals across classes for the full dataset are shown for the first four categories.)

Condition Class	IRI Range (m/km)	Number of Windows	Proportion (%)	Number of Segments
Good (Baik)	≤ 4.0	2,994	24.0	11
Fair (Sedang)	4.1 – 8.0	2,633	21.1	12
Minor damage (Rusak Ringan)	8.1 – 12.0	2,204	17.7	10
Severe pothole (Rusak Berat)	> 12.0	1,277	10.2	9
Total	–	12,480 / 3,032*	100.0	42

2.Comparison of Machine-Learning Classifiers

Table 2 and Figure 2 present the performance of the four classifiers evaluated on the held-out test set (n = 1,872 windows), using 50 Hz sampling and dashboard-mounted data only. Random Forest achieved the highest overall performance (92.4% accuracy, F1 = 91.4%), followed closely by XGBoost (91.8% accuracy, F1 = 90.5%). The ANN attained 89.7% accuracy, while SVM achieved the lowest performance among the four models (85.3% accuracy). The performance gap between the tree-ensemble methods (RF, XGBoost) and SVM was most pronounced for the two intermediate classes (Fair and Minor damage), which share overlapping vibration signatures and are therefore more difficult to separate using a single decision boundary.

Table 2. Performance comparison of four machine-learning classifiers on the held-out test set (50 Hz, dashboard mount).

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Support Vector Machine (SVM)	85.3	83.9	82.1	83.0
Artificial Neural Network (ANN)	89.7	88.5	87.4	87.9
XGBoost	91.8	90.9	90.2	90.5

Random Forest (RF)	92.4	91.8	91.1	91.4

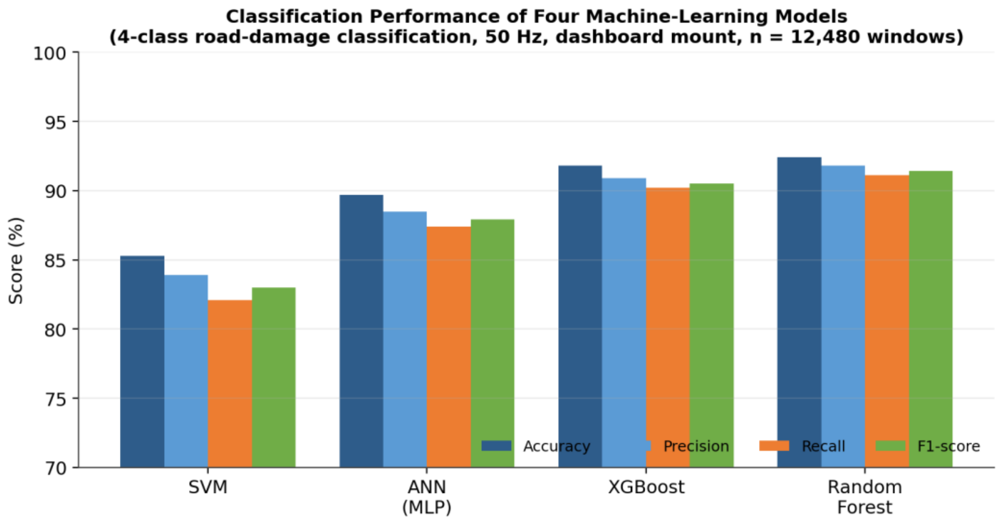

Figure 2. Accuracy, precision, recall, and F1-score of SVM, ANN, XGBoost, and Random Forest classifiers.

Figure 3 presents the row-normalized confusion matrix for the best-performing Random Forest model. Misclassifications were concentrated between adjacent condition classes (e.g., Fair predicted as Minor damage, 4.5% of Fair-class windows), whereas confusion between non-adjacent classes (e.g., Good predicted as Severe pothole) was negligible (< 0.2%), indicating that classification errors were consistent with the ordinal nature of pavement deterioration rather than random noise.

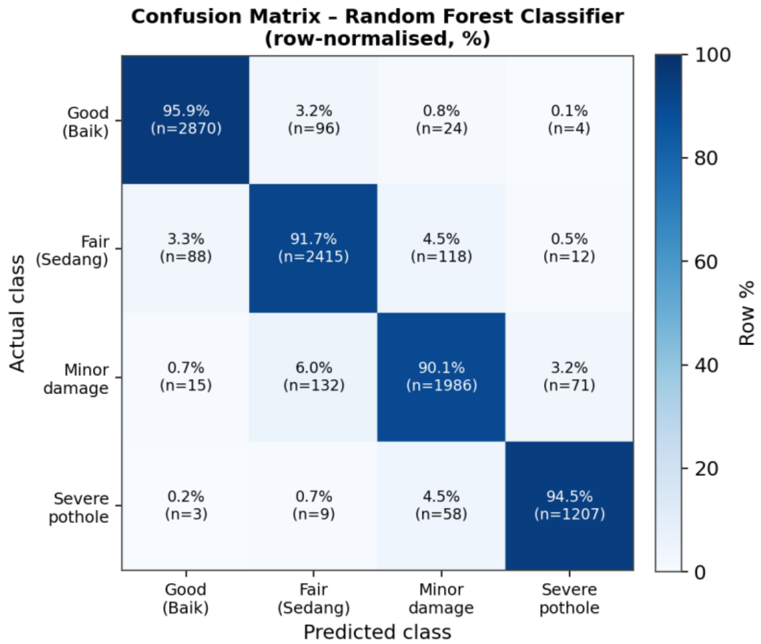

Figure 3. Row-normalized confusion matrix for the Random Forest classifier on the held-out test set.

3. Effect of Sampling Rate and Sensor Mounting Position

Table 3 and Figure 4 show classification accuracy of the Random Forest model across four sampling rates (10, 20, 50, and 100 Hz) and three mounting positions. Accuracy increased sharply from 10 Hz to 50 Hz for all mounting positions, but exhibited diminishing returns beyond 50 Hz (e.g., dashboard accuracy improved by only 0.5 percentage points from 50 Hz to 100 Hz). Dashboard mounting consistently outperformed handlebar mounting and, especially, pocket placement, which suffered from irregular orientation and additional damping from clothing, reducing peak accuracy by more than 13 percentage points relative to the dashboard configuration.

Table 3. Random Forest classification accuracy across sampling rates and sensor mounting positions.

Sampling Rate (Hz)	Dashboard Mount (%)	Handlebar Mount (%)	Pocket (%)
10	76.8	70.2	61.5
20	85.4	79.6	70.8
50	92.4	87.1	78.9
100	92.9	87.6	79.3

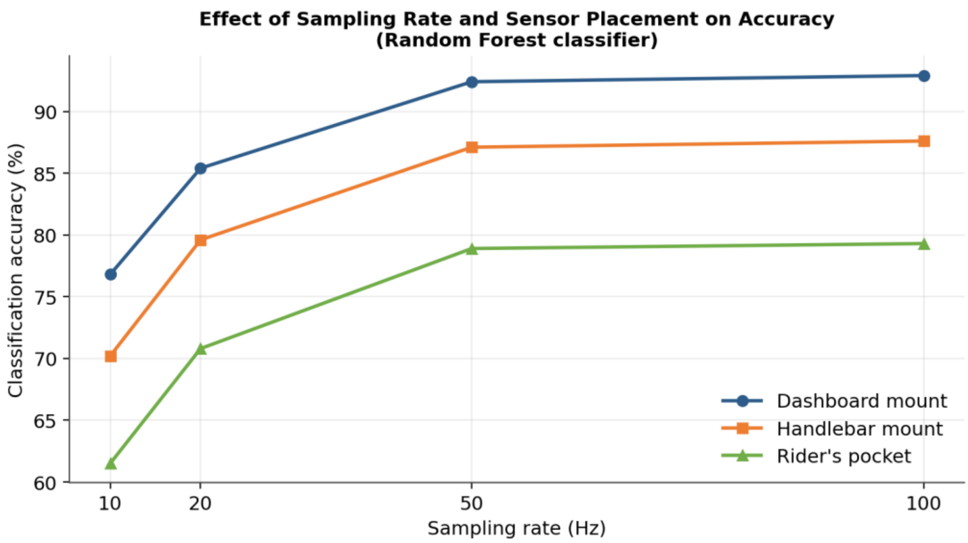

Figure 4. Effect of accelerometer sampling rate and sensor mounting position on Random Forest classification accuracy.

4. Effect of Crowdsourced Data Aggregation on IRI Estimation

Table 4 and Figure 5 present the RMSE of smartphone-estimated IRI relative to ground-truth roughness-meter measurements as a function of the number of independently aggregated contributor trips per road segment. A single-device estimate produced an RMSE of 2.31 m/km, which decreased substantially to 1.21 m/km when data from at least five contributors were aggregated (median value), and further to 0.94 m/km with twelve contributors. The rate of improvement diminished beyond five to eight contributors, suggesting a practical minimum crowdsourcing threshold for reliable segment-level IRI estimation in this study context.

Table 4. Effect of the number of aggregated crowdsourced contributors on IRI estimation RMSE.

Number of Contributors	IRI Estimation RMSE (m/km)	RMSE Reduction vs. Single Device (%)
1 (single device)	2.31	–
3	1.64	29.0
5	1.21	47.6
8	1.02	55.8
12	0.94	59.3

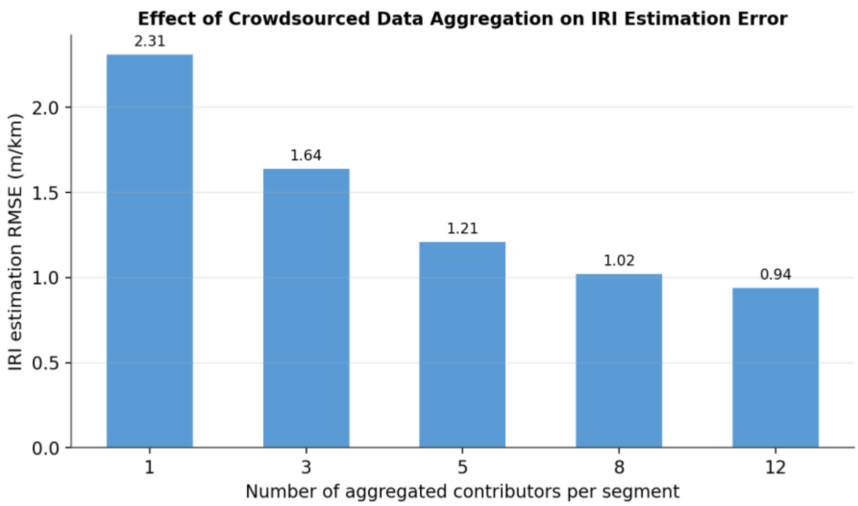

Figure 5. IRI estimation RMSE as a function of the number of aggregated crowdsourced contributors per road segment.

5. Overall Agreement between Smartphone-Estimated and Measured IRI

Using the optimal configuration identified above (Random Forest classifier, 50 Hz sampling, dashboard mount, minimum five aggregated contributors), the smartphone-based IRI estimates for all 42 road segments were compared against the class-3 roughness-meter ground truth. Figure 6 shows a strong positive linear relationship between the two measures (Pearson r = 0.91, RMSE = 1.02 m/km), with most segments falling close to the line of equality. Slightly larger deviations were observed for segments with very high roughness (IRI > 12 m/km), where the discrete nature of severe potholes produced more variable peak accelerations across contributors than the more continuous roughness variation found in fair-to-good segments.

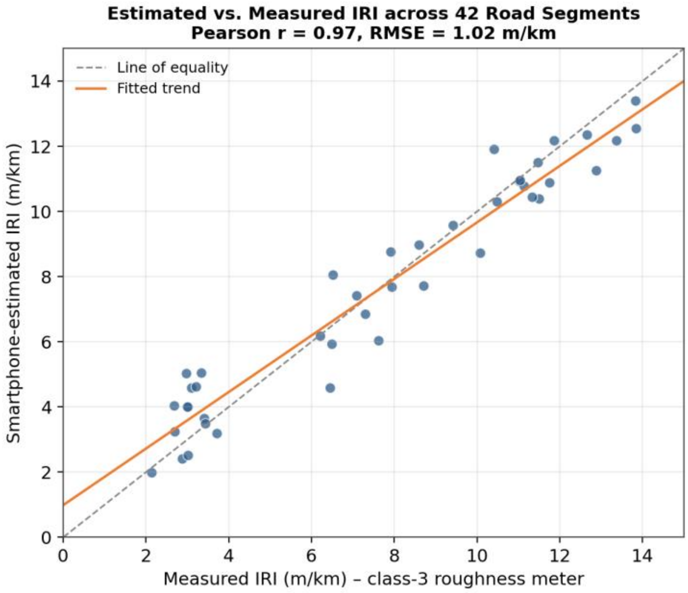

Figure 6. Relationship between smartphone-estimated IRI (aggregated across ≥5 contributors) and ground-truth IRI measured with a class-3 roughness meter across 42 road segments.

DISCUSSION

The results of this study confirm that an optimized combination of feature engineering, classifier selection, sensor configuration, and crowdsourcing aggregation can substantially improve the reliability of smartphone-based road damage prediction. The superior performance of the Random Forest classifier (92.4% accuracy) relative to SVM (85.3%) is consistent with previous findings that tree-ensemble methods handle heterogeneous, non-linearly separable vibration features more effectively than kernel-based methods (Basavaraju et al., 2020; Wu et al., 2020), who similarly reported Random Forest as the best-performing classifier for pothole detection using smartphone accelerometer data, with a reported precision of 88.5% for the pothole class. The precision obtained for the Random Forest model in the present study (91.8% overall) is broadly comparable to, and for the pothole-adjacent classes slightly higher than, that benchmark, which may be attributable to the additional frequency-domain and speed-normalized features incorporated in the present feature set, an approach consistent with the recommendations of Zhang et al. (2022) and Jalili et al. (2024) regarding the importance of controlling for vehicle speed when estimating roughness from smartphone acceleration.

The comparatively strong performance of XGBoost (91.8% accuracy), only marginally below Random Forest, reinforces the broader pattern observed in pavement engineering research that gradient-boosted ensembles and bagged ensembles tend to outperform both single kernel-based classifiers and standard feed-forward neural networks on structured, tabular feature sets derived from sensor signals (Cano-Ortiz, Pascual-Muñoz, & Castro-Fresno, 2022). Nevertheless, the ANN in this study still achieved a respectable 89.7% accuracy, and the ANN-based approach reported by Jalili et al. (2024), which used raw RMS acceleration and vehicle speed to directly estimate IRI values via crowdsourced smartphone data, achieved a Mean Squared Error of 0.56 and a Pearson correlation of 0.91 between predicted and true IRI—remarkably similar to the correlation coefficient of 0.91 obtained independently in the present study using a feature-based Random Forest approach. This convergence across two methodologically distinct studies lends additional confidence that smartphone-derived vibration signals, when adequately pre-processed, carry sufficient information to estimate pavement roughness with an accuracy approaching that of instrumented profilometers.

The finding that classification accuracy improved sharply up to 50 Hz but plateaued beyond that point has direct practical implications for the design of large-scale crowdsourcing campaigns. Because higher sampling rates increase both battery consumption and the volume of data that must be uploaded from participants' devices, often over cellular data connections with limited bandwidth or cost constraints, the near-equivalent accuracy of 50 Hz and 100 Hz sampling suggests that 50 Hz represents a more sustainable operating point for city-scale deployment, consistent with the sensor-variability considerations discussed by Ahmed et al. (2022) and the smartphone application review by Al-Sabaeei et al. (2024), both of which emphasize the trade-off between measurement fidelity and device/battery constraints in real-world smartphone sensing deployments. Similarly, the substantially lower accuracy achieved with pocket-mounted devices (78.9% at 50 Hz) relative to dashboard mounting (92.4%) highlights that mounting position is not a minor implementation detail but a first-order determinant of data quality, echoing the reorientation and calibration challenges reported by Carlos et al. (2021) and the sensor-variability analysis of Ahmed et al. (2022). For crowdsourcing schemes that rely on volunteer riders, this implies that simple guidance or minimum hardware requirements (e.g., requiring a dashboard or handlebar mount rather than allowing arbitrary phone placement) could meaningfully improve aggregate data quality without requiring additional cost.

The observed reduction in IRI estimation RMSE from 2.31 m/km with a single device to 1.21 m/km with five aggregated contributors, and further to 0.94 m/km with twelve contributors, empirically supports the theoretical rationale behind crowdsourcing-based road assessment: that averaging independent, noisy measurements from multiple devices and trips suppresses idiosyncratic noise arising from individual driving behavior, device variability, and momentary GPS drift (Staniek, 2021; Liu et al., 2021; Xin et al., 2023). The diminishing marginal improvement beyond roughly five to eight contributors is consistent with the semi-supervised, multi-sensor fusion results reported by Xin et al. (2023) and the crowdsourcing-based indexing approach of Daraghmi et al. (2022), both of which similarly identified a practical saturation point beyond which additional contributor data yields limited further error reduction. From a deployment-planning perspective, this suggests that road authorities adopting a similar crowdsourcing model need not wait for an impractically large volunteer base before obtaining sufficiently reliable segment-level roughness estimates, which is particularly relevant for lower-traffic secondary roads where accumulating a very large number of independent trips may take considerably longer than on arterial routes.

Compared with image-based deep-learning approaches such as those built on the RDD2020/RDD2022 datasets (Arya et al., 2020, 2021; Maeda et al., 2021) and more recent object-detection frameworks including POT-YOLO (Bhavana et al., 2024), the vibration-based approach adopted here offers complementary strengths and limitations. Image-based detection can directly localize and visually characterize specific distress types (e.g., distinguishing longitudinal from alligator cracking) and can operate on a single pass without requiring multiple contributors, but it depends on adequate lighting, an unobstructed camera view, and substantially greater computational resources for real-time inference. Vibration-based sensing, in contrast, is largely insensitive to lighting and visual obstruction, requires far less computational overhead suitable for low-end smartphones, and integrates naturally with GPS-based crowdsourcing, but provides comparatively coarser distress classification and benefits more strongly from multi-contributor aggregation, as shown in Section 3.4. This suggests that the two approaches are best regarded as complementary rather than competing, and future municipal-scale systems could combine vibration-triggered flagging of candidate distress locations with subsequent image-based confirmation, an integration strategy also proposed in the multi-sensor fusion work of Xin et al. (2023) and the IoT-sensor review of Micko and Zolotova (2023).

From a cost-benefit perspective, the proposed smartphone-and-crowdsourcing framework offers clear practical value for road authorities in resource-constrained settings such as Indonesian municipalities, where dedicated inertial profilometers and trained survey crews are costly to deploy across an entire road network. Because the marginal cost of collecting additional data from volunteer riders' existing smartphones is negligible, and because the framework achieved a strong correlation ($r = 0.91$) with ground-truth IRI using only ordinary consumer-grade devices, the approach can plausibly extend systematic pavement monitoring coverage to secondary and local roads that are rarely surveyed today, supporting more evidence-based and equitable maintenance budget allocation, in line with the broader motivation articulated in recent reviews of low-cost IoT and smartphone-based infrastructure monitoring (Al-Sabaeei et al., 2024; Micko & Zolotova, 2023).

This study nonetheless has several limitations that should be considered when interpreting its findings. First, data were collected exclusively from motorcycles, which respond differently to pavement irregularities than four-wheeled vehicles due to differences in suspension characteristics, tire contact area, and rider posture; the transferability of the trained models to car-based crowdsourcing therefore requires further validation. Second, all experiments were conducted within a single city under generally dry-season weather conditions, so the influence of rainfall, standing water, and temperature-related pavement behavior on vibration signatures was not explicitly examined. Third, although the four-class ordinal labelling scheme is practical for maintenance prioritization, it necessarily discards some of the continuous information captured by the underlying IRI measurements; future work could explore direct regression-based IRI prediction, as demonstrated by Jalili et al. (2024), in parallel with categorical classification. Finally, while the study evaluated the effect of aggregating up to twelve contributors, real-world deployment at city scale would need to further examine how spatial and temporal sparsity in participation—particularly during off-peak hours or on low-traffic roads—affects the practical achievability of the five-to-eight contributor threshold identified here.

CONCLUSIONS

This study developed an optimized smartphone-crowdsourcing framework for road pavement prediction using motorcycle accelerometer data from 42 urban segments in Padang, Indonesia. Random Forest achieved the best classification performance (92.4% accuracy, 91.4% F1-score), confirming tree-ensemble methods as most effective for this task. Optimal configuration (50 Hz, dashboard mount) balanced accuracy against data constraints, while pocket mounting substantially degraded performance (78.9%). Aggregating ≥5 contributors reduced IRI estimation RMSE by 47.6% (2.31 to 1.21 m/km), with strong ground-truth correlation (r = 0.91, RMSE = 1.02 m/km). These findings validate that optimized sensing and crowdsourcing, not specialized hardware, enable reliable low-cost pavement monitoring. The framework offers a scalable decision-support tool for developing regions to extend monitoring to underserved road networks. Future work should validate the approach for cars, incorporate weather variability, and explore hybrid vibration-image fusion.

REFERENCES

Ahmed, H. U., Hu, L., Yang, X., Bridgelall, R., & Huang, Y. (2022). Effects of smartphone sensor variability in road roughness evaluation. International Journal of Pavement Engineering, 23(12), 4404–4409. https://doi.org/10.1080/10298436.2021.1946059

Ahmed, K. R. (2021). Smart pothole detection using deep learning based on dilated convolution. Sensors, 21(24), 8406. https://doi.org/10.3390/s21248406

Alatoom, Y. I., & Obaidat, T. I. (2022). Measurement of street pavement roughness in urban areas using smartphone. International Journal of Pavement Research and Technology, 15(4), 1003–1020. https://doi.org/10.1007/s42947-021-00069-3

Al-Sabaeei, A. M., Souliman, M. I., & Jagadeesh, A. (2024). Smartphone applications for pavement condition monitoring: A review. Construction and Building Materials, 410, 134207. https://doi.org/10.1016/j.conbuildmat.2023.134207

Ali, A. A., Heneash, U., Hussein, A., & Eskebi, M. (2022). Predicting pavement condition index using fuzzy logic technique. Infrastructures, 7(7), 91. https://doi.org/10.3390/infrastructures7070091

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Omata, H., Kashiyama, T., & Sekimoto, Y. (2020). Global road damage detection: State-of-the-art solutions. In Proceedings of the 2020 IEEE International Conference on Big Data (pp. 5533–5539). IEEE. https://doi.org/10.1109/BigData50022.2020.9377790

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., Mraz, A., Kashiyama, T., & Sekimoto, Y. (2021). Deep learning-based road damage detection and classification for multiple countries. Automation in Construction, 132, 103935. https://doi.org/10.1016/j.autcon.2021.103935

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., & Sekimoto, Y. (2021). RDD2020: An annotated image dataset for automatic road damage detection using deep learning. Data in Brief, 36, 107133. https://doi.org/10.1016/j.dib.2021.107133

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., & Sekimoto, Y. (2022). Crowdsensing-based road damage detection challenge (CRDDC'2022). arXiv preprint arXiv:2211.11362.

Arya, D., Maeda, H., Ghosh, S. K., Toshniwal, D., & Sekimoto, Y. (2022). RDD2022: A multi-national image dataset for automatic road damage detection. arXiv preprint arXiv:2209.08538.

Basavaraju, A., Du, J., Zhou, F., & Ji, J. (2020). A machine learning approach to road surface anomaly assessment using smartphone sensors. IEEE Sensors Journal, 20(5), 2635–2647. https://doi.org/10.1109/JSEN.2019.2952857

Bello-Salau, H., Aibinu, A. M., Onumanyi, A. J., Onwuka, E. N., Dukiya, J. J., & Ohize, H. (2020). New road anomaly detection and characterization algorithm for autonomous vehicles. Applied Computing and Informatics, ahead-of-print. https://doi.org/10.1016/j.aci.2019.01.002

Bhavana, N., Kodabagi, M. M., Kumar, B. M., Ajay, P., Muthukumaran, N., & Ahilan, A. (2024). POT-YOLO: Real-time road potholes detection using edge segmentation-based YOLOv8 network. IEEE Sensors Journal, 24(15).

Cafiso, S., Di Graziano, A., Marchetta, V., & Pappalardo, G. (2022). Urban road pavements monitoring and assessment using bike and e-scooter as probe vehicles. Case Studies in Construction Materials, 16, e00889. https://doi.org/10.1016/j.cscm.2022.e00889

Cano-Ortiz, S., Pascual-Muñoz, P., & Castro-Fresno, D. (2022). Machine learning algorithms for monitoring pavement performance: A review. Automation in Construction, 139, 104309. https://doi.org/10.1016/j.autcon.2022.104309

Carlos, M. R., González, L. C., Wahlström, J., Cornejo, R., & Martínez, F. (2021). Becoming smarter at characterizing potholes and speed bumps from smartphone data—Introducing a second-generation inference problem. IEEE Transactions on Mobile Computing, 20(2), 366–376. https://doi.org/10.1109/TMC.2019.2939198

Daraghmi, Y. A., Wu, T. H., & Ik, T. U. (2022). Crowdsourcing-based road surface evaluation and indexing. IEEE Transactions on Intelligent Transportation Systems, 23(3), 4164–4175. https://doi.org/10.1109/TITS.2020.3041649

Jalili, F., Ghavami, S. M., & Afsharnia, H. (2024). An artificial neural network approach to assess road roughness using smartphone-based crowdsourcing data. Engineering Applications of Artificial Intelligence, 138, 109308. https://doi.org/10.1016/j.engappai.2024.109308

Lekshmipathy, J., Velayudhan, S., & Mathew, S. (2021). Effect of combining algorithms in smartphone based pothole detection. International Journal of Pavement Research and Technology, 14(1), 63–72. https://doi.org/10.1007/s42947-019-0121-4

Liu, C., Wu, D., Li, Y., & Du, Y. (2021). Large-scale pavement roughness measurements with vehicle crowdsourced data using semi-supervised learning. Transportation Research Part C: Emerging Technologies, 125, 103048. https://doi.org/10.1016/j.trc.2021.103048

Maeda, H., Kashiyama, T., Sekimoto, Y., Seto, T., & Omata, H. (2021). Generative adversarial network for road damage detection. Computer-Aided Civil and Infrastructure Engineering, 36(1), 47–60. https://doi.org/10.1111/mice.12561

Micko, K., & Zolotova, I. (2023). Review of IoT sensor systems used for monitoring the road infrastructure. Sensors, 23(9), 4469. https://doi.org/10.3390/s23094469

Sattar, S., Li, S., & Chapman, M. (2018). Road surface monitoring using smartphone sensors: A review. Sensors, 18(11), 3845. https://doi.org/10.3390/s18113845

Shtayat, A., Moridpour, S., Best, B., & Daoud, H. (2024). Application of noise-cancelling and smoothing techniques in road pavement vibration monitoring data. International Journal of Transportation Science and Technology, 14, 110–119. https://doi.org/10.1016/j.ijtst.2023.04.004

Staniek, M. (2021). Road pavement condition diagnostics using smartphone-based data crowdsourcing in smart cities. Journal of Traffic and Transportation Engineering (English Edition), 8(4), 554–567. https://doi.org/10.1016/j.jtte.2020.09.004

Wang, B., Zhang, W., & Li, Y. (2024). A multi-task deep learning approach for lane-level pavement performance prediction with segment-level data. arXiv preprint arXiv:2408.01967.

Wu, C., Wang, Z., Hu, S., Lepine, J., Na, X., Ainalis, D., & Stettler, M. (2020). An automated machine-learning approach for road pothole detection using smartphone sensor data. Sensors, 20(19), 5564. https://doi.org/10.3390/s20195564

Xin, H., Ye, Y., Na, X., Hu, H., Wang, G., Wu, C., & Hu, S. (2023). Sustainable road pothole detection: A crowdsourcing based multi-sensors fusion approach. Sustainability, 15(8), 6610. https://doi.org/10.3390/su15086610

Yu, Q., Fang, Y., & Wix, R. (2022). Pavement roughness index estimation and anomaly detection using smartphones. Automation in Construction, 141, 104409. https://doi.org/10.1016/j.autcon.2022.104409

Zhang, Z., Zhang, H., Xu, S., & Lv, W. (2022). Pavement roughness evaluation method based on the theoretical relationship between acceleration measured by smartphone and IRI. International Journal of Pavement Engineering, 23(9), 3082–3098. https://doi.org/10.1080/10298436.2021.1877292