

IoT-Based Multi-Sensor Fusion and Machine Learning Framework for Geo-Tagged Road Anomaly Classification

Akhilesh Kumar Singh¹, Pawan Kumar Verma^{2,3}

¹ Lincoln University College, Petaling Jaya, Malaysia ; pdf.akhilesh@lincoln.edu.my;

² Lincoln University College, Petaling Jaya, Malaysia; abes.pawan@gmail.com;

³ Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, Maharashtra, India;
pdf.akhilesh@lincoln.edu.my

Abstract: Road condition monitoring in real-time is important for safety, maintenance, performance, and smart transport systems. The conventional approach to road inspection involves manual inspections, inspection vehicles, cameras, or expensive sensors that prevent regular and large-scale monitoring. This research provides an IoT-based, multi-sensor fusion, and machine learning framework to detect and geographically locate road anomalies. The system employs an MPU6050 accelerometer and gyroscope, Global Positioning System (GPS), and Raspberry Pi 4 as a processor. Vehicle response data are collected in a continuous manner and further filtered, segmented, normalized, analyzed in terms of orientation, and features extracted. A Random Forest classifier is employed to classify roads into categories such as normal road, pothole, bump, and rough roads. Anomalies detected by the model are mapped to GPS coordinates for creation of geo-referenced road conditions data. Unlike vision-based monitoring, the proposed framework focuses on the analysis of vehicle dynamics and, therefore, is less dependent on lighting conditions and image processing requirements. Local processing also facilitates sustainable edge AI development by minimizing constant sending of data collected by sensors. The framework provides a practical approach to road health monitoring and further connected intelligent transport solutions.

Keywords: IoT, Road Anomaly Detection, Random Forest, MPU6050, Raspberry Pi, GPS, Multi-Sensor Fusion, Intelligent Transportation System.

Introduction

The road infrastructure plays a crucial role in transport safety, economy, accessibility, costs of operation of vehicles, and sustainable development of cities. Potholes, bumps, cracks, depression, and unevenness of road cause unnecessary vibration, reduce comfort for passengers, increase the wear of cars, and decrease the safety of driving. Therefore, detection of such features of road is important for maintenance and intelligent transport systems. Traditional methods for road inspection include manual inspection, special road surveying cars, imaging processing, and special devices for road surface measurement. The manual inspection is a time-consuming process and is difficult to be used continuously. Imaging processing provides a detailed description but is sensitive to changes in lighting, shadow, weather conditions, camera position, and requires heavy computation. Measurement using LiDAR and special road profile measurement systems give precise results but increase the cost significantly.

Modern achievements in IoT technologies, embedded computing, machine learning, and cheap sensing devices allow another solution to the problem. A car moving on an irregular road makes specific variations of acceleration and angular velocities. Linear vibrations are measured by accelerometers; angular movements are calculated by gyroscopes. Using this data together with GPS and machine learning algorithms one can detect road anomalies. This research work follows previous research efforts on Edge AI, inertial sensors, IoT, machine learning, and connected vehicle for anomaly detection [1] [10], filling a gap in the literature by proposing a methodology to implement a framework for multi-sensor fusion of accelerometer and gyroscope measurements, embedded processing, multi-class classification of road conditions, and geographical localization.

The proposed framework is IoT-based Multi-Sensor Fusion and Machine Learning for Geo-Tagged Road Anomaly Detection. The processing device used is Raspberry Pi 4 with an additional MPU6050 inertial sensor that provides acceleration and angular velocity. Anomalies are geotagged based on recorded GPS coordinates. The framework utilizes Random Forest to classify road conditions as either normal, pothole, bump, or rough. Contributions of this paper include: (i) an inexpensive IoT sensing architecture design; (ii) a fusion of data from an accelerometer, gyroscope, orientation sensors, and GPS; (iii) a pre-processing and statistical features generation; (iv) multi-class road condition classification based on Random Forest; and (v) road anomalies geotagging based on GPS coordinates.

Related Works

Road-anomaly detection has increasingly moved from manual inspection toward automated sensing and machine-learning-based methods. Vibration-based detection is particularly attractive because road irregularities produce changes in the vertical and longitudinal motion of a vehicle. The accelerometers sense the translation vibration, while the gyroscopes provide information about the angular motion associated with pitch, roll, etc. The smartphone-based sensing system is a good example of an economical way of implementing sensing in mobile devices. Zareei et al. used data from the accelerometers and the gyros in combination with machine learning and got 98% accuracy for the pothole and speed bump detection experiment [1]. This result demonstrates that inertial sensing is a feasible solution, but the efficiency depends on the sensing equipment, installation, car, road, sampling, and the data set used.

Traditional algorithms for machine learning continue to perform well when the input is made up of a small set of hand-crafted features. Decision trees, KNN, SVM, logistic regression, and Random Forest have relatively low computational needs. Random Forest is especially useful in situations where there are non-linear dependencies between the sensors' features, since several decision trees are able to model various decision boundaries. The inclusion of GPS adds the element of location to the vibration detection process. An anomaly detector, which outputs an anomaly label, informs us about the occurrence of the event, but not of its location. Adding latitude, longitude, and timestamp information to the predictions will enable us to create a series of location-tagged road observations.

The above-mentioned system involves four key competences: inertial sensing from mobile devices, feature extraction locally, machine learning classifiers, and geographic location detection. This system is not intended to replace but rather supplement camera-based and deep-learning systems.

Proposed Methodology

Pipeline processing procedure: Sensor acquisition -> Signal pre-processing -> Features extraction -> Random Forest classifier -> Geo tagging using GPS. The system design is centered around Raspberry Pi 4 installed onto the four-wheeled smart vehicle. Three-axis linear acceleration and three-axis angular

velocity are obtained from the MPU6050 sensors, while the geographical coordinates are determined by the GPS (latitude and longitude). Define acceleration vector at time t as $a(t)=[a_x(t), a_y(t), a_z(t)]$ and angular-velocity vector as $\omega(t)=[\omega_x(t), \omega_y(t), \omega_z(t)]$. The windows of observations are used to take into account the temporal properties of the signal, but not instantaneous values. Inertial signals can suffer from high-frequency noise and other disturbances caused by vehicular motion. Hence, pre-processing is required prior to features extraction. The filtering stage is developed to retain only major motion components associated with road surface roughness and to remove the undesirable high-frequency variations of signals. In terms of each window, the features extracted will be the following: Mean value, Standard deviation, Minimum value, Maximum value, RMS value, Acceleration magnitude, Angular velocity magnitude, Orientation variation. They provide a compact description of the motion profile and decrease the amount of data transferred to the classifier.

The random forest algorithm is used due to its ability to establish non-linear relationships between the selected features along with compatibility with embedded system implementation. Each tree generates a class label, and a class label of the ensemble is established based on majority voting. In case of anomalies, the predicted class label is correlated with the nearest GPS reading. Hence, each event is characterized by class label, location, and timestamp. This enables representation at the event level without sending all sensor sample data continuously to the remote site/server. The experimental setup includes the Raspberry Pi 4 board attached to the four-wheel smart rover. The smart rover uses the MPU6050 board with three axis accelerometers and three axis gyroscopes for motion sensing and GPS for localization. Wi-Fi is used for communication during remote monitoring. The proposed processing involves local buffering and then feature extraction and classification.

The process involved in carrying out the experiment involves the following six stages:

- (1) Collecting inertial and GPS measurements on various road surfaces;
- (2) Segmenting the data collected using the sensor into constant time intervals;
- (3) Extracting features from the data;
- (4) Labeling the data with the label of the type of road surface;
- (5) Training the Random Forest model on the labeled dataset; and
- (6) Testing the model using the test dataset.

Result and Discussion

Table I shows the example confusion matrix for the case of binary classification. The number of normal-road windows and anomaly windows in the independent test set equals 100 each. In the classification, 96 normal windows and 93 anomaly windows are correctly classified, whereas four normal windows are misclassified as anomalies, and seven anomaly windows are missed.

Based on the example confusion matrix, the accuracy equals $(96 + 93) / 200 = 0.945$ or 94.50%. The Precision, Recall, and F1-Score for the Road-Anomaly class equal $93 / (93 + 4) = 95.88\%$, $93 / (93 + 7) = 93.00\%$, and approximately 94.42%, respectively. These figures represent the right way of computing and reporting the results; the measured performance of the prototype should not be considered as an example.

Example system-level measurements given in Table II demonstrate the proper way of reporting latency, communication reduction, and energy. For example, assuming that some hypothetical inference takes 82 ms, and Raspberry Pi consumes 6 W during the inference time, the energy will equal $6 \times 0.082 = 0.492$ J per inference. Assuming that the raw transmission requires 2.80 MB for the observation period and event-level transmission – 0.24 MB, the communication reduction is approximately 91.43%. The main benefit of the architecture consists of the combination of motion sensing with the geographic localization. Acceleration-based detector may inform about disturbance, but the use of GPS transforms the information into road-referenced observation, which can be repeatedly done and used to identify the anomalies.

The Random Forest algorithm will be a good fit for a feature-engineered edge pipeline because it can learn non-linear interactions between acceleration, rotations, and the state of the roads without requiring an extensive deep neural network for inference. Nonetheless, the richness of the training dataset is critical. An algorithm trained on a limited number of surfaces or velocities might not perform well on other cars and suspensions.

Table 1. Confusion Matrix

Actual \ Predicted	Normal Road	Road Anomaly	Total
Normal Road	96	4	100
Road Anomaly	7	93	100
Total	103	97	200

Table 2. Edge System Performance

Metric	Illustrative value	Measurement basis
Inference latency	82 ms/event	Raspberry Pi 4 inference interval
GPS localization	Implemented	Latitude/longitude attached to event
Communication delay	35 ms/event	Event packet transmission
Raw data volume	2.80 MB / observation period	Hypothetical raw stream
Event data volume	0.24 MB / same period	Hypothetical event packets
Communication reduction	91.43%	From Eq. (16)
Device power during inference	6.0 W	Hypothetical measurement
Energy per inference	0.492 J/event	From Eq. (17): 6×0.082

Conclusion

This paper presents an IoT-based multi-sensor approach for the geo-location of road anomalies using the Raspberry Pi 4, MPU6050 sensor for inertial measurement, GPS positioning, signal processing, and Random Forest classifier. Data on vehicle motion is collected, physical properties associated with acceleration and angular motion are calculated, relevant statistical features are extracted, road anomalies are classified, and geographic location information is assigned to each event.

This approach presents a real-world solution for the affordable road condition monitoring and smart transportation systems at the edge. The main novelty of the paper is based on the fusion of sensing, local learning, and localization capabilities into the embedded system.

In order to provide a more complete picture, this updated version features placeholders for numerical results and a confusion matrix that is filled. These results have been created purely for illustration purposes and cannot be considered actual experimental results. In the future version of research, these results will need to be replaced by real measured values taken from the built robot. Once they are known, the very same formulas can be used.

References

1. V. Katariya et al., "VegaEdge: Edge AI Confluence Anomaly Detection for Real-Time Highway IoT Applications," 2023.
<https://dx.doi.org/10.2139/ssrn.4830704>
2. A. Solaas et al., "Anomaly Detection in Connected Vehicles: A Systematic Review," IEEE Access, 2024.
<https://doi.org/10.48550/arXiv.2405.02731>
3. Rachid, T., Abdellah, O., Abdellatif, L., Jarrar, A., & Youssef, B. (2025). Hybrid AI framework for anomaly detection and root cause analysis in multi-agent systems. IAES International Journal of Artificial Intelligence (IJ-AI), 14(6), 5290–5302.
<https://doi.org/10.11591/ijai.v16.i6>
4. M. J. C. S. Reis et al., "Lightweight Signal Processing and Edge AI for Real-Time Anomaly Detection," MDPI Sensors, vol. 25, no. 3, 2025.
<https://doi.org/10.3390/s25216629>
5. Wong, W.K., Baskar, S., Abubeker, K.M. et al. Sustainable cyber-physical VANETs with AI-driven anomaly detection and energy-efficient multi-criteria routing using machine learning algorithms. Sci Rep 15, 40068 (2025).
<https://doi.org/10.1038/s41598-025-28212-1>
6. D. Das, S. Banerjee, P. Chatterjee, U. Ghosh and U. Biswas, "Blockchain for Intelligent Transportation Systems: Applications, Challenges, and Opportunities," in *IEEE Internet of Things Journal*, vol. 10, no. 21, pp. 18961-18970, 1 Nov.1, 2023
[doi: 10.1109/JIOT.2023.3277923](https://doi.org/10.1109/JIOT.2023.3277923)
7. M. Zareei, C. A. L. Castañeda, F. Alanazi, F. Granda and J. A. Pérez-Díaz, "Machine Learning Model for Road Anomaly Detection Using Smartphone Accelerometer Data," in *IEEE Access*, vol. 13, pp. 122841-122851, 2025
[doi: 10.1109/ACCESS.2025.3586812](https://doi.org/10.1109/ACCESS.2025.3586812)
8. Singh, A.K., Edla, D.R. Auto-GeRo: An IOT based Geo Spatial Model for Real-Time Road Condition Detection. SN COMPUT. SCI. 5, 999 (2024)
<https://doi.org/10.1007/s42979-024-03362-7>
9. Prabha, M., Nandhini, S., Dayanidhy, M. et al. Edge-AI integrated secure wireless IoT architecture for real time healthcare monitoring and federated anomaly detection. Sci Rep 16, 574 (2026)
<https://doi.org/10.1038/s41598-025-30150-x>
10. S. K. Oruganti, A. Khosla and T. G. Thundat, "Wireless Power-Data Transmission for Industrial Internet of Things: Simulations and Experiments," in *IEEE Access*, vol. 8, pp. 187965-187974, 2020
[doi: 10.1109/ACCESS.2020.3030658](https://doi.org/10.1109/ACCESS.2020.3030658)