

PAPER • OPEN ACCESS

The Vehicle as a Mobile Sensor Network base IoT and Big Data for Pothole Detection Caused by Flood Disaster

To cite this article: A Mochamad Rifki Ulil *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **239** 012034

View the [article online](#) for updates and enhancements.

The Vehicle as a Mobile Sensor Network base IoT and Big Data for Pothole Detection Caused by Flood Disaster

Mochamad Rifki Ulil A, Fiannurdin, Sritrusta Sukaridhoto*, Anang Tjahjono, and Dwi Kurnia Basuki

Politeknik Elektronika Negeri Surabaya, Jl Raya ITS, Surabaya, Indonesia 60111

*Corresponding Author: dphoto@pens.ac.id

Abstract. The condition where Indonesia is a tropical country and has two seasons where in April to September is the dry season while October to March is the rainy season where during the rainy season it causes flooding in several big cities in Indonesia so that puddles meet the road segments resulting in roads become damaged and perforated where the problem arises starting from a traffic accident, the exhaust gas CO₂ on the vehicle does not burn completely, so that adds to air pollution. To answer this problem, we built a solution to make a platform to detect potholes and road monitoring. The research carried out the implementation of road observers that utilized IMU sensor that attached to the embedded system, send the location information to Smart Environment Monitoring and Analytics in Real Time System, is a monitoring platform system based on the Internet of Things, and Big Data. This system is an integration of several sensor devices connected to the embedded system and communication devices attached to the vehicle.

Keywords: Vehicle as a Mobile Sensor Network, Flooding, Pothole Smart Environment Monitoring, IoT-Big Data.

1. Introduction

Transportation networks support economic activities by enabling the movement of goods and people. As long as there are severe weather events, transportation infrastructure can be directly or indirectly damaged, pose a threat to human safety, and cause significant disturbances related to the economic and social impacts of flooding. Especially the high rainfall is the main cause of disruption related to fur in the transportation sector. The existing approach to assessing the impact of road transport disruptions failed. The condition of hollow roads due to flooding in Indonesia is one of the causes of traffic accidents. Both the driver and road management who want to improve as soon as possible, there are several problems how to monitor hole conditions in all cities [1].

One of the approaches to the detection of road damage is to use a manual that is humans and report to the central authority. Despite having the highest accuracy, It is assumed that human beings are fair



that judging that the road is a hole or not have required a lot of time, cost and difficulty in clustering data. Road level (International Roughness Index, IRI) is one factor/service function (functional performance) of the effectiveness of very high pavement damage to the driver's comfort (driving quality). Existing road quality or what have to be built must be by applicable standards and provisions. Main requirements a good road is strong, flat, airtight, durable and economical for a lifetime Potter. To fulfill these requirements, monitoring and periodic or online evaluation proper construction[2].

To answer these problems, in this study the implementation of integration with Road and Environmental Monitoring as a Supporter of road monitoring or Smart Environment Monitoring and Analytics in the Real-time System is a monitoring system based on the Internet of Things, Learning and Big Data. [3] This system is an integration of several sensor devices connected to embedded systems and communication devices that are attached to the vehicle. Data is sent to the data center and analyzed using Learning machine learning that can analyze the collected data. This system also has a visualization in the form of graphs, and map [4].

2. Related Works

The Smart Environment System is currently developing a system to detect potholes and assess road conditions in real-time. Our solution is a mobile application that captures data on a car's movement from gyroscope and accelerometer sensors in the phone. Using this sensor data, we trained SVM models to classify road conditions with 93% accuracy and potholes with 92% accuracy, passing the base rate for both problems. As the user drives, the models use the sensor data to classify whether the road is good or bad [5].

The Big Data Platform is a data retrieval process that focuses on more and more on streaming from multiple sensor data. MapReduce algorithms and machine learning are used in Big Data analysis methods. In this study, we also researched the IoT-Big Data platform [6].

3. Proposed Method

3.1. IOT Architecture

The system is built using a seven-layer model of IoT with grouping according to the role and function as in Table 1. The built systems are grouped according to the IoT reference model consisting of physical devices and controllers (1), connectivity (2), edge computing (3), data accumulation (4), data abstraction (5), application (6), collaboration and process (7). The overall system design is shown in Figure 1.

3.2. Physical Devices Controller and Connectivity

This section is the part used to detect road surface conditions using the gyro sensor, accelerometer, and GPS on the Head Unit. The application works starting from the sensor reading and sent to the server via the internet. Through 4G / 3G signals, this section is used to connect physical devices and controller layers with edge computing layer or internet in data transmission. The 4G modem is also used to connect between the Head Unit on layers the edge computing and cloud computing. The throughput of this 4G modem is about 20 Mbps.

3.3. Edge Computing and Head Unit

Sensor data is sent, then the data is received by the MQTT Broker which have been continued to other sub-systems. Figure 2 shows the edge computing or the head unit that has been attached to the vehicle and have to distribute data accelerometer, gyro and GPS to cloud computing database server (SEMAR).

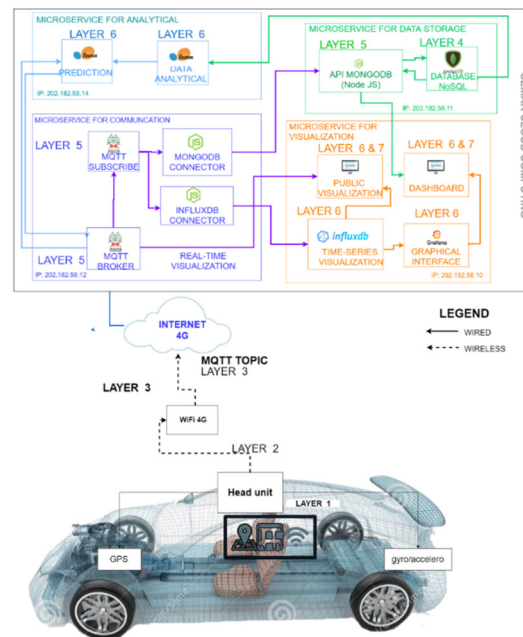


Figure 1: Design System of VaaMSN.

Table 1. IOT Architecture

No	Name Layer
7	Collaboration & Processes (website)
6	Applications (Website, Grafana, influxDB and Mosca)
5	Data Abstraction (MQTT Broker, Node JS)
4	Data Accumulation (MongoDB)
3	Edge computing (ProxmoxVE, MQTT Broker Node JS)
2	Connectivity (4G/3G, wifi)
1	Physical Devices and Controller (Raspberrypi, IMU, GPS)

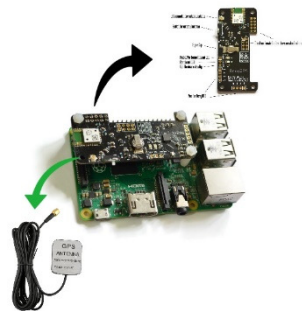


Figure 2: Physical Devices.

3.4. Accumulation data and Data abstraction

This process store the data that have been sent to cloud computing to the NoSQL database, MongoDB, as the Big Data platform. The data are used as a training data to be analyzed so that it produces a model that will be used in the prediction process. This section serves to regulate the data flow that exists in cloud computing; there is a connector application that distributes data flow from MQTT to Big Data through the MongoDB RESTful API on the Big Data server. The next connector serves to distribute data to InfluxDB. Also, by using MQTT communication the data received can be displayed directly to public visualization in real time, with the aim of reducing delivery time [7].

Algorithm.Accumulation Data

1	Begin:
2	MQTT Connect
3	Loop:
4	Read Sensor Data
5	Read sensor IMU
6	Read GPS sensor
7	Add Imu sensor data and GPS
8	Send data to the server using MQTT (topic,data)
9	End.

Scheme 1: Algorithm Accumulation Data

3.5. Applications

This section consists of machine learning, prediction and visualization processes, and IoT Cloud platform website dashboard results of cloud computing systems from the road surface damage classification system.

3.5.1. Learning Proses and prediction

The Machine Learning process is needed to build a classification model. This process is needed to provide a classification model before performing a real-time classification process. Thus, the level of confidence in the classification results can affect the level of accuracy of the model produced.

In the training process of the dataset, we used Scikit-learn for conducting the training process of the dataset. Support Vector Machine[8] and Decision Tree are used as classification algorithms for this research. The best method between the two would be selected by comparing the results of the algorithm[9].

Support Vector Machine (SVM) used for regression and classification algorithm, this algorithm has been implemented for big data classification. In SVM, this classification is performed by giving a training vector $y_i \in \mathbb{R}^n$, $i = 1, \dots, l$, with an indicator vector $y_i \in \mathbb{R}^1$ such that $y_i \in \{1, -1\}$ to solves the following primal optimization problem:

$$\begin{aligned}
 & \min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \\
 & \text{subject to : } y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \\
 & \quad \xi_i \geq 0, i = 1, \dots, l,
 \end{aligned} \tag{1}$$

Due to the high dimensional possibilities of the vector variables w , where $C > 0$ is the parameter of regularization and $\phi(x_i)$ map x_i to the higher dimension. We can solve the following double problem:

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \alpha^T Q \alpha + e^T \alpha \\ \text{subject to} & y^T \alpha = 0 \\ & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l, \end{aligned} \quad (2)$$

After problem (2) is solved, using the primal-dual relationship, the optimal w satisfies:

$$w = \sum_{i=1}^l y_i \alpha_i \phi(x_i) \quad (3)$$

The decision function of the classification becomes

$$f(w^T \phi(x) + b) = \text{sign} \left(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b \right) \quad (4)$$

The actual SVM can be classified into two classes. A correct multiclass method is some classification problems. In this case, combine several binary classifiers with two methods. The first method is 'One on one' means applying a comparison of inter-class pairs. The second method is 'One against the other' means comparing one class with all other classes.

Decision trees are machine learning algorithms that utilize tree decisions such as impact probabilities, which involve the results of events, resource costs, and utilities. The Decision Tree is one of the best classifiers when considering classification accuracy. This algorithm studies the classification function which includes the dependent attribute (variable) given by the attribute value (input) (variable).

Some of the most popular decision tree algorithms are C4.5, CART, and Naive Bayes Tree. This research uses CART which stands for Classification and Regression Trees. CART analysis is a form of binary recursive partition and can handle numerical and categorical variables. The error rate of the data received can be measured by CART, it can also build a binary tree where each internal node produces two classes for the attributes received. The way in which trees are built by selecting attributes recursively uses the attributes they have the lowest Gini Index. Attribute with the lowest Gini Index value is obtained by calculating the Gini Index value in each attribute. Gini Index is calculated based on the formula below, where the probability of the i^{th} the class for c target classes of a given attribute is P_i . Meanwhile, P_i is the probability of class i .

$$Gini = 1 - \sum_{i=1}^c (P_i)^2 \quad (5)$$

The accuracy of the classifier algorithm is evaluated by dividing the dataset into two subsets of about 70% for the training set and the remaining 30% for the test set. The training set is intended to build a classification model. While the performance measurement of the classification model is built using a test set. This method can be called the hold-out method. This learning procedure is shown in algorithm

3.5.2. Visualization and Dashboard

Figure 3 is a web visualization that represents the gyro value and visualizes a pothole map. In Figure 3a, graphical form displays surface quality parameter values based on the time taken. In the view provided, data are retrieved for the last 1 minute, 5 minutes, and all data shares the image at the bottom which shows all road quality data that can be obtained. The graph was made using the Highchart library.

Figure 3b shows the visualization of data on the map. Maps are built using the Leaflet.js library; there are markers that represent the location of data collection. On the sidebar there is a panel that shows the description of the marker color, the color of the marker is based on the color of the road quality monitor used by the government, the marker color represents the condition of the road by the PSI rules. The next panel shows the latest data on the quality of roads detected.

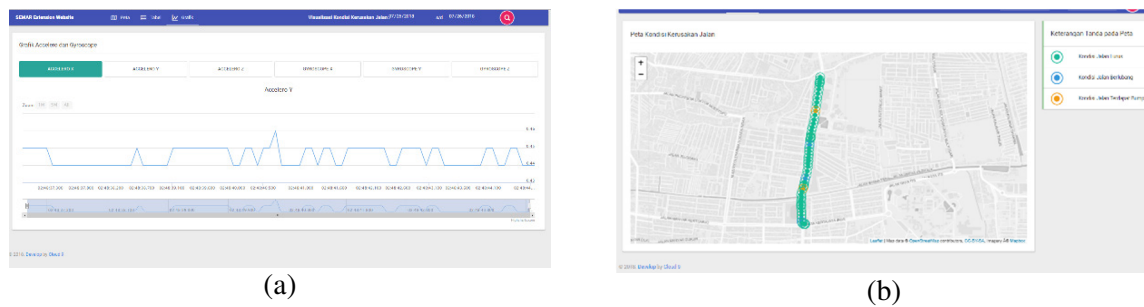


Figure 3: (a) Display of Akselero Graph Visualization and (b) Map Visualization Display

3.6. Collaboration and Processes

This section serves to provide a feedback to the users by providing information about parameters in the data derived from measurement data obtained. The system will send a notification if there are pothole and bumps. The notification about holes and protrusions in a visual form is a good feedback for the government to immediately repair the hole, and for the riders to avoid damaged roads.

4. Result and Discussion

Learning system has been applied to determine the accuracy of the proposed machine learning. First, data retrieval is done to build the data set from the machine learning.

4.1. Learning data retrieval

4.1.1. Data collection Pothole

Data retrieval set is done by retrieving the pothole data by attaching the device to the running car holder. Retrieving data sets is carried out using different vehicles at different speeds. Pothole data is obtained using Avanza car with a speed of 20 km/h, 40 km/h, and 50 km/h, and running location on the road Kertajaya Timur V and Jalan Gebang Lor.

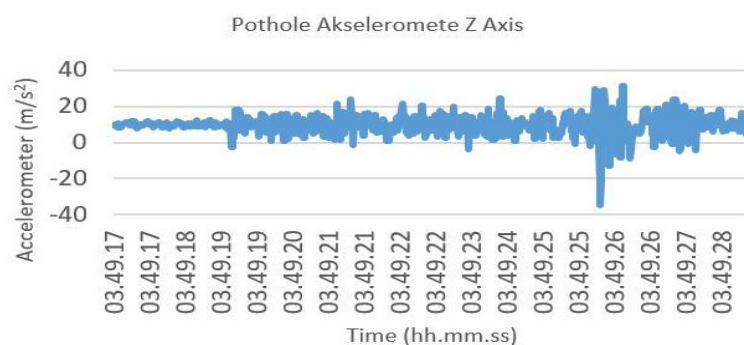


Figure 4: Pothole Accelerometer Z-Axis

4.1.2. Data collection set Bump

Bump testing is done to determine the characteristics of the data. Testing is done by taking the data set on the vehicle running and passing through the bump. Retrieval of data sets using different vehicles at different speeds is the testing environment for retrieving data set machine learning. Bump data is obtained using Avanza car with a speed of 20 km/h, 40 km/h, and 50 km/h, and running location on the road Kertajaya Timur V and Jalan Gebang Lor.

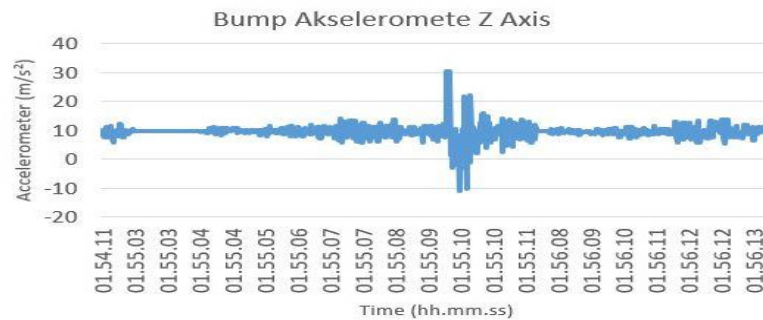


Figure 5: Bump Accelerometer Z-Axis

4.2. Results of Confusion Matrix

The confusion matrix is a method that is usually used to calculate accuracy in the concept of data mining. This formula calculates with three outputs, namely recall, precision, accuracy and error rate. In this test is done to test a machine learning algorithm that has been made. From the table of the confusion, matrix results can be seen errors from the predicted levels of the SVM and DT systems.

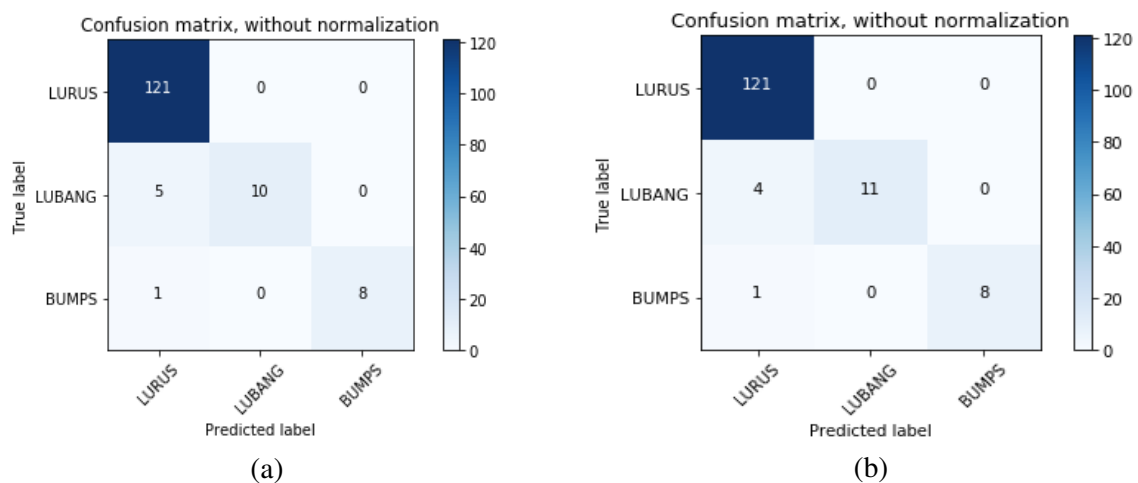


Figure 6: (a) Confusion Matrix from SVM and (b) Confusion Matrix from D3

The classification of SVM and DT in the calculation using Confusion Matrix shows that SVM has an accuracy rate of 98% with an error rate of 10 hole holes and eight bumps, while DT has an error value of 4 for holes and 1 for bump and eight bump readings and 11 holes.

Table 2. Classification Result

Features	Algorithm	MSE
Pothole and Bump	Support Vector Machine	0.06206896551724138
	Decision Tree	0.05517241379310345

4.3. Classification Result

Our second experiment measured the acceleration of the classification results. To measure acceleration, we calculate the accuracy and MSE (Mean Squared Error).

5. Conclusions

In this study, we built vehicles as a mobile sensor network (VaaMSN) and intelligent and analytic pothole detection in Real-Time Systems (SEMAR) for pothole monitoring systems. VaaMSN consists of modem accelerometer, gyro, GPS, 4G WiFi which is connected to the Head Unit and connected to the vehicle. Data from VaaMSN is sent via the MQTT protocol to the Big Data platform and analyzed by decision trees (DT) and machine support vector algorithms (SVM). Data is sent and visualized in real time. We conducted experiments by combining the functions of analytic data using the SVM Linear algorithm and also the decision tree algorithm. MSE values from the decision tree can be 6.2% and 5.5% for the SVM algorithm.

References

- [1] M. Pregnotato, A. Ford, S. M. Wilkinson, and R. J. Dawson, "The impact of flooding on road transport: A depth-disruption function," *Transp. Res. Part D Transp. Environ.*, vol. 55, pp. 67–81, 2017.
- [2] R. Bajwa, R. Rajagopal, P. Varaiya, R. Kavaler, and N. Street, "In-Pavement Wireless Sensor Network for Vehicle Classification Categories and Subject Descriptors," pp. 85–96, 2011.
- [3] A. F. Abdillah *et al.*, "Design and Development of Low Cost Coral Monitoring System for Shallow Water Based on Internet of Underwater Things," *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 2–5, pp. 97–101, 2017.
- [4] R. Arridha, "Classification extension based on IoT-big data analytic for smart environment monitoring and analytic in real-time system," 2017.
- [5] U. Bhatt, S. Mani, E. Xi, and J. Z. Kolter, "Intelligent Pothole Detection and Road Condition Assessment," 2017.
- [6] Z. Lv, H. Song, P. Basanta-Val, A. Steed, and M. Jo, "Next-Generation Big Data Analytics: State of the Art, Challenges, and Future Research Topics," *IEEE Trans. Ind. Informatics*, vol. 13, no. 4, pp. 1891–1899, 2017.
- [7] F. Panduman, Yohanes Yohanie Fridelin, Besari, Adnan Rachmat Anom, Sukaridhoto, Sritrusta, Budiarti, Rizqi Putri Nourma, Sudibyo, Rahardhita Widyatra, Nobuo, "Implementation of Integration VaaMSN and SEMAR for Wide Coverage Air Quality Monitoring," *Telkomnika Indones. J. Electr. Eng.*, vol. Vol 16, No, 2018.
- [8] C. Fernandez, R. Izquierdo, D. F. Llorca, and M. A. Sotelo, "A Comparative Analysis of Decision Trees Based Classifiers for Road Detection in Urban Environments," *2015 IEEE 18th Int. Conf. Intell. Transp. Syst.*, pp. 719–724, 2015.
- [9] A. Mednis, G. Strazdins, R. Zviedris, G. Kanonirs, and L. Selavo, "Real time pothole detection using Android smartphones with accelerometers," *2011 Int. Conf. Distrib. Comput. Sens. Syst. Work. DCOSS'11*, 2011.

Acknowledgments

This research is funded by KEMERISTEKDIKTI from "Penelitian Terapan Unggulan Perguruan Tinggi" in 2018 scheme with the number: 09/PL14/PG.1/SP2P/2018 and the title is "Implementation Vehicle as a Mobile Sensor Network terintegrasi dengan Smart Environment Monitoring and Analytics in Real-time(SEMAR) system sebagai Pemantauan Permukaan Jalan dan Lingkungan untuk mendukung Smart City".