Classification of Road Surface Quality Based on SVM Method

Adhelinia Afenika¹, P. H. Gunawan²*, D. Tarwidi³

¹,²,³School of Computing, Telkom University, Jalan Telekomunikasi No. 1 Terusan Buah Batu, Bandung 40257, Indonesia.

E-mail: ¹adheliniaaf@student.telkomuniversity.ac.id; ²phgunawan@telkomuniversity.ac.id; ³dedetarwidi@telkomuniversity.ac.id

*Corresponding author

Abstract. Damaged and potholes roads can occur due to rain puddles, too many heavy vehicles, poor asphalt quality, or maybe long road life. Damaged roads can hamper activities and endanger the safety of road users. It is necessary to monitor road quality periodically which is conducted by government, so that roads improvement can be done quick as possible. The aim of this study is to build a system that can classify roads surface quality. Support Vector Machine (SVM) classifications method is used to classify roads based on roadworthiness. In this study, 300 road surface data which contains good/smooth and damage quality of road are used. The simulation results show that SVM model can classify road surface data into two classes with average accuracy of 93%. The results can be a recommendation for government to prioritize which roads need to be improved.

1. Introduction
The road is one of the most important public facilities because it is always used for daily activities. As time goes by, the quality of the road will decrease and it can cause road damage. Road damage that can cause various disadvantages such as causing inconvenience and insecurity for road users. According to the data from World Health Organization (WHO), most occurred accidents is caused by poor road conditions [1]. Furthermore, bad impact of damaged roads can cause damage to vehicles, traffic jams, even road-traffic accidents. Therefore, it is important for the government to record road conditions routinely so that road damage can be handled properly in right away. One of the road condition records is classification of road surface quality.

Some researchers have been developed classification and roadworthiness using several ideas. For instance, Allouch [2], discussed machine learning algorithms that can predict the quality of roads. Data collection was performed using accelerometer sensor and gyroscope. In this study, support vector machine (SVM) method will be used to classify the condition of roads. In [3, 2, 4], the SVM method is easy and not expensive to implement. Moreover, it is known that SVM can be a good decision maker in two classes classification and has low error value.

In this study, data set were collected by using the Phypos application to gain road surface profile. Here, two classes will be identified for quality of road, which are good/smooth and damaged road. Roads that have holes and have uneven surfaces are labeled as damaged roads, while roads that have flat surfaces and no holes are labeled as good/smooth roads. There is also
mixture of some good and bad roads, and the data is always dominated by bad roads. Then the data is labeled as bad road because the road needs repairment. Examples of good and damaged road conditions can be seen in Figure 1.

![Figure 1: The good/smooth (left) and damaged (right) road illustration.](image_url)

In classifying road surface quality, several steps are elaborated, including data collection using the accelerometer sensor, characteristics finding of road surface using the value of the size of data centralization, implementation of SVM method, and performance analysis.

2. Classification method and its measurement

In this section, a brief discussion of binary SVM and measurements tool to see the performance analysis are presented.

2.1. Binary class of SVM

SVM (Support Vector Machine) is one of several methods in machine learning that can classify data sets into several classes in linear high-dimensional space [5]. Basically, the concept of SVM is trying to find the best hyperplane to linearly separating data. SVM has several models based on the number of classes in the classification process, one of them is Binary SVM Class. The SVM Binary Class is basic model SVM to solve two class classification problems. In the SVM Binary Class, there are two important concepts that must be understood, which are hyperplane and the concept of kernel tricks [6].

2.2. Performance analysis

Confusion matrix is a matrix that can compare the results of predictions with the actual conditions of the data and is used to analyze the performance of the implemented algorithm of Machine Learning. In this case, confusion matrix can be used to see the accuracy of the designed SVM model. Confusion matrix can be seen in Table 1.

| Predicted Positive Class | Actual Positive Class | Actual Negative Class |
|--------------------------|-----------------------|----------------------|
| Predicted Negative Class | TP (True Positive)    | FP (False Positive)  |
|                          | FN (False Negative)   | TN (True Negative)   |

where

- TP (True Positive) = the amount of positive data classified correct
- TN (True Negative) = the amount of negative data that is classified correctly
• FP (False Positive) = the amount of positive data classified incorrectly
• FN (False Negative) = the amount of negative data classified incorrectly

The accuracy formula of two classes using confusion matrix [7] [8] is

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100\% \quad (1)
\]

Generally, confusion matrix is used to measure the data testing. Data testing will be tested using the SVM model that has been built, then do a comparison between prediction classes and actual classes. Moreover, accuracy values were calculated using (1).

3. Methodology
3.1. Data Collecting

Road surface data was measured using the accelerometer in the phypox application on the smartphone, and the data was stored in Excel format. Phypox (Physical Phone Experiments) is an application which can be used in physics experiments[9]. Phypox has some sensor features, one of the sensor is known as accelerometer.

Each data is treated the same so that it has the same length and they are measured with the same speed and time. Data was taken by putting smartphone on the motorbike, then the accelerometer starts up, and the motorbike was driven on the road under study with intermediate speeds 25-27 km/h in 60 seconds. The collected data add up to 300 data consisting of 200 data of good roads and 100 data of damaged roads. The raw data sample is shown in Table 2.

![Table 2: Road surface data measured by the accelerometer](image)

| Time (s) | Linear Acceleration x (m/s²) | Linear Acceleration y (m/s²) | Linear Acceleration z (m/s²) |
|---------|-----------------------------|-----------------------------|-----------------------------|
| 0       | 0.6312                      | 1.7139                      | -3.2325                     |
| 0.0052  | 0.3743                      | -0.6714                     | -3.9856                     |
| 0.0104  | 0.2947                      | -0.5954                     | -3.2631                     |
| 0.0156  | 0.0741                      | -0.8343                     | -0.1597                     |
| 0.0208  | 0.1015                      | -0.4170                     | 1.3062                      |
| 0.0260  | 0.0593                      | -0.4888                     | 0.3853                      |
| 0.0312  | 0.4888                      | 0.3485                      | -0.3607                     |
| 0.0363  | 0.1500                      | -0.6875                     | -2.1655                     |

Figure 2: The coordinate system on the supine (left), upright (middle) and oblique (right) position of smartphone
Accelerometer is a sensor that can be used to measure acceleration. An accelerometer can be found easily, for example on a smartphone application. Process measurement using the accelerometer is fairly easy for people in general, users only need to put the accelerometer sensor on the object to be measured. The coordinate system of \(x\), \(y\), and \(z\)-axis on accelerometer changes according to the position of the used device [10, 11, 12, 13]. The coordinate system of \(x\), \(y\), and \(z\)-axis on the accelerometer can be illustrated in Figure 2.

Meanwhile, the example result of Phypox application in \(x\), \(y\) and \(z\)-axis charts on good/smooth and damaged roads are shown in Figure 3.

![Figure 3: Graphic of good/smooth (left) and damaged (right) roads on \(x\), \(y\), and \(z\)-axis](image)

### 3.2. Data processing

A centralized measure was searched for each data to find out the characteristics of each data. The characteristic of data is given by finding the statistical descriptive of data. Here, the centralized value size that was searched are the value in quartile 1, quartile 2, quartile 3, mean, minimum, maximum, standard deviation, and variance for each axis on the accelerometer.

### 4. Evaluation

#### 4.1. Analysis of result

An experiment was conducted on the SVM model that has been built by looking at the accuracy values in 4 scenarios which depend on axis. Here, it was done to find out which axis is the most influential in increasing the accuracy value. The results of the experiments are shown in Table 3. The accuracy is taken by average value in 10 times experiments.

| Scenario | Axis          | Accuracy |
|----------|---------------|----------|
| 1        | all axis \((x,y,z)\) | 0.90     |
| 2        | \(x\)         | 0.89     |
| 3        | \(y\)         | 0.91     |
| 4        | \(z\)         | 0.92     |

Based on the comparison of the accuracy values in Table 3, it can be acknowledged that the \(z\)-axis has an highest average score with accuracy value is 92\%. It is confirmed that the \(z\)-axis is the most influential axis among the other axis. Furthermore, the features that will be used in the SVM model is only the features on the \(z\)-axis.
4.2. Table of correlation

As mentioned in Section 3.2, that the features are some statistical descriptive measurements. Therefore, the correlation between each features should be investigated to increasing the accuracy by eliminating the lowest correlation effect. Correlation table can be used to see which features has an effect on improving accuracy value. Correlation table on road surface data is shown in Table 4.

| min z | q1 z | median z | mean z | q3 z | max z | sd z | var z | kelas |
|-------|------|----------|--------|------|-------|------|-------|-------|
| 1     | 0.66 | 0.3      | -0.13  | -0.89| 0.05  | -0.92| 0.63  | -0.79 |
| q1 z  | 0.88 | 1        | 0.47   | -0.064| -0.98| 0.82 | -0.95 | -0.86 |
| median z | 0.3  | 0.47     | 1      | 0.7   | 0.31  | 0.42 | -0.28 | -0.19 |
| mean z | -0.13| 0.064    | 0.7    | 0.21  | -0.053| 0.19 | 0.25  | 0.03  |
| q3 z  | -0.89| -0.98    | 0.31   | 0.24  | 1     | 0.79 | 0.97  | 0.89  |
| max z | -0.85| -0.82    | -0.42  | -0.053| 0.79  | 1    | 0.79  | 0.62  |
| sd z  | 0.92 | 0.95     | 0.29   | 0.19  | 0.97  | 0.79 | 1     | 0.96  |
| var z | -0.83| -0.86    | 0.19   | 0.25  | 0.89  | 0.62 | 0.96  | 1     |
| kelas | -0.79| -0.76    | -0.39  | -0.03 | 0.73  | 0.75 | 0.75  | 0.68  |

Two scenarios are conducted based on the correlation values in Table 4 to find the most influential features. The result of this experiment can be seen in Table 5.

| Scenario | Feature         | Information                                  | Accuracy |
|----------|-----------------|----------------------------------------------|----------|
| A        | min, q1, q3, max, sd, var | Correlation value with other features ≥ 0.80 (very strong correlation level) | 0.91     |
| B        | min, q1, mean, med, q3, max, sd, var | Correlation value with other features ≥ 0.60 (strong correlation level) | 0.92     |

* According to Table 5, it can be seen that scenario B has the highest accuracy value, which is 0.92. Moreover, all features used in scenario B are features in the z-axis and the accuracy is obtained from the average of 10 times experiments.

4.3. Tuning parameters

Looking for the most proper parameters can be done by tuning parameters. Tuning parameter was done by comparing three basic SVM kernels that are popular and often used are linear, polynomial, and RBF. Implementation of parameter tuning is done by comparing each kernel with three parameters, namely Degree (1, 2, 3, 4, and 5), Gamma (0.1, 1, 10 and 100), and Cost (0.1, 1, 10 , 100 and 1000). The parameter tuning program is designed to be able to compare all kernels with their parameters in one run program. The parameter tuning program is run 10 times and produces the conclusions shown in Table 6.

Based on the comparison of the accuracy values in Table 6 it can be seen that the most proper parameter in this experiment is the kernel polynomial with degree 1, gamma 10, and $C = 10$. 
Table 6: Tuning parameter

| Kernel Type | Degree | Gamma | C(Cost) | Accuracy |
|-------------|--------|-------|---------|----------|
| Poly        | 1      | 100   | 100     | 0.92     |
| Poly        | 1      | 100   | 1       | 0.92     |
| Poly        | 1      | 10    | 10      | 0.95     |
| Poly        | 1      | 10    | 1000    | 0.93     |
| Poly        | 1      | 10    | 100     | 0.93     |
| Linear      | 1      | 0.1   | 100     | 0.93     |

4.4. K-fold cross validation

In this Section, k-fold cross validation is applied to test whether the results obtained are representative or not. The steps taken are dividing the data into several folds of the same size, then 1 fold is used for testing and the rest is used for training. In general the cross validation value used ranges between 3-5. This study has been conducted by using 3-fold cross validation because it was considered the most representative with high accuracy is around 93%. A cross validation value that is too large causes the data per fold to be smaller and cross validation score to be very sensitive and it may not provide representative results.

5. Conclusion

Classification of road quality has been done by using SVM method with high accuracy is around 93%. Classification was divided into two classes to find out whether the road surface is worth using or need a reparation. Data sets was collected by using the accelerometer sensor to measure road surface vibrations on the x, y, and z-axis to get the profile of road surface shape. The classification process was conducted by using 300 data to optimize the performance of the SVM model. From the simulation, it is found that the axis on the accelerometer with the most influential on the performance results was the z-axis. All features on the z-axis have an important influence in determining performance results. Moreover by tuning parameter, the best parameters are found that polynomial kernel with degree 1, gamma 10, and $C=10$.

References

[1] Organization W H 2015 Global status report on road safety 2015 (World Health Organization)
[2] Allouch A, Koubia A, Abbes T and Ammar A 2017 IEEE Sensors Journal 17 4231–4238
[3] Nakano T, Nukala B T, Tsay J, Zupancic S, Rodriguez A, Lie D Y, Lopez J and Nguyen T Q 2017 International Journal of Software Innovation (IJSI) 5 17–29
[4] Mathur A and Foody G M 2008 IEEE Geoscience and remote sensing letters 5 241–245
[5] Andrew A M 2000 Robotica 18 687–689
[6] Suyanto 2018 Machine Learning Tingkat Dasar dan Lanjut (Informatika Bandung)
[7] Townsend J T 1971 Perception & Psychophysics 9 40–50
[8] Octaviani P A, Wilandari Y and Ispriyanti D 2014 Jurnal Gaussian 3 811–820
[9] Staacks S, Hütz S, Heinke H and Stamper C 2018 Physics Education 53 045009
[10] Fernandez A and Dang D 2013 Getting started with the MSP430 launchpad (Newnes)
[11] De Marsico M and Mecca A 2017 Human Recognition in Unconstrained Environments (Elsevier) pp 177–195
[12] Mednis A, Strazdins G, Zviedris R, Kanonirs G and Selavo L 2011 2011 International conference on distributed computing in sensor systems and workshops (DCOSS) (IEEE) pp 1–6
[13] Brezmes T, Gorricho J L and Cotrina J 2009 International Work-Conference on Artificial Neural Networks (Springer) pp 796–799