Automatic labeling of road quality using machine learning

STEVEN TAZELAAR, University of Twente, The Netherlands

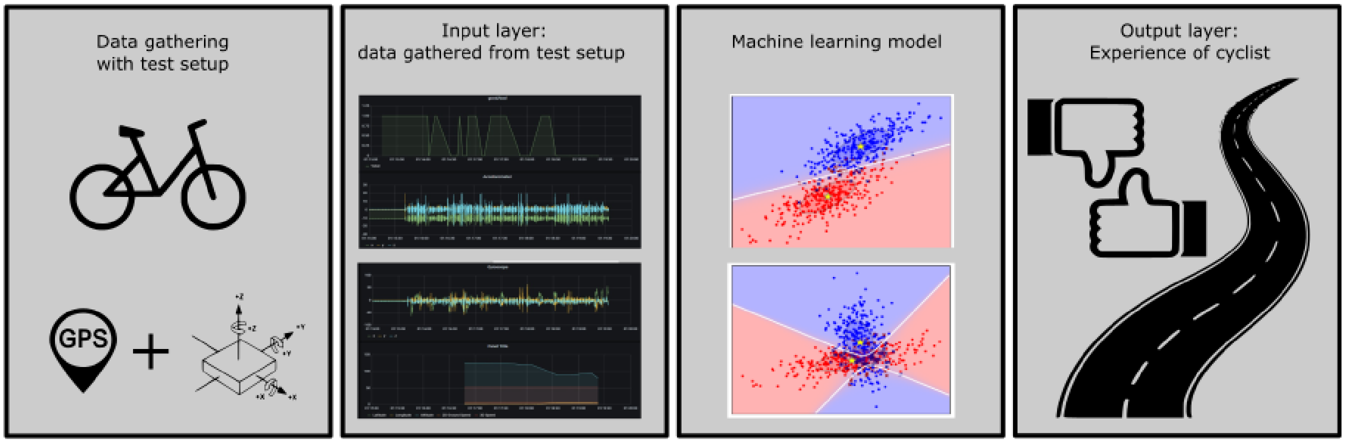

Fig. 1. Schematic overview of the research. First a data collection setup consisting of GPS and IMU sensors will be implemented on a bicycle. Then data will be obtained by the sensors and this will be gathered for analysis. On this data two machine learning classifiers, Linear Discriminant Analysis and Quadratic Discriminant analysis, will be trained to classify road quality, on which finally, an evaluation and validation will be performed.

The goal of this research is to create a software that uses machine learning to accurately label road quality when cycling to get an overview of the roads that are in need of improvement or are being experienced as a good road.

A data collection setup was made, using IMU and GPS sensors connected to a Raspberry Pi with a connection to bucket. The user labels roads as good or bad using two buttons on the bike handlebars. Two machine learning algorithms are applied to the data: LDA and QDA. A validation was performed on a road of which the model did not get a label as input.

LDA ad QDA lead to an accuracy of 0.83 and 0.82 respectively. A specificity of 0.66 and 0.72 was reached respectively and both had a sensitivity of 0.92. During the validation, all three types of roads that were measured got a classification that matched their label, validating the model.

Further research could improve the model, for instance use of more sensors, more data to feed to the model, application of other machine learning models and other methods of labeling data.

The LDA and QDA machine learning algorithms both have the potential to automatically label road quality in cycling.

Additional Key Words and Phrases: Cycling, Road Quality, Machine Learning, IoT, Application

1 INTRODUCTION

In the Netherlands 28% of the travel movements were made by bike. On average 1098 KM was cycled per person and these numbers are only expected to grow[2]. The goal of this research is to create a software that uses machine learning to accurately label road quality when cycling to get an overview of the roads that are in need of improvement or are being experienced as a good road.

One of the influences on cycle experience is the quality of the surface that is biked on. [1] To help the government and municipalities determine which roads to improve, a repeatable study on the current road quality situation is needed. Currently, methods to determine road quality are performed manually. To save time and effort, an automatic method to determine the road quality would be of value. This would enable further research into specific situations, like where to improve roads or what actually makes a road better or worse to cycle on.

This leads to a more concrete and specific goal of this study: Enabling research into improving the experience of future cyclists. To achieve this goal a study on automatic labeling of road quality is needed.

1.1 Research questions

The main question of the research is: How can automatic labelling of road quality be applied to a cyclist based on sensor data?

To answer this question the following sub questions were devised:

(1) How can road quality be measured while cycling?
(2) what is a possible method of collecting labeled data on road quality?
(3) How can machine learning play a role in analyzing the measurements?
   (a) What are relevant machine learning methods?
   (b) Is the automatic labeling of cycling experience using machine learning a useful method?

TScIT 37, July 8, 2022, Enschede, The Netherlands

2 RELATED WORK

Earlier research has been conducted on cyclists in a similar fashion as proposed for this study. Erdei et al. [3]. created a similar labeling setup to the one proposed in this research in the way that they are also making use of button presses. Also a GPS is used as one of the data sensors. A difference with this study is that it consists of a 'classical analysis' while the proposed setup proposes a focus on labeling instead of a repeatable experience/reaction measurement study.

Machine learning has been applied to many different medical fields. The book created by Paliouras et al. [4] shows an overview of current machine learning methods and techniques. The section by Magoulas on Machine Learning in Medical Applications is interesting because of the link between psycho-metrics and machine learning that might also be applied in this study.

The study on bike road conditions conducted by Peng et al. [5] uses shared bikes fitted with an Inertial Measurement Unit (IMU) to gather data that is then used to train a machine learning algorithm to detect different road surfaces (Asphalt, Pebbles, Bumpy path). Since this paper uses machine learning on similar data, e.g. IMU data, a similar algorithm is used in this research to see if it is relevant for this purpose as well.

3 METHODS

3.1 Data collection

3.1.1 *What to measure.*

Both Peng et al and Calvey et al state that vibration is a major indicator of road quality. [5][1] This can be measured using an accelerometer for measuring change in velocity and a gyroscope for measuring angular velocity. Both will be used for determining the quality of the road.

A GPS sensor will also be used as an additional layer of validation because this enables a link between gathered data and a physical location that can be manually checked.

3.1.2 *Data collection setup.*

To be able to conduct a relevant analysis on cycle data, data is needed. So a data collection setup was built. The requirements of the setup are based on necessities for being able to derive relevant road quality data while cycling:

- The setup should be mobile
  - No external power connection should be needed.
  - The setup should not rely on a wired or WiFi based internet connection.
- The setup should collect relevant data
  - The setup should collect rotational data (gyroscope)
  - The setup should collect acceleration data (Accelerometer)
  - The setup should collect locational data (GPS)
- The setup should upload collected data to an external location
- The setup should be able to be used with minimal understanding or explanation of the setup
- The setup should be easy to reproduce

To meet these requirements, a Raspberry Pi based IoT solution was build. This enabled the use of compatible sensors such as an IMU (Inertial Measurement Unit) which contained both a gyroscope and an Accelerometer and a GPS sensor. Data read out from these sensors is pre-processed by the Raspberry Pi before being send over the internet to a cloud data collection platform developed by the TUDelft called Bucket. Bucket then stores the data to be analysed later. The whole setup is powered by a mobile phone powerbank and has an internet connection based on the mobile network (4G). Because the setup is all based on an existing platform and supplies for mobile devices, it easily reproducible.

This all is attached to a bike as shown in figure 2.

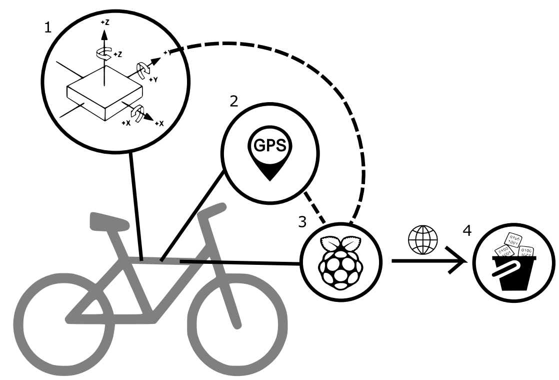

Fig. 2. A schematic image of the set up. An IMU (1), GPS (2) and Raspberry Pi (3) are mounted to a bike. The IMU and GPS send sensordata to the Raspberry Pi, which preprocesses the data before being send over the internet to a cloud data platform called Bucket (4).

The setup described in 3.1.2 collects all the useful data but misses the ability to label said data.

3.1.3 *Data labeling.*

The functionality of labeling the sensor data is added by attaching two buttons that can be easily pressed when biking to the bike handlebars. They are connected by wires to the Raspberry Pi. One for labeling the current data as a good road surface and one for bad road surfaces. The event of a button press triggers the Raspberry Pi to send a corresponding data point to Bucket that can later be used for analyzing the gathered data.

3.1.4 *Data collection.*

The setup was then used to collect data. To combat user-bias, seven people were asked to cycle to help collect data. All students or researchers at the University of Twente The participants were asked to bike like they normally do and to press a button when the road quality was good or bad in their opinion. By the use of multiple participants this should average out differences in opinion and thus eliminate bias. The participants cycled in Enschede, mostly near the city center and in nearby neighbourhoods.

3.2 Machine learning

3.2.1 *Feature extraction.*

To be able to generate features, a subset of the data is needed. For training purposes each data slice should be associated with a label. While the final goal is to have no need for given labels and have labeling based on sensor data. To meet both these requirements, a sliding window approach was chosen. For the training data each window is selected based on a label event while the resulting algorithm will use a normal sliding window strategy. The size of this window is 2 seconds of collected data. The window size is chosen based on an average reaction time following an event for which the cyclist might decide to label data. For the training data, the labeling event will signal the end of the window

From the data collected in each window. features are extracted. These are all extracted on all axis of both the gyroscope and the accelerometer. This results in six axis. Based on Peng et al the feature selection for each axis is: mean, median, standard deviation, maximum value, minimum value, skewness, kurtosis, slope sign change, mean of frequency and median of frequency. Both mean of frequency and median of frequency are derived from analysis with the Fast Fourier Transformation (FFT). [5]

For evaluation, three metrics will be used: accuracy (the probability of a correct prediction), specificity (the probability of a negative prediction, given that the road quality is negative) and sensitivity (the probability of a positive prediction, given that the road quality is positive). The accuracy gives a more general view of how well the algorithm can predict a label. The specificity and sensitivity give a view on how well the bad and good roads are being classified respectively. In an ideal situation, all approach 1. These three metrics provide an overview of the quality of the predictions.

3.2.2 *Data analysis.*

For analyzing the data, a comparison between classic classifiers is made. These classifiers are Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA), which are illustrated in figure 3. These are chosen because of their relatively low amount of data points needed (about 50) and because they handle many features quite well. LDA is based on projecting all measurements into a plane which is then reduced such that the distance between labels is maximized. In this space a linear line is then trained to differentiate between 'good road quality' and 'bad road quality' data points. This classifier can then be tested against a set of data not used for training. This process is the same for a QDA but the decision line is not linear but quadratic. The difference between LDA and QDA is also visualised in figure 3 This will probably result in a line that fits the data better but is thus also more prone to over fitting. [6]

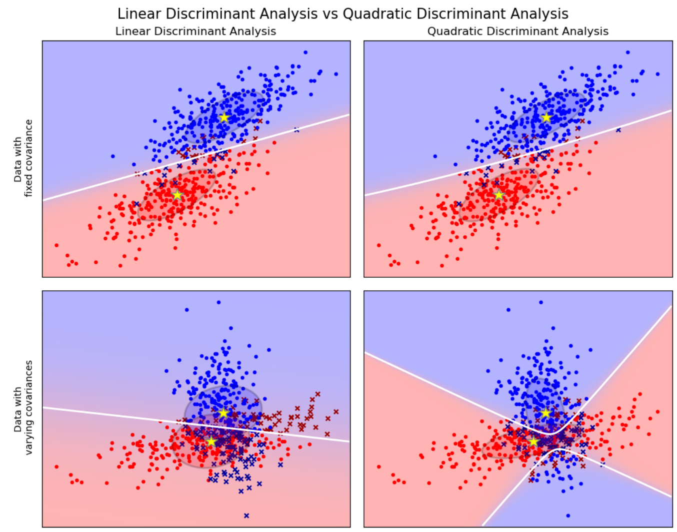

Fig. 3. Machine learning algorithms based on Bayes theorem. The machine learning model learns to find the function that can be used to classify data points, in this case sensor data to a certain class, in this case good or bad road quality. [6]

3.3 Validation

This classifier can be used to determine any further data collected. This will be done to both test and validate the classifier. This will be done by manually checking the classifier on my personal opinion and the opinion of others. The classifier can also be compared to official road quality data from the municipality.

4 RESULTS

4.1 Data collection

The collection of data was a success, in total seven participants have all collected a similar amount of data with one participant collecting significantly more (being the author of this research) but not more then half of the collected data. In total 20.000 accelerometer and gyroscope data points were gathered. This totals in approximately 120.000 measurements. These points were labelled by 583 collected labels. In total 317 windows were constructed to train and test the machine learning algorithm. Of these 327 were positive while 255 were negatively labelled. Not all labels were useful for training and testing because they were to close to another label or had less than 10 data points on each axis and were thus not information dense enough to apply a frequency analysis on.

4.2 Machine Learning

Both the LDA and QDA algorithm were trained using the features described in section 3.2.1. A training/test ratio of 0.8/0.2 was chosen. This resulted in 263 training windows and 54 test windows. After training of the 263 windows, a test was conducted on the 54 test windows. The outcome for LDA is provided in table 1 with the same for QDA in table 2. The same windows were used for both algorithms

		Ground truth	
		Positive	Negative
Prediction	Positive	33	6
	Negative	3	12

Table 1. Results of an LDA machine learning algorithm visualized in a confusion matrix.

		Ground truth	
		Positive	Negative
Prediction	Positive	33	6
	Negative	3	12

Table 2. Results of an QDA machine learning algorithm visualized in a confusion matrix.

	LDA	QDA
Accuracy	0.83	0.85
Specificity	0.66	0.72
Sensitivity	0.92	0.92

Table 3. Metrics comparing the LDA and QDA algorithm

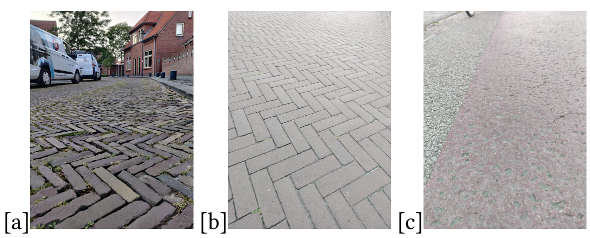

Fig. 4. (a) An old paved road (b) A recently paved road (c) Asphalt

With these numbers, the metrics shown in table 3 can be calculated. This shows that both algorithms are quite similar in accuracy and sensitivity but differ in specificity.

4.3 Validation

To validate the algorithm, three locations with different road surfaces were selected: An old paved road, a recently paved road, and an asphalt road as can be seen in figure 4. On all surfaces data was collected and classified by the before trained algorithms. The classifiers both marked the old paved road as bad road quality and both of the others as good quality. With this I completely agree. The asphalt and recently paved road were both quite smooth and did thus not generate many vibrations or big shocks while this could not be said for the old paved road.

5 DISCUSSION

While the study conducted was a success, there are some possibilities for further research. For instance, the use of more sensors for data gathering could add more reliability to the outcome. For example: the speed of the cyclist might influence the level of shock or the amount of vibration but this is now not possible to be taken into account by the algorithm.

Also, a more advanced method of gathering labels might be looked into. The researched setup allowed for 'good' and 'bad' but the actual experience of road quality could be more defined.

Like many experiments, the gathering of more and more varied data would have helped make the algorithm less biased and also combat over fitting. This could be done by asking more people to cycle on the bike for longer distances and in other places than Enschede.

Further research could also apply the methods described in this paper to more than road quality. For instance, the experience of the cyclist in general.

Finally, more algorithms could be looked into. With more data, linear regression, decision trees, and nearest neighbours could both be applied to this type of data.

6 CONCLUSION

The application of machine learning on the gathered cycle data was successful but can also be improved. The build setup proved able to collect and label data as stated in the requirements. Both researched machine learning algorithms were able to produce an accurate validated classifier. If the focus of the application is to correctly classify bad road conditions, an application of QDA is advised because of its higher specificity while maintaining a high accuracy and sensitivity

REFERENCES

[1] J Calvey, J Shackleton, Mark Taylor, and Richard Llewellyn. 2015. Engineering condition assessment of cycling infrastructure: Cyclists' perceptions of satisfaction and comfort. *Transportation Research Part A Policy and Practice* 78 (06 2015), 134–143. https://doi.org/10.1016/j.tra.2015.04.031

[2] CBS. 2019. *Hoeveel fietsen inwoners van Nederland?* Retrieved May 8, 2022 from https://www.cbs.nl/nl-nl/visualisaties/verkeer-en-vervoer/personen/fietsen

[3] Elke-Henriette Erdei, Jochen Steinmann, and Carmen Hagemeister. 2020. Comparing perception of signals in different modalities during the cycling task: A field study. *Transportation Research Part F: Traffic Psychology and Behaviour* 73 (2020), 259–270. https://doi.org/10.1016/j.trf.2020.06.011

[4] George D. Magoulas and Andriana Prentza. 2001. *Machine Learning in Medical Applications.* Springer Berlin Heidelberg, Berlin, Heidelberg, 300–307. https://doi.org/10.1007/3-540-44673-7_19

[5] Daiyan Peng, Zach Strout, Shuo Jiang, and Peter Shull. 2019. A Road Condition Classifier via Lock Embedded IMU on Dock-Less Shared Bikes. In *Proceedings of the International Conference on Industrial Control Network and System Engineering Research* (Shenyang, China) *(ICNSER2019).* Association for Computing Machinery, New York, NY, USA, 32–36. https://doi.org/10.1145/3333581.3333597

[6] scikit-learn developers. 2022. *Linear and Quadratic Discriminant Analysis.* Retrieved june 26,2922 from https://scikit-learn.org/stable/modules/lda_qda.htmllda-qda