Cyclist Weight Inference from Bicycle-Mounted Sensor Data

Remy Benitah
r.j.benitah@student.utwente.nl
University of Twente
Enschede, The Netherlands

ABSTRACT

As crowd-sensing infrastructure becomes increasingly widespread, researchers are developing technologies such as bicycles with sensors for providing information about road quality. These technologies can accomplish this by processing data collected by the sensors attached to the bicycles. However, while location sensors and cameras are widely considered as privacy-sensitive data sources, seemingly innocuous sensors like accelerometers might also leak sensitive information such as the cyclist's weight. This research aims to investigate if sensitive data, such as the cyclist's weight, can be extracted from these seemingly innocuous sensors. First, it will consider the positioning of the sensing hardware devices on the bicycle. Next, users with varying weights will test the bicycle under controlled conditions. Finally, it will analyze the data and implement machine learning solutions to determine if the user's weight can be inferred. This research is expected to contribute to the knowledge of privacy considerations in the field of opportunistic and pervasive sensing. This especially true as crowd-sensing infrastructure becomes increasingly widespread and technologies such as these are developed.

CCS CONCEPTS

• **Computer systems organization** → **Sensors and actuators**; • **Security and privacy** → *Privacy protections.*

KEYWORDS

privacy, sensitive insight, bicycle, accelerometer, weight, inference, data analysis, road quality

1 INTRODUCTION

Crowd-sensing is a type of data collection that relies on the participation of a large number of individuals to gather data from various sources. This data is then used to gain insights into various phenomena such as traffic patterns, environmental conditions, and even human behavior. Bicycles can be considered as a mobile sensor platform, as they are able to move through different areas, collecting data on various aspects of the environment. Crowd-sensing applications can be beneficial for a wide range of stakeholders, including governments, businesses, and researchers. However, it is important to ensure that the data collected is used in a responsible and ethical manner. This includes considering the privacy and security of personal data, as well as addressing any potential biases in the data collection process. As the use of bicycles as a data collection platform becomes more prevalent, it is important to investigate the potential implications of this trend and to develop methods for ensuring the responsible and ethical use of this data.

In 2016, about 25% of daily transportation occurred via bicycle in the Netherlands [7]. Thus, we must maintain the infrastructure supporting cyclists to ensure safety, efficiency, and comfort. However, it is difficult and timely to manually measure the quality of roads. This is because professional inspectors must perform a check according to a strict manual [1]. To solve this, we can collect data about road quality by using sensors placed on bicycles instead of manually inspecting the roads. This solution for road quality monitoring is a form of a crowd-sensing application. With this, municipalities can gain insights into road quality conditions through people traveling via bicycle in an opportunistic way [2].

To provide some key background information, a distinction between personal information and sensitive information [2] will be provided. Personal information is quite a broad term which includes any information relating to an individual or someone who can be identified without difficulty. Personal information includes name, date of birth, and address. Sensitive information refers to any personal information that can harm an individual if not handled properly; for example religion, political beliefs, and sexual preferences. This is due to the fact that if others learned that info, then it could be used in some manner to bring harm.

Value Sensitive Design (VSD) is a paradigm which can help in safeguarding the private and sensitive information of an individual. It calls for human values to be translated into design requirements. Often times, designers already include in this implicitly, even when it may not be their main focus [13]. However, VSD is a framework that would benefit all stakeholders if it becomes more of a conscious decision. Certainly not all end users share the same values, or even if they do, they may value them to varying degrees. But when a user's values have clearly been considered, then that can be reflected in the success of a product.

The main objective of this study is to conduct proof-of-concept research to investigate the feasibility of inferring the weight of a cyclist from IMU bicycle-mounted sensors. We will create machine learning algorithms to test the limits of inferences with this sensor data. This will ensure the protection of cyclists using bicycles with sensors. The paper has been structured in nine different sections: Section 3 presents the related work, Section 4 the research background, and Section 5 highlights the methodology. In Section 6, the

TScIT 38, February 2023, Enschede, NL

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

[1] Inspection Manual
[2] Bicycle Lights with sensors

different experiments conducted will be explained; of which the results can be found in Section 7. Finally, in Sections 8 and 9, the discussions, conclusions, and future work can be found.

2 PROBLEM STATEMENT

The current state of research on the collection of data using bicycles is limited. And we need to investigate the use of motion sensors mounted on a bicycle to estimate the weight of a cyclist. Meanwhile, researchers have conducted research on sensitive insights from smartphones [3, 8]. The use of accelerometer sensors on bicycles has the potential to provide a wide range of insights based on the movement of the bicycle. These include the cyclist's weight, speed, and tire pressure. However, there is currently no known literature that can demonstrate a correlation between this data and the cyclist's weight. Or even to infer sensitive information from the data. Protecting an individual's privacy can increase user satisfaction and prevent GDPR regulation problems.

From research, we suspect that accelerometer sensors provide a wide range of insights based on the bicycle movement. Empirically, we should correlate the data with the cyclist's weight, speed, and tire pressure. There is no known literature that can demonstrate this correlation, or even to infer sensitive information from the data. We will investigate this problem by trying to estimate the cyclist's weight. This will further motivate the need for more privacy-sensitive methods in crowd-sensing applications.

One issue related to the use of motion sensors on bicycles is the issue of accuracy. While the data collected by the sensors may provide insights into the cyclist's weight and other metrics, it is important to ensure that the data is accurate and reliable. This is especially important in a scenario where the data is used for medical or insurance purposes. This is because inaccurate data could lead to incorrect conclusions about an individual's health or insurance rates. Additionally, it's important to consider the potential for bias in the data collection process. Factors such as the type of bicycle or the rider's physical characteristics may affect the accuracy of the data. Thus, it's necessary to investigate and validate the accuracy of the sensor data before using it to infer sensitive information about the cyclist.

2.1 Research Question

In order to solve the issues mentioned in the problem statement; investigate if accelerometer data can be used to infer sensitive information of a cyclist, a research question has been constructed, that will be the basis of this research work:

How can the cyclist's weight be inferred from IMU sensors mounted on a bicycle?

This leads to the following sub-questions:

(1) In which positions/orientations should the sensors be fastened to the bicycle?
(2) How can the cyclist's weight be extracted from the sensors?
(3) How does the system perform in terms of accuracy?

3 RELATED WORK

In this section we will go over some of the related work in the area of sensitive information inference. We will also discuss some literature that uses sensors on bicycles for various purposes.

Kroger et al. [6] introduce accelerometer data collection and the accompanying potential for inference. They discuss the possible damaging implications of tracking factors such as identification, health, and even activity. In a more specific paper from 2022, Naval et al. use smartphone sensors to reveal personal four-digit PIN numbers by password input movement [8]. They accomplished by using a machine learning algorithm and the motion sensors which give insight into how the phone shakes, tilts, and moves. In a 2021 paper, Tahir et al. researched how wearable sensors on the wrist can be used to recognize human activities [12]. These publications are examples of inference attacks using accelerometer data. They can be used to inspire the data processing and model creation methodology of this paper.

Table 1: Inference using related sensors in literature

Sensor(s)	Information	Source
Accelerometer+ Motion Sensor	PIN of Smartphone	Naval et al. [8]
Accereometer+ Speaker	Speech	Anand et al. [1]
Accelerometer+ Gyroscope	Human Activity	Hernandez et al.[5]+ Tahir et al. [12]

Research has also been conducted into the placement of sensors (such as an IMU) on a bicycle. Springer et al. [11] have also conducted a similar study, measuring road quality. In their paper, they deeply consider the architecture of the sensing device used to gather information. They design an IMU sensor to measure vertical acceleration at a rate of 50Hz, and place it on the handlebar of the bicycle. In addition, they mount a sensor box to the frame of the bicycle (between the seat and handlebars) that contains an additional IMU sensor among others. We will use these positions in the design of the experimental methodology to discover the most appropriate placement for the IMU sensor. A similar paper by Owens et al. [9] employs a sensor package over the front wheel attached to the bicycle head. One of the goals of that paper was to measure kinematic road data, which is the same type of data that we will use in this paper. Finally, Patil et al. use an IMU sensor with both an accelerometer and gyroscope to measure the roll of a bicycle [10]. While this paper will focus on vertical acceleration, it is useful to know how the sensor package was attached to the system in related work.

In summary, we will use previous examples of sensor usage to extract information as a guide for our methodology. In deciding how to place sensors for this research, we will consider similar studies of bicycle-mounted sensor placement.

4 METHODOLOGY

This section will detail the methods used in order to perform this research project. In order to address the main research question– How can the cyclist's weight be inferred from IMU sensors mounted on a bicycle?–the following steps were taken. First, we performed

Table 2: Relevant Thingy:52 Specifications

Device	Thingy:52
Data streaming	BLE
Sampling rate (hz)	5-200
Powered	1440 mAh rechargeable battery
Sensor parameters	x,y,z acceleration axes

a literature review in order to find the optimal placement for the sensor used. We followed this closely with an investigation on collecting and pre-processing the sensor data. Two main approaches explored the second sub-question proposed in Related Work. First, we performed a manual inspection of the collected data in order to find ways to estimate the weight of a cyclist. Next, we explored a machine learning approach through literature and tuning for the purpose of proof of concept. Finally, we evaluated the success of these approaches in order to find the most accurate and feasible approach to weight inference.

4.1 Sensor Configuration

In this section, we will explain the methodology we followed to find the best sensor placement, configure it, and collect/pre-process data.

4.1.1 ***Sensor Placement.*** The sensing device chosen for this project is the Nordic Thingy:52. See this Table 2 for some specifications of this sensing device.

As mentioned in Related Work, similar research using bicycles with sensor for various crowd-sensing purposes have been conducted. Sensing devices have been placed on the handlebar [11, 9] or on the front wheel [4]. Each of these studies yielded promising results, justifying their reasons for sensor placement. However, a key difference is that those studies typically focus on road measurement data, while the purpose of this research is on the cyclists themselves. We found that most of the weight of a cyclist is located in the back of the bicycle; on the back wheel in particular [3]. Thus, we decided to place the sensor on the back wheel under the cyclist as that is where the weight should have the most impact on the measured acceleration. Figure 1 shows the end configuration of the sensor on the bicycle. The sensor is attached next to the back wheel. We place a phone on the bicycle that is easily accessible to the cyclist, and is also close enough in proximity for BLE to transfer captured data.

4.1.2 ***Data Collection & Pre-processing.*** Because the Thingy uses BLE, we decided to create an android application in order to facilitate the data collection process. The first step was to scan for the Bluetooth device and establish a connection. This required usage of Nordic Semiconductor's Android Library[4]. After this step, we sent configuration metrics such as the desired sampling rate of 200Hz from the android application to the sensing device. Because a Bluetooth connection can sometimes be dropped, we displayed the status of the connection.

Once we had established a secure connection, we then collected data with some pre-processing. The first step of this pre-processing

[3]Investigating Weight Distribution on a Bicycle
[4]Android Nordic Thingy Git Repository

Figure 1: Bicycle System with Thingy Sensor and Phone

was calibration of the accelerometer sensor. We observed that the accelerometer data had noise (or fluctuation), even when it was in a standstill with no motion. Therefore, before collecting data, it is necessary to keep the sensor still and allow it to collect a noise average for each axis. This will then subtract from any values observed in the data.

Additionally, we calculated the magnitude of the acceleration vector during collection and sent to the mobile device alongside the raw data. The magnitude was calculated using Formula 1. The magnitude of the acceleration allows for an observation of the overall amount of acceleration that is acting on the sensor. This is a valuable feature for this type of data. This is calculated using the x, y, and z axes from the accelerometer.

$$|\overline{\alpha}| = \sqrt{x^2 + y^2 + z^2} \tag{1}$$

After pre-processing, the BLE connection constantly sent data to the mobile device with the timestamp. However, the device only saves data from when the rider signals the beginning to the end of data collection. When the rider presses the button to end data collection, the data from this interval compiles into a .csv file and saved to the storage of the android device.

This section discussed the methodology for where the sensor was placed. We conducted a literature review after weighing the benefits of using the Nordic Thingy:52 device. We then decided that based on previous research and information about the weight distribution on a bicycle, the sensor should be placed on the back wheel under the cyclist. This placement should optimize the influence the rider has on the measured accelerometer data. Additionally, we've shown the methods used for collecting data with the sensor and pre-processing data were also shown.

4.2 Weight Inference

In this section we will discuss the two methods through which we attempted to estimate the weight of a cyclist using the collected data. The results obtained from these two methods will show whether it is indeed possible to infer the weight of a cyclist using mounted sensors.

Table 3: Weight ranges corresponding to classes

Class	Weight (Kg)
low	63-79
avg	80-86
high	87-110

4.2.1 ***Data Inspection & Analysis.*** First, we performed an inspection and analysis of the data. This approach allows us to investigate whether it is possible to infer weight without machine learning and instead by doing it in an algorithmic fashion. We created graphs and charts using parts of the data set and looked for relationships and correlations. For example, by plotting the x (vertical) axis of the accelerometer data and the weight of several participants, we may be able to observe a trend that could be used to create a function for weight estimation. However, even if it is not possible without machine learning, this method will give us insights into which parts of the data set can be used as valuable features for building the models.

4.2.2 ***Machine Learning.*** The first step towards implementing a machine learning algorithm was deciding which type and model to use. There are many applications of machine learning, such as: regression, classification, and clustering. For the purpose of this research, the task of regression would be likely to output the most accurate estimation of an individual's weight (assuming that it is possible). However, due to the limited scope of this project, we decided to instead use classification. This is because the purpose is not to create a highly accurate estimation of a cyclist's weight, but rather to investigate if it is at all possible.

The classes that were used can be seen in Table 3. We chose the ranges of weights such that each class had a similar amount of samples based on the final data set. We decided to use three classes as that should be sufficient to show that there is a relationship between the accelerometer data and weight (if the model can accurately classify the weight of an individual in these classes). After identifying which classes to use, we chose four classification models from literature that have distinct strengths and weaknesses that can apply to this data set. First, we implemented logistic regression[5]. Logistic regression is a very strong classifier for time series data, and when we extracted more quality features that are independent, the classifier performed very well. Second, we implemented the K-Nearest Neighbours classifier. This model is attractive due to its simplicity and ability to handle large amounts of training data. The drawback of KNN having a high computation time can be disregarded in this research, as there is no time constraint for this computation. Next, we implemented a Support Vector Machine using a linear function. Finally, we implemented a decision tree. We implemented them in Python using the scikit-learn library.

After selecting the classification models, we split the data into features which we tuned through manual inspection. We also split the data into two-second windows of 400 points (200 points per second) with step sizes of 200 points. We labeled these windows by the most-occurring category of weight present in that interval. By splitting the data set into windows, we artificially created a larger sample of data for training the models. This allowed for higher performance in terms of accuracy by these classifiers.

Finally, we also split the data into a training set and a testing set. To maximize performance, we split this data set such that there was an equal proportion of similar data in each set. We split the data into 3:1 training to testing data. We did this by using the first eight participants for training, and the last four for testing.

To summarize, we used two methods to estimate the weight of a cyclist. The manual inspection of the data had the potential to infer weight without machine learning. It also provided necessary information for tuning and increasing the performance of the ML that we implemented. We chose classification due to the scope and purpose of this paper as a proof of concept.

[5]Example logistic regression implementation using time series data

Table 4: IV/DV

Independent Variable	Unit	Dependent Variable	Unit
Weight of Cyclist	Kg	Acceleration	g (9.8m/s^2)

4.3 Inference Evaluation

To determine if we can accurately obtain insights into this sensitive data, we must show consistent performance in realistic conditions. We evaluated each of the models against one another to see which one performed best. The metric we used to compare the models was their classification accuracy. As previously mentioned, there was no time constraint for this calculation. So we did not consider the memory and time costs of the models in this evaluation.

5 EXPERIMENT

This section will describe the experiment that we designed and executed to answer the second sub-research question: How can we extract the cyclist's weight from the sensors? It is important to include a detailed outline of the experiment to ensure that the results can be replicated and verified. We will first outline the variables that we considered, followed by the hypotheses that we tested. Next, we will describe the setup of the experiment and the tools we used. Finally, we will explain how we gathered participants and how we treated their data.

5.1 Variables

Here, we outline the variables that we explored in the experiment. First, we derive the independent and dependent variables from the sub-question based on the relationship between weight and sensor data. These variables can be seen in Table 4.

The next set of variables that should be discussed are the extraneous variables. It is key in an experiment to consider any external factors which may have an effect on the results of the experiment. Table 5 shows those selected for this experiment, along with a plan for this.

5.1.1 ***Hypotheses.*** We outlined the hypotheses that we considered for the experiment. Both the null and alternate hypotheses can be seen in Table 6. It is important to include a null and alternate hypothesis to ensure that the experiment is not one-sided or biased. As this research is exploratory in nature, the outcome can be difficult to predict.

Table 5: Extraneous Variables

Extraneous Variable	Plan to mitigate / control it
Speed	Display speed being traveled on the app so that the cyclist can self regulate their speed to a set bound.
Type/quality of road	All data will be gathered in the same stretch of the same road.
Distance cycled	The participants will be informed to cycle a set distance between two marked locations
Tire pressure	The pressure of the tires will be measured before each data collection session to ensure it does not vary between tests.
Weight distribution	Participants will be instructed to straighten their backs and bend forward only as necessary to use the handlebars. This is to keep their weight mostly centered on the seat of the bicycle.
Temperature	Temperature can have an effect on the cyclist and the tires, and we will measure this for consideration each time.

Table 6: Hypotheses

Null Hypothesis (H_0)	Alternate Hypothesis (H_1)
The weight of a cyclist and the sensor data measured do not have a correlation.	As the weight of the cyclist increases, the magnitude of the vertical acceleration will decrease.

Table 7: 3x3 Factorial Design

Weight/Speed	Slow	Normal	Fast
+0 Kg	3 trials	3 trials	3 trials
+ 5 Kg	3 trials	3 trials	3 trials
+10 Kg	3 trials	3 trials	3 trials

5.1.2 Treatment. The Table below shows a 3x3 factorial design in which we modify both the weight and speed of the participants. We modify the weights by adding weight to the participants with a backpack, and the participants control the speed on the bicycle. We measure three trials for each variation of the weight and speed modifications, for a total of 27 measurements per participant (see Table 7).

5.2 Tools

This section highlights the tools that we used in this experiment and, where necessary, their specifications.

We developed an android application for the collection and preprocessing of data. The application had two main interfaces, labeled "Scan BLE Devices" and "Data Collection". The first interface featured a button that would list any broadcasting devices that the mobile phone could detect. Once the participant selected the Thingy from this list, the second interface would appear. This interface included a text box displaying the status of the Bluetooth connection, a button for calibrating the sensor, a text view showing the cycling speed (in km/h), a button for starting and stopping data recording, and a text input box for naming the data file. After collecting and pre-processing the data for each trial, the system exported a .csv file with the provided name to the storage of the mobile device. For the sake of organization, this file-naming convention was as follows:

<*participant-no*>_<*added-weight*>_<*speed-cat*>_<*timestamp*>.csv

Table 8: Bicycle Specifications

Brand/Model	Riverside 120
Weight	14.6 Kg
Frame	Steel
Tire width	28 mm
Tire pressure	2-4 BAR
Tire size	700x19C
Suspension	None

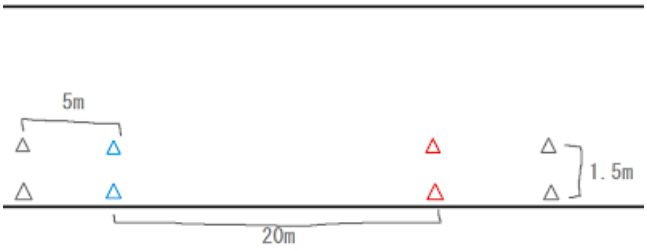

Figure 2: Cone setup overview

The specifications of the bicycle used can be seen in Table 8.

5.3 Setup

This section describes the steps which we followed prior to performing the experiment.

The researcher will firmly attach the sensor to the frame of the bicycle by the back wheel, using zip ties and duct tape. This is because most of the weight is distributed toward the back wheel of the bicycle[6]. Therefore, this area is where the weight will have the most impact on any readings from the Thingy device.

We will use a device that can attach to the handlebars to mount the smartphone that will connect to the thingy sensor and display the speed to the cyclist. This can be seen in Figure 1 holding the smartphone on the handlebars.

Measure tire pressure using a tire pump before each session to ensure consistency. For the bicycle in this experiment (See Tools), the tire pressure was 3 bar as measured by the pressure gauge attached to the pump.

The participant collected data in a 20m segment of a lane with a width of 1.5 meters. See Figure 2 for an overview of this design.

5.4 Participation

The experiment requires human participation. We needed to collect the weight of each participant at the time of participation to establish a ground truth for comparison. Therefore, following the guidelines of the TCS ethics committee [7], we created a participation

[6]Investigating Weight Distribution on a Bicycle
[7]TCS Ethics Committee

consent form. The form explains the procedure and purpose of the experiment to participants, provides a means for them to withdraw consent by contacting the researcher, and assures them that their data will be anonymous and secure.

5.5 Procedure

The procedure of the experiment for each participant was:

(1) Measure tire pressure of bicycle and adjust as necessary.
(2) Use a scale to measure the weight of the participant and note the ground truth.
(3) Connect the smartphone application to the sensor and calibrate it.
(4) When the participant reaches the first set of cones, they will press the start collecting button, and between the inner pair and outer pair of the last set of cones they will press the stop collecting button. During this moment they should ensure that their speed is consistent and within the set category boundaries.
(5) The participant will repeat this procedure three times for each speed category. Once this has been completed with no additional weight, weight, the participant will wear a 5Kg weight (in a backpack). And they will repeat the procedure.
(6) Finally, after another nine trials, another five kilograms will be added to the backpack and the procedure will be repeated.

In conclusion, we designed an experiment to collect data from participants at varying speeds and weights under controlled conditions.

6 RESULTS

In this section, we will present an overview of the data set that we collected. Afterward, we will present and discuss the results of the experiment. Our findings will determine whether we can use our method to infer the weight of a cyclist.

6.1 Data set

Twelve participants volunteered for this research. We conducted tests with additional weight (five and ten kilograms) for each participant, resulting in a data set of 36 weights (3 per person). The distribution of the weights can be seen in Figure 3. Although the range of nearly 50 kilograms is satisfactory, the data was not evenly distributed. This resulted in uneven splits for the classes in terms of the range for each class. This means that the model will be most accurate for weights near the mean (82.1Kg), and less accurate near the maximum and minimum.

6.2 Data Inspection & Analysis

Using the data, we investigated the relationship between the magnitude of the acceleration and the weight of the participants during the experiment. The magnitude of the acceleration vector gave us insight into a general relationship, which we later focused by looking at the specific axes of acceleration. Our investigation produced promising results. We observed a trend by plotting the magnitude of the acceleration vector over time for each participant. The magnitude of the acceleration consistently decreased as the weight of the participant increased (via added weight), as seen in Figure 4.

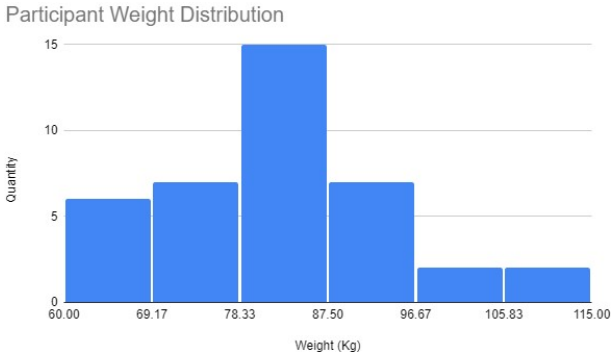

Figure 3: Data Histogram

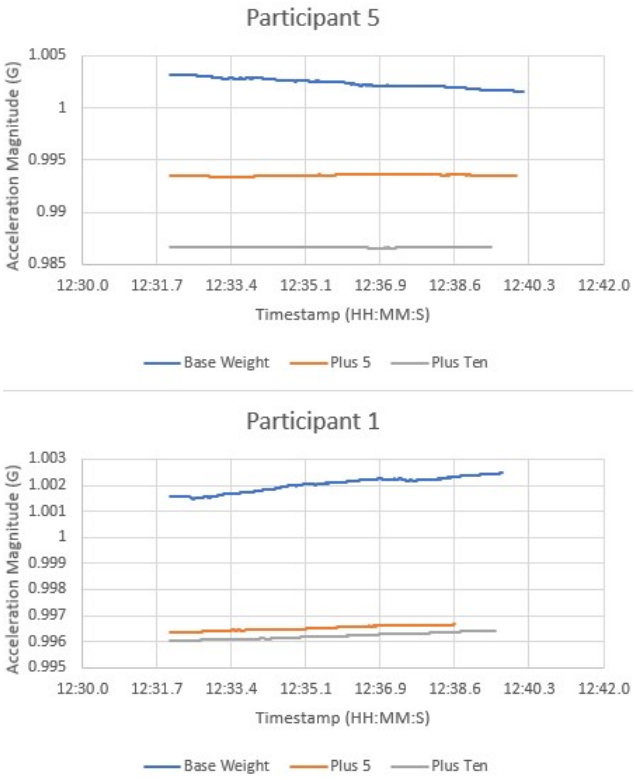

Figure 4: Magnitude as weight increases for two participants

This trend held true for the slowest speed category and for higher speeds as well.

While this seems promising, there are some caveats in this trend as a whole which we should note. First, the actual value of the magnitude does not seem to correlate with the participant's actual weight. Figure 4 shows that if we compare the data for Participants 1 and 5, there is over a twenty kilogram weight difference between them. But the acceleration magnitudes that we observed are quite similar to start. The next detail that requires attention is that the change in magnitude with added weight is not consistent. On average, the magnitude for Participant 5 decreased by 0.7%, while for Participant 1, this decrease was around 0.3%. We found the mean change in magnitude with added weight to be 0.2%, which

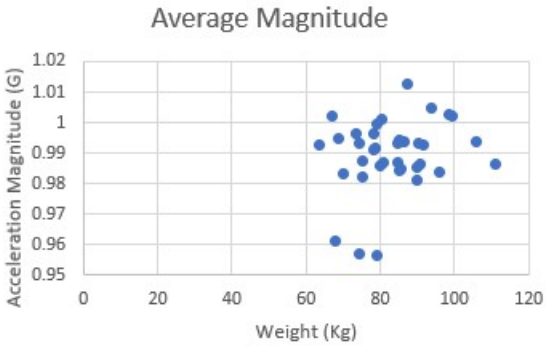

Figure 5: Average Magnitude for all participants at a single speed

also makes Participant 5 an outlier in this regard, according to the 1.5*IQR rule.

While there seems to be a clear relationship between weight and measured acceleration for a single participant, viewing the data set with all participants in mind makes it much less clear. Figure 5 shows this, where we have plotted the average acceleration for all participants at the slowest speed. We also found randomness for the other speeds.

6.3 Weight Classification Model

After conducting a manual inspection and some preliminary analysis of the collected data, we decided to explore a machine learning solution. As previously mentioned in Related Work, the state of the art for inferences such as this paper feature machine learning implementations. Using the previously outlined methods, we extracted key features from the data.

In total, we identified 64 features that provided useful information for training the models. We carefully refined these features by plotting them and testing whether a human could observe differences for the three classes. Additionally, we extracted features using reasoning. For example, we found that the z axis of the acceleration was not very useful, as the vertical (x axis) and–to a lesser degree, the horizontal-acceleration (y axis) were the most useful for the information that they provided. In addition to using time domain features, we performed a Fourier transformation and extracted some additional features in the frequency domain.

We focused on accelerometer data as the main source of motion data collected from the Thingy. But we also collected gyroscope data. We experimented with using the gyroscope data as a feature to improve accuracy. However, it did not provide any additional information, and instead lowered accuracy by confusing the models.

We observed the resulting accuracy for each model in Table 9. Both the KNN and decision tree models performed poorly, with accuracies below 50%. We should note that the baseline accuracy for this research is 33%, or 1 in 3. This is because any accuracy above guessing randomly for the three classes adds value to this system. Therefore, we see that both the SVM and logistic regression models significantly improve upon this baseline.

In Figure 6, you can observe the confusion matrix for the best logistic regression model. It performed at the highest accuracy. We observe that the medium weight class often gets confused with higher and lower weights. To correct this, we tried creating gaps in

Table 9: sklearn accuracy score for each model

Model	Accuracy Score
KNN	~0.33
Decision Tree	~0.43
SVM	~0.68
Logistic Regression	~0.72

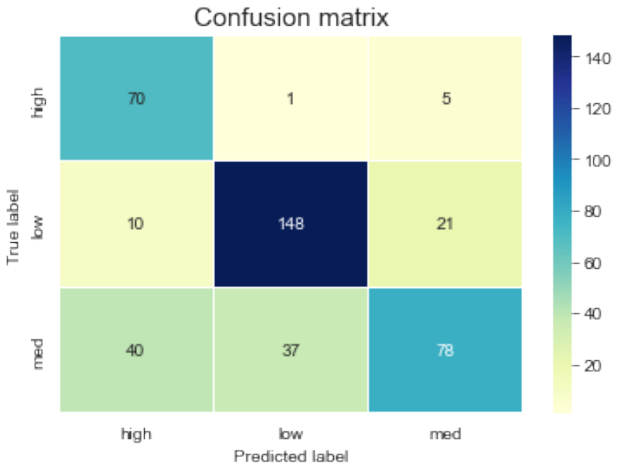

Figure 6: Logistic Regression confusion matrix

the weights between the classes to remove confusion on the boundaries between them. However, this approach caused the accuracy to drop significantly, likely due to the loss of data for training. As a small test, we retrained the models with the weights separated into only two classes: low and high. In this case, the models performed much worse, and the results were discarded.

Empirically, we have found that these models perform better when they are trained on larger sets of data. This is not in the sense of over-fitting. However, while preserving the 75/25 split and simply reducing the amount of data available to the models, they perform significantly worse.

In this section, we conducted two different methods to estimate a cyclist's weight. Initially, we performed an inspection and analysis of the data. This check suggested a relationship between the acceleration and the weight of a participant. Next, we implemented some machine learning classification models using some of the knowledge from the previous method. After extracting and refining features from the collected data, we trained the models and presented the results. The highest accuracy attained was roughly 72% from the logistic regression model.

7 DISCUSSION

This section will explain and evaluate our findings.

Based on the results from the previous section, we have found a relationship between the weight of a cyclist and accelerometer data. The performance of the models did not reach levels such as those observed in either state-of-the-art literature or some of the research from Related Work. However, it is not so low as to discard. Being able to do a ballpark estimate of a cyclist's weight with even a 72%

accuracy is already concerning for people who value the protection of their privacy.

We controlled many variables which could have otherwise had a large impact on the results. Such variables would not be controlled, as they were in the context of a real world cyclist. On the other hand, we drew the results using only one type of sensor data: motion. Some electric bicycles also include a pedal pressure sensor, which could be used in addition to motion sensor data.

8 CONCLUSIONS & FUTURE WORK

In modern western society, people are becoming increasingly aware of how companies handle their data. Companies must respect the potential harm that sensitive data can bring if mishandled. This is also combined with the fact that the field of crowd-sensing technologies is expanding. While there have been other studies inferring sensitive information, none so far have focused on cycling. Considering how crucial bicycle infrastructure is to the Netherlands, and that experts are developing new crowd-sending solutions to road quality measurement, we must incorporate privacy into the design of these systems. This paper aimed to prove the existence of a relationship between a cyclist's weight and accelerometer data from a sensor attached to the back wheel of the bicycle. We conducted an experiment with participants of varying weights under controlled conditions. Then, we analyzed the data and implemented machine learning solutions.

An investigation revealed that placing the Nordic Thingy:52 near the axle of the back wheel is appropriate. This sensor allowed for BLE communication that we paired with an android phone for data collection and pre-processing. The results proved the existence of a relationship between acceleration and weight on a bicycle. The nature of this relationship also showed that as the weight increases for a participant, the overall acceleration felt by the back wheel decreases. This verifies our alternative hypothesis: As the cyclist's weight increases, the magnitude of the vertical acceleration decreases. Our research could serve as a basis for improving these findings and creating a more accurate estimation of a cyclist's weight.

Due to time constraints and the nature of our experiment, the amount of data was rather low. We conducted the experiment outdoors, in the winter, when it often rains. This makes volunteering to participate in the experiment quite daunting. Participants also had to consent to having their weights recorded, which could have turned away some participants and affected the distribution of their weights. We should consider that someone insecure about their weight may be less likely to volunteer for research. Given more time, we would prioritize a larger amount of more diverse data. This would allow the model to perform more reliably for a random person and potentially more accurately with more data upon which to train.

Another consideration for future work is how we added weights to participants. We only had two 5 kilogram weights available to use for the experiment, but lower weights such as 2.5 kilograms would have been more ideal. This would have allowed for an even greater amount of artificial data by adding both weight and more subtle weight increments for the participants, rather than large 5Kg intervals.

In conclusion, we have proven the existence of a relationship between the weight of a cyclist and accelerometer data. Motion sensors such as those found in the Thingy:52 can provide insights into sensitive information when placed on a bicycle. These findings highlight the importance and need for value-sensitive design in crowd-sensing technologies that have the potential to collect personal sensitive information.

ACKNOWLEDGEMENTS

I am very grateful to Deepak Yeleshetty, who supervised this research project. He provided valuable advice throughout the process and helped me to stay on track and motivated. I also want to thank all of the participants. There were no incentives to participate in this research, and it was unpleasant at best in the cold weather. But without the data that was collected with their support, this could not have been possible.

REFERENCES

[1] Anand, S. A., Wang, C., Liu, J., Saxena, N., and Chen, Y. Spearphone: A lightweight speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers. *WiSec 2021 - Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks 1* (6 2021), 288–299.

[2] Christopher Lichtenberg. What's the Difference Between Personal and Sensitive Information?, 5 2020.

[3] Delgado-Santos, P., Stragapede, G., Tolosana, R., Guest, R., Deravi, F., Vera-Rodriguez, R., Vera, R., Stragapede, . G., Tolosana, R., and Vera-Rodriguez, R. A Survey of Privacy Vulnerabilities of Mobile Device Sensors. *ACM Computing Surveys 54*, 11s (1 2022), 1–30.

[4] González-novo López, F. Bike lane quality estimation under variable speed conditions using off-the-shelf motion sensors.

[5] Hernandez, J., McDuff, D. J., and Picard, R. W. BioInsights: Extracting personal data from 'Still' wearable motion sensors. *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2015* (10 2015).

[6] Kröger, J. L., Raschke, P., and Bhuiyan, T. R. Privacy implications of accelerometer data: A review of possible inferences. *ACM International Conference Proceeding Series* (1 2019), 81–87.

[7] Ministerie van Algemene Zaken. Cycling Facts 2018, 4 2018.

[8] Naval, S., Pandey, A., Gupta, S., Singal, G., Vinoba, V., and Kumar, N. PIN Inference Attack: A Threat to Mobile Security and Smartphone-Controlled Robots. *IEEE Sensors Journal 22*, 18 (9 2022), 17475–17482.

[9] Owens, J. M., Alden, A., Antin, J., and Gibbons, R. B. Development and Testing of an Integrated, Versatile, Bicycle-Based Data Acquisition System.

[10] Patil, O., Jadhav, S., and Ramakrishnan, R. Development of Reaction Wheel Controlled Self-Balancing Bicycle for Improving Vehicle Stability Control. *Lecture Notes in Mechanical Engineering* (2021), 187–195.

[11] Springer, M., and Ament, C. A Mobile and Modular Low-Cost Sensor System for Road Surface Recognition Using a Bicycle. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems 2020-September* (9 2020), 360–366.

[12] Tahir, S., Raheel, A., Ehatisham-Ul-Haq, M., and Arsalan, A. Recognizing Human-Object Interaction (HOI) Using Wrist-Mounted Inertial Sensors. *IEEE Sensors Journal 21*, 6 (3 2021), 7899–7907.

[13] van de Poel, I. Translating Values into Design Requirements. *Philosophy of Engineering and Technology 15* (2013), 253–266.