School of Electronic Engineering and Computer Science

Interim Report

Programme of study:
Electrical and Electronic Engineering

Project Title:
Transportation Mode Detection In Smartphones

Supervisor:
Lin Wang

Student Name:
Nader Akmel

Final Year
Undergraduate Project 2020/21

Date: 4/05/2021

Queen Mary
University of London

Acknowledgements

Firstly, I would like to thank and show gratitude to my supervisor Lin Wang for his support and guidance throughout the completion of the project.

I would also like to thank my friends and family, who helped support me through the progression of this project and academic year.

Abstract

The exponential growth in smartphone technology has brought forth many potential applications, coupled with the development of sensor technology, we can accurately determine the user's mode of transport. This application can help provide alternatives to the current data collection methods that use transport data commercially, which in effect could benefit health and transport sectors in their practices. This report describes a method used to classify various transportation modes with a machine-learning algorithm. Using 21 hours of user collected data, the modes: Still, Drive, Walk, Run, Bike were classified. With a subset of selected features, the Decision Tree classifier proved to be most efficient, producing an average accuracy of 98.0159%. The results show that the modes classified by the model were well recognised and could therefore be visualised using a web application feature, in addition to a real-time mode recognition task that tested the machine learning model in real-time.

Contents

Chapter 1: **Introduction**

1.1 Background

Currently, we see frequent innovations in technology. For instance, in Mobile phone services, there are developments in GPS based technology which has changed the way we travel. In effect, this provides Stress-free experiences to the users travelling via different transport mediums. However, there are limitations in GPS-based technology to where commercially, the mode of transport cannot be determined during travel. Many positive effects come with the prospect of transportation mode detection: For example, Hedemalm (2017) explains how the global climate change is not improving, and how the study of transport habits can be used to influence changes to transport selection, “Assuming this is successful, it could be one of many steps to try and decrease our total carbon emissions and sustain the planet.”.

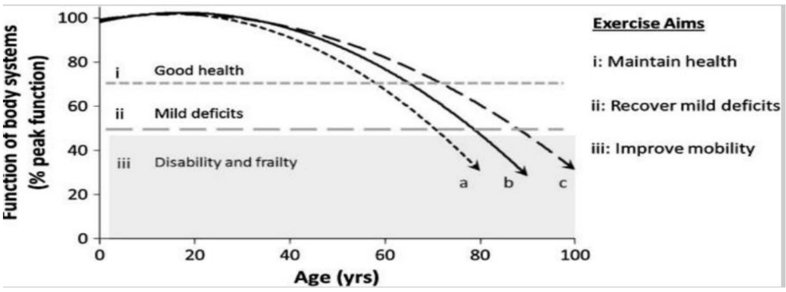
Currently, there are implementations of surveyed data collection that learn about travel behaviour. However, these are never completely accurate since they depend on a study of individuals recollection which may lack substance. Transportation mode detection generates concerns when it comes to accuracy, though with developments in technology and research in this field this becomes less of a concern, thus allowing sectors other than transportation to use this application efficiently. Looking into the medical sector, we see wearables used to detect data such as steps taken and physical activity levels. However better communication can be achieved if the mode of transport is detected since transport selection in most cases can be influenced for positive health effects.

1.2 Problem Statement

Transportation mode detection has impacts on everyday life, from traffic management to routing recommendations affected by congestion, accidents, and roadworks. The use and evolution of smartphones have grown rapidly to where data collection from smartphones can provide useful information. This in effect enables possible applications like the monitoring of activity and health.

The motivation behind this project is to determine whether the transportation data collected by the physically impaired, such as the elderly and disabled will be accurate enough to benefit the medical sector in diagnosis. **Figure 1** below shows the effect of age on the body's peak function, in addition to how physical activity can affect this correlation. Although exercise and recreational activities have positive effects on health, we understand that daily transport choices make up most of the physical activity completed by an average person. Cooper et al (2019) reinforces this point with a study showing how increased car usage can have negative impacts on physical health. Mode detection is a beneficial way of learning the habits and behaviour of the client in question. In effect, this would allow doctors and medical staff to advise on health conditions and issue changes to daily transport and physical activity. If effective, major organisations such as TFL and the NHS could produce real-life implementations that can use this data effectively.

Figure 1 Representation of ageing trajectories effect on physiological function (McPhee et al., 2016)



1.3 Aim

The project aims to develop a machine learning algorithm that can detect the different modes of transport using sensor data provided by the smartphone user. The implementation of the machine learning code will measure the smartphones: Accelerometer, Gyroscope and Magnetometer sensor values to determine the effect each of these will have on the precision of the system. Therefore, the effectiveness of the Decision Tree classifier in differentiating between the data will also be evaluated. This classifier will be used on a 21-hour Dataset containing Still, Drive, Walk, Run and Bike data.

1.4 Objectives

- To achieve a working algorithm that can detect the mode of transport
- To determine the effectiveness of the decision tree classifier compared to other classifiers
- To determine which of the smartphone sensors has a noticeable effect on precision
- To investigate existing transportation mode detection projects to determine the effectiveness of my system
- To gain a greater understanding of how this application can benefit the data collection in the medical sector
- To implement real-time mode detection using the classifier model
- To design a Web application to display the results of the project

1.5 Research Questions

How can transportation mode detection benefit the lives of the physically impaired?

How the different sensors present on a smartphone affects the success rate?

How effective machine learning is in detecting the mode of transport?

How the classifier features affect the precision?

1.6 Report Structure

The report is structured in the following chronological format

Introduction:

In this chapter of my report, I elaborate on the background of my project idea, in addition to my project aims and objectives.

Literature Review:

In this chapter of my report, the research currently available in Transportation mode detection and its associated areas Is displayed, in addition to the project background and lessons that can be learned.

Design and Implementation:

In this chapter of my report, I will discuss the methodology behind the transportation mode detection task, in addition to a design overview of the tools used in the production of the algorithm. This section will also explain the reasons behind the use of each component.

Software and Hardware:

This chapter describes the different hardware and software used in this project, along with the reasoning behind their use.

Evaluation:

In this chapter of my report, the methods used to evaluate the results of this project will be elaborated, it is here where the web app and real-time mode detection applications will be explained. Furthermore, the limitations present with the current design will be also be discussed

Conclusion:

In this chapter of my report, the achievements and challenges of this project will be reviewed, this is also where future applications/work are presented and justified.

Risk Assessment:

In this chapter of my report, I will evaluate the risks associated with my project and how I can prevent them from occurring.

Time Plan:

This chapter contains a timeline of the different stages present in this project

References:

This chapter contains a list of references used in the project

Chapter 2: **Literature Review**

2.1 Location-Based applications

In recent years, location-based services have grown rapidly allowing users to travel efficiently using the GPS navigation technology built into their smartphones. This has opened doors in technology allowing for different sectors such as health, commercial and food to take advantage. Schiller and Voisard(2004) explained the importance of location-based applications in the relevant industries, “Every day, 170,000 emergency calls are made in the United States. Of those, one-third originate from mobile phones, and, in most cases, people do not know where they are precisely in order to guide support to the correct location”.

The information provided by the real-time geodata provides us with the ability to accurately capture a user’s trajectory and speed. When looking at the commercial use of location-based data, we see the implementation of this data collection in the industry is important in providing useful data. Mobility and transport research are examples of influential data that can provide great insight for many companies in addition to society. However, the accuracy and the limitations provided by current implementations noticeably becomes an issue. Schiller and Voisard(2004) further explained this “For some time, marketers have been unsure whether lower levels of accuracy as they are obtained from Cell-ID would be sufficient to launch compelling consumer and business services.”.

Filtering out other means of transport, can be beneficial to the accuracy and broadens the use of this data in a grand scheme of things. Traffic updates and hot path data are examples widely used by the general public, which are being limited because transportation mode detection has not been properly commercialised yet.

2.2: Machine learning approach

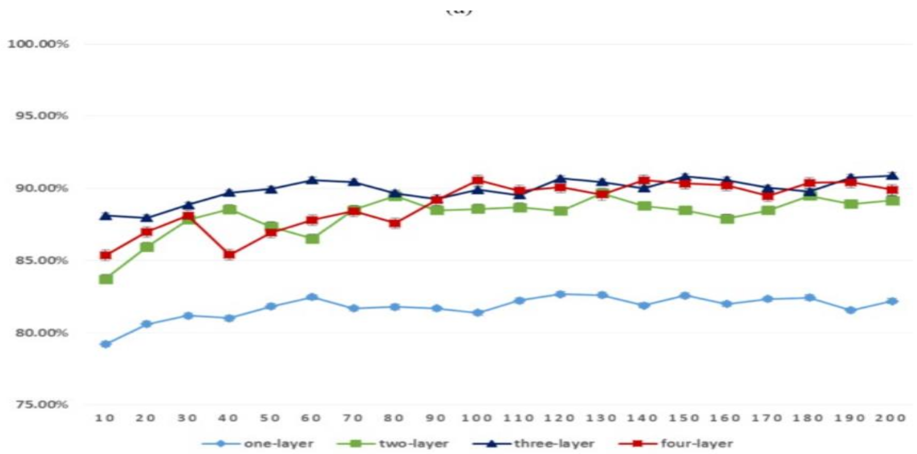
The machine learning approach uses an automated algorithm that takes in and learns from data to make a decision based on what has been learned. Generally, feature extraction is used along with a classification method to differentiate between the different transportation medians. Lara and Labrador (2013) states that “Acceleration signals are highly fluctuating and oscillatory, which makes it difficult to recognize the underlying patterns using their raw values.” so systems are required to employ a feature extraction method to extract necessary data. Generally, as part of feature extraction, time and frequency domain methods are used to extract the necessary sensor data. Patterns and trends would then be visible allowing for the classification method to make the necessary decisions.

Table 1 below of relevant feature extraction and classification methods used in classifying sensor data (Lara 2013)

Classification Methods	Feature Extraction Methods
Decision Tree	Time Domain:
Bayesian	Mean, standard deviation, variance, interquartile range (IQR), mean absolute deviation (MAD), correlation between axes, entropy, and kurtosis
Neural Networks	Frequency Domain:
Fuzzy logic	Fourier Transform, Discrete Cosine Transform
SVM	Other:
Classifier ensembles	Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Autoregressive Model (AR), and HAAR filters.
Makov Model	

In contrast to the well-known machine learning algorithm, more recently the Deep Learning approach has been used in many studies. Deep learning works like machine learning in that it comes to a decision after parsing data; it works similar to how a human would come to a decision and uses a layered structure with numerous algorithms. However, the problem with the Deep learning approach is that it requires powerful hardware to run since there tends to be numerous algorithms and a lot of data being processed continually. My task of transportation mode detection would not 100% capitalise on the deep learning layering approach, since our dataset will not include a large enough amount of data.

Figure 2 shows the effect of layering in recognising human activity (Ronao et al 2016, p.26)



2.3 Assistive technology in medical data collection

Data collection allows us to accurately analyse different sources of information, this is effectively used by many sectors including the health industry. “Data collection in healthcare allows health systems to create holistic views of patients, personalize treatments, advance treatment methods, improve communication between doctors and patients, and enhance health outcomes.” (Sakovic 2019). To date, there are numerous technologies in place to gather this information, such as electronic questionnaires that survey the population, or mobile applications and wearables that collect physical data This information is then be used by medical practices such as local GP’s, which innovates the diagnosis process and grants doctors the ability of predictive analysis.

Advancements in technology also have impacts on convenience to the user. We see that “66 percent of the population is willing to use wearables and mobile apps to monitor their health and improve the treatment of conditions” (Sakovic 2020). This statistic shows that simple tasks such as downloading an application are not tedious to the patient.

In application Mandolesi (2018) explains how physical activity “determines positive biological and psychological effects that affect the brain and the cognitive functioning and promote a condition of wellbeing.”. Therefore, if medical practitioners take accurate physical data collection, the prevention of cognitive deterioration is possible. **Figure 1** in the problem statement further reinforces this notion.

2.4: Related Works

In the table below, several related studies investigate the effectiveness of various sensors in detecting the user's mode of transport. In this section, I will review the related pieces of work.

Table 2 Related Work on Transportation Mode Detection

Publisher	Transportation modes	Sensors	Participants	Accuracy (%)
Gonzalez, P. Weinstein, J. Barbeau, S. Labrador, M. Winters, P Georggi, N. Perez, R (2020)	Car, Walk, Bus	GPS	7	87.36
Miluzzo, E., Lane, N.D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., Eisenman, S.B., Zheng, X. and Campbell, A.T. (2008)	Sitting, Standing, Walk, Run	Accelerometer	8	77
Feng, T. and Timmermans, H.J.P. (2013)	Walk, Run, Cycling, Bus, Motorcycle, Car, Train, Metro, Tram, Light Rail	GPS	2	75
Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M. and Srivastava, M. (2010)	Still, Walk, Run, Bike, Motor	GPS, Accelerometer	16	93.65
Zhao, H., Hou, C., Alrobassy, H. and Zeng, X. (2019)	Still, Subway, Bicycle, Walk, Run, Bus			92.8

Weinstein et al (2008) uses GPS data along with a neural network classifier to determine the different modes of transport presented. This method showed effectiveness in the conservation of power, one of its main objectives was to create a data collection method that would reduce costs. To achieve this critical points were used, “using a critical point algorithm to reduce the size of required GPS datasets obtained from GPS-enabled mobile phones, thus reducing data collection costs while saving mobile phone resources such as battery life.” The method in place works well in classifying walking from the other modes of transport. However, this task is simpler since the modes were not motor-based, the problem came when differentiating between bus and car, since only GPS data was being used it would become more difficult to accurately differentiate between the two so more sensor data could be used.

Another similar project was by Miluzzo et al (2008) where accelerometer data is used to determine the different modes of transport on a Nokia N95. This data was then classified using a decision tree classifier which has shown effectiveness in numerous reports. Miluzzo et al (2008) showed the method was efficient in detecting walking from running, however, the classifier had difficulty differentiating sitting and standing given the similarity in the raw accelerometer data. This is believed to be because the only sensor data used is from an accelerometer.

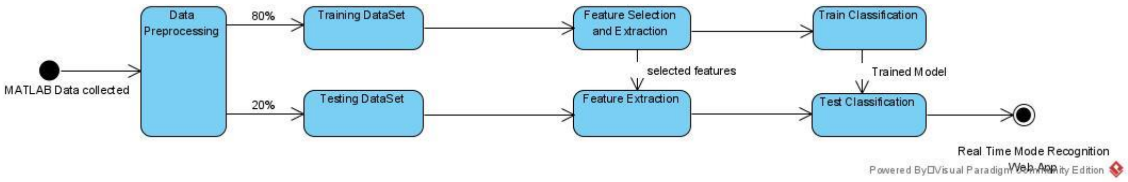
Miluzzo et al (2008) states “After interviewing the participants in our user study, we conjecture that most of people carry their phones in their pants pockets, clipped to a belt or in a bag.”. Although this brings convenience to the study since only accelerometer data is being used, placement becomes of larger importance. To improve the reliability of this report, more device placements could be investigated, to improve the model's ability to analyse the vibration characteristics associated with using the accelerometer sensor.

A more effective approach by Reddy et al (2010), uses accelerometer data in addition to GPS data to determine different modes of transport. The method in place shows success with a detection rate of 93.65%, which tells us that the “decision tree followed by a firstorder discrete Hidden Markov Model place”, used by Reddy et al (2010) can accurately differentiate between the selected modes. Since the data contains the switching of Transport modes in a sliding window Reddy et al (2010) smartly uses DHHM to apply the common sense of what are the likely transitions between the modes of transport. It is apparent that this has made a noticeable difference in the precision, especially with the walking mode. This approaches success was useful as background research in this project.

Chapter 3: **Design and Implementation**

Using the research obtained from my literature review of the background of this project, along with knowledge gained from similar projects I can split my project plan into different stages.

Figure 3 Flowchart depicting the different stages in the implementation of my Project.



MATLAB Data Collection

As a pre-requisite to data pre-processing, sensor data is collected via the MATLAB mobile software, this is done using multiple participants with data being collected for each mode in phases.

Data Pre-processing

Using the data collected via MATLAB, the data can be pre-processed to get the Dataset into an 80-20 Train-Test data split form from which features can be extracted. This stage is also where anomalies/outliers are removed.

Feature Selection and Extraction

Using the train data, supervised analysis can be completed to determine which features can be used to improve the predictive power of the model, whilst accounting for overfitting and computing time.

Classification

The extracted features can then be applied to the Classification algorithm, which builds a model that can learn from the input data and make predictions.

Application

The classification model built, can then be used in applications such as real-time mode recognition and web applications, that can evaluate the model's efficiency.

3.1 Sensor data collection

To collect data for the model implementation, 3 participants were used. Mobile sensor data was collected from 2 male and 1 female participant, amounting to 21 hours' worth of data. the collection was split across the 5 modes of transport being investigated(Still, Drive, Walk, Run, Bike). The use of multiple participants reaffirms the reliability of the collection, which allows for the different patterns of user movement to be taken into account, this ensures that the model is not fitted to just a single users movement pattern. This is useful when training the model, which would improve the reproducibility when testing on new sets of data and live sensor readings.

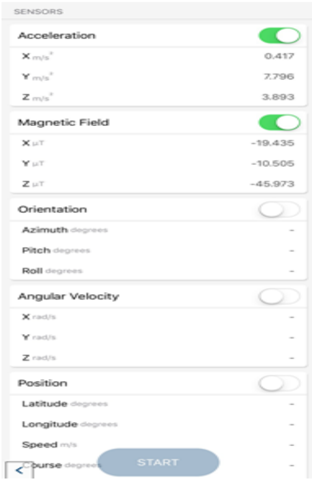
Figure 4 shows how much data was collected for each mode of transport

The hardware used by each of the users to collect data was an iPhone 11, with the phone placed downright in the trouser front pocket, each of the users mimic the same position to ensure the 3 axis sensor values will be similar across the 3 users. This position also simulates a realistic phone placement, which is conveniently placed on the body of the user. So in the application, the user's everyday life will not be impacted.

On the iPhone, the IOS app MATLAB mobile was used as the data logger app of choice. Although there may have been more accurate alternatives with additional sensor options, MATLAB is the software where the data was processed, so there is convenience in using these together. The standout feature allowed sensor data to be uploaded directly to MATLAB drive where the data can be logged into a .mat file or live streamed directly to the workspace for applications such as real-time mode detection. In comparison, applications such as HASC(sensor data logger) required a server for the sensor data to be uploaded.

The sensors Accelerometer, Magnetometer and Gyroscope were determined via test data collections and research into the sensors general effect in similar projects. With each sensor being compatible with MATLAB, I was able to use them in the larger data collection where participants were involved. The sampling frequency by the MATLAB app was left at its default 10hz, which was found to be accurate leaving no gaps in data.

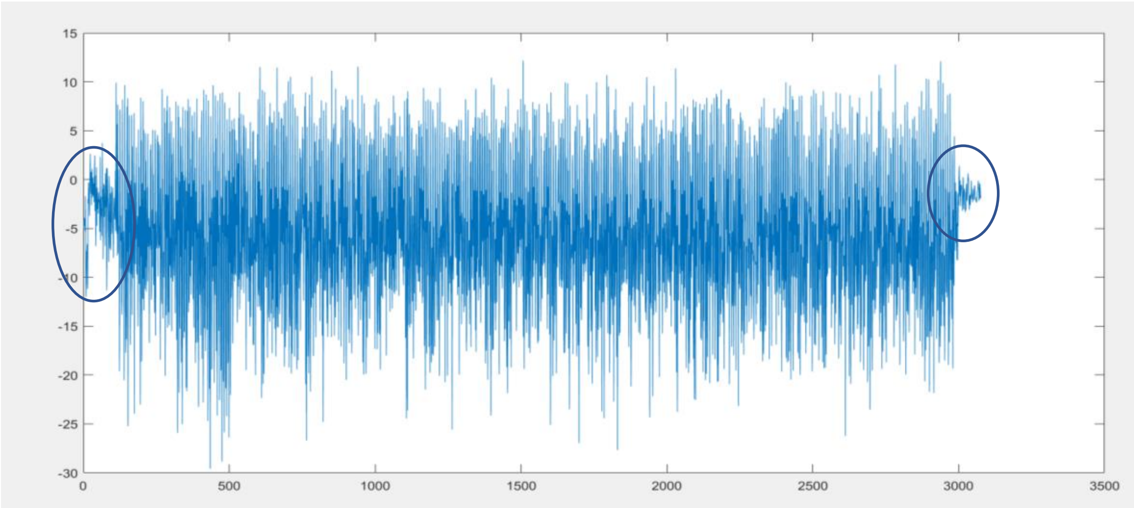
Figure 5 (Mathworks.com, 2020) example showing sensor selection on MATLAB mobile



3.2 Pre-processing

For the acquired data to be used in the training and testing of the model, a pre-processing step must be applied. It is here where the raw sensor data is processed and formatted to become suitable for the application of the machine learning model. Before analysis the data must be loaded and visualised, to do this the raw sensor was loaded and plotted to analysed for outliers and missing values. Since the data was not collected in one sitting, the participants were required to stop and start the sensor recordings on the mobile device, so in the plots of the raw-sensor data, irregularities were visible at the start and end of the collection periods. Each data collection was manually edited removing these outliers along with any gaps in data, so that the data could be labelled and assembled in a single text file, where each mode consisting of X hours was grouped.

Figure 6 shows a collection of a 5 minute run by a participant with irregularities annotated



Before initiating collection for the larger datasets, various sensor data samples were collected using MATLAB mobile to familiarise myself with the software and implementation of the sensor data. Figures 7-11 show the magnitude of the Accelerometer data of each of the modes of transport. This data had provided an insight into the variations in Acceleration that occurs during travel. These figures along with research and further analysis are examples of how data was used to support the feature extraction process.

Figures 7-11 below shows raw sensor data collected from Accelerometer values

Figure 7 Still Accelerometer Data

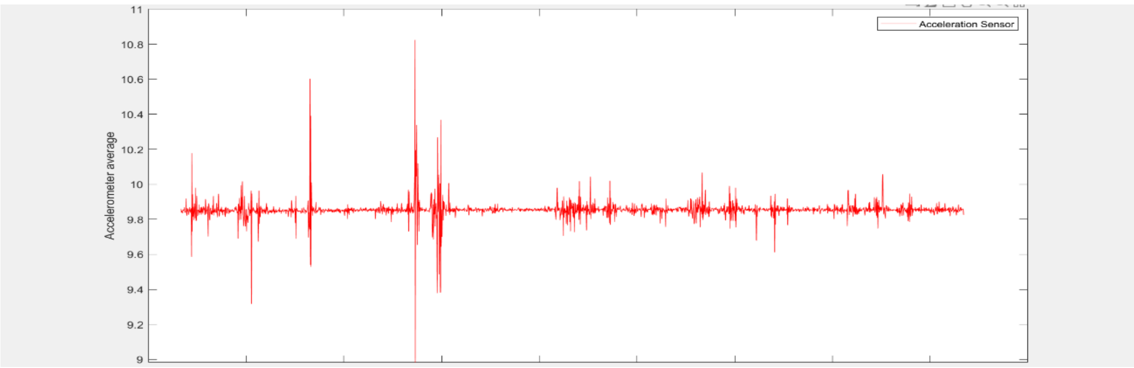


Figure 8 Run accelerometer Data

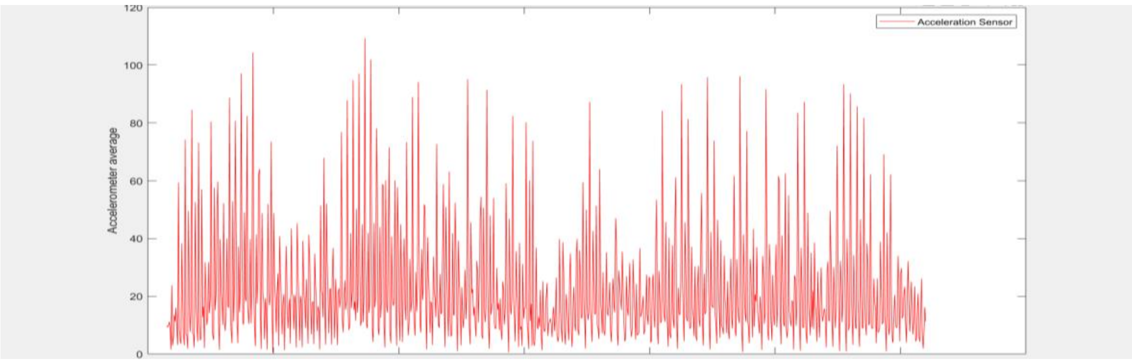


Figure 9 Walk accelerometer Data

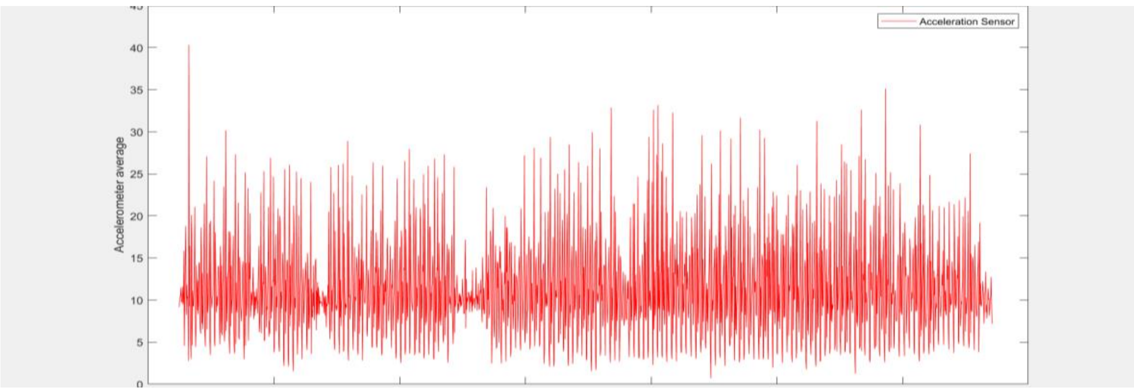


Figure 10 Drive Accelerometer Data

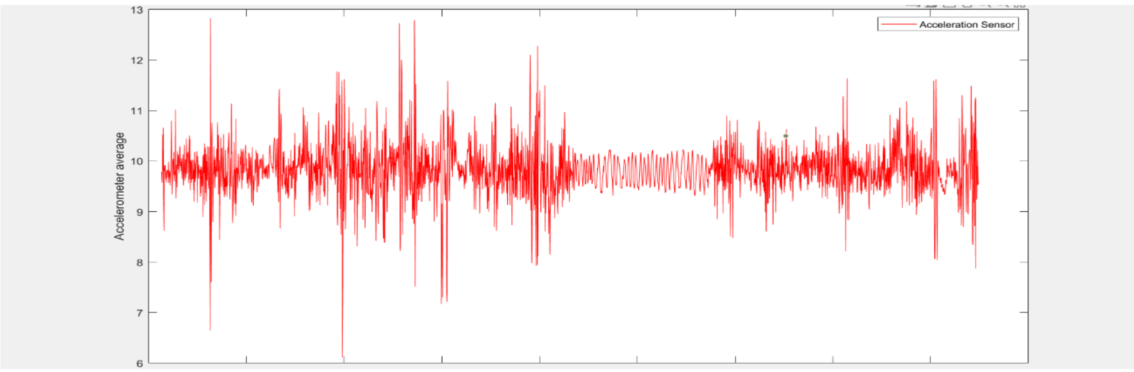
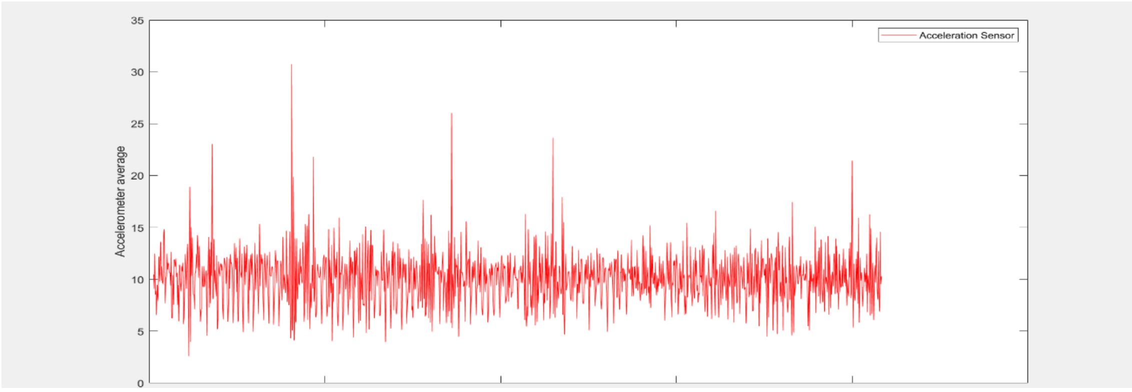


Figure 11 Cycle Accelerometer Data



Since data for each mode were grouped, the effectiveness of the model could not be correctly determined in this form. Instead, the sensor data must be grouped in a single table labelled by the modes of transport 1-5, which allows for the 5 second time windows in the table to be randomised using the MATLAB Rand function. Given that the activities running, and cycling are investigated we know the orientation of the mobile phone may change during the collection period. To remove this variable when observing the accelerometer data, the magnitude of the acceleration is determined for its use in the feature extraction step. This also helps when identifying the variations present in the still mode, where we classify standing, sitting, and laying in the same model. In practice, they would have distinctive XYZ values, so when investigating the frequency and time domain features using magnitude is beneficial.

Figure 12 shows the equation used to calculate the magnitude of 3 axis acceleration values

$$\text{Magnitude} = \sqrt{x^2 + y^2 + z^2}$$

Before features were selected, the data must be reshaped to suit the desired time window. 5 seconds was the window used for the period of classification with 50 samples in each window; this window size was a median where enough data was present to accurately determine the mode of transport. Since a sampling frequency of 10hz was used, a smaller window size was not chosen, this is due to there not being enough data to capture all the necessary features and trends in the designated time window. In addition, a large time window was not used due to noise being introduced from there being multiple activities present in the same window Reddy et al (2010). This is also reinforced in the application of real-time mode detection, where a longer window size would not be viable since the user would want to know the mode predictions in a shorter time frame.

In terms of feature extraction, classifier features that help distinguish the different modes for the classifier must be carefully selected. The features are given in the frequency or time domain, which are obtained via MATLAB tools from the Accelerometer, Gyroscope and Magnetometer sensor data. To extract the frequency domain features a Fast Fourier Transform(FFT) must be applied to let us know the Discrete Fourier Transform of signal, this gives us an idea of the frequency distribution of the signal which can then be used to extract features. Time-domain features do not require any kind of conversion and are computed in relation to the sample size of each window.

For the time domain features: Mean, Standard Deviation and Median were used to analyse the characteristics in the time domain. When looking at the magnetometer since magnetic field levels are being measured, Byon et al (2009) explains how the mean magnetic field value can be used in detecting transport modes. However, this task is difficult since magnetic field values are not distinct in most modes of transport. Instead, standard deviation features are used to help when determining features such as bike and drive whose magnetic field value does not vary as much unless the direction is changed. It was therefore noticeable that the Magnetometer Z-axis values provided better predictive power to the model. For the Gyroscope, we look specifically at the X(vertical) component of the sensor with the features mean and standard deviation being extracted. These features are useful when looking at the periodic rotational movement present in the activities of running, cycling and walking. Lastly looking at the Accelerometer magnitude the features mean, standard deviation and median were used. We see that this set of accelerometer features were the most effective at distinguishing between the transport modes, this comes down to the patterns in the raw accelerometer data being the most distinctive.

For the frequency domain features: Mean, Energy and Max Frequency value were used to analyse the characteristics in the frequency domain, for this function the accelerometer in specific is the sensor that was being evaluated. For the max frequency(FFT) value, it is understood that certain activities with better periodicity will have a more dominant frequency component. As explained by Bao et al (2004), we know cycling has a more consistent circular motion pattern, so in the vertical component, we are more likely to see a more dominant frequency component since phone placement is in the front pocket region.

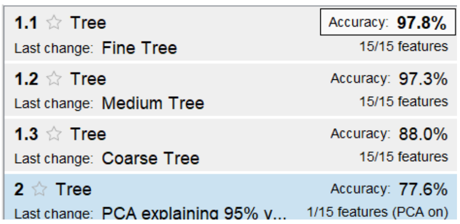
Before selecting which features were necessary, research was conducted into similar projects and its associated sensors choices to determine what features were effective in the mode recognition task. In addition to this, plots of the raw data allowed visualisation of the problem at hand. This allowed me to infer which features would initially be effective.

3.3 Classification/Result

Using the features mentioned above, a classifier can be trained to learn from the data provided by the sensors, which can then be used to test the efficiency of the algorithm in the test data. For this transportation mode recognition task we see the various classifiers: Decision Tree, Naïve Bayes, Neural Networks and K-Nearest Neighbour used frequently when investigating the machine learning problem. In my project, the Decision Tree classifier is used to build a Decision Tree based on the features extracted from the training dataset; this is then used to determine the mode of transport in a test dataset. Related research using Decision Trees achieved good results (Bao et al 2004) (Kurnia et al 2018). In comparison to traditional Neural Networks, Decision Trees provide a final model that can be easily understood by the user. When using decision trees, the precision of the algorithm can be visualised using confusion matrices which provide insight into where the algorithm lacks and can influence changes to be made. To test the accuracy of the classifier the dataset must be split into training and testing, 80% of the data is used to train the classifier, with the remaining 20% being used to validate the trained classifier's effectiveness on new data. This test data is further split into 3 separate datasets for analysis of the web application.

A general problem with Decision Trees is that they tend to overfit, so they may account for variations present in the training data that may not be present in other datasets applicable to the classifier. When building a Decision Tree accuracy is important. However, we don't want the model to overcomplicate the problem at hand with too many features. Principal component analysis(PCA) is an unsupervised technique used to remove unnecessary features whilst simplifying the Decision Tree, in the implementation PCA is tested with components that can explain 95% of the variance, with this 1 out of the 15 features were selected. Approximately 20% accuracy is lost since features were removed and instead of simplifying the model, we can see that the decision tree output was more complex. For these reasons, PCA was not used in the final implementation since the trade-off was not beneficial. The manual supervised feature extraction completed was sufficient for the mode recognition task. With the classification being completed on MATLAB, Fine, Medium, and Coarse Decision Trees were available for use. Upon testing Fine tree was chosen, since it provided the best model flexibility and accuracy whilst not impacting computing time.

Figure 13 shows the accuracy of the fine tree model compared to the other Tree models along with the impact of PCA



In Decision Trees, the method generally used to prevent overfitting is tree pruning, which is where the size of the Decision Tree is reduced by removing/merging unnecessary branches. This method was not adopted since there was a negative effect on the classifier accuracy on new datasets, which gave the impression that underfitting was instead present which indicates the model has not learned enough from the train data.

Cross-validation is the method used to optimise the classifier and prevent overfitting, this is an effective method where the train data generates multiple splits of test data that will help fine-tune the model. Specifically, 5-fold cross-validation was used which split the data into 5 subsets, therefore leaving 1-fold (20%) of the data out for testing. Since the dataset is relatively large, 5 folds were used instead of 10 folds. The chosen value of k-folds is proportional to the size of a dataset: for this project, a very larger dataset is not simulated, so a larger value of k is not necessary. Although cross-validation was used in overfitting prevention, the visible discrepancies between training and testing accuracy are still a problem, in addition to a large number of nodes in the Decision Tree. The design was too complex, and this was due to the number of splits in the Decision Tree, to combat this the max number of splits was altered from 100 to 20. 20 was the perfect balance between underfitting and overfitting where the discrepancy between testing and training accuracy amounted to approximately 0.4%.

After analysis and optimisation of the trained model, as noted earlier we must train on new unseen data. Using the MATLAB predictfcn function, The trained model's efficiency could be tested on the 20% test data split. The output of this function gives us the test data predictions of the model. These predictions give us the ability to evaluate the model further.

Looking at **Figure** 7, the Decision Tree produced is shown. With this we can understand the predictive power of each of the sensors in use, this in effect aided the process of feature selection in this project. It is visible that Accelerometer and Gyroscope features were more useful to the Decision Tree classifier since more decision nodes use the Accelerometer and Gyroscope sensors features to make decisions. In comparison, there were fewer decision nodes that use the Magnetomer sensor and although fewer magnetometer features were in use, we still understand through the analysis of the magnetometer that the patterns associated with each mode were less recognisable. This lets us know that the predictive power of the Magnetometer was smaller.

Figure 14 below shows a Decision Tree with 189 nodes that was experiencing overfitting

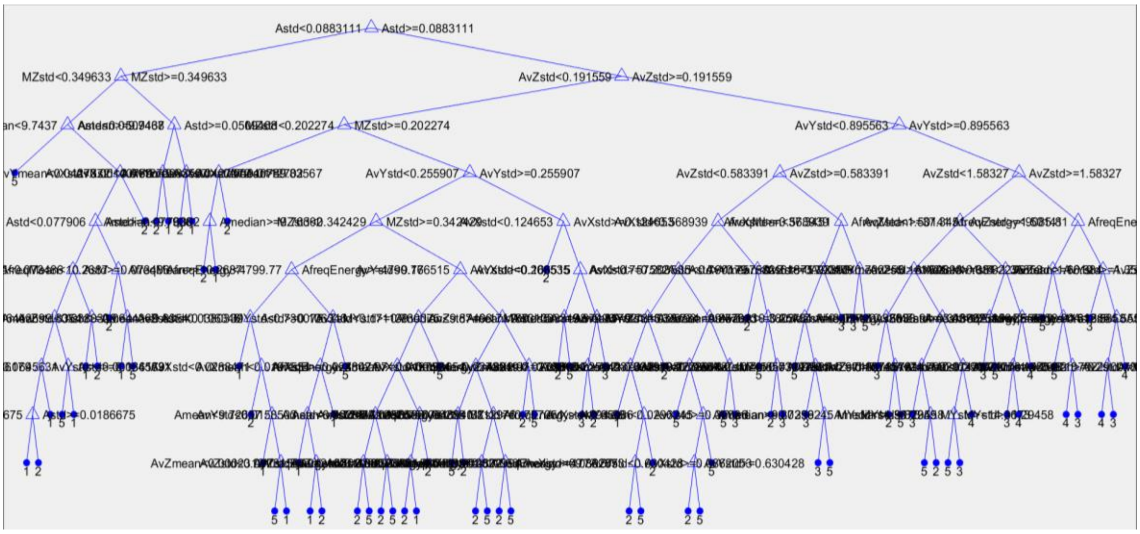
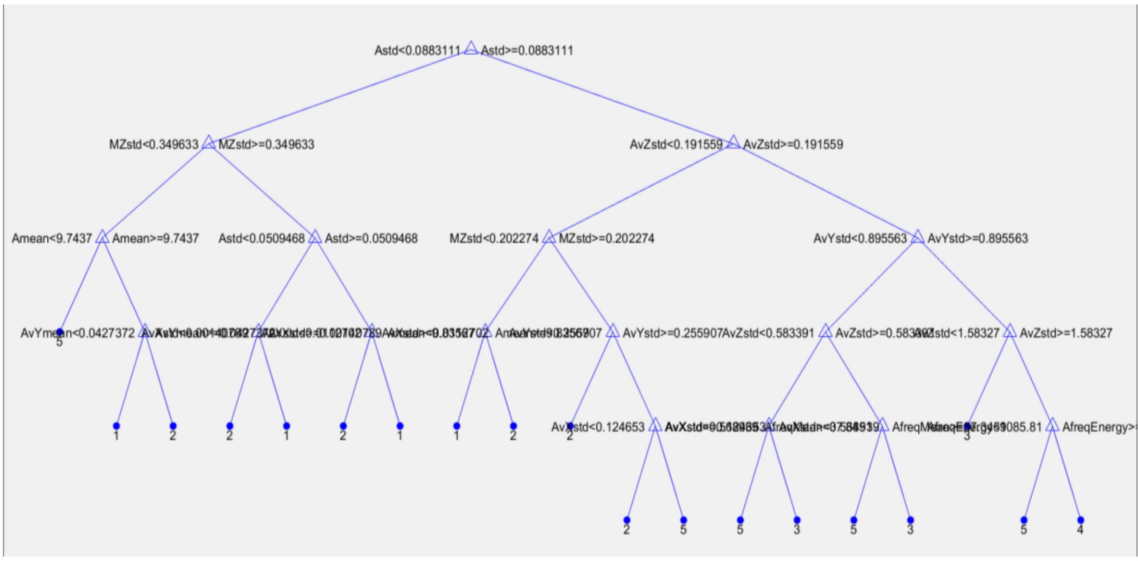


Figure 15 below shows the Decision Tree used in the final classification with the max number of splits changed to 20 to prevent overfitting



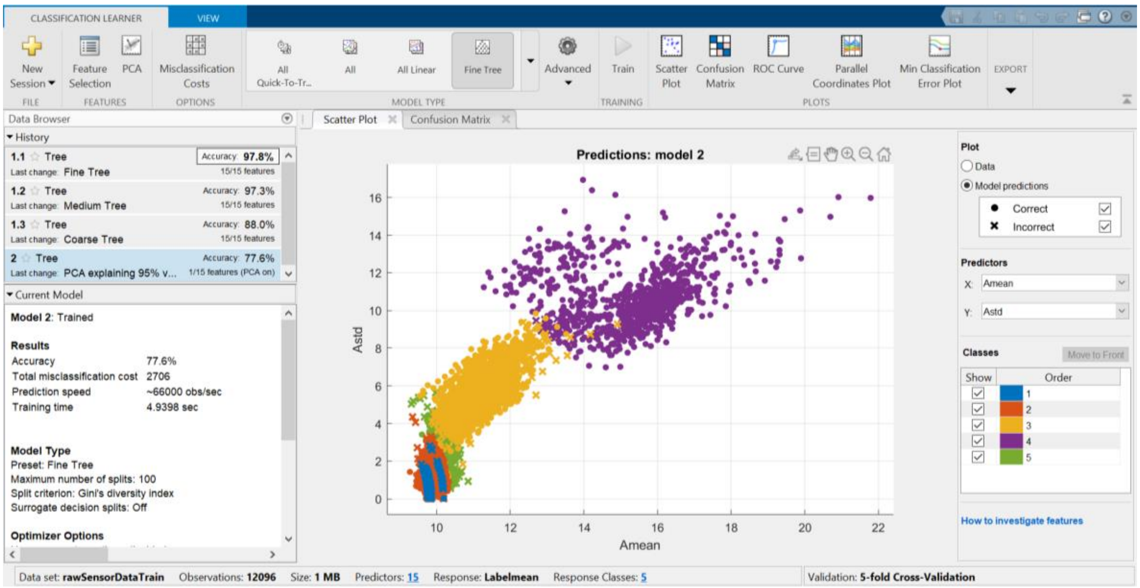
Chapter 4: **Software and Hardware**

This section details the software and hardware being used, and the reasoning behind the use of these features in the implementation of the design.

4.1 MATLAB

To implement the machine learning algorithm, the software I will be using is MATLAB which provides various tools that make the implementation of algorithms simple and time-efficient. The MATLAB software works with two applications being the MATLAB mobile IOS app on a mobile device and MATLAB R2020B software on a PC. As mentioned in previous sections MATLAB mobile is used to make the process of data collection for MATLAB very efficient. MATLAB R2020B was used, which provided a variety of tools that were useful to the transportation mode detection project. These tools are applied via the MATLAB app feature, which allows us to download toolboxes that provide us with additional features; an example is the signal processing toolbox which provides us with functions such as FFT which is used in frequency domain analysis. In addition, the software has access to a Machine learning toolbox, which is key to solving my project's problem. In specific the classification learner app is used which provides ease of use when interacting with the classification model, allowing the problem to be visualised, whilst also providing various features that can improve the model's efficiency.

Figure 16 shows the MATLAB Classification learner used in the approach to my problem



4.2 iPhone

To collect data for the implementation, a iPhone 11 was used. Evolutions in smartphone technology have made it possible to conveniently collect and use sensor data. The data collection uses a 3-axis Gyroscope, Accelerometer and Magnetometer sensors that measure the X, Y and Z values of the 3 axes. It uses a 3110 mAh Li-Ion battery that improves reliability, allowing participants to collect data over longer periods. The main reason as to why the iPhone 11 was chosen is down to the participants, since all 3 participants had an iPhone 11 as their current mobile device, there is convenience in the use of the iPhone 11 where no new sensors or mobile devices would have to be acquired. The specification listed in Figure 17 also supports the use of the mobile device: the parameters such as sampling frequency and standard deviation in noise levels let us know it is more than capable of collecting accurate sensor data.

Figure 17 shows the third party recorded sensor specifications of an iPhone 11 (Phyphox.org, 2021)

iPhone 11

Device	
Manufacturer	Apple
Brand	Apple
Sample size	217
Variants	1

Accelerometer	
Available	✓
Name	
Vendor	
Range	
Resolution	
Rate	100.0 Hz
Average	9.822 m/s²
Standard deviation	0.0095 m/s²

Acceleration (without g)	
Available	✓
Name	
Vendor	
Range	
Resolution	
Rate	100.0 Hz
Standard deviation	0.0088 m/s²

Gyroscope	
Available	✓
Name	
Vendor	
Range	
Resolution	
Rate	100.0 Hz
Standard deviation	0.0026 rad/s

Magnetometer	
Available	✓
Name	
Vendor	
Range	
Resolution	
Rate	100.0 Hz
Standard deviation	0.24 μT

Pressure	
Available	✓
Name	
Vendor	
Range	
Resolution	
Rate	1.0 Hz
Standard deviation	0.011 hPa

Temperature	
Available	✗

Humidity	
Available	✗

Light	
Available	✗

Proximity	
Available	✓
Name	
Vendor	

Figure 18 shows the front and back view of the iPhone model used in this project. (Backmarket.co.uk, 2020)

4.3 Vemont action camera

One of the applications of the classification model was in the real-time mode detection task. To visualise and evaluate the participant's movements in real-time, a camera is required. The Vemont HD 1080p action camera in practice provides a clear view of the user's movements, allowing us to easily see which transportation mode is being used. The dimensions being 59.27 * 41.13 * 29.88mm also provided ease of use when strapped to a chest mount on the user.

Figure 19 shows the Vemont action camera used in project recording (Amazon.co.uk, 2021)

4.4 Laptop

To process the data on the MATLAB software was run on a personal computer, with an Intel i5-7200U CPU 2.50GHz processor with 8 GB of RAM, the tasks run did not use too much CPU so the personal computer was fit for the project.

Figure 20 shows Laptop used to operate the MATLAB software (Krosseyed, 2017)

Chapter 5: **Evaluation**

This project aimed to develop a machine learning algorithm that can detect the different modes of transport using sensor data provided by the user. In this section, I will be presenting the results achieved from this project, whilst evaluating the effectiveness of the chosen classification model. Additional applications in Web Application and Real-Time Mode Detection used in the evaluation will also be explained in this section.

5.1 Web app

When testing the model on new data the results produced by the MATLAB code was inconvenient to display on the interface, with 3 new test datasets being used in testing the model, a Web application seemed practical in displaying features such as the Confusion Matrix, F1 score and Accuracy that were used in evaluating the model. This application allowed the user to interact with the web application, to pick and choose which dataset they would like to display the results of, along with the ability to display a Confusion Matrix of the true and predicted classes of the model.

Figure 21 shows the interface of the web application

The web application is embedded into the main code after predictions are made on new datasets. This is necessary since the workspace of the code is linked to the application so any changes made in the code will automatically update the web applications workspace where the results are displayed.

To design the web application, the MATLAB toolbox Web App was required, which provided access to the app designer feature. Due to limited experience in application development, this feature was very useful as it strongly links the Layout and code view so code changes can be easily seen by the user. The environment works hand in hand with the MATLAB editor workspace so there is convenience in using the MATLAB web app. In comparison, Python was an option generally used in application development, however, due to limited experience in python programming, MATLAB was the fitting choice.

5.1.1 Accuracy and F1 score

Figure 22 shows the formula used to determine Accuracy and F1 score (Packtpub.com, 2021)

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

$$Precision = \frac{T_p}{T_p + F_p}$$

$$Recall = \frac{T_p}{T_p + T_n}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

In the interface of the web application, the parameters Accuracy and F1 score were displayed. These values allow us to determine the effectiveness of the model by providing numerical values that can evaluate success. The accuracy is computed by summing the values returned by the comparison of prediction labels against the actual label, this value is then divided by the number of labels to provide an average for accuracy. Initially, 31 features were extracted from the data producing an accuracy of 98.21%, however, some of the features were deemed unnecessary by the model and provided a cost in computing time in addition to increasing the energy footprint of the classifier. In comparison, when only 15 features were selected more effective results were produced, with an accuracy 98.0159%, the difference in accuracy however, was not enough to outweigh the negative computational effects of the 31 features, so the 15 selected features were deemed better for use in the final model.

F1 score is also used to measure the correctness of the model, it is a weighted average of precision and recall that can be determined using a confusion matrix. A combination of the true positives(correctly predicted positive values), false positives(incorrectly predicted positive values) and false negative(incorrectly predicted negative values) is used to determine the recall and precision. F1 score is very suited for real-life classification problems such as transportation mode detection since imbalanced class distribution generally exists and the miscalculation costs are considered important. For the selection of 15 features, a reading of 0.9795 for the model is received which is similar in ratio to the accuracy value, which lets us know that the model is still successful.

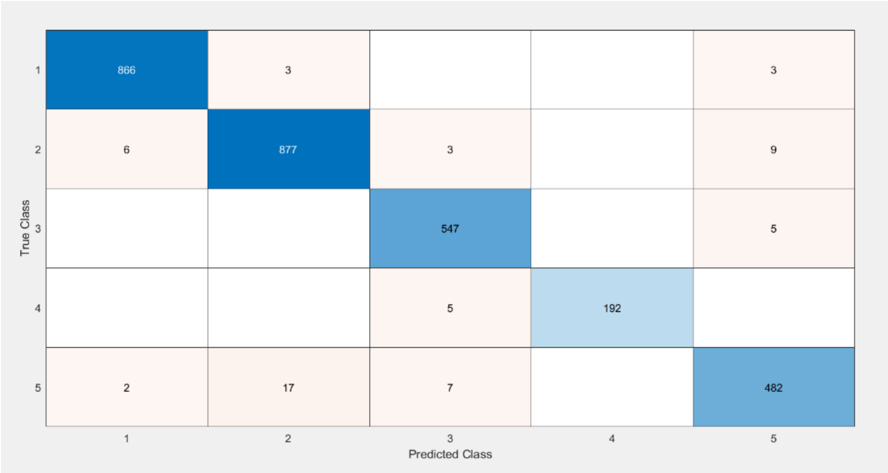
Figure 23 shows the computational time for the testing and training of 31 features in comparison to 15, attained via the Tic Toc MATLAB function

Feature No	Training time(seconds)	Testing time(seconds)
15	0.567	0.0382
31	1.543	0.132

5.1.2 Confusion Matrix

In the interface of the web app, the confusion matrix of the true and predicted classes of the model can also be displayed. A confusion matrix for each of the datasets can be produced, that provide a visual example of how effective the model is at classifying each mode of transport. This was crucial in the feature selection process, where it is possible to visualise where the model lacks and what class the model is ineffective on.

Figure 24 shows the confusion matrix for the total test dataset



Looking at the confusion matrix, the overall classification accuracy is quite high, however, when looking at individual modes we can see that the number of misclassifications was larger for the Bike and Drive modes. Due to the layout of London, Bike and Drive modes often got involved stoppages at traffic lights which may have been the reasoning behind the minor confusions between the different modes. However, these misclassifications do not have a drastic effect on the overall accuracy so these errors can be ignored.

Figure 25 shows the model accuracy for each mode

Mode	Accuracy
Still	99.08%
Drive	97.77%
Walk	97.33%
Run	100%
Bike	96.59%

5.2 Real-Time mode detection

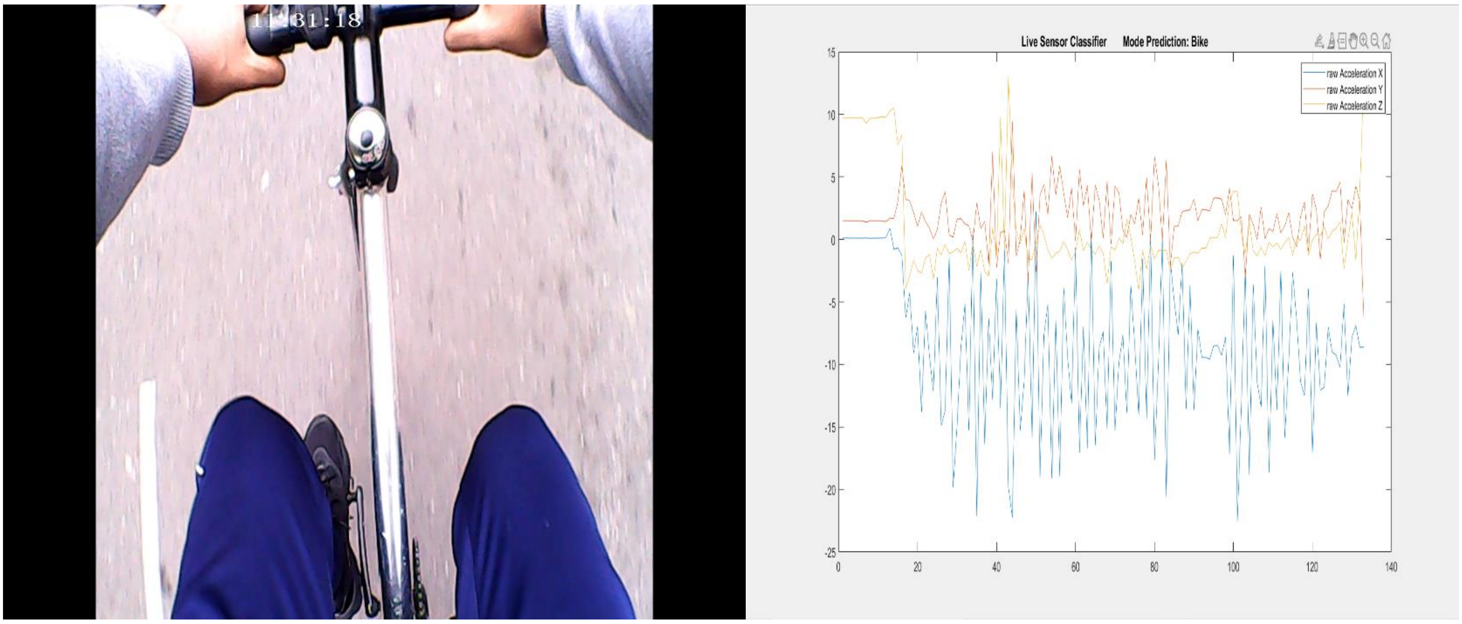
Real-time mode detection is an additional objective of the project, using the same trained model, the effectiveness of the model in detecting transport mode in real-time was tested. This application gives us an insight into how this transportation mode detection task could be applied in real life by larger organisations. Byon et al (2009) used GPS sensor data to determine the mode of transport in real-time and explains how this task could be used for niche applications such as the forecasting of travel times on transit vehicles.

MATLAB Mobile iOS app and MATLAB software were used in Unisom for this application. The mobiledev function in MATLAB is used to create an object that can communicate wirelessly with the iOS device. With this, the logging of sensor data in real-time can be initiated. The classifier used in the previous application of sensor detection was used here as well by loading the saved classifier file, the data from the live sensor readings can then be classified by applying the same 15 features used in previous applications. Using the tic toc function in MATLAB, a period of classification of 5 seconds was achieved. Although a shorter period would look better visually. In application, during testing, the accuracy saw a drastic drop at time windows of 1 and 2 seconds. This lets us know that at shorter time windows, there is not sufficient data for the classifier to be successful.

The effectiveness of the application is shown in the project video, where the process of logging and classification of the sensor data is shown. The activity in real-time is captured using the Vemont action camera, where the activity taking place can be seen. In addition, the plot feature is used to produce graphs that can show the predictions in real-time along with the live accelerometer values. Using the camera footage and MATLAB screen captures, the recordings can be compared side by side to see how the mode prediction is affected in real-time, along with how accurate these predictions are compared to the actual mode.

Figure 26 shows the placement of the Vemont action camera on participant

Figure 27 shows a screen capture of the real-time classification of the drive mode, along with the graph used to display the predictions of the classifier and variations in live accelerometer data.



Looking back at the video footage for the live classification of each mode, the application is generally successful at identification. With the processing delay and window time of 5 seconds taken into account, the data is correctly classified. Although there were minor misclassification errors with the Run mode being incorrectly classified as the Walk mode. In the classification of the 21-hour dataset, the same misclassifications are not widely present, so we can assume this error is down to the running pattern recorded during this real-time recognition task, where the participant was not running at a fast enough speed.

5.3 Limitations

The transportation mode detection project was successful in achieving the aims set out initially, however as I progressed through the project time plan, limitations became apparent in certain areas.

The first limitation was revealed in sensor data collection when GPS data was initially analysed. MATLAB mobile offers the ability to collect GPS sensor values in Latitude, Longitude, Speed, Course, Altitude and Horizontal Accuracy. However, there were issues in the sampling frequency of the sensor. Although the listed sampling frequency of GPS was 1Hz, in application big gaps in the collected data were visible. So, in addition to the low sampling frequency, in practice, there would be time windows without GPS data present. This lets us know that a larger time window would be required for classification which would impact the application of real-time mode recognition. With the data being collected in London, it is also understood that signal strength may vary which can produce these noted anomalies. Stenneth (2011) explains that to analyse the GPS data, a Kalman filter can be used to remove these anomalies. However, the positive effect of the GPS sensor in classification was not deemed big enough to warrant use in this project.

Figure 28 shows a test data collection with the anomalies in GPS(Position) data

Name	Value
Acceleration	*29630x3 time...*
AngularVelocity	*29630x3 time...*
MagneticField	*29628x3 time...*
Orientation	*29628x3 time...*
Position	*2759x6 timet...*

An additional limitation was present with the participants of the data collection. The Vemont action camera used in the real-time mode detection task could be used in the collection of sensor data, where the collection of data would be compared with the camera recording to determine when the change of transport occurs. In effect, this would allow the data to be processed into the different mode classes. Various limitations were present in this application, the data collection aimed to make the data collection simple for the participants in this project and to simulate its use in its potential application for the physically impaired, but the use of an action camera whilst travelling was not practical for the participants in everyday travel. In addition, the data would have to be processed frame by frame to determine the changes of transport, in contrast to the current method which allows the participant to label which mode of transport they are taking which makes it practical and easier to process.

The sampling frequency chosen was left at the default 10hz as mentioned previously to prevent gaps in data. In similar applications 100hz was popular in its use, however, in application on MATLAB mobile gaps in data was frequent. The default 10hz however never experienced any anomalies whilst providing enough data for the pre-processing step to be effective. Issues were also present when larger frequencies were used in the live mode detection task where the classifier would fail to detect the correct mode in addition to delays caused by the large amount of data being processed in real-time. The default sampling frequency provided few issues and was determined to be sufficient for its application.

For the web app application, a connection between the app and live mode detection task was attempted where the live mode predictions could be presented on the web application in real-time, though there were limitations in its implementation. It was found that the workspace of the web application could not be updated while its process is running, which prevented live predictions from being continuously displayed.

Chapter 6: **Conclusion**

This chapter concludes what went well in this project in its achievements, along with the challenges faced throughout the project's implementation. Future work that overcomes limitations in this project will also be discussed along with any possible implementations that coincide with the problem statement.

6.1 Achievements

Overall the project was a success, the targets set out in chapter 1 of my project have been achieved. The project aimed to develop a machine learning algorithm, that can detect the different modes of transport using sensor data provided by the smartphone user. Using 21 hours' worth of transportation data, in addition to real-time sensor data, the most effective frequency and time domain features were selected allowing us to investigate the effects of the different smartphone sensors. In the proposed method the Decision Tree classifier was used to evaluate the chosen data, which allowed me to compare its effectiveness in comparison to other commonly used machine learning classifiers. Not only was the aim of this project achieved, but I was also able to implement additional features such as live mode detection, which gives us a clearer idea of the possible future applications of this recognition task.

Upon finishing the project, research questions set out initially can now be answered from the various stages of progress in my project. Investigating the modes: Still, Walk, Run, Drive and Bike, the design achieved a relatively high classification accuracy of 98.0159%, whilst battling the impacts of factors such as overfitting and misclassifications that arose from the traffic-heavy manner of London transport. The focus was also put on user convenience, where the user requirements required no hardware installation and the data collection allowed participants in this project to go about their everyday activities undisturbed.

The success and user convenience associated with the project ties in with the potential application in health monitoring mentioned in the problem statement. We understand in a real-life scenario, the application must remain simple and convenient whilst maintaining a high success. So the methods used in this project simulate potential applications.

The project has allowed me to gain knowledge and develop skills in areas of personal interest. Having no prior experience in machine learning applications, I was able to use the knowledge gained from the literature reviews of similar projects and practice of the MATLAB software to make this project a success. Furthermore, my experience in Application development was limited, thus the Web Application task was new and allowed me to understand the mechanics behind a functioning web application. Developing experience in these different fields has made the project enjoyable and heightened my interest in the mode recognition area.

6.2 Challenges

Throughout the project, there were challenges that I faced that impacted progression in my time plan. The major and initial challenge I faced was down to my experience in MATLAB, although I had previously used MATLAB in various applications the machine learning task required a lot of time to initially grasp, this impacted my ability to progress swiftly through the project. Since I was constantly researching and learning it was difficult to anticipate upcoming challenges which led to standstills in the project progression, where alternatives had to be investigated.

An additional challenge came during the implementation of my model when investigating how to improve its performance. Although there were countless ways to improve the operation of the model, I lacked knowledge of the different features and applications. Hence, extensive research was required on how to apply certain features and whether they will be advantageous for use in the project.

Upon beginning the project, Covid-19 made it difficult to collect sensor data due to lockdown and travel restrictions. This posed a challenge since I was unable to investigate collected sensor data and its patterns at the early stages of this project. Instead, research was done into readily available sensor data to initially understand the project and its challenges. In addition, due to the length of the lockdown restrictions in place, this challenge, therefore, increased the time taken to complete the final sensor collection causing setbacks to the time plan.

6.3 Future work

The application of transportation mode detection is not yet properly established, which leaves room for various future improvements. In my case, I would intend to implement additional modes of transport, in specific Train, Motorbike and Bus. These modes would make the mode detection task more realistic for a real-life application since the classifier would cover all the generally used modes of transport.

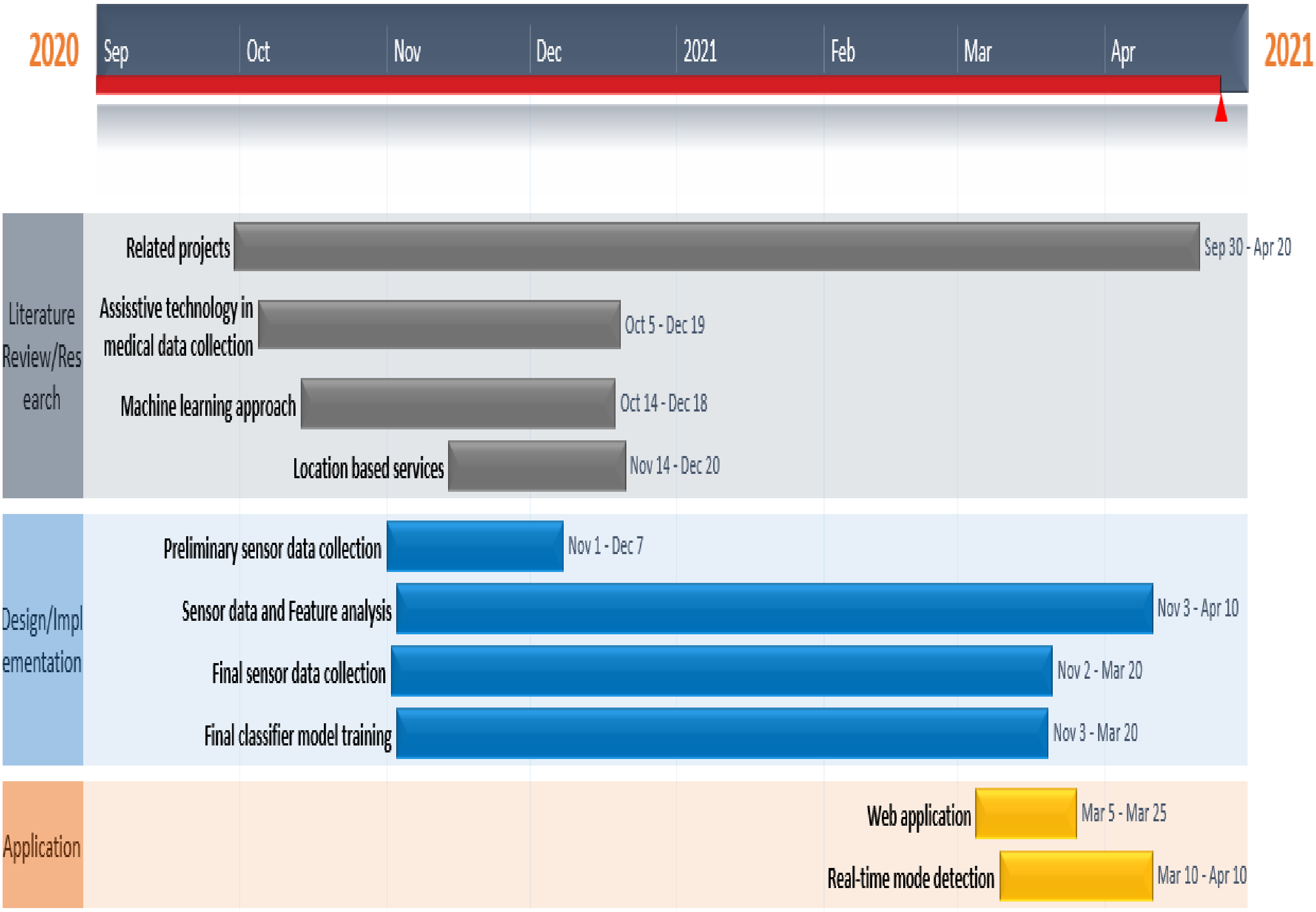
In my project, the iPhone 11 is placed in the front pocket of the user as a single sensor device for convenience and practical reasons. For future work as an objective, I would also like to investigate different sensor device placements to determine the effect sensor placement has on the accuracy of the classifier.

The problem statement for this project details the potential impact of mode detection in the monitoring of health and activity. This is an area I am interested in and for future work, I could use more participants with different physiological health levels to investigate the correlation between cognitive function and transportation mode choices.

Chapter 7: **Risk Assessment**

Description of risk	Impact of risk	Likelihood rating	Impact rating	Preventive Actions
MATLAB mobile stops functioning efficiently	-Prevents the collection of sensor data -Sensor data can no longer be uploaded to MATLAB drive	Low	Low	-Motion Data Logger IOS app as an alternative
Machine learning algorithm too hard to implement	-Project not completed in time The project not completed to sufficient quality	Medium	Medium	-Find alternative solutions to problems -Alter features and sensors used
Bad time management	-Project not completed on time -Project not completed to sufficient quality	Low	High	-Follow updated time plan -delegate time frequently spent on each stage -minimise procrastination
smartphone Hardware malfunction	-Prevents the collection of sensor data	Low	Low	-use a backup smartphone to collect data -cautious when using hardware
The sudden growth of requirements	-Project not completed on time -Time plan affected	Medium	Medium	-anticipate worst-case scenarios and plan for them
Participants drop out	-reliability of experiment effected - longer period required to collect data	Low	Medium	-backup participants can be screened for participation -start data collection early

Chapter 8: **Time plan**



References

Hedemalm, E. (2017). Online Transportation Mode Recognition and an Application to Promote Greener Transportation. [online] Available at: https://www.diva-portal.org/smash/get/diva2:1136300/FULLTEXT01.pdf.[Accessed 19/11/2020]

McPhee, J.S., French, D.P., Jackson, D., Nazroo, J., Pendleton, N. and Degens, H. (2016). Physical activity in older age: perspectives for healthy ageing and frailty. Biogerontology, [online] 17(3), pp.567–580. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4889622/figure/Fig1/ [Accessed 10/11/ 2020].

Schiller, J. Voisard, A. (2010). Location-Based Services. [online] Available at: https://books.google.co.uk/books?hl=en&lr=&id=wj19b5wVfXAC&oi=fnd&pg=PP1&dq=location+based+services&ots=leLo8qxiOp&sig=edJo5jp-kMx78yht5vYSXC_9ywQ#v=onepage&q&f=false [Accessed 15/11/2020].

Lara, O. Labrador, M (2017). IEEE Xplore Full-Text PDF: [online] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6365160 [Accessed 22/11/2020].

Ronao, C.A. and Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, [online] 59, pp.235–244. Available at: https://reader.elsevier.com/reader/sd/pii/S0957417416302056?token=67969B12B9C2E07B349D891EF7D0D7B2277570721CE2754F8A2D5D1CA34A1E5C9601083D37E02AE5AE0997F73392F924 [Accessed 22/11/2020].

Sakovich, N (2019). The Importance of Data Collection in Healthcare and Its Benefits | SaM Solutions. [online] SaM Solutions. Available at: https://www.sam-solutions.com/blog/the-importance-of-data-collection-in-healthcare/#:~:text=The%20importance%20of%20data%20collection%20and%20its%20analysis,is%20greatly%20influenced%20and%20altered%20by%20Big%20Data. [Accessed 22 Nov. 2020].

Sakovich, N (2020). Digital Reality of the Healthcare Industry: the Ultimate Guide | SaM Solutions. [online] SaM Solutions. Available at: https://www.sam-solutions.com/blog/digital-transformation-in-healthcare-why-is-it-important/ [Accessed 22 Nov. 2020].

Mandolesi, L., Polverino, A., Montuori, S., Foti, F., Ferraioli, G., Sorrentino, P. and Sorrentino, G. (2018). Effects of Physical Exercise on Cognitive Functioning and Wellbeing: Biological and Psychological Benefits. Frontiers in Psychology, [online] 9. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5934999/#:~:text=It%20has%20been%20showed%20that,Chang%20and%20Etnier%2C%202009 [Accessed 22 Nov. 2020].

Gonzalez, P. Weinstein, J. Barbeau, S. Labrador, M. Winters, P Georggi, N. Perez, R (2020). (PDF) AUTOMATING MODE DETECTION USING NEURAL NETWORKS AND ASSISTED GPS DATA COLLECTED USING GPS-ENABLED MOBILE PHONES. [online] Available at: https://www.researchgate.net/publication/251758946_AUTOMATING_MODE_DETECTION_USING_NEURAL_NETWORKS_AND_ASSISTED_GPS_DATA_COLLECTED_USING_GPS-ENABLED_MOBILE_PHONES [Accessed 10/11/2020].

Miluzzo, E., Lane, N.D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., Eisenman, S.B., Zheng, X. and Campbell, A.T. (2008). Sensing meets mobile social networks. Proceedings of the 6th ACM conference on Embedded network sensor systems - SenSys '08.[Accessed 05/11/2020]

Feng, T. and Timmermans, H.J.P. (2013). Transportation mode recognition using GPS and accelerometer data. Transportation Research Part C: Emerging Technologies, [online] 37, pp.118–130. Available at: https://reader.elsevier.com/reader/sd/pii/S0968090X13002039?token=CCA01C7C49E5A82F0F627CF3C1AFF55AEBA904A5AAA72C6358A7A44FF1912C6CC9AB1706B4ACCA707AD62E8C6B90C623 [Accessed 08/11/2020].

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M. and Srivastava, M. (2010). Using mobile phones to determine transportation modes. ACM Transactions on Sensor Networks, 6(2), pp.1–27. [Accessed 16/11/2020]

Zhao, H., Hou, C., Alrobassy, H. and Zeng, X. (2019). Recognition of Transportation State by Smartphone Sensors Using Deep Bi-LSTM Neural Network. Journal of Computer Networks and Communications, [online] 2019, pp.1–11. Available at: https://www.hindawi.com/journals/jcnc/2019/4967261/#conclusion [Accessed 16/11/2020].

Stenneth, L., Wolfson, O., Yu, P. and Xu, B. (n.d.). Transportation Mode Detection using Mobile Phones and GIS Information. [online] Available at: https://www.cs.uic.edu/~boxu/mp2p/acmgis11-mode-detection.pdf [Accessed 20/11/2020].

Bao, L. Intille, S.S (2004). Download Limit Exceeded. [online] Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.839.8212&rep=rep1&type=pdf [Accessed 19/11/2020].

Kurnia, Y. and Kusuma, K. (2018). Comparison of C4.5 Algorithm, Naive Bayes and Support Vector Machine (SVM) in Predicting Customers that Potentially Open Deposits. bit-Tech, 1(2), pp.40–47.[Accessed 20/11/2020]

Mathworks.com. (2020). VisibleBreadcrumbs. [online] Available at: https://uk.mathworks.com/help/matlabmobile/ug/sensor-data-collection-with-matlab-mobile.html [Accessed 22 Nov. 2020].

Byon, Y.J. Abdulhai, B. Shalaby, A. (2009). Real-Time Transportation Mode Detection via Tracking Global Positioning System Mobile Devices. [online] Available at: https://www.tandfonline.com/doi/full/10.1080/15472450903287781?scroll=top&needAccess=true [Accessed 24 Apr. 2021].

Phyphox.org. (2021). [online] Available at: https://phyphox.org/sensordb/ [Accessed 24 Apr. 2021].

Backmarket.co.uk. (2020). iPhone 11 64 GB - Black - Unlocked. [online] Available at: https://www.backmarket.co.uk/second-hand-iphone-11-64-gb-black-unlocked/290057.html?shopping=gmc&gclid=Cj0KCQjw9_mDBhCGARIsAN3PaFNdjtn

HGjbkqp7z2YnxVHNTE-bcpxPgMuKUuaISOBad1skSty5GzPoaAt70EALw_wcB [Accessed 24 Apr. 2021].

Amazon.co.uk. (2021). VEMONT Full HD 2.0 Inch Action Camera 1080P 12MP Sports: Amazon.co.uk: Camera & Photo. [online] Available at: https://www.amazon.co.uk/gp/product/B06Y65B76R/ref=ppx_yo_dt_b_asin_title_o02_s00?ie=UTF8&psc=1 [Accessed 24 Apr. 2021].

krosseyed (2017). [Guide] Dell Inspiron 13 5378 (2-in-1) - MacOS 10.12.6. [online] tonymacx86.com. Available at: https://www.tonymacx86.com/threads/guide-dell-inspiron-13-5378-2-in-1-macos-10-12-6.230009/ [Accessed 24 Apr. 2021].

Packtpub.com. (2021). {{metadataController.pageTitle}}. [online] Available at: https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781785282287/10/ch10lvl1sec133/computing-precision-recall-and-f1-score [Accessed 24 Apr. 2021].

Cooper, E., Gates, S., Grollman, C., Mayer, M., Davis, B. and Bankiewicz, U. (2019). Transport, health, and wellbeing: An evidence review for the Department for Transport Prepared for: Department for Transport Prepared for: The Department for Transport. [online] . Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/847884/Transport__health_and_wellbeing.pdf.