Master Computer Science

Augmenting existing activity recognition classifiers for inactivity detection of the elderly

Name:	Frederick Philippe van der Meulen
Student ID:	s2294273
Date:	[dd/mm/yyyy]
Specialisation:	Advanced data analytics
1st supervisor:	Dr. Matthijs van Leeuwen
2nd supervisor:	Dr. Arno Knobbe

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

This computer science research focuses on the field of activity recognition, black-box classifiers and machine learning. These topics are relevant to the goal of this research, which is to determine whenever improving black-box classifiers with sensor data gathered from participants with mobility issues would provide an increase in accuracy of detecting activities for mobility impaired humans. The three research fields in this research, activity recognition, black-box algorithms and machine learning, are popular fields of computer science research and contain works that are either related to activity recognition for elderly or mobility impaired people. There are also many researches related to opening, understanding and mimicking black-box classifiers. The goal of this research is to analyze and improve a black-box algorithm with additional data. This would increase the performance of activity recognition and will also prove that black-box classifiers can be improved, opening up the possibility to increase the performance of existing classifiers for research purposes. Improving a black-box algorithm has been researched before, however, it has not been researched in combination with activity recognition. This research was affected by the Covid-19 pandemic in that the sensor data could not be collected from the elderly, due to the Covid-19 regulations set by the Dutch health Institute. Instead, the data was collected from a smaller group of participants with ages under 30. The results from this research are affected by this.

This research started with creating a dataset containing sensor data from the accelerometer and gyroscope, which was obtained from performing several activities that elderly could perform. A publicly available dataset was considered, however, there are not any publicly available datasets that contain the desired sensor data that fits the participant category. Once the dataset was created and several analyses had been performed, it was possible to build the mimicked black-box classifier. The mimicked classifier was then augmented using gyroscope sensor data. Using an incrementally trained baseline classifier, the performance of the three classifiers was compared based on their accuracy, precision/recall and AUC values. Since the classifiers are multi-class models, an One vs Rest approach was used to calculate the precision/recall and AUC values. An N-fold approach was used to test the three classifiers. The experiments reveals that the classifier that was augmented with the sensor data was performing slightly better than the baseline classifier and the black-box classifier in some cases, but the baseline classifier is performing better in the other evaluation metrics. The conclusion therefore was that the performance of a black-box classifier can be increased by augmenting the black-box classifier with additional sensor data.

Contents

Chapter 1

Introduction

Park Vossenberg is a revalidation center for short- and long term health care and a nursing home for elderly located in Kaatsheuvel. The revalidation center offers apartments for people that need light or heavy care. The apartments are connected with hallways that lead to meeting spaces and other relevant zones. Park Vossenberg has started renovations in 2017. These include the renovations of the apartments and the addition of a park, which were completed in 2021.

The goal of these alterations was to give more freedom to the elderly with dementia, so that the number of incidents and negative side effects of dementia are reduced [5]. The alterations include a garden that the elderly have access to and two open living rooms in which the elderly can walk and talk. For researchers, it could be interesting to analyze the activities that these elderly are performing to learn more about dementia. For instance, what activities do these elderly perform in different stages of dementia? How active are they? Activity recognition is the field of using classifiers and/or other tools to detect activities from sensor data [49]. Most of the research papers in this field are using two different kinds of sensors for recognizing activities: external sensors and wearable sensors. External sensors are sensors that do not have a direct physical contact with the user, such as RFID tags, motion sensors, camera's etc. Wearable sensors are sensors that are worn by the user. These wearable sensors often have an accelerometer, a sensor that measures the acceleration, a rotation vector sensor and a sensor that detects angular velocity called the gyroscope [26]. This research focuses on wearable sensors for recognizing activity.
For research groups, such as Project Wearables (a project dedicated to using smart bands on elderly for research purposes), or for the Computer Science Institute of the University Leiden, (LIACS), the sensor data from the fitness bands is of great value. Based on this sensor data, behavioral studies, activity recognition studies or classification related studies could be performed. LIACS is collaborating with Zorggroep Elder Maasduinen, the owners of Park Vossenberg and Nivel, the Dutch institute for healthcare research. LIACS uses the Samsung fitness bands on the residents of Park Vossenberg to gather their data for research purposes. The sensor data from the Samsung wristband, such as the accelerometer, gyroscope and the rotation vector sensor, can be accessed with the Tizen API, the operating system of Samsung [16]. The classifiers from the Tizen API use this raw sensor data to detect activities.

1.1 Problem statement

The Tizen classifiers are made for people that have no problem performing physical activities. However, in the case of the elderly from Park Vossenberg, these problems have more difficulty performing those activities. This in turn causes that the Tizen classifier will identify that these elderly haven't done enough physical activities on a day, which isn't true in the case of the elderly.
By augmenting the existing Tizen Activity recognition with additional sensor data, these activities could potentially be detected more accurately, which would help with identifying how active the elderly were. However, due to their patented usage, the Tizen classifiers are not internally accessible, therefore, another way to verify that the augmentation works needs to be investigated.

The research question can be formulated as follows: How can existing Activity Recognition classifiers be augmented using raw sensor data to detect the activity of elderly?

1.2 Approach

The easiest way to answer the research question would be to retrain the Tizen classifiers with additional sensor data. However, Samsung will not share the source code or how the classifiers are trained since the Tizen software is confidential information according to Samsung. Even so, by approaching the Tizen classifiers as black-boxes, it is still possible to work with the Tizen classifiers.

The first step of this research is to create a custom dataset containing sensor data from activities that elderly humans could perform during the day. This sensor dataset will contain sensor data from the accelerometer, since that is the sensor that is currently used by the Tizen classifiers, and the gyroscope sensor, since that sensor is often used in other activity recognition machine learning studies [34]. The dataset will also contain the labels that are generated from the Tizen Activity Recognition API. Participants will be asked to perform the activities, with tools for recording the sensor data and the labels. This process is explained in Chapter 3. With the dataset created, analysis on the dataset will be performed. This analysis will help with removing the noise from the dataset and finding out if any characteristics from the data will aid with creating the mimicked black-box machine learning model or the augmentation of the machine learning model. Chapter 4 will investigate which sensor features could be important.

Using the sensor features from Chapter 4, an approach for the mimicked black-box machine learning model, which is based on the Tizen classifiers can be created. As earlier mentioned, it is needed to create this mimicked black-box machine learning model due to the inaccessibility of the Tizen classifiers. The process of creating the mimicked black-box machine learning model is visualized in Figure 1.1 and will be further explained in Chapter 5.
The creation of the augmented classifier consists of augmenting the mimicked black-box classifiers. This process is visualized in Figure 1.2. In short, by retraining the black-box classifier with new sensor data (in this case from the gyroscope sensor), the performance of the augmented classifiers should be more accurate with detecting the different activities, since more

data is used to get a more accurate activity estimation. The augmented classifiers are based on incremental learning using sensor fusion techniques. The incremental learning is explained in Chapters 2 and 5, while the sensor fusion is explained in Chapter 5. The Tizen platform was able to record the Accelerometer and Gyroscope sensor data, and based on the classifiers, was able to generate the Tizen labels. Relevant works for this research can be found in Chapter 2. In addition of augmenting the classifier, steps for creating the training and test data will be analyzed and performed, as well as which evaluation metrics will be used to evaluate the performance of the mimicked black-box machine learning model and the augmented machine learning model. The last step in Chapter 5 consists of explaining how the baseline model will be created to compare the results of the black-box classifier and the augmented classifier.

The setup of performed experiments, the experiments itself and the results from the experiments will be discussed in Chapter 6. The entire research process will be discussed in Chapter 7 and based on the findings that this research has gathered, a conclusion will be formed in Chapter 8.

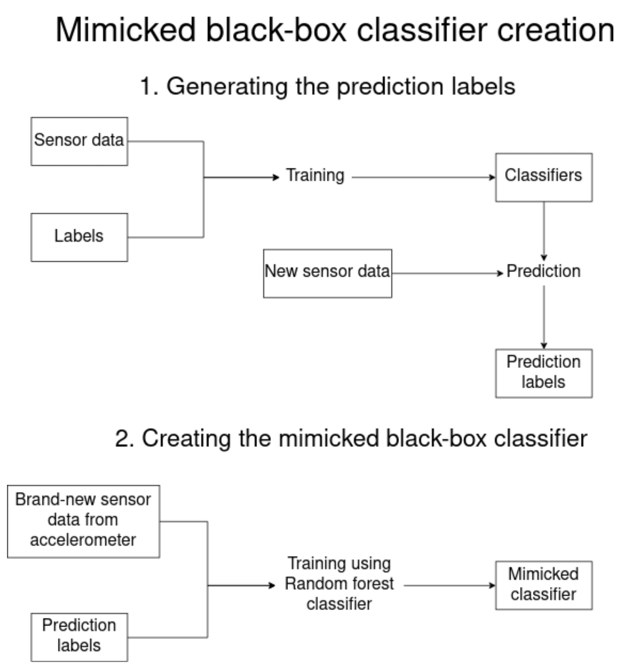

Figure 1.1: The process of creating the mimicked black-box classifier

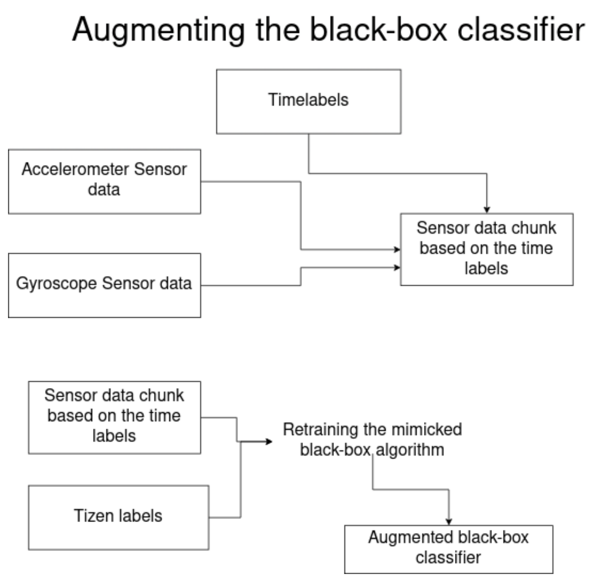

Figure 1.2: The inner workings of the augmentation of the black box classifier

Chapter 2

Related Work

This chapter discusses the related areas to this research. Much of the research that has been performed in activity recognition, sensor data analysis and black-box machine learning models can be applied to this research. Several state-of-the art papers regarding the aforementioned topics will be analyzed and whenever their mentions can be applied for this research.

2.1 Activity Recognition

This section discusses how related human activity recognition studies can contribute to finding a solution to the earlier described problem statement. This section discusses the state-of-the-art studies that have researched Human Activity Recognition and created classifiers for their studies.

2.1.1 Human Activity recognition classifiers

Human Activity Recognition, HAR in short, is a field of research that holds a diverse number of possible research topics, varying from sensor placement to analyzing sensor data to which sensor features will yield the highest accuracy from the sensor data. We use the surveys of Shoaib et al. [40], Lara et al. [26], and Wang et al. [49] for identifying the different activities, learning which sensors are useful for recognizing activities.
Shoaib et al. [40] created a Table in their study which describes the number of studies that have implemented a type of classifier on mobile phones for online activity recognition, which can be found in Table 2.1. According to this table, most of the classifiers that were created were based on Decision trees. As stated in Chapter 1, the goal was to mimic the black-box classifier using a Random Forest Classifier, which is based on the decision tree classifier type. The Random Forest Classifier will be further discussed in Subsection 2.3.3.

2.1.2 Used activity recognition platforms

The survey from Shoaib et al. [40] revealed that, based on the 35 studies they have researched, the majority uses the Android platform, followed by the now obsolete platform Symbian, iOS and finally Debian Linux. Their studies do not include the Tizen platform, which this study uses.
When looking at activity recognition studies that utilize the Tizen platform in any way, be it using the Tizen platform directly or using a smart device that use the Tizen platform, only

Implemented classifier type	Total relevant studies
Decision Tree	11
SVM	6
KNN	5
Naive Bayes	4
Multi-layer classifiers	3
Probabilistic Neural Networks	1
Rule-based Classifier	1
Quadratic Discriminant	1
Analysis	1
Decision Table	1
Fuzzy Classification	1

Table 2.1: Implemented classifiers for activity recognition from the survey of Shoaib et al. [40].

the studies from Srinivasan et al. [44] and Al-Naffakh et al. [3] use the Tizen platform. Both studies created their own application on the Tizen device that records the sensor data from the devices and uses that data for their machine learning activity recognition software. This study also uses an application to record the sensor data from the smart devices. What is different from this study with the aforementioned studies, is that this study also generated the labels from the Tizen activity recognition API, while the studies from Srinivasan et al. [44] and Al-Naffakh et al. [3] do not use these at all. Therefore, these studies are interesting in how they have gathered their sensor data.

2.2 Sensor dataset

As discussed in Chapter 1, this research wishes to create its own sensor dataset. This section will discuss how the sensor placement and the sensor feature analysis was determined.

2.2.1 Determining which activities to use

In order to create the sensor dataset containing sensor data from human activities, it needs to be determined which activities the elderly can perform and also which activities are used by state-of-the-art HAR systems. The survey of Lara et al. [26] created groups for the activities that are recognized by state-of-the-art Human Activity Recognition, which can be seen in Figure 2.1. Based on these groups of activities, the groups of Phone Usage, Military and Upper body are not interesting for this research. This is either due to activities that are to intense for elderly (Military) or activities that the sensors within the Wristband cannot pick up (Phone Usage and Upper body). The other activity groups have interesting activities that can be used to create a set of activities for the sensor dataset.

2.2.2 Sensor placement

As stated earlier in this chapter, we are looking at inertial sensors, sensors such as the accelerometer and the gyroscope, and mobile sensors in the Samsung Galaxy Fitband 2 Pro. The

Group	Activities
Ambulation	Walking, running, sitting, standing still, lying, climbing stairs, descending stairs, riding escalator, and riding elevator.
Transportation	Riding a bus, cycling, and driving.
Phone usage	Text messaging, making a call.
Daily activities	Eating, drinking, working at the PC, watching TV, reading, brushing teeth, stretching, scrubbing, and vacuuming.
Exercise/fitness	Rowing, lifting weights, spinning, Nordic walking, and doing push ups.
Military	Crawling, kneeling, situation assessment, and opening a door.
Upper body	Chewing, speaking, swallowing, sighing, and moving the head.

Figure 2.1: Types of activities recognized by state-of-the-art HAR systems, from the survey of Lara et al. [26]

sensors on the Fitband consists of a Bluetooth sensor, an accelerometer sensor, a gyroscope sensor, which are found on mobile devices [49]. Research by Wang et al. [49] mentions types of activity that can be recognized with the used sensors, the features that are extracted from these sensors and finally training a model that is based on a deep learning algorithm. Figure 2.2 visualizes this process.

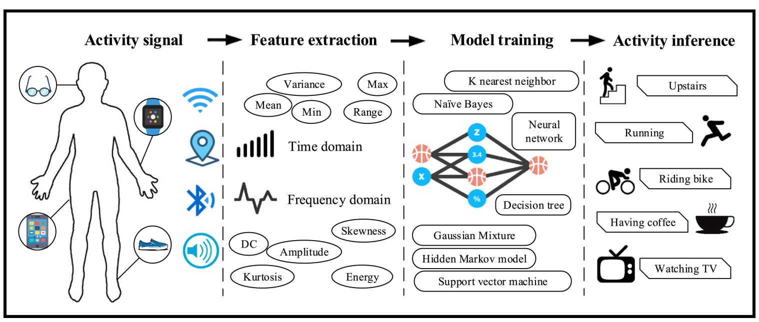

Figure 2.2: An illustration of sensor-based activity recognition using conventional pattern recognition approaches, from Wang et al. [49].

2.2.3 Sensor feature analysis

Using sensor features to help with recognizing an activity or to build an activity recognition model has been performed in various studies, most notably for this research the studies of Pirttikangas et al. [34] and Wang et al. [50]. According to these studies, there are several types of sensor features which can be categorized in the sensor feature domains of time, frequency or wavelet. These sensor feature domains are the common used sensor feature domains within the sensor data related studies [15] and [34]. This also includes the fields of activity recognition. These studies proved to contain much information on how to apply the sensor feature analysis on this research.

The study of Pirttikangas et al. [34] used the mean, standard deviation, correlation and the

mean crossing from the accelerator sensor. After applying a forward-backward sequential search algorithm, which is an algorithm that tests every feature for the classification one by one and uses that to create a subset of the best features and then removing the worst performing feature [34], Pirttikangas determined that the mean features were the most important features. This research also uses the mean sensor features.
The research from Wang et al. [50] uses, in addition to the features used by Pirttikangas et al. [34], also the Mode (the value with the highest frequency), Mean Crossing rate (Rate of times signal crossing mean value) and the DC (Direct component) to name a few.

2.3 Black-box machine learning models

This section will discuss the papers that were used for understanding black-box machine learning models, how they can be understood and how the Random Forest Classifier will help with understanding the black-box machine learning models.

2.3.1 How to understand black-box machine learning models

Understanding Black-Box classification models is an upcoming field of research, which is explored in a survey by Guidotti et al. [21], and more recently in the research of Burkart et al. [11]. Both studies make various points on why black-box machine learning models need to be explained. According to Burkart et al. [11], the main reasons to explain black-box classifiers are trust, causality and transferability to name a few. Guidotti et al. [21] describes various use cases where black-box machine learning models made mistakes in the areas of ethics, but also on safety, industrial liability and countability.
Interpretability is a major topic for the two surveys [11] [21]. Guidotti et al. [21] used this as one of the three requirements needed to define an interpretable model. Burkart et al. [11] goes into more depth by creating a problem statement called the explanation generation and applying it on machine learning models that are either easy to understand for humans and on difficult models. Both surveys note that using an interpretable model comes at the cost of flexibility, accuracy and usability. Guidotti et al. [21] states this as well, and notes that the accuracy of the interpretable machine learning model needs to be as close to the original model as possible. This is noted as their second requirement for creating an interpretable model. When applying this knowledge on this research, it means that comparing the improved black-box machine learning model with the interpretable machine learning model would not be a fair comparison, since the accuracy of the black-box model isn't taken into account. Both studies have divided the different machine learning models into several types. These are based on Decision Trees (which is a commonly used machine learning model for HAR [40]), Linear models and rule-based models to name a few.

2.3.2 LIME

Another method that was discussed in the works of Burkart et al. [11] and in the survey of Guidotti et al. [21], and which is interesting to apply on this research, is using the Local Interpretable Model agnostic Explanations (LIME) algorithm [36]. The goal of the LIME algorithm is to explain the predictions of any classifier using the input data, by approximating it locally with an interpretable model. LIME uses an Interpretability / fidelity trade-off while the model is still interpretable for humans. One advantage of using LIME, is that it is model-agnostic, meaning that it can be applied to any black-box classifier.

2.3.3 Random Forest

Random Forest is a type of ensemble classifiers that produces multiple decision trees, based on a randomly selected subset of training data [9]. An advantage of using Random Forests for activity recognition is easily combining the data from several sensors into one classifier [29] [12]. However, before this can be done, it is required that the sensors are on the same frequency, as otherwise you will get inconsistencies with the results of the algorithm. Random Forests are also useful to interpret black-box algorithms [21]. Guidotti et al. included a paper of using a random forest on a dataset, then using a decision tree on the random forest to analyze the internal behaviour of the random forest. They were in the end able to understand how the black-box labels were generated by the black-box algorithm [18]. A research from Tamagnini et al. [45] was able to analyze their random forest algorithm, which used a random forest on their dataset and then analyzed the Random Forest classifier using their own tool.

To verify the results from the black-box classifier and the augmented classifier, a baseline classifier will be created. This baseline classifier is based on an incremental learning model, which is a model able to continuously learn from new samples and is capable of containing most of the previously learned knowledge [28]. This makes that an incremental learning model is suitable for Human Activity Recognition, since the behaviour of a human is changing frequently over time. This is verified in by the research of Xiao et al. [51]. Using a incremental learning model has the advantage that, according to Xial et al. [51], it is flexible with learning new activities based on the sensor data, which makes the augmented model flexible as well. The incremental learning will be used in conjunction with the sensor fusion, which will be discussed in Chapter 5.

Chapter 3

Data Collection

This chapter discusses the creation of the dataset, why the sensor data from specific sensors such as the accelerometer and gyroscope were used and what their possible advantages for the augmentation could be. The remainder of this chapter discusses which activities were used for creating the dataset, and how the data gathering protocol was created. The data gathering protocol can be found in Appendix B.

3.1 Existing activity recognition datasets

There are several online datasets that are related to activity recognition. Some of the more popular datasets include the OPPORTUNITY [37], the USC-HAD [6] and the PAMAP2 [35] datasets. Comparing these three datasets reveals an overlap in the activities recorded for the datasets. The overlapping physical activities are walking, running, lying down, and ascending or descending staircases. These are common activities that are performed by humans, according to the creators of the datasets [6] [35] [37].

There are two reasons why these datasets were deemed unusable for this research. The first reason lies with the age group of the participants. The age group of the USC-HAD [6] and the PAMAP2 [35] dataset is 24 - 41 years old (the OPPORTUNITY dataset [37] did not provide an age group), which makes the participants too young to be considered elderly. Thus, this does not represent the target age group that this research wants to study. Using an age group that does not represent the target age group could result into wrongly augmenting the Tizen classifiers, since these younger participants overall have less difficulty in performing those physical activities, which could result into accelerometer and gyroscope sensor values that are not compatible with expected sensor values from an elderly group.

The second reason why the three online datasets are deemed unusable for this research, is that the activities were performed by healthy participants without the use of any equipment, such as a walking stick or walking with grocery bags for instance. This could affect the augmentation of the Tizen classifiers for the same reasons as mentioned above, namely that the participants do not match the desired age group.

Although there are more datasets about physical activity than the three described earlier, they only recorded data from the accelerometer, and not from the gyroscope or other sensors. Some of these datasets were interesting for this research, since these datasets either had an elderly age participant group, or had activities that matched what the elderly could perform. However, due to technical incapability's of the Tizen platform, these datasets were deemed unfit (further explanation in Chapter 7). Characteristics of the datasets can be found in Table 3.1.

The Opportunity dataset consists of sequences of activities instead of stand-alone activities.

Dataset	**Age group**	**Used sensors**	**Performed activities**	**Sample rate**
USC-HAD	21-49	Accelerometer, Gyroscope, Magnetometer, Orientation	Walking, running, jumping, sitting, standing, sleeping	100hz
PAMAP2	27-30	Accelerometer, Gyroscope, Magnetometer	Walking, sitting, Lying, Cycling, Nordic Walking, Rope Jumping	104hz
OPPORTUNITY	-	Accelerometer, Gyroscope, Magnetometer, Orientation	Prepare setting coffee, Drinking coffee, preparing a sandwich, eating a sandwich, cleaning	

Table 3.1: Characteristics of the three datasets. Six activities have been put in the table.

The accelerometer by itself is the most used sensor for recognizing activities [26]. Accelerometers are inexpensive sensors that are easy to use, since they are built in most cellphones. There are several papers that have reported high accuracy's for activities such as walking, running, cycling, etc [13] by using activity recognition based on accelerometers. However, for activities such as brushing teeth, writing and other 'smaller physical activities', the accelerometer-based classifiers were less effective [26].
Devices such as smartphones contain more sensors than the accelerometer. For instance, smartphones also have magnetometers and gyroscope sensors. According to the research from Shoaib et al. [41], combining the gyroscope sensor data and the accelerometer sensor data proved more accurate results in recognizing activities in general. Buenaventura et al. [38] came to the same conclusion. Using only the gyroscope for activity recognition had significantly less accurate results than using only the accelerometer. However, a combination of the accelerometer and the gyroscope proved to have a higher classification rate than using a single sensor [38].

3.2 Creating the Elderly Activity dataset

Before the dataset can be created, it is important to look at the existing Tizen classifiers, so that they match. This will make it easier to augment the classifiers, since the dataset has activity data that matches the currently existing classifiers.
The Tizen classifiers were based on running, walking, vehicle and stationary movement. One notable observation that can be made is, when looking at the names of the three classifiers, they seem rather generic. For instance, when talking about stationary activities, this can either mean that the user is standing still, lying down or sitting on a chair. The Tizen documentation [16] does not define the classifiers. To make sure that the Tizen classifiers are properly

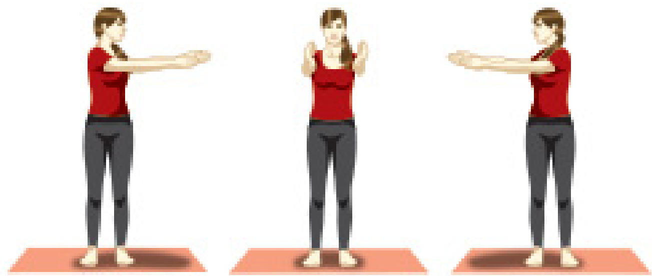

Figure 3.1: The spinal twist yoga move

augmented, there must be multiple activities per Tizen classifier, to make sure that most of the possible definitions of the Tizen classifiers are covered. Therefore, in order to improve the existing classifiers, it is important to add multiple activities that can refer to a single Tizen classifier.
This information gives criteria that the new dataset must fulfill: an activity in the new dataset must be related to a Tizen activity classifier, in order to be usable.
In addition to using multiple activities that correspond with the three Tizen classifiers, the activities that are performed must also contain specific activities performed by the elderly. For instance, when referring to walking, it is important to include whenever an elderly is walking with a walker or a walking stick. Based on this, an extra criterion can be made for an activity on the new dataset: an elderly person is able to perform one of these activities in their daily lives.

In Table 3.2, the activities that will be performed for creating the dataset, alongside a description, the duration for each activity and finally on which Tizen classifier the activities are based are described. These nine activities are selected because they fulfill the two criteria that the new dataset must fulfill: an activity must have a connection with a Tizen classifier and an elderly person must be able to perform one of these activities in their daily lives. The entire data gathering consent, is added in Appendix C. The data gathering protocol is added in Appendix B.

Name	**Description**	**Duration**	**Corresponding Tizen class**
Sitting	The participant is sitting on a chair	60s	Stationary
Lying down	The participant is lying down on the back on a yoga mat	60s	Stationary
Cutting	The participant is cutting a cucumber using a knife while sitting down	60s	Stationary
Yoga	The participant is performing the standing spinal twist yoga exercise (see Figure 3.1)	60s	Stationary
Walk with walking stick	The participant is walking with a walking stick, using that for support	120s	Walking

Walk groceries	The participant is walking with heavy grocery bags, with a weight of 4.3 kg per bag, on both hands	120s	Walking
Jogging	The participant is performing jogging movements	60s	Walking or Running
Sprint	The participant is running for 25m.	60s	Running
Cycling	The participant is cycling on a bike in the second gear	120s	In-Vehicle movement

Table 3.2: An overview of the activities that will be used to gather data

3.3 Using the dataset with the Tizen classifiers

The dataset cannot be used in its current form for the classification of the activities. This is due to the fact that there are only four kinds of Tizen Labels [16] (further described in Table 3.3), against the nine activities which are described in Table 3.2. This means that the data entries from the activities cannot be used as the ground truth for the dataset, since they differ from the Tizen labels. The option to use the labels from the nine activities instead of the Tizen labels is ruled out, since that would mean that the black-box Tizen activity recognition classifier is no longer mimicking the behaviour of its original counterpart. The problem is how to use the Tizen labels and the activity labels from the dataset together in such a way that the activities can be used in the classification of the Tizen labels and still create a ground truth.
The Tizen documentation [16] can help with solving the problem. Even though the documentation of the Activity Recognition is limited, the descriptions given by the labels provides some information about the activities. For instance, the website describes that whenever the user is in a stationary state, the stationary label is assigned, for walking the label gets assigned, for running the running label gets assigned and for movement happening with using a vehicle, the in-vehicle movement label gets assigned. Based on these descriptions, assumptions can be made on which activity from the dataset can be assigned to the Tizen labels. This results that an additional column is made in Table 3.2, with the possible corresponding Tizen label. This then reveals that there are four activities determined as a stationary activity, two as a walking activity, one that could be a walking or a running activity, one running activity and lastly an in-vehicle movement Tizen activity. This research will use this information as our ground truth for the classification problem, therefore solving the issue that the activities from the dataset cannot be used for the Tizen classifiers.
One of the activities, jogging, received the walking or running label in Table 3.2. This is since jogging could either be perceived as walking or running. When recording this activity, the label 2, running was the most appearing in the data. Therefore, for the rest of this research, the running label will be used with the Jogging activity.

When creating the dataset, one of the labels, the in-vehicle movement label, did not show up during the creation of the dataset. This gives the problem on what to do with the activity where this label would, in theory, be recorded: the cycling. Either the activity gets omitted from the dataset, the missing label will still be used, or the label gets omitted from the dataset.

Omitting the cycling activity from the dataset seems like a logical step. The label was theoretically assigned to this activity and since there was no recording of the Tizen label, then it wouldn't matter that the activity doesn't show up in the Tizen classifier. However, the sensor data that came out of this activity is valid data. Omitting this sensor data from the dataset does not seem logical, since it is still valid data. And even when the expectation does not match the reality, then it should still be used.

Using the in-vehicle movement label for the rest of the research has the advantage that the classifier is mimicking the Tizen classifiers. Even if there is no mention of this label, it can still be used for this research to compare against the other labels. However, this will affect the accuracy of the classifiers in a negative way. Since one label never gets 'recognized', it will never be detected and therefore the classifier will be less accurate with detecting the Tizen labels.

The last option, omitting the label from the dataset comes with the advantage that the sensor data of the cycling activity can still be used, even though they have a different label. Furthermore, there is less confusion regarding the evaluation of the classifiers, since the evaluation metrics only look at the three Tizen labels instead of four. There is also a disadvantage, in that it is not mimicking the Tizen classifier accurate anymore, since one label is missing.

Based on the three options, the decision was made to still use the in-vehicle movement label for the results. This way, the data from the cycling activity can still be used to train the classifiers. For the remainder of this research, the research will mention the three labels when creating the classifier and the sensor feature analysis. The missing label will be used in the results to discuss the evaluation metrics.

Tizen Label	**Definition**
0	Stationary movement
1	Walking movement
2	Running movement
3	In-Vehicle movement

Table 3.3: The labels from the Tizen classifiers

Chapter 4

Dataset Analysis

This chapter explains preprocessing the sensor data from the gathered dataset. The preprocessing of the sensor data consists of preparing the dataset by filtering sensor noise from the dataset, and discussing the effects of the sensor noise filtering. Sensor noise refers to unexpected values within the data, and can affect the training of an activity recognition model, since the model takes data that has outliers to use. To prevent the chance that the model is over fitting, feature selection will be applied on the available sensor data [52].

4.1 Preparation for sensor feature analysis

The most common way to filter out sensor noise is by using a noise filter. There are several sensor noise filters, such as a median filter, wavelet filter, and a Kalman filter [7] [14].
A Kalman filter uses a series of mathematical equations that estimate the state of a process (in the case of this research, an activity) recursively and effectively by minimizing the average square root error [4]. Kalman filters are commonly used in sensor related fields, such as activity recognition, due to the ability of handling noise on sensor data well and the minimal amount of performance needed when applied to the dataset [4]. Due to the minimal computational performance when applied on a dataset and how the filter handles noise from the sensor data well, it was decided to use a Kalman filter to filter out noise on the dataset, since the other noise filters are either to computational expensive [14] or could not handle sensor data as well as the Kalman filter [7].
A decision that had to be made was if the Kalman filter should be applied to the full dataset, the data of one participant (all performed activities of one participant), or for each activity separate performed by a participant. The first option, applying the Kalman filter on the entire dataset at once was ruled out, since there were different activities performed by the participants and predicting the next value of a sitting activity in comparison with a running activity could cause false noise to be filtered. This also rules out the second option, leaving applying the Kalman filter for each activity a participant performed as the chosen technique.
The next step of the preparation consists of coming up with an approach to implement a time window on the data. According to Chapter 3, the duration of the majority of the activities were 60 seconds long, with the exception of the cycling activity, which is 120 seconds long. By following the start and end times of each recorded activity, it is possible to segment the data in several segments.

Filtering the noise and segmenting the data based on their activity resulted into a cleaner

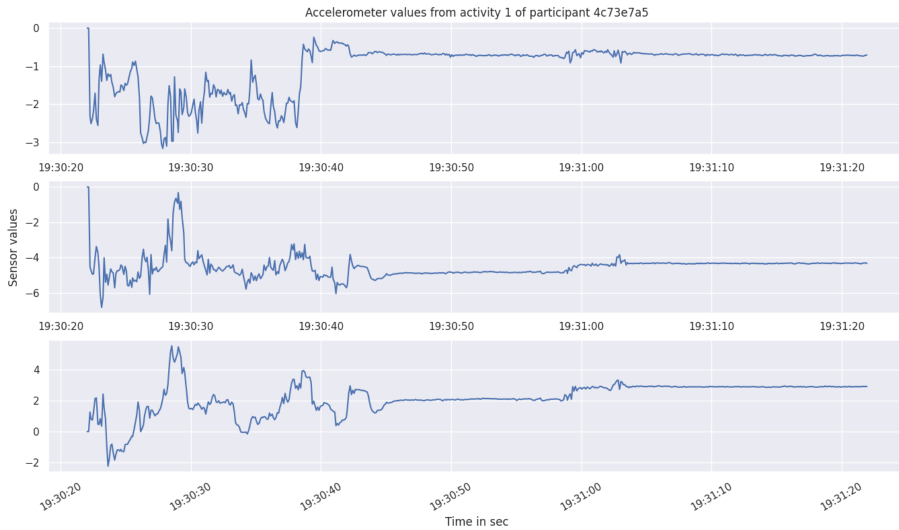

Figure 4.1: The filtered accelerometer data of an activity

dataset. With the clean dataset, sensor features can be extracted with reduced sensor noise.

4.2 Results of applying Kalman filter to the dataset

Figure 4.1 shows the results of applying a Kalman filter to a single activity of one participant. Looking at the differences between the two graphs, the first noticeable difference between the two graphs, is that the sensor values of the filtered activity have a lower range than the non-filtered activity, with the exception of the Z-axis of the Accelerometer. Especially the y-axis on the filtered activity has a lower sensor value range than the non-filtered.
However, in terms of noise reduction, there is little noise removed from the data, less than one would expect when applying the Kalman filter. This can also be seen when comparing Figures 4.2 and 4.3. There is a possibility that there was already a noise filter applied on the data, when the Tizen system was recording the data. This would make the Kalman filter less effective, since there was already a noise filter applied on the sensor data. However, the documentation itself does not refer to applying a denoising filter on the data. Even if it is unknown whether the data has already been denoised or not, it would still be useful to apply the Kalman filter applied on the data, due to the small amount of performance loss.

4.3 Sensor feature extraction and selection

This section aims at understanding sensor features, and how these could be applied to the black-box algorithm or the improvement of the augmentation of the black-box classifiers. It contains a short introduction to sensor feature extraction, the features that are going to be used for the black-box classifiers and the explanation of how sensor features will help with mimicking the black-box classifiers. Initially, the sensor features from the accelerometer will be taken, analyzed and determined if they will provide a benefit when creating the mimicked black-box classifier. For upgrading the black-boxes, the sensor features that were selected for the accelerometer can also be applied to the gyroscope sensor data. The sensors gather the data real-time at a set interval of 50-ms. This interval will be used to calculate the sensor features.

The common sensor features within the time domain are the mean, standard deviation and the min-max [14]. The mean is, in this case, the direct component of the signal, the standard deviation is the insensitivity of the signal (in the case of this research, that is the intensity of the activity) and the min-max feature shows the changing signal of the sensor.
Frequency related sensor features involve the Fourier Transformation and the Fast Fourier Transform. Fourier Transformation is an algorithm that changes the domain of a signal from a frequency-based signal to a time-based signal [39]. The Fourier Transformation algorithm has the disadvantage of taking a substantial amount of time to compute, since it computes the signal change for the entire amount of data. The improved version of the Fourier transformation is the Fast Fourier Transformation algorithm [39], which offers a higher computational performance when compared with the Fourier Transformation algorithm. Instead of computing everything at once, it computes the Fourier transformation in chunks of data instead of all the data at once. The chunks are then combined with each other.

The sensor features were calculated after preparing the data. The result of these calculations can be found in Table 4.1. Due to the size of the table, an exert of the sensor feature data is given. The remaining values can be found in Appendix A.
Observations of these data include the comparability of the x-axis medians of walking with a crane and walking with grocery bags. This makes sense, because the participants were making the same wrist-movements when performing these activities. Based on this data, the question would rise whenever it was useful to walk with the different objects, and instead have the participants perform another activity. Another observation is that the values y-axis mean, median and the z-axis median are close to each other. This is explainable, because the wrist of the participant does not move when performing the sitting and cycling. Since the participant is moving at a higher velocity when cycling, the x-axis mean value is higher than with the sitting activity.

One challenge that this research has, is that it is not possible to determine which sensor features are useful to create the mimicked classifier, due to the fact that it is unknown how the black-box classifier was build. In order to determine which sensor features would provide a benefit to the mimicked black-box classifier, papers in the field of activity recognition will be analyzed to find out which sensor features were used for training their classifier. After reading the studies performed in [14], [15], [50], and [7], the commonly used time sensor feature domains are the mean of the x-, y- and z-axis of the accelerometer and the overall standard deviation. This research has decided to use the means and the standard sensor feature domains of the accelerometer data for each of the performed activities and the standard deviation of the sensor values. This can provide additional information about the performance of each activity and whether the selected activities for the dataset would either be beneficial or a disadvantage for creating the mimicked black-box algorithm.

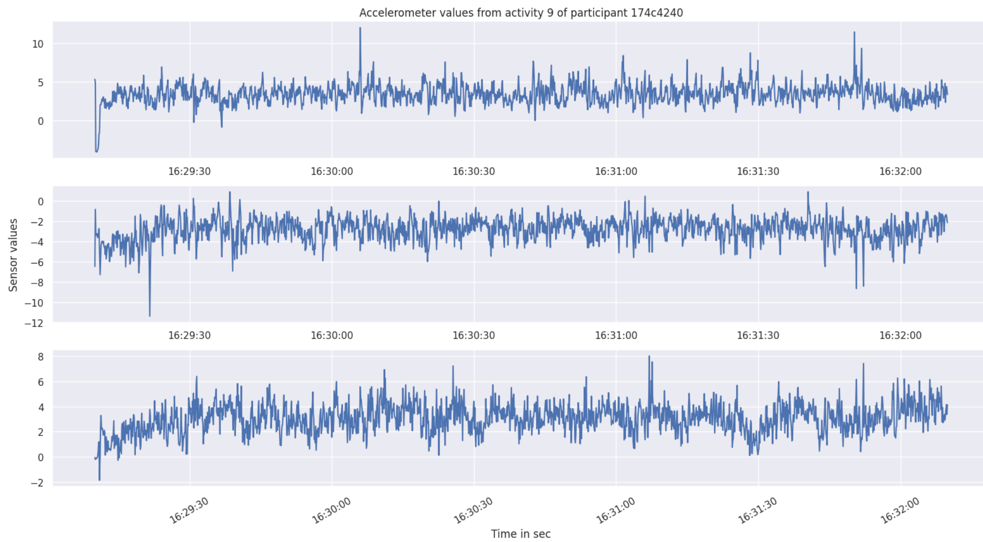

Figure 4.2: filtered

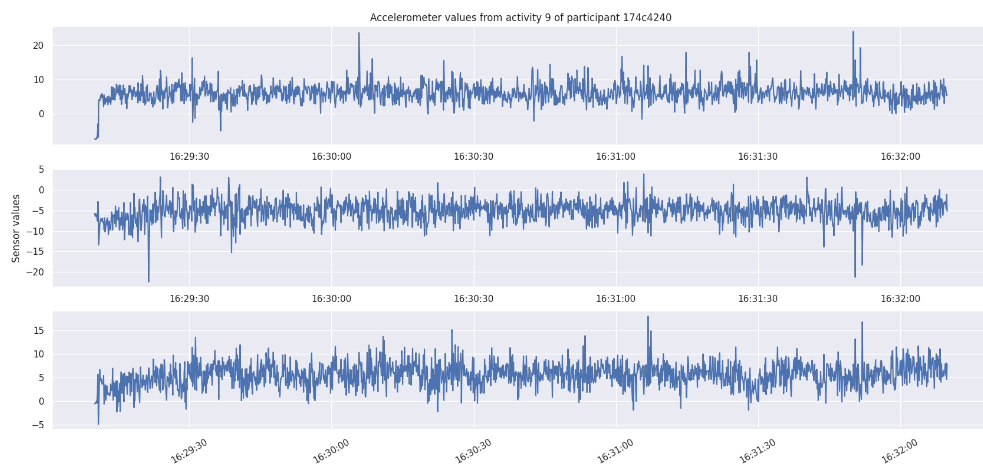

Figure 4.3: non-filtered

Activity	x-axis mean	y-axis mean	y-axis median	z-axis median
Sitting	0.524	-2.757	-2.582	3.936
Cutting	-0.104	-1.931	-1.826	4.876
Lying down	-2.087	-1.423	-1.435	4.396
Yoga	-0.028	-4.483	-5.016	2.817
Walking with walking stick	5.204	-1.714	-1.597	0.570
Walking with grocery bags	5.266	-1.266	-1.024	0.653
Jogging	1.775	-5.936	-5.283	-0.386
Running	6.860	-7.805	-7.221	-0.811
Cycling	2.974	-2.631	-2.517	3.466

Table 4.1: A sample of the sensor features from the accelerometer that were calculated

Chapter 5

Method

This chapter discusses how the algorithm for augmenting the black-box classifiers was created. The algorithm consists of creating an interpretable model from the Tizen black-box activity recognition classifiers. This interpretable model is then augmented using the additional sensor data, and finally the performances of the augmented and non-augmented black-box classifier are evaluated based on evaluation metrics. However, before the interpretable model of the black-box data can be created, it is important to make an attempt at understanding the black-box classifiers.

5.1 Understanding the black-box classifiers

Understanding the Tizen classifiers is important in order to augment them. Black-Box classifiers are sometimes seen as hard to comprehend, but, it is not impossible to understand them. For instance, taking the dataset from Chapter 3, the activity according to the Tizen classifiers is noted with the corresponding accelerometer values. Using this information, it is possible to understand that, given a certain accelerometer value, an activity gets assigned. Knowing when a certain activity gets assigned based on the input of the Tizen classifiers will help with augmenting the classifiers, since this knowledge can be used to improve the decision making.

As earlier discussed in Chapter 2, in order to understand the decision making process of the black-box classifier, one of the methods that was considered was the LIME algorithm [36]. However, using LIME in order to interpret the Tizen classifier does come with a disadvantage: using a local method to interpret a black-box classifier cannot yield an overall proxy of the black-box classifier according to Pedreschi et al [32]. This means that the LIME model of the Tizen black-box classifier is not a complete proxy of the Tizen classifier. In the end, the decision was made to not make use of the LIME algorithm for understanding the black-box classifier.

5.2 Base transparent classifier

In their survey, Guidotti et al. [21] discussed transparent classifiers. The term transparent classifier is also referred to as an interpretable classifier in their paper. This research will use the term transparent classifier to avoid confusion. A transparent classifier is, according to Guidotti et al. [21], either a Decision Rules based classifier or a Prototype Selection classifier. A Decision

Rules based classifier, which will be further referred to as DR, is based on a set of rules that are used to understand the black-box model. These rules are mined from the available data and can be based on the values from the accelerometer for instance. A Prototype Selection based classifier, further referred to as PS, returns a created example of an earlier classified record. Even though these created examples are not perfect matches of said classified record, they can be used to clarify based on what criteria the prediction was made. The example itself is also based on the rules from the dataset [21].
Guidotti et al. [21] has done multiple state of the art studies that are related to creating the two types of transparent classifiers. Based on the two types of transparent classifiers, a DR based transparent classifier will be used for this research, since this research uses sensor data, which makes it easier to use a DR based transparent classifier. A way to analyze the decisions from a black-box classifier is by using Random Forest classifiers, which is mentioned by Guidotti et al. [21] [20], Ahmed et al. [2] and Singh et al. [43]. In the mentioned studies, the authors have used a Random Forest classifier on their data from a black-box classification algorithm (varying from medical data [43] to educational data [2]) to create a Random Forest Black-box classifier, in which they then used different approaches in order to understand the decisions made by the algorithm.

For this research, a random forest classifier based on the Tizen labels will be trained on supervised data. This random forest will have the accelerometer sensor data and the corresponding Tizen label as input. This random forest is based on a DR classifier, since the random forest will look at the sensor data and attempt to find why Tizen recognized it as a specific activity based on the accelerometer sensor values.

5.3 Improving the mimicked black-box classifier

As mentioned earlier, the black-box classifier is improved using the gyroscope data that was generated alongside the accelerometer data. Obviously, the gyroscope sensor data lacks the labels generated from Tizen.

A field of research that could be interesting to investigate is the field of sensor fusion, which focuses on combining the output of several sensors into one dataset [48]. Sensor fusion is used often in the field of activity recognition to, for instance, improve the accuracy of an activity recognition classifier [15]. In the research of San et al. [38], several fusions of sensors were tested, such as the fusion of accelerometer with gyroscope sensors or gyroscope with the magnetometer sensor. They concluded that a fusion of the accelerometer and the gyroscope sensor offers a higher classification rate than using a single sensor.
By fusing the data of the accelerometer with the gyroscope data from the dataset, a higher classification rate based on the sensor data should be possible. However, there are several questions that first must be answered. For instance, the overall question is; why would sensor fusion be a good choice for augmenting the black-box classifier? Additional questions would be; when should the sensor data be fused? Which sensor fusion techniques exist and finally; which can be applied on the data from the dataset? This research must first answer these questions before sensor fusion can be considered to be used.

5.3.1 Using sensor fusion

It is logical to use sensor fusion techniques to fuse the accelerometer and gyroscope sensor for this research, since sensor fusion is generally used in the field of activity recognition. As stated earlier, there were several studies that concluded that sensor fusion improved the accuracy and the overall performance of the classifier. However, the situation for this research, is different, since the accelerometer data was already used to create the mimicked black-box classifier.
A possible solution for making the sensor fusion work is to retrain the mimicked classifier with the addition of the gyroscope data. Retraining a classifier with new or additional data is something that is used in the works of Hu et al. [24], Park et al. [31] and Adler et al. [1]. The advantages they mention of retraining an existing classifier, such as [33], and [47], seem applicable on this research. Retraining a black-box algorithm is a concept that is mostly used in the field of security [24] to improve upon an encryption algorithm or on a decryption algorithm. That said, the same concepts can be applied on this research.
The research of Arvidsson et al. [8] used a technique that involves combining two events of sensor data based on the timestamps, which could be applied here. The fusion of the accelerometer and gyroscope data should happen before the retraining of the black-box classifier. However, there is a minor timestamp difference between the sensor recordings. This difference can be migated by applying a rolling time window on the data. This is done already in the research of Peng et al. [33], where they fused their data before starting the training.

Improving the black-box algorithm will be done by retraining the black-box classifier that was made based on the accelerometer. First, a time window will be applied on the dataset in order to synchronize the times of the accelerometer and the gyroscope entries, which will result into a fusion of their data. Once the sensor data fusion is completed, the algorithm will be retrained with the fused data. Since using additional sensor data improves the black-box algorithm, it will improve the black-box algorithm in this case as well.

5.4 Evaluating the performance from the classifiers

In order to evaluate the performance of the classifiers, a baseline classifier, evaluation metrics, and training test data will be analyzed and used. The first part focuses on how a baseline classifier will help with measuring the performance of the black-box classifier and the augmented classifier. The second part will focus on how the training and test data are created and applied on the classifiers, and the third part will focus on the evaluation metrics how they will be utilized in order to measure the performance of the classifiers.

5.4.1 Baseline classifier

This research uses two classifiers: the black-box classifier and the augmented classifier. One of the main themes of this research is to proof that improving an existing classifier yields a better performance than the original classifier. This theme does raise the question whenever it would be beneficial to improve an existing classifier or to simply rerun it with the additional data. Furthermore, a comparison between the black-box classifier and the augmented classifier yields insufficient data on whenever it is an actual performance increase, since there is only one other classifier to compare it against. To verify that the augmented classifier is an actual improvement over the black-box classifier, a simplified classifier that is based on the other two

classifiers needs to be created that can be used to get a clearer view of the performance of the two other classifiers, a baseline classifier.
The baseline classifier is based on a random forest classifier [28], like the black-box classifier and the augmented classifier. This is done in order to create an accurate estimation of the black-box classifier and the augmented classifier. Unlike the black-box classifier, the input data is similar to the augmented classifier, which is the fused accelerometer sensor data and the gyroscope sensor data. This makes the baseline classifier an unique classifier: it uses the same steps as the augmented classifier, however, it is trained only one time, whenever the augmented classifier is trained twice, with the first time it created the black-box classifier, and the second time it takes the gyroscope sensor data as well. Table 5.1 gives an overview on the main differences between the three classifiers.

	Baseline classifier	**Black-box classifier**	**Augmented classifier**
With what data were they trained?	Accelerometer and gyroscope	Accelerometer	Accelerometer and gyroscope
How were they trained?	Random Forest classifier	Random Forest classifier	Retrained black-box
How many times was the classifier trained?	Two	One	Two

Table 5.1: An overview on the differences between the three classifiers

5.4.2 Creating the training and test data

In order to create the training and test data, the data from Section 3 will be used. As described earlier, the dataset consists of the sensor data from ten participants, which have performed the nine activities described in Chapter 3. Since this dataset is smaller than other activity recognition datasets, such as OPPORTUNITY [37] and USC-HAD [6], it is hard to segment the data into a training and test data segment. Therefore, a different approach on segmenting the data will be used. This subsection will discuss the matched-pairs and the cross-validation approach for creating the training and test data. A comparison will be made between the two approaches and based on their ad- and disadvantages, an approach will be selected.

A matched-pair dataset consists of splitting the data into several pairs [46]. The pairs are created based on the characteristics of either the data and/or the participants. A matched pair approach is often used in statistics. This research, the pairs could be created based on age, gender, whenever the person had the wristband on the left or right hand and other possible relevant characteristics. The next step consists of selecting a sample from one half of the pair for the training data and a sample from the other half for the test data. In reverse, the results of the classifier are compared with each half of the matched-pair to evaluate the performance of the classifier.
One advantage of using a matched-pair setup is that it can be applied on a dataset containing a small number of participants, and can still provide realistic results. However, in order for a

matched-pair dataset to succeed, it is important that there are enough metrics to create the pairs. Since most of the additional information about the participants is similar, it is difficult to create matched-pairs.

Cross-validation consists of re sampling the data in such a way that the generalization ability of a predictive model can be accessed, while preventing the occurrence of over-fitting [10]. In most cases, **n**-fold cross-validation is used, which means that **n**-folds are made from the dataset. In the case of this research, it means that the data from the ten participants will be re-sampled. This results into ten folds, each fold representing the sensor data of a participant. In order to be sure that the process is more realistic, Mannini et al. [30] proposed to use LOSO cross-validation, Leave One Subject Out. In a LOSO cross-validation, the folds consist of the subjects. Additional advantages of using LOSO cross-valiation for data consisting of subjects is that LOSO cross-validation gives more realistic accuracy estimates, as it uses different subjects for evaluations in different folds, and LOSO in the model selection phase should lead to more robust models.
One possible disadvantage of using LOSO, when using multiple age groups for the cross-validation, is that the accuracy of the model can be negatively affected, due to the great differences in performance [30]. Even though this can't be applied to the dataset that is gathered for this research, it is something to keep in mind for possible future studies based on this research.

Taking the advantages and disadvantages of cross-validation and matched-pair approached into consideration, it is preferable to use LOSO cross-validation on the model.

5.4.3 Evaluating the performance

In order to evaluate the performance of the mimicked black-box classifier and its improved counterpart, two evaluation metrics will be used to evaluate the performance of these classifiers [22]. Evaluation metrics can be divided into evaluation metrics based on threshold, probability and ranking. Each of these metrics have different aims and goals of evaluating a classifier [22]. In order to evaluate the augmented black box classifiers, it is important to evaluate them based on multiple metrics instead of just one. This is done in order to prevent any bias that can be generated within the experiment. Aside from taking the most common evaluation metric for evaluating classifiers, the accuracy, other evaluation metrics will be taken into account as well.

The accuracy of a classifier means how accurate the classifier is on identifying the correct label given a certain value. For this research, the average accuracy of the labels from the three classifiers will be calculated. The accuracy has the advantage that it is easy to compute, however, the accuracy itself is insufficient to determine the performance of a classifier, since it does not describe the actual performance of the classifier. Another disadvantage of using accuracy is that the accuracy is not reflective of the performance of the classifier whenever the data itself is imbalanced [17]. The accuracy of the three classifiers will be calculated and compared with each other. Even though it won't describe the actual performance of the classifier, it could be used to get a view on how the classifiers are performing.
The first evaluation metric, or evaluation metrics since these evaluation metrics are connected with each other, are the precision and recall of the classifier. The precision of a classifier is the ratio of correctly classified positives to the total number of classified instances as positive,

while the recall of a classifier is the ratio of correctly classified positive instances to the total number of positive instances [19]. The precision and recall metrics are calculated with the following four values:

1. True Positives (TP): The number of positive instances that were classified as positive
2. False Positives (FP): The number of negative instances that were classified as positive
3. True Negatives (TN): The number of positive instances that were classified as negative.
4. False Negatives (FN): The number of negative instances that were classified as negative.

The precision of a classifier, also known as the true Positive Rate, is calculated using the TP and the FP:

$$Precision = \frac{TP}{TP + FP}$$

While the recall is calculated using the TP and FN:

$$Recall = \frac{TP}{TP + FN}$$

The precision and recall values are often used as evaluation metrics for evaluating the performance of a classifier. The precision and recall values are a more accurate representation of the performance of a classifier than the accuracy of the classifier [17]. The TP, FP, TN and FN values will also be used to calculate a confusion matrix for each of the three classifiers. Confusion Matrix is a popular measure for solving classification problems [25] and consists of a table that is used to define the performance of the classification algorithm [42]. Usually, the precision and recall scores are calculated for binary classifiers which are classifiers with two classes. The classifiers that are used for this research however have three classes (stationary, walking, and running, as described in section 3.3), which means that the usual way to calculate the precision and recall scores of an classifier cannot be used. To overcome this, there are two different ways to calculate the ROC of a multiclass classifier: One vs Rest and One vs One.
One vs Rest means that one class, which will be labeled as the positive class, is compared against the other classes, which are labeled as negative classes, therefore transforming the multiclass classifier to a binary class [19]. This process is repeated for each of the other classes, which results into having a precision and recall score for each class, which means that there will be three precision and recall scores for each classifier.
One vs One works on the same principal as the One vs Rest method, however, instead of classifying the other two classes as negative, only one class is labeled as negative [23]. This results into six different precision and recall scores for each classifier, one for each possible combination. This research will use the One vs Rest method, since by comparing the stationary behaviour class, zero, against the active behaviour classes, one and two, will reveal whenever stationary behaviour in general is easier to detect or active behaviour.

The second evaluation metric is the Area Under the ROC Curve, (AUC). AUC is a ranking-based measure of performance, which is mostly used in binary classification problems [27]. The AUC value can be interpreted as the probability that a randomly selected positive sample will rank higher than a randomly selected negative sample.
Before the AUC of a classifier can be calculated, the ROC needs to be calculated first. The

ROC is calculated using two metrics: Precision and False Positive Rate. The precision is already calculated and used as a metric, the False Positive Rate is calculated using:

$$FalsePositiveRate = \frac{FP}{FP + TN}$$

An advantage of using the AUC is that it is a more discriminating performance measure than accuracy, therefore giving a more accurate view on how the classifier is performing than the accuracy alone. A disadvantage of using the AUC metric would be that it has a high computational cost [27] [22]. However, this disadvantage is negligible for this research. Usually, the ROC is calculated for binary classifiers. As stated earlier, the method of One vs Rest was used for calculating the different precision and recall scores for each label for each classifier. The One vs Rest method was also applied on the ROC curves. This resulted into three ROC curves for each classifier, one for each of the available classes. This makes for a total of nine curves.

This subsection looked at how to evaluate the performance of the augmented black-box classifier. This was done by looking at the accuracy, the precision and values, and the AUC of the augmented classifier. This was compared it with the black-box classifier to verify that augmenting the black-box classifier improves the original black-box classifier.

Chapter 6

Experiments and results

In order to verify that the improved black-box classifier is performing better than the original black-box classifier, several experiments will be performed to verify these claims using the earlier determined evaluation metrics. These experiments are divided into evaluating the performance of the black-box classifier, improved black-box classifier and the baseline classifier separately from each other and then divided into comparisons with the other classifiers. First, this chapter will discuss the setup of the experiments, followed by the experiments that are conducted for this research including their results. As discussed in Chapter 5, the results of the experiments are based on the accuracy, precision/recall and the AUC curve from each classifier.

6.1 Experiment setup

The experiments will use the data collected for this research. Since **n**-fold cross-validation is used for creating the training- and the test datasets, the experiments will use the test set that was generated from the cross-validation. In order to achieve the highest possible performance, while reducing over fitting, three different numbers of estimators will be used. These values can be found in Table 6.1. In the end, 100 estimators were used, since this number of results delivered the best performance on the metrics, while not causing over-fitting to take place.
In the end, a fold size of 10 was used, since the sensor data from 10 participants was gathered. The training of the model was operated on a computer with a AMD-Ryzen 3800X processor and 32GB RAM. The algorithm itself used four processor cores, and took two hours in order to train the model.

Number of estimators
10
50
100

Table 6.1: The number of estimators used for the algorithm

6.2 Results

Each classifier, the baseline classifier, the mimicked black-box classifier and the augmented black-box classifier, were tested based on the evaluations metrics that were determined in

Chapter 7.3. These were the accuracy, precision/recall values, and the AUC of each classifier type. For consistency, the augmented black-box classifier will be referred in this section as augmented classifier. The results make use of the True Positives, True Negatives, False Positives and False Negatives that were generated from the three classifiers.

6.2.1 Accuracy

The accuracy's of the three classifiers can be found in Table 6.2. The closer the accuracy is to 1.0, the better the classifier is at correctly classifying the performed activity. The majority class of the labels is the stationary label, 0. This label makes for 69.4 percent of the entries in the dataset, which means that the dataset is imbalanced. As stated in section 5.4.3, this means that the accuracy of the three classifiers is not reflective of the performance of the classifier.

The accuracy from the three classifiers is poor. However, for the three classifiers they are above average. The black-box classifier has a mere zero point two percent higher accuracy score than the augmented classifier. The baseline classifier is performing better than the augmented classifier and the black-box classifier. Based on these accuracy scores, a black-box classifier has a positive effect on the accuracy. The results were in line with the expectation that the augmented classifier would perform better than the black-box classifier, however, the augmented classifier is falling behind the baseline classifier, although this difference is small.

Classifier type	Accuracy
Black-box classifier	0.62
Augmented classifier	0.64
Baseline classifier	0.71

Table 6.2: Accuracy results

6.2.2 Precision/Recall

The precision and recall scores from the three classifiers can be found in Tables 6.3, 6.4 and Table 6.5. This is made into three tables due to the usage of the OvR method, since the classifiers are multi-class classifiers (This was stated in section 5). As stated earlier, the higher the precision is of a classifier, the higher the chance that whenever it predicts something correct. Furthermore, as stated, the higher the recall is of a classifier, the higher the chance that the relevant labels are returned.

Looking at the precision scores from the three classifier, there is a notable difference between the precision scores of label 0 and the other two labels, where label 0 outperforms the other two labels. This difference is most notable with the black-box classifier. The differences between the three classifiers is difficult to determine, since the 0 label is performing better with the black-box classifier, but the 1 and 2 labels are performing better with the baseline classifier. One thing that is noticeable, is that the 1 label is missing from the precision scores for the black-box and augmented classifier. The baseline is therefore the classifier with the best precision scores, since all the scores are present.

The recall scores shows the trend of the precision scores, meaning that there is again a difference between label 0 and the other two labels. However, where there was a difference present with the precision scores, this difference with the recall scores is even more present, with label 0 being even more present than the other labels. Between the classifiers, the baseline classifier is barely outperforming the augmented classifier, with having a higher recall scores on the 1 and 2 label, although these scores are minimal. As with the precision scores, for the black-box classifier and augmented classifier, the recall scores for the 1 label are not present at all, making the baseline classifier the classifier with the best recall scores.

	Label 0	**Label 1**	**Label 2**	**Label 3**
Precision scores	0.70967742	0	0.71052632	0.
Recall scores	0.84615385	0	0.71052632	0.

Table 6.3: Precision and recall values from the baseline classifier

	Label 0	**Label 1**	**Label 2**	**Label 3**
Precision scores	0.64788732	0.	0.61538462	0.
Recall scores	0.80701754	0.	0.51612903	0.

Table 6.4: Precision and recall values from the black-box classifier

	Label 0	**Label 1**	**Label 2**	**Label 3**
Precision scores	0.66666667	0.	0.57142857	0.
Recall scores	0.84210526	0.	0.51612903	0.

Table 6.5: Precision and recall values from the Augmented classifier

In addition to the precision and recall values from Tables 6.3, 6.4 and Table 6.5, confusion matrices for the baseline, black-box and the augmented black-box classifiers have been made from the TP, FP, TN and FN values. The confusion matrix for the black-box classifier can be found in Figure 6.1, for the baseline classifier in Figure 6.3 and the confusion matrix for the augmented black-box classifier can be found in Figure 6.2.

6.2.3 AUC curve

The last performance metric that will be used to evaluate the performance of the three classifiers is the Area Under the ROC curve metric, AUC in short. As written earlier, since the ROC can only be used on classes within a binary classifier. Therefore, the OvR method will be used to get the ROC of the three different classes, of the three classifiers. This makes a total of nine graphs, one for each class of the classifier. The figures will also include the probability distribution of each class within the classifier. These ROC curves are grouped per classifier in Figures 6.4, 6.5 and 6.6. Finally, Table 6.6 contains the AUC scores from the nine ROC curves, while Table 6.7 shows the average AUC scores from each classifier.

The closer the AUC value is to 1.0, the better the model is at predicting what kind of activity took place at the time of recording. Only the running label is the most close to 1.0, while the other two labels are falling behind. The worst performing label is the stationary label, 0, while the best performing label is the running label, 2. Between the classifiers is the augmented classifier performing slightly better on average than the black-box classifier and the baseline classifier.

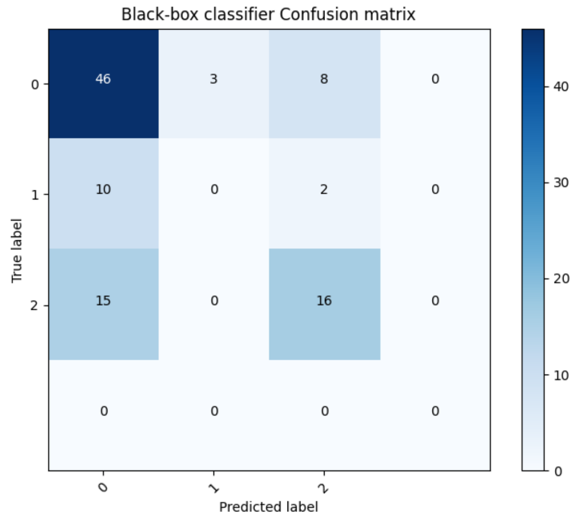

Figure 6.1: The confusion matrix of the black-box classifier

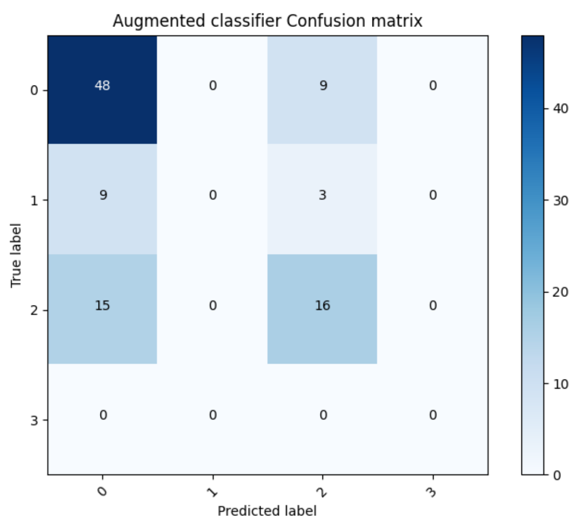

Figure 6.2: The confusion matrix of the augmented black-box classifier

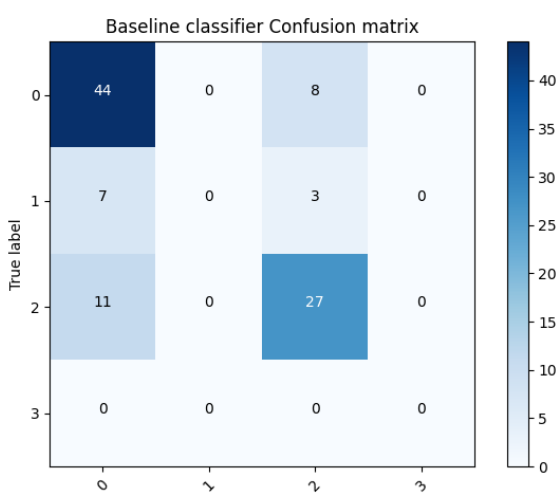

Figure 6.3: The confusion matrix of the baseline classifier

Classifier type	Class	AUC score
Black-box	0	0.376
Black-box	1	0.432
Black-box	2	0.675
Black-box	3	0.
Augmented	0	0.358
Augmented	1	0.483
Augmented	2	0.671
Augmented	3	0.
Baseline	0	0.317
Baseline	1	0.426
Baseline	2	0.746
Baseline	3	0.

Table 6.6: The AUC scores from the ROC curves, rounded down on three digits

Classifier type	Average AUC score
Black-box	0.371
Augmented	0.378
Baseline	0.372

Table 6.7: The average AUC scores from the ROC curves, rounded down on three digits

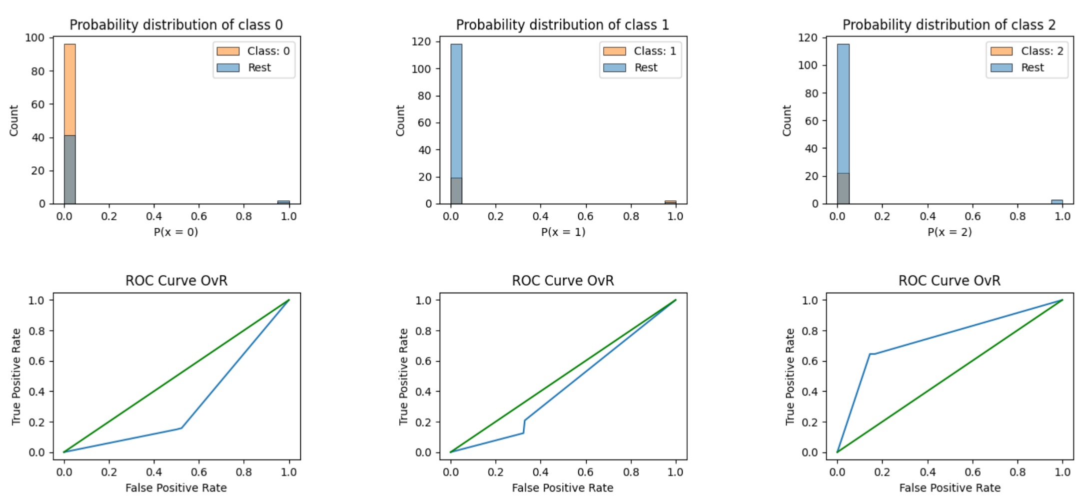

Figure 6.4: The ROC curve and the ROC class separation distribution of the baseline classifier

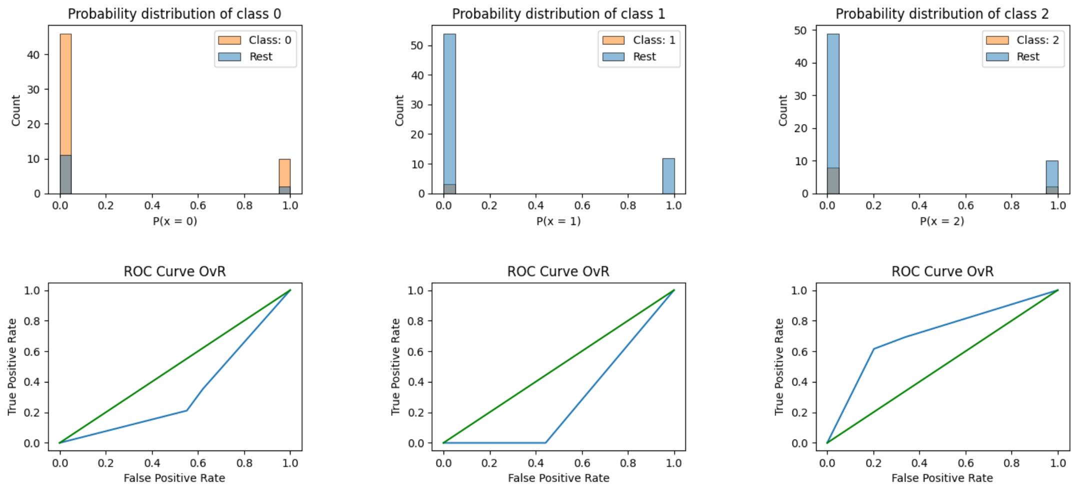

Figure 6.5: The ROC curve and the ROC class separation distribution of the black-box classifier

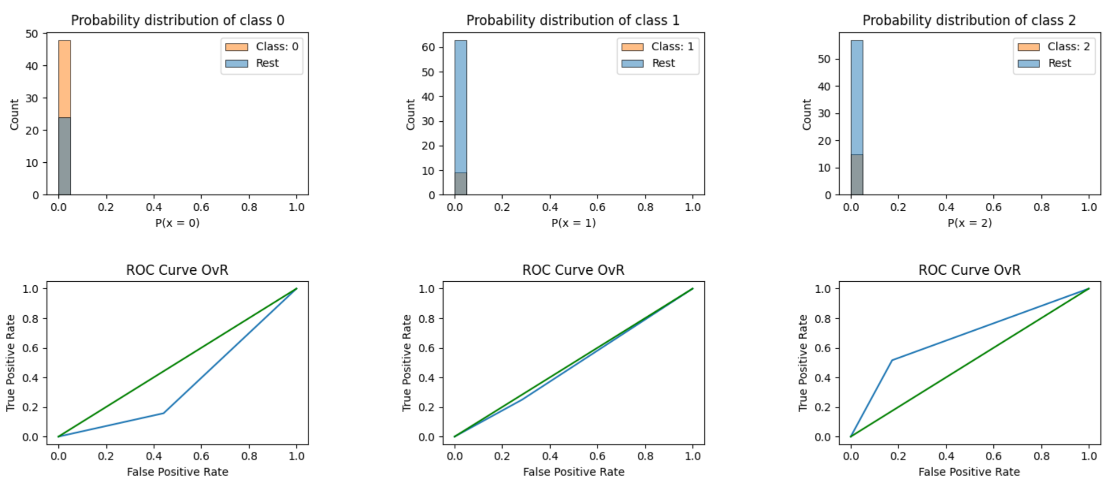

Figure 6.6: The ROC curve and the ROC class separation distribution of the augmented classifier

Chapter 7

Discussion

This research was started to investigate whenever existing Activity Recognition classifiers could be augmented using raw sensor data to detect the inactivity of elderly. The expectation was that by improving the existing activity recognition classifier with additional sensor data, the classifiers would have an increase in performance and accuracy. There was a small increase in performance. However, this research had some limitations, which are discussed in this chapter, as well as the results of this research.

7.1 Creating the dataset

As stated in Chapter 1, the algorithm was planned to be used on elderly people and in order to create the dataset, the data should be gathered from elderly doing the described activities. However, the participants that were used for creating the dataset were under 30 years old instead of the desired elderly. This change in the target group was due to the Covid-19 pandemic and its contact limitations, nullifying the possibility of creating the dataset based on elderly people. Due to this change of the target group, it is likely that the results presented in this research are not as expected, since the data is retrieved from younger individuals that are likely healthier and different in mobility and movement than the intended target group. This research is therefore presented as a proof of concept concluding that, altering existing black-box activity recognition software will reduce a more accurate activity recognition prediction.
Another problem that appeared with this younger participant dataset, is that there are fewer participants, The original plan was to retrieve the data from at least twenty elderly participants from Park Vossenberg. Even though more data is desirable, the current solution was the best available solution at the time of the data gathering. For future studies, it would be interesting to see if more participants and within the desired age group would affect the results of this study.

During the gathering of the data by the participants, it was discovered that the sensor data from the final activity, cycling, was inconsistently recorded. Only the sensor data from five participants was deemed complete. For the other five participants, either one minute or half a minute of cycling sensor data was recorded. A possible cause of the inconsistently recording of the cycling data is that the watch wasn't close enough to the computer when the data was recorded. Even though the data should be recorded regardless of an internet connection, a flaw within the application could be the cause of this. As a result, the five participants that gathered all the sensor data were used as training data, so that the algorithm was trained with the complete part of the dataset. This also means that the test set contains incomplete data. This can affect the results from this research, since there isn't enough cycle data to verify the Tizen labels. This should be verified in a future study.

Finally, an interesting observation is that there are only three labels visible as results even though Table 3.3 refers to the four Tizen labels. This is because the in-vehicle movement label, number three, wasn't detected when creating the dataset. Even letting some participants perform activities that aren't present in Table 3.2, such as sitting in a car and recording their data, just to identify the criteria that the Tizen Label gets assigned, did not result into a recording of the in-vehicle movement label. Since it is unknown on what criteria the labels are generated (due to the inaccessibility of the Tizen classification software), it is difficult to identify which sensor movement values would be sufficient to create this fourth label. There is also the possibility that the Tizen software is incapable of detecting the in-vehicle movement label. For related future work, it could be interesting to use another open source activity recognition platform, to identify if this is a Tizen only topic, or that more activity recognition's platforms are affected by this.

7.2 Tizen platform related issues

The original approach on gathering the activity recognition labels from the Tizen platform was by using the emulator provided by the Tizen platform. The Tizen emulator had the functionality to simulate events to which the to be developed application could respond. The original idea was to insert the Accelerometer sensor data and create an application that would be used on the emulator to retrieve the labels. However, as stated in the Tizen documentation [16], the activity recognition functionality does not work on the emulator due to technical reasons. The aforementioned issue wouldn't have taken place on another mobile platform, such as Android or Apple, due to the technological support that they offer and the extended functionalities of their emulators. For future research, it would be interesting to redo the research using a different software platform, so that the algorithm can also be ran on the watch and show while the user is wearing the smartwatch how active someone is.

7.3 Creating the Algorithm

Creating the interpreted black-box classifier came with several challenges. The initial idea was to create an interpretable algorithm based on the works of Yan et al. [52], which was described in the survey of Guidotti et al [21]. The algorithm that this research used proved that it was possible to mimic the black-box classifier, with the addition that the algorithm was interpretable. Even though the research claimed that the algorithm would work on multi-class label data, due to logistic issues, it was not possible to study this in this study. It could be interesting to look whenever the quality of the mimicked black-box classifier would improve with using the algorithm from the studies of Yan et al [52].

7.4 Results

Three different metrics were used to calculate the performance of the classifiers: the accuracy, the precision and recall scores, and the AUC curve.

Looking at the results from the accuracy metric, the difference between the three tested was expected to be more substantial, since the difference between the black-box classifier and the augmented black-box classifier is just 0.02. The expectation was that there would be a greater difference between the augmented black-box classifier and the black-box classifier and that the augmented classifier would perform better than the baseline classifier.
A notable observation of the accuracy metric reveals that the baseline classifier has a higher accuracy than the other two classifiers. One possible explanation could be that the baseline classifier had earlier access to the sensor data chunks (since the gyroscope sensor is fused with the accelerometer sensor data), but there is no hard proof on that.
Another possibility lies with the imbalance of the dataset. According to Section 6.2.1, there is an imbalance within the dataset, with the stationary movement label having significantly more entries in the dataset than the walking and running labels. Perhaps the baseline classifier is less capable of handling this imbalance than the other classifiers. It could be interesting to run the experiments again with a more balanced dataset to see whenever the baseline would perform better in this scenario. When looking at the confusion matrices in Figures 6.1, 6.2 and 6.3, this data imbalance becomes even more clear, with a lot of labels assigned towards the stationary activity when compared with the black-box and augmented classifier.

When looking at the precision and recall scores of the three classifiers and labels, the imbalanced dataset problem is even more present, since the differences between the labels is more present. The 0 label has higher precision and recall scores in comparison with the other two labels. This was especially the case with the recall scores, where the recall scores of the 0 label was significantly higher than the other two labels. This further indicates that the imbalance of the dataset affects the outcomes of this research.
When comparing the precision and recall scores of the three classifiers with each other, the baseline classifier has higher precision and recall scores when compared with the black-box classifier and the augmented classifier. This is an interesting observation, since the augmented classifier should be performing better than the other two classifiers. The differences between the scores of the labels is skewered, since the 1 label lacks a precision and recall scores from the black-box classifier and the augmented classifier. It is unknown what causes this, a possibility could lie with the data imbalance, since there are significantly more data samples of the 0 label in comparison with the 1 and 2 labels. The confusion matrices that were made show this even more.

The baseline classifier is outperforming the black-box and augmented classifiers could mean that the sensor features from the dataset was insufficient for the training of the classifiers. The baseline classifier outperforming the other classifiers could also means that the sensor features lack any predictive power.
Another observation is that the baseline classifier is performing as well as the other classifiers when classifying 0 label entries. It is difficult to determine why the black-box classifier is performing on-par with the other classifiers. It was expected that the augmented classifier would perform better than the black-box classifier. A possibility could lie within the fact that

retraining the black-box classifier did not had the desired effect of improving the black-box classifier. This would be a case where retraining the black-box classifier is not effective.

The goal of the ROC curve is to have an line which represents the TPR = FPR, and another line that represents the ROC curve itself. The nine ROC curves that were generated from the reveal that the three classes from the three classifiers have a poor to medium fit. This means that they are in most cases unable to correctly guess the activity class based on the input data. On average has the augmented classifier a better fit than the black-box classifier and the baseline classifier, with two out of the three classifiers having an ROC curve that is either close or further away from the TPR=FPP line in comparison with the others. The probability distributions reveal that the data imbalance discovered with the confusion causes that the distributions will not help much with determining their capability of predicting the classes correctly.
The last criteria was the AUC scores from the three classifiers. As earlier stated, the closer the AUC score is to 1.0, the higher the chance that the predictions are correct. Two tables with the AUC score from each class from the classifier and one table containing the average AUC scores were created to determine their scores. The classifiers are acting poorly, with the lowest score being 0.317 and the highest score 0.746. Looking at their average scores, the augmented classifier is barely outperforming the other two classifiers. Another observation is that the average AUC scores are also very low. This is caused by the fact that the label 3 did not appear when recording the data, as stated in Chapter 3. This is the main cause for the lower average AUC scores.
The table further indicates that the running Tizen class, 2, is significantly outperforming the other two classes. It is unclear what would cause that the running label is outperforming the other classes. One possibility is that, since there is a significant amount of 0 labels due to the data imbalance, it is more difficult to determine the correct label. However, this would mean that the walking label would perform equally with the running label, which is not the case.

Chapter 8

Conclusion

This research aimed to augment black-box classifiers by adding additional sensor data, gathered from participants doing activities for people with disabilities, mostly elderly in nursing homes. This research was able to create a mimicked machine learning model from the black-box classifier using a custom created sensor dataset with labels that were generated from the black-box classifier. The performance was evaluated using the accuracy, precision and recall values, and the Area Under the Curve metrics.

The evaluation metrics showed that the accuracy of the Baseline classifier was 69 percent, of the mimicked black-box algorithm 62 percent, and finally the accuracy of the augmented black-box classifier was 64 percent, which makes the baseline classifier the more accurate of the two classifiers. The other evaluation metrics showed that the baseline classifier was performing better on the precision and recall scores, but the augmented classifier was performing better on the AUC values. Therefore, the augmented classifier is performing better in some cases than the baseline and black-box classifiers. The hypotheses that augmenting an black-box classifier would yield an performance increase is proven by the experiments. That said, this improvement is minimal. Depending on the use, it could be beneficial to improve the black-box classifier, but overall, it is better to retrain the classifier, unless the classifier is not accessible.

8.1 Future work

The most desired next step is to substitute our dataset with a dataset in which the participants are actually elderly, as that was the required study subject. As described in Chapter 7, this unfortunately was not possible for this research due to the Covid-19 Pandemic. A comparison study can be made to see if there is a change in the results and whenever this research still upholds when set in the desired scenario. This dataset should also be more balanced to prevent imbalanced classification.
Another next step could be to use activity recognition software from the Android Library or Apple IOS. This activity recognition is not only supported better, but it is also more advanced. In addition to this, these classifiers can be used with the Android / IOS emulators, therefore the experiments can be operated on the devices, which could be interesting to look at. It could be interesting to see whenever using the different software libraries will positively affect these classifiers. Again, a comparison study can be made to see if there are any significant changes in the results.

Chapter 9

Appendix

A Sensor feature extraction values

Mean values

Participant ID	Activity Number	xvaluesMean	yvaluesMean	zvaluesMean
0750bbbb	1	-0.042627998125197	-2.26310984575936	4.83912432765037
	2	-0.737039215736884	-2.06899621044639	4.8108001944894
	3	-1.29614298191878	-0.977762062903543	5.13004996910752
	4	3.9452673316994	-2.00676425282127	4.7717379190766
	5	5.14430058530165	-1.83293295658299	0.467217940882856
	6	5.42436908195634	-1.25998251391558	-0.078836221167949
	7	5.31255028479992	-4.6613559485546	-0.961656609547452
	8	15.1246957491017	-11.2476822854561	-5.54028770889346
	9	4.49823052528428	-2.43755381235264	2.71960612169966
174c4240	1	-0.02923284611524	-2.36677358396179	4.8125494000639
	2	-0.399523748614275	-3.92356743145901	3.34055777803843
	3	-2.0120593194478	-1.55923551569571	4.57343700189282
	4	-3.38884564947362	-6.27199853853999	-0.648020840665495
	5	5.48507011655563	-1.58048106326797	0.479708764703376
	6	5.16752767858617	-1.61012229811292	1.25353525513185
	7	0.099859887685326	-6.25785175028331	-1.03097721314786
	8	1.20599298415394	-7.7745443026323	-0.580422025066069
	9	3.40813621371701	-2.81317417692735	3.07277353667245
2838o0a97	1	2.45750025903865	-2.22640899842353	3.9789420576032
	2	-0.591995824240793	-2.29810829752707	4.73186703171129
	3	-2.41305205116817	-0.403465113433696	4.67044193971283
	4	-1.74021821531009	-4.36294547529118	2.97974048962826
	5	4.5564802426269	-3.03362036202666	0.229995689070236
	6	4.85185079571553	-1.92318138722075	0.736701451791461
	7	1.10059724776003	-6.33769443567432	-1.68440100440639
	8	10.6665311973459	-3.9391542420454	-1.80318228927523
	9	2.22831952191294	-3.19495113274899	3.5731106756902
4c73e7a5	1	-1.02335654397436	-4.52423516355586	2.17433141580317
	2	2.65857877028636	-3.0452873901819	2.99870904331216
	3	-0.538819330356034	0.265077321283394	5.47380398657749
	4	2.4218443169414	-3.57969609349642	4.07125586357623
	5	5.2113331539163	-1.75975684588041	0.844576137862659
	6	5.41315640764812	-1.01675911868102	0.928564824638084
	7	6.08365070642896	-5.14433245134426	-0.258749905918926
	8	11.9712632875894	-6.78900999867743	-1.832772445114
66eeee48	1	0.815168982751337	-3.67275524580592	3.63376459420871
	2	1.40540544575577	-0.896042716818586	5.23076798264897
	3	-2.7856547498885	-2.09193190519386	3.75548667262942
	4	1.82429061538403	-4.56171924440753	3.49834684669192
	5	5.40938387736018	-1.27033216200717	1.38272736257471
	6	5.40794045543231	-1.01824311363416	-0.427627314646491
	7	4.31024076865413	-6.23142310225867	-3.45214101575843
	8	8.77625167766239	-9.40707923465593	-8.33218575601766
	9	1.43125214108214	-1.86926580136965	4.59818921527414
7ace4d32	1	0.569600685229325	-4.7560573889117	1.69495819342391
	2	0.413980114306861	-1.58494934741118	5.26072163532333
	3	-3.74790950116741	-1.27156177420792	3.15288143625379
	4	-0.102279171863271	-5.50032327399327	3.23660978819472
	5	4.89493164573557	-1.98302535441294	1.1605516621835
	6	4.54894494470561	-2.62463430120037	0.760284293110357
	7	1.25873485335597	-7.55674077616105	-0.792070591694864
	8	4.92489565384637	-7.31301362956612	1.70002686645011

Page 1

Mean values

	9	2.80145641844463	-1.4588183164064	4.47206674996801
85baca89	1	0.690751742774888	-1.14999413751774	5.24312038179576
	2	-1.56339958463977	-0.930130359929882	5.09547983998579
	3	-3.55504545223975	-0.772433445975985	3.59677296899159
	4	-2.19843828010895	-1.71245040688034	5.81687101611745
	5	5.05376184429206	-1.8829984194792	0.763794149683702
	6	5.63970616320476	-0.72661548324851	0.081776756424192
	7	-0.931008020487716	-6.14928587480407	1.52943674979334
	8	3.0337791698307	-9.81049986762535	2.28438787781573
86cf2ef9	1	0.419472842339677	0.273256743175732	5.52288049425839
	2	-0.239811994953618	0.093134906032263	5.63031275652182
	3	0.000860574965696	-4.13445692946007	3.31974193678219
	4	-0.796647863124151	-5.47910505507308	0.307410329002376
	5	5.48061437381347	-1.12012225386728	0.254791680261742
	6	5.54782307766733	-0.444333243404154	0.880973425459637
	7	-2.38373240561885	-5.1480732899459	0.917825266868829
	8	0.300827097461125	-6.94831002852038	0.943647642289662
b99c6603	1	-0.130300844823959	-3.95207274065531	3.60715040556618
	2	-0.822381058390176	-2.77229245084923	4.55085982818216
	3	-2.0536217211697	-1.88400512503638	4.42265089005085
	4	-0.135615325344573	-5.6184975181865	1.06326253189911
	5	5.65494909450609	-0.873030647057355	0.176761674298216
	6	5.43823833086691	-0.887057711405153	0.811657523301803
	7	0.647581460080458	-6.41296399648627	-0.982803943397422
	8	4.53227602961342	-7.6886449245915	-1.18147824512062
	9	3.86668104304533	-2.38703012181361	2.96952401708445
e3b8753	1	1.51341571812666	-2.91347969873898	3.8153310370138
	2	-1.18009151750674	-1.87831124755351	4.84359474540678
	3	-2.51014123624551	-1.4090622550447	4.35308243169973
	4	-0.139593729504602	-5.7064957052801	-0.563342682120705
	5	5.14804452632571	-1.78854936992817	1.55744821868225
	6	5.22066024970499	-1.14531032866886	1.37624746066319
	7	3.06314624484534	-5.55002740307855	0.117161644790051
	8	8.10715765549594	-7.16990173228205	-1.18781096031158
	9	2.90392385304343	-3.4946577965376	2.70996306311974

Median values

Participant ID	Activity Number	xvaluesMedian	yvaluesMedian	zvaluesMedian
0750bbbb	1	0.012141156113297	-2.56399236932141	4.73801957623597
	2	-0.706423131884393	-1.89969892485596	4.95003318503597
	3	-1.18653450824635	-0.900581130865939	5.21534382069895
	4	4.74260159794508	-2.50154489939918	4.60889761302326
	5	5.14867672792549	-1.68140278915154	0.318245264105592
	6	5.30905790952342	-1.22571853584448	-0.073946845398024
	7	5.36023545776956	-4.55279704831759	-1.04239165920603
	8	11.2149413121841	-9.88587982349947	-3.59916489971869
	9	4.35945320566072	-2.2507370718717	2.70676839289458
174c4240	1	-0.019806337729701	-2.43410682040099	4.7947769357087
	2	-0.383893242092706	-4.0020749563971	3.42656719053088
	3	-1.9958958436045	-1.53412824673541	4.59780822172661
	4	-3.25735853654169	-6.32139292118804	-1.07324533864901
	5	5.29787364137469	-1.38498115598714	0.480507056182603
	6	4.9790066462373	-1.47383860547117	1.15176125340769
	7	0.183213071258774	-5.62028733477632	-0.847454706691691
	8	1.61859750600612	-7.41883198704567	-0.53797302075172
	9	3.36553440035671	-2.69853155480684	3.09461904980521
2838o0a97	1	2.47606573476451	-2.23030598443451	4.01306701321884
	2	-0.613591694152355	-2.18451851273297	4.8588435098791
	3	-2.56866412265387	-0.223624971763756	4.71067071545175
	4	-1.521013937163	-4.48147536947427	2.90537612915495
	5	4.40940607340797	-2.92925602074189	0.190199669706823
	6	5.01513285245783	-1.82319666755579	0.686172649412503
	7	2.01240928714298	-5.35989296870456	-1.52901441102098
	8	10.3381358788998	-5.23582141262488	-1.59527862612218
	9	2.18309586640061	-3.1506767793529	3.6338545791657
4c73e7a5	1	-0.713345085331302	-4.47828513137282	2.12891466514519
	2	2.59270967551464	-2.94898372819761	3.10694807135068
	3	-0.561875251564285	0.345092824245804	5.51721896847003
	4	2.35985111197266	-3.57068063735222	4.2535761939195
	5	5.19089904289307	-1.71868630418789	0.811673543228791
	6	5.26924935296551	-0.965304088923046	0.90021860978021
	7	6.30913698829186	-5.62540337549241	-0.315911884870935
	8	13.5182876269417	-9.56995600471749	-1.92305752230828
66eeee48	1	0.824910971641809	-3.69749168818416	3.64291998501692
	2	1.31696204968696	-0.868786153376265	5.17270851864675
	3	-2.79636652335657	-2.12133075964227	3.73602895223739
	4	1.89103864449385	-4.48993869393095	3.54594798318133
	5	5.20593370964516	-1.13718468527662	1.18417010657737
	6	5.38232441091038	-0.828286587660366	-0.260886786924412
	7	5.31110535595787	-6.65148867590313	-3.06206880546457
	8	7.85533763022975	-9.51278176484697	-7.5318619822577
	9	1.17044980859429	-1.68564737955044	4.94668515497991
7ace4d32	1	0.133029799102219	-5.1633923870116	1.53442315810871
	2	0.401292469944204	-1.47708975289483	5.2112270925357
	3	-3.71982359667865	-1.41588545176847	3.30214327115786
	4	0.07567031138022	-5.96704284402485	3.13429266189163
	5	4.80938120999505	-1.97014456614484	1.06502095251088
	6	4.5350133416071	-2.56754209608881	0.713852668234906
	7	2.08048891985965	-6.74024103562203	-0.510986099438345
	8	5.75440516620587	-6.9194534995402	1.76783618818697

Median values

	9	2.67193117867838	-1.31057158199907	4.43334948309122
85baca89	1	0.691562157506176	-1.15024323007082	5.25501576705548
	2	-1.53523263493413	-0.729100165934042	5.0729807240153
	3	-3.70144692609034	-1.31722893184953	3.39624794423043
	4	-2.61342478986406	-1.801364622216	6.01354832803129
	5	4.91012739506521	-1.85966790929594	0.718756369787575
	6	5.56458713215194	-0.675442798242912	0.073449376118791
	7	-1.49553344125034	-5.55517658793178	1.42004620399447
	8	3.02183015708382	-8.52803090313805	1.8184200781895
86cf2ef9	1	0.398054003675854	0.256424952177478	5.55671905135799
	2	-0.237636703692242	0.10887528403077	5.58538462396802
	3	-0.002674904257248	-4.12695036541422	3.35303549567784
	4	-0.689681999308047	-5.73586942944296	-0.109233037009347
	5	5.46076743312258	-1.08683243047388	0.23663966956595
	6	5.43725734524143	-0.428027549055507	0.836404314575664
	7	-2.3143681806708	-4.54735000509739	0.604810818281135
	8	0.362161266542458	-5.75694520871799	0.653841221327696
b99c6603	1	-0.128839556066145	-3.97471793841127	3.60984718191272
	2	-0.831265928651291	-2.82560772492351	4.43694671280898
	3	-2.04748118419347	-1.88377278846914	4.42426994170818
	4	-0.016027940400362	-5.61636290818085	0.895107821849208
	5	5.53854045201446	-0.787279960813316	0.199819787725654
	6	5.41153078392866	-0.747540679601156	0.729288725545348
	7	1.071430528934	-5.42636842016879	-0.73310925776081
	8	4.76476255591692	-6.60831995123012	-1.08617612042292
	9	3.88489805687535	-2.19760139547352	2.96125329323167
e3b8753	1	1.10654631065387	-2.86243816172447	3.75327687706484
	2	-1.21024900655984	-1.75776126040074	4.93285375533506
	3	-2.5138279841565	-1.37389264886296	4.33936169413098
	4	0.070402633549183	-5.73317447869761	-0.612878769377579
	5	5.03006941957947	-1.70023463477607	1.45520801876784
	6	5.09620463820914	-1.09780270899073	1.32959202033849
	7	3.7987844923532	-3.52334494720496	0.010414670924821
	8	8.18196547594557	-6.70509319919057	-0.918252067923934
	9	2.62630309315008	-3.43458908439107	2.74273604479464

Std values

Participant ID	Activity Number	xvaluesStd	yvaluesStd	zvaluesStd
0750bbbb	1	0.296198849595283	0.601289273569521	0.44207851443943
	2	0.519171657327089	0.79753243060052	0.777093859688131
	3	0.377494760410498	0.298074004651893	0.278868440923708
	4	2.02923103163649	1.80704600340287	0.752928329861487
	5	0.84096867319536	1.21451233551584	0.891487800517956
	6	0.607226988327302	0.466202635486013	0.309018259861479
	7	4.65344569460528	4.68276524020086	1.79686506475971
	8	14.8311571100872	13.5348967851036	7.59169894109403
	9	2.74615184328344	1.98381560699748	1.79822420096909
174c4240	1	0.093803429047539	0.243417812638015	0.308135837308404
	2	0.708841920846175	1.00486791673494	1.3078902831281
	3	0.14015012356288	0.238894599542925	0.144769610035427
	4	2.10561196370909	2.13655397262206	2.4258806308995
	5	1.65318859469659	1.66646523808268	1.4228477106846
	6	0.861632459730855	1.04432053833353	0.605032763473836
	7	2.48129839118934	2.69633081948807	1.05739616468231
	8	5.36823745712923	4.38857504521651	1.71330342606853
	9	1.29316099549092	1.16005822906005	1.16432808102036
2838o0a97	1	0.165165226270573	0.196686614857396	0.247565370697426
	2	0.565163649422972	0.921961104568965	1.03140726207797
	3	0.715418551187952	0.71549994111356	0.428216298895117
	4	1.17267128496528	1.04095323862966	1.00535374905514
	5	1.42971360773508	1.40309723528856	0.672413842073007
	6	1.86883404163318	0.981188137936513	0.519605359149803
	7	2.9620179030997	4.08217783284125	1.52803903050916
	8	6.8411635771267	7.68964868740657	2.24282311923652
	9	1.66905481235719	1.31863856077867	1.32708731297333
4c73e7a5	1	0.635178699589239	0.663546825435663	1.05046920770038
	2	0.943846227635936	1.00952796186226	1.25857531161872
	3	0.286952593807378	0.647496145469932	0.342666898835658
	4	1.88933362066696	1.54307839445212	1.0278781341319
	5	1.00874733657772	0.925276708525675	0.858247329349973
	6	0.908984936492467	0.638588228942754	0.469835256137996
	7	6.23453024965063	5.84709320874851	1.6415120753071
	8	10.2052790406463	10.2982189142886	3.44631393192747
66eeee48	1	0.105112653911624	0.226573734467477	0.216474116825626
	2	1.10981779179444	0.348533565690631	0.595148291609681
	3	0.139019897560194	0.169502332656455	0.150105398295089
	4	1.3592163280286	1.97720140919308	1.29010146223496
	5	1.96430327124579	1.91757047891391	1.31589591848272
	6	1.34589708316118	1.08716060369909	1.02952068998023
	7	5.2294809938256	6.26267683344543	2.87425579705154
	8	10.7101479342314	10.5541774768362	5.65719148361273
	9	1.07117703381177	1.15278479843217	1.28377925045622
7ace4d32	1	0.733079632096008	1.20023218071308	1.29084617501688
	2	1.25211521259968	0.864088795040953	0.717478582438874
	3	0.393681458731474	0.755344847966538	0.951529175356913
	4	1.49858610801215	1.7961376564445	1.2048470923468
	5	1.49716721397617	1.26517806088228	1.06461529843534
	6	1.35414038184807	1.56464105922722	0.85224332207023
	7	5.04510576779281	5.30000066577967	2.04884723117592
	8	6.4393713342377	5.9635258754172	2.94996651893822

	9	1.31837914968122	1.37312554289749	0.958259231816537
85baca89	1	0.072570509596911	0.141451113056396	0.312117397726948
	2	0.532642580768152	0.682187267626607	0.708933351349716
	3	0.678076458314921	0.831358577622758	0.602131527057459
	4	2.19785110852455	1.88786090221899	0.720529648694006
	5	1.23011989994649	1.00694830926452	0.639912733846879
	6	1.1737278907799	0.487727807400663	0.321474616650415
	7	3.02495747715248	3.90302171329961	1.59306372677736
	8	8.89653633029015	7.05600447182237	4.41666651911625
86cf2ef9	1	0.225143744053245	0.279282818353038	0.354817384519244
	2	0.283937081519197	0.276659866621551	0.616235826729795
	3	0.091143279728322	0.283692379196579	0.311246472209293
	4	1.24522013418522	1.07242449543936	2.16975497822571
	5	0.654925873229639	0.791899587958794	0.387741110740461
	6	0.978528016508949	0.331901580879187	0.374004081682327
	7	1.86679454253423	2.71651655997482	1.33127294279505
	8	3.52317070004586	4.34285093787906	2.4667758984315
b99c6603	1	0.021702136109175	0.242285834463891	0.22418896099469
	2	0.651873622162124	0.686305665198104	0.988599892322985
	3	0.067038027685762	0.086281644532756	0.051239054775393
	4	0.655080979247234	0.493273482744326	0.758647330284459
	5	1.1184322815978	1.76969526784725	1.1652216722325
	6	0.895269048647399	1.03261747870083	0.585874328093225
	7	2.79287561273804	4.97235458371793	1.57744995508847
	8	8.46597134508075	7.68351597589617	2.5549000190729
	9	1.03519131876097	1.37643967628926	1.12450993470713
e3b8753	1	1.06698667313005	0.815709595924395	0.478338739727195
	2	0.732915861222831	0.906468784936593	0.860338373934103
	3	0.112625306082644	0.150000187995562	0.104170401562372
	4	0.659812804257342	0.293232129593897	0.526560091884673
	5	2.03519129083287	1.46984123344682	1.46759507231197
	6	0.802182381250151	0.638812796703002	0.515744801133952
	7	4.03798059924778	5.89275440023524	1.36374008410094
	8	8.29306235851576	8.13264462872957	2.85777536130026
	9	1.96527694284966	1.94947268516927	1.57322340562363

Min values

Participant ID	Activity Number	xvaluesMin	yvaluesMin	zvaluesMin
0750bbbb	1	-2.51622805238804	-3.85806254956104	0
	2	-2.43703619969471	-5.79892740254678	1.12626454635626
	3	-2.95483078368669	-2.582026201823	3.47271960518768
	4	-2.16515783063032	-6.15260354960414	3.49486453577959
	5	-1.19758678753567	-13.2454200720728	-5.87761079407599
	6	4.09757366412379	-2.77561827649699	-1.04669018994022
	7	-4.56448834379464	-16.7280974787268	-8.08293357496403
	8	-11.7737424969413	-45.8744541334823	-35.3867271820938
	9	-3.12324932368911	-9.71613462691611	-5.42655647691599
174c4240	1	-1.08787172745673	-3.38120691673351	0
	2	-2.44459518723647	-9.98945645410761	0.049271058543403
	3	-3.70077193085278	-2.75604399726855	3.93250544231906
	4	-10.8778329777945	-11.6558844563184	-8.24078981130801
	5	-7.20376331915996	-18.4161555705967	-13.3950709506106
	6	3.659358112998	-4.46097354747347	-0.977375112812088
	7	-6.01202996349314	-14.9356931666567	-5.11148415315888
	8	-11.2339081438955	-24.252920802868	-6.39758111051764
	9	-4.1222618678892	-11.40078762776	-1.86981174898827
2838o0a97	1	0	-2.64635352237519	0
	2	-2.51045504087406	-5.53733491531691	1.62340001968635
	3	-4.13450159749129	-6.03796131970219	1.06743326499605
	4	-5.82113671403485	-6.70296710403629	0.556243460355386
	5	-1.57804860063935	-16.2912877585883	-4.46293940144719
	6	-6.00712075171956	-7.02186038984952	-1.29511450472994
	7	-7.0055730466061	-17.4223719491547	-9.00255254536778
	8	-6.76494515682531	-20.7130853298236	-12.9407842662089
	9	-8.95816601236168	-9.3535790877185	-5.26176406888887
4c73e7a5	1	-3.16446569340802	-6.82530627577232	-2.24329940629772
	2	-0.501052041963402	-6.4543498902394	-1.01349961010061
	3	-2.68625967724989	-5.36895720788368	1.9579048670856
	4	-3.54584410128872	-8.2526560289057	1.10790515384377
	5	1.37786035383308	-8.29053737523146	-5.09227886723064
	6	3.17791732049569	-3.37722061752639	-0.130425489906436
	7	-5.3744424372926	-20.4290588148323	-6.7069326635432
	8	-10.1214452937932	-27.5214661297941	-13.9908669555632
66eeee48	1	0	-3.89737532299263	0
	2	-0.854886599274873	-3.4418602630934	3.6851640330289
	3	-3.20421231900539	-2.28163940704258	3.62842938206981
	4	-2.75406481324099	-9.16065459281625	0.11758028081078
	5	-1.09026446645624	-26.9790732533162	-5.57614480285714
	6	0.200427244594872	-5.46300641566029	-6.51300035670177
	7	-8.55400729813584	-22.9103708778546	-25.0373703651195
	8	-22.5502463160798	-43.7002600013321	-29.623076432695
	9	-0.019754334362073	-6.39351504989398	-4.00576711140737
7ace4d32	1	-1.56367877619047	-6.74326475986557	0
	2	-3.36682284872378	-5.59420610788486	1.8937550568621
	3	-6.37719373887609	-4.75001643890839	-4.46073605669876
	4	-5.13847184671776	-9.89147109496575	0.224359949529496
	5	0.341302127665151	-11.2463543184354	-4.01495806799935
	6	0.816092004917743	-8.13426666511568	-9.88198203624881
	7	-10.1988166759613	-24.941185344016	-14.6985130628872
	8	-10.5458707619186	-31.6435083374903	-8.6930487826419

Min values

	9	-1.72564980186224	-17.508204055945	0.220309911004181
85baca89	1	0	-1.41669789999377	0
	2	-4.05335945875492	-3.62482570903994	2.32748156347053
	3	-5.8137720249453	-2.6928454575199	0.910845369579494
	4	-6.09685872337284	-8.76538669031809	3.36557126235175
	5	0.879294723170423	-9.56946345743544	-4.00590290330394
	6	1.66511325773139	-4.15125397493651	-1.24964815236508
	7	-7.50025563423756	-16.4237416685499	-2.3982796388079
	8	-29.1442654636103	-35.7235842456696	-12.585166885321
86cf2ef9	1	-1.1491586198193	-2.07797192558874	0
	2	-1.28861665170609	-0.92009963408232	3.87900682876793
	3	-0.365644354084385	-5.5908197205289	0.924731729376008
	4	-4.57738695219512	-7.97999979494895	-4.32449703307489
	5	-1.08980554850033	-7.63571343515128	-3.82395339130374
	6	3.50464390946191	-2.74278695837372	-0.110353737579249
	7	-8.86564728047645	-13.2392733135391	-2.5469326336783
	8	-11.029846425631	-20.6857429844027	-9.5631511525517
b99c6603	1	-0.254274974790184	-4.88371533690421	0
	2	-2.86308838872076	-5.39076709843908	2.0510201774856
	3	-2.24963906744025	-3.01915206417507	4.05057214204696
	4	-4.38647935693193	-7.62079563370073	-0.236738337125398
	5	0.944526837372288	-25.3236588647727	-14.8875592746252
	6	-2.24758483070707	-15.0389926524593	-1.69793991282011
	7	-5.93359803420225	-21.8575796718461	-10.3013034151481
	8	-12.6298877879681	-34.3727658397121	-11.6934239116265
	9	-1.82246340626649	-18.8180884905594	-1.22531126503871
e3b8753	1	-1.91459707118026	-8.10828483249022	0
	2	-3.42271282601769	-4.62734498085821	2.33655031035178
	3	-2.7754968125622	-1.95109055341113	4.10576937377879
	4	-2.41445705730635	-6.47968003266251	-1.54977044006665
	5	-1.33019641011572	-14.6333810214542	-4.86392322871842
	6	2.48644714357988	-7.65004685988544	-1.77884934060425
	7	-6.01144841263774	-30.7332469160969	-5.13273179713552
	8	-9.89772158582248	-30.4081337297772	-10.3195103211528
	9	-3.59474240638127	-9.07182860573083	-1.97556828279707

Max values

Participant ID	Activity Number	xvaluesMax	yvaluesMax	zvaluesMax
0750bbbb	1	0.545690775318153	0	5.61331445763688
	2	1.19427834608741	-0.535085043398914	6.9494913510344
	3	1.24800657973496	-0.04303970244409	6.63099988426922
	4	6.82851492094134	3.26478325032161	7.83161227398064
	5	10.600624726603	3.27114921751136	9.30123583125446
	6	8.26102809485283	-0.100245341567996	1.43014350852269
	7	17.8966591139825	5.84799578358455	3.38690517714789
	8	48.1235398641644	29.4093996243604	10.3993067321693
	9	11.7604230135157	3.75574093927403	9.90384065541196
174c4240	1	0.079100501508534	0	6.33633399847901
	2	2.13763226293569	0.078100649446661	8.75206545616103
	3	-0.240877721712094	0.211925903069312	6.23029643237335
	4	0.673488442888709	-0.895650070194415	7.0850035151755
	5	19.7230676507092	8.77073066622588	11.9431561977685
	6	8.09018521583805	0.710737352654823	3.98993856805586
	7	8.04577298393921	-1.28415852799517	1.01986599664421
	8	13.6211464863077	0.527263211216182	7.46152999051823
	9	12.0191152123179	0.920690452503782	8.0289688934767
2838o0a97	1	2.78582522735825	0	4.26277551329842
	2	2.19464512375675	-0.191185904170728	16.3286610002272
	3	3.67719470716471	1.48524270018308	7.15440283642614
	4	0.613419258877213	-1.33829484156513	5.84368853779729
	5	13.7173545585	5.34491049938646	5.2924527138982
	6	7.98102927140615	0.252218617215744	3.18718703016368
	7	12.5369280774062	0.220846851781858	2.31623330295623
	8	34.7500119558156	17.1086966750023	4.65642050490023
	9	25.5816788136293	9.09680597043832	9.96899824631146
4c73e7a5	1	0	0	5.55235337474155
	2	7.41011139422095	-0.062526427013635	8.86516714503338
	3	1.75702988196536	4.53871735497122	7.40342993901269
	4	7.80049823480715	0.322281677705076	6.78842756066134
	5	11.0544424776212	2.82401276221355	6.09406411058655
	6	7.75924708637652	0.247524173401421	2.94494905861502
	7	28.8979149487653	11.8836805935107	8.76220232376929
	8	38.21362723468	24.2531227488293	9.97728365337673
66eeee48	1	1.02238051480565	0	3.84991252713499
	2	4.13763568457483	-0.163194131525779	6.98876267401988
	3	-0.341242505156659	-0.194750540383201	5.83648946030654
	4	5.2817578687557	0.549530039770958	7.09514359636842
	5	22.694302752278	7.26003841956567	16.1946628134518
	6	8.6112004092459	1.53200207305179	3.71575740886511
	7	30.2306691817461	9.10471478410933	3.0363364619426
	8	43.4083541511308	18.5198965670987	2.74917597372379
	9	12.3721627687724	7.22599089680346	6.80132329171751
7ace4d32	1	3.10307336219394	0	5.91335283670423
	2	5.01674668295983	2.26929707679381	9.08776434925662
	3	0.200372158986758	4.53227013461006	5.64488816295369
	4	3.51554674644556	0.089832376383775	6.46060337865816
	5	24.4441084125737	2.81991131771367	10.7778575974491
	6	8.90012272559259	1.12306342134834	5.39076393039517
	7	12.6681181195846	2.55751614849697	5.20068735558808
	8	18.8967903752662	8.28387115507721	10.5591385912335

Max values

	9	9.23026753801685	3.21605591425773	9.34930528276169
85baca89	1	1.75435031306566	0	6.3176473683835
	2	0.852668367618307	0.197782945354071	7.97611246240781
	3	-0.863708215943518	1.97274000499363	5.69106361695672
	4	2.40734181131496	2.66957769227249	7.28707947697081
	5	11.5104632489145	6.38535926333134	3.81916502323733
	6	8.83090703702806	1.14061930229793	1.79171493430159
	7	6.30252207981835	2.81755545529715	8.32007909696076
	8	30.0894991359634	4.22252108218699	17.7578487868272
86cf2ef9	1	2.33816822171922	1.58369693104214	8.07701155245508
	2	0.701761698888657	1.39194994729696	8.00734981510363
	3	1.60924527917069	0.524071934381325	4.80185505208238
	4	2.99198239707104	-0.978152473548557	6.18103227235974
	5	9.30122092994005	3.60748951570914	3.02417611159613
	6	8.94848989501566	0.576109869957104	2.47065823382829
	7	4.59599408488431	-0.773904011041817	8.19407394869027
	8	10.8116579052875	0.246310474815516	12.8846947586212
b99c6603	1	0	0	3.83360305448869
	2	2.32514259646643	-0.115497033259471	9.10322636398316
	3	-1.15301511908507	-1.61490934873531	5.0011655753139
	4	2.11875205184173	-2.33832068158903	4.31224812881398
	5	15.5611686141858	12.2943325757472	13.0990030767593
	6	8.27171265861313	1.289638667017	4.79121677022199
	7	10.7269144911807	1.96498613013354	3.91182733671757
	8	35.0672594010935	10.9216675884369	9.46794736403229
	9	11.9527389163325	1.32036877291877	7.58523443394441
e3b8753	1	4.00553360252401	0.378484929267122	6.60539659381886
	2	2.07566769109742	0.300122613064671	8.69912553925463
	3	-1.26504831588946	-1.06851059568364	5.16304297501003
	4	0.976238922804033	-2.38588613881857	3.94079724376586
	5	18.5639294060875	6.0322568216892	14.7170932029988
	6	9.05798837955971	0.707288810426827	8.06181310075312
	7	18.9229264112361	2.81075348775159	8.24288083282643
	8	34.163909497013	8.78623918831695	8.36790184476578
	9	8.80993163698143	5.52274407577822	6.84185064165196

B Data gathering protocol

Thesis - Data Collection Protocol

f.ph.vandermeulen

September 2020

1 Introduction

The research that needs this data is about detecting inactivity from elderly with dementia by augmenting the Tizen black-box classifiers on a Samsung Galaxy fitband Pro 2, which results into the calculation of an inactivity score, so that the physical activity of an elderly on a day to day base can be detected.
This document contains which tools are needed for the data collection, which materials will be used for the data collection, what kind of activities are performed and finally how the data will look like when it is saved. Before that however, it is important to discuss which covid-19 related measurements are taken in order to make sure that the data gathering process is according to the rules set by the Dutch government.

2 Covid related measures

Due to the pandemic, there are several changes in terms of how the data will be gathered:

1. The first and primary change is that it would be irresponsible to gather the data from the desired target group, eg, elderly in the age range between 75-85. Since the health risks are to significant for this age group, a younger group that is more resilient against the covid crisis will be selected and used as participants.
2. In addition to selecting the participants, a screening questionnaire from the National Institute for Public Health and the Environment will be sent around 24h before data collection and just before data collection to check for COVID 19 symptoms. Participants who indicate to have symptoms will be asked to refrain from taking part in the data collection.
3. Additional hygiene related measurement will be taken: all materials will be cleaned with sterile (are they sterile? Alcohol wipes? Disinfectant wipes?), participants will clean their hands before data collection and the room will be ventilated at regular intervals. Furthermore we will adhere to the universities rules on how many people are allowed per room to maintain a safe distance of 1,5 mt. Participants will be registered with the administration of the building for contact tracing.

3 Data collection tools

The data will be collected with the Samsung Galaxy Pro 2 Fitness bands. A special application is developed that reads the raw sensor data from the fitness bands and stores the data on the fitness bands. Once the fitness bands makes a connection with a host PC, the data gets transferred and stored as a CSV file.
In addition of gathering the data from the wristbands, video recordings are made of the participants when they are performing their activities. These recordings will be used to determine when an activity has started, and when an activity has ended.

4 Preparation for the data gathering

In order to make the gathering of the data a success, several materials are required in order to be able to perform the activities needed for the research. In addition to the materials that are required to perform the activities, hygiene related materials are also included according to the safety guidelines[1]. The materials can be found in table 1.

MaterialID	Item
0	Samsung Galaxy Fitness Pro 2 bands
1	Bike
2	Walking stick
3	Chair
4	Table
5	Grocery bag filled with grocery related items
6	Cutting board
6a	Knife
6b	Cucumbers
7	Tape that can be placed on the ground
8	A yoga mat
9	A stopwatch

Table 1: The materials that are required for the activities

The age group of the participants will be 20-30. Even though an older age group is desired, it is not possible to gather data from them due to the pandemic. Each of the participants will be given a randomly generated identification number. The participants will be also asked to hand in their email address at least 48 hours before the start of the test. This email address will be used for the administration of the building, so that they know who are entering the building.

5 The performed activities

In order to gather the data needed for the research, multiple activities that are related to the four classifiers will be performed by the participants. The classifiers are ambiguously described, for example, does a stationary activity mean that the user is standing still, laying down or sitting in a chair? These activities mimic the activities that elderly could be performing. Table 2 contains an ID number of an activity, the name of the activity, an description of the activity, the duration of the activity and the ID's of the materials that are required. The stopwatch is used to make sure that the activity takes the amount of time that was determined in table 2.

ActivityID	Name	Description	Duration	MaterialID
1	Sitting	The participant is sitting on a chair	60s	3
2	Laying	The participant is laying down on a yoga mat	60s	8
3	Cutting	The participant is cutting a cucumber using a knife	60s	4, 6, 6a, 6b
4	Yoga	The participant is performing the standing spinal twist yoga exercise (see fig 1)	60s	none
5	Walk with stick	The participant is walking with a walking stick, using that for support	120s	2, 7

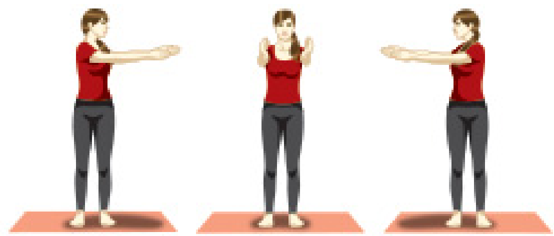

Figure 1: The yoga pose that the participants will perform

6	Walk groceries	The participant is walking with heavy grocery bags on both hands	120s	5, 7
7	Sprint	The participant is sprinting from one spot to the other	60s	7
8	Jogging	The participant is performing jogging movements	60s	7
9	Cycling	The participant is cycling on a bike in the second gear	120s	1

Table 2: An overview of the activities that will be used to gather the data from

Between each of the described activities, the participant is requested to clean their hands using hand gel, in order to conform with the covid-19 regulations that are determined by the government. In addition, the materials are cleaned between every time a participant is performing the activity.

6 Data collection

Before the start of the gathering of the data, each participant is asked to fill in an information consent form. This form contains the information that is written in the data collection protocol, while also going more in-depth on how the data is stored, what data is required from the user etc. Finally, an example of how the data will be saved can be found in table 3.

UID	**time**	**accX**	**accY**	**accZ**	**gyrX**	**gyrY**	**gyrZ**	**labels**
2342	1449044888500	0.0123	0.1432	0.2347	0.7262	0.1842	0.8100	Yoga

Table 3: An example of how an data entry will look like. UID refers to the user identification, acc to the accelerometer sensor and gyr to the Gyroscope sensor.

References

[1] Dutch national institute for Public health and environment. Health checklist, 2020.

C Data gathering consent

Information sheet and informed consent form

Data collection – Benchmark Data for Inactivity Detection

1. General information about the research and the collected research data

- The goal of this research is to detect inactivity of elderly with dementia using a wearable device. The data we are collecting today will serve as a training and benchmark data set to develop better computational methods to detect inactivity.
- Participation in the data collection is voluntary. Participants can stop their participation at any time by indicating this to the data collection assistant and returning the sensors used for the data collection
- Two types of data will be collected during the data collection:
  - Raw sensor data (accelerometer and gyroscope) from participants while doing a set of physical activities which are related to activities that elderly could perform.
  - Observations of the performed activities (labels of the data set)
- The data will be made available to the research community at large and along with this a meta data set, which includes a description of the set-up of the data collection.
- Please feel free to contact Daniela Gawehns, gawehnsd@liacs.leidenuniv.nl with any questions you might have regarding the data collection. If you have questions or complaints about the researcher, you can turn to the Scientific Director of LIACS, Prof. Aske Plaat via the secretary of the computer science institute.

2. Personal information

- The sensor data are not considered personal information, but might be identified as personal information when combined with knowledge about when individuals participated in the research.
- All collected data is stored during the data collection and post-processing on laptops used by research staff and. The sensor data will subsequently (about one month after data collection) be made available online in a research repository. The timestamps will have randomized start times such that identifying individuals from the shared data is almost impossible.

3. General data Protection Regulation

- Personal Data will be processed based on the consent of the participants.
- Questions regarding data privacy can be addressed at: privacy@BB.leidenuniv.nl
- Participants have the right to request access to their personal data. They also have the right to have their personal data erased from the database of this research. This can be requested till one month after the data collection. The final data set is not considered personal data, once the data collection is over and the timestamps have randomized start times, individuals cannot be identified by the researchers anymore and individual's data can hence also not be removed anymore.
- Data is retained indefinite to be reused for future research.

Leiden, 30. 10. 2020

Informed Consent for

"Data Collection – Augmenting black-box classifiers"

Please tick the appropriate boxes

	Yes	**No**
1. Taking part in the study		
I have read and understood the study information dated 30/10/2020, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.		
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and that I can withdraw from the study at any time, without having to give a reason.		
2. Use of information in the study		
I understand that information I provide will be used to develop and improve on computational methods to detect physical activity of geriatric patient populations.		
I understand that personal information collected about me that can identify me, such as my name, will not be shared beyond the study team.		
3. Future use and reuse of the information by others		
I give permission for the **sensor data** that I provide to be deposited in a research repository (server based in the EU) so it can be used for future research and learning.		

________________________ ________________________ ____________

Name of participant[IN CAPITALS] Signature Date

[In case the information sheet is read out loud:] I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

________________________ ________________________ ____________

Name of researcher[IN CAPITALS] Signature Date

Bibliography

[1] Philip Adler, Casey Falk, Sorelle A Friedler, Tionney Nix, Gabriel Rybeck, Carlos Scheidegger, Brandon Smith, and Suresh Venkatasubramanian. Auditing black-box models for indirect influence. *Knowledge and Information Systems*, 54(1):95–122, 2018.

[2] Nawzat Sadiq Ahmed and Mohammed Hikmat Sadiq. Clarify of the random forest algorithm in an educational field. In *2018 International Conference on Advanced Science and Engineering (ICOASE)*, pages 179–184. IEEE, 2018.

[3] Neamah Al-Naffakh, Nathan Clarke, Paul Dowland, and Fudong Li. Activity recognition using wearable computing. In *2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)*, pages 189–195. IEEE, 2016.

[4] Rio Ikhsan Alfian, Alfian Ma'arif, and Sunardi Sunardi. Noise reduction in the accelerometer and gyroscope sensor with the kalman filter algorithm. *Journal of Robotics and Control (JRC)*, 2(3):180–189, 2021.

[5] Hester Anderiesen, Erik JA Scherder, Richard HM Goossens, and Marieke H Sonneveld. A systematic review–physical activity in dementia: the influence of the nursing home environment. *Applied ergonomics*, 45(6):1678–1686, 2014.

[6] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *Esann*, volume 3, page 3, 2013.

[7] Anindya Das Antar, Masud Ahmed, and Md Atiqur Rahman Ahad. Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: a review. In *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pages 134–139. IEEE, 2019.

[8] Simon Arvidsson, Marcus Gullstrand, Beril Sirmacek, and Maria Riveiro. Sensor fusion and convolutional neural networks for indoor occupancy prediction using multiple low-cost low-resolution heat sensor data. *Sensors*, 21(4):1036, 2021.

[9] Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016.

[10] Daniel Berrar. Cross-validation. *Encyclopedia of bioinformatics and computational biology*, 1:542–545, 2019.

[11] Nadia Burkart and Marco F Huber. A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70:245–317, 2021.

[12] Pierluigi Casale, Oriol Pujol, and Petia Radeva. Human activity recognition from accelerometer data using a wearable device. In *Iberian conference on pattern recognition and image analysis*, pages 289–296. Springer, 2011.

[13] Cagatay Catal, Selin Tufekci, Elif Pirmit, and Guner Kocabag. On the use of ensemble of classifiers for accelerometer-based activity recognition. *Applied Soft Computing*, 37:1018–1022, 2015.

[14] Yufei Chen and Chao Shen. Performance analysis of smartphone-sensor behavior for human activity recognition. *Ieee Access*, 5:3095–3110, 2017.

[15] Seungeun Chung, Jiyoun Lim, Kyoung Ju Noh, Gague Kim, and Hyuntae Jeong. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors*, 19(7):1716, 2019.

[16] Han Daeh and Technical Steering Group. Tizen, 2019.

[17] Peter A Flach and Meelis Kull. Precision-recall-gain curves: Pr analysis done right. In *NIPS*, volume 15, 2015.

[18] Robert D Gibbons, Giles Hooker, Matthew D Finkelman, David J Weiss, Paul A Pilkonis, Ellen Frank, Tara Moore, and David J Kupfer. The cad-mdd: A computerized adaptive diagnostic screening tool for depression. *The Journal of clinical psychiatry*, 74(7):669, 2013.

[19] Margherita Grandini, Enrico Bagli, and Giorgio Visani. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*, 2020.

[20] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Dino Pedreschi, Franco Turini, and Fosca Giannotti. Local rule-based explanations of black box decision systems. *arXiv preprint arXiv:1805.10820*, 2018.

[21] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42, 2018.

[22] Mohammad Hossin and MN Sulaiman. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2):1, 2015.

[23] Yen-Chang Hsu, Zhaoyang Lv, Joel Schlosser, Phillip Odom, and Zsolt Kira. Multi-class classification without multi-class labels. *arXiv preprint arXiv:1901.00544*, 2019.

[24] Weiwei Hu and Ying Tan. Generating adversarial malware examples for black-box attacks based on gan. *arXiv preprint arXiv:1702.05983*, 2017.

[25] Ajay Kulkarni, Deri Chong, and Feras A Batarseh. Foundations of data imbalance and solutions for a data democracy. In *data democracy*, pages 83–106. Elsevier, 2020.

[26] Oscar D Lara and Miguel A Labrador. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, 15(3):1192–1209, 2012.

[27] Erin LeDell, Maya Petersen, and Mark van der Laan. Computationally efficient confidence intervals for cross-validated area under the roc curve estimates. *Electronic journal of statistics*, 9(1):1583, 2015.

[28] Yong Luo, Liancheng Yin, Wenchao Bai, and Keming Mao. An appraisal of incremental learning methods. *Entropy*, 22(11):1190, 2020.

[29] Hadi Tabatabaee Malazi and Mohammad Davari. Combining emerging patterns with random forest for complex activity recognition in smart homes. *Applied Intelligence*, 48(2):315–330, 2018.

[30] Andrea Mannini, Mary Rosenberger, William L Haskell, Angelo M Sabatini, and Stephen S Intille. Activity recognition in youth using single accelerometer placed at wrist or ankle. *Medicine and science in sports and exercise*, 49(4):801, 2017.

[31] Hosung Park, Gwonsang Ryu, and Daeseon Choi. Partial retraining substitute model for query-limited black-box attacks. *Applied Sciences*, 10(20):7168, 2020.

[32] Dino Pedreschi, Fosca Giannotti, Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, and Franco Turini. Meaningful explanations of black box ai decision systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9780–9784, 2019.

[33] Jing Peng, Ping Zhang, Lanxiang Zheng, and Jia Tan. Uav positioning based on multi-sensor fusion. *IEEE Access*, 8:34455–34467, 2020.

[34] Susanna Pirttikangas, Kaori Fujinami, and Tatsuo Nakajima. Feature selection and activity recognition from wearable sensors. In *International symposium on ubiquitious computing systems*, pages 516–527. Springer, 2006.

[35] Attila Reiss and Didier Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th International Symposium on Wearable Computers*, pages 108–109. IEEE, 2012.

[36] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1135–1144, 2016.

[37] Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Kilian Förster, Gerhard Tröster, Paul Lukowicz, David Bannach, Gerald Pirkl, Alois Ferscha, et al. Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh international conference on networked sensing systems (INSS)*, pages 233–240. IEEE, 2010.

[38] Charlene V San Buenaventura and Nestor Michael C Tiglao. Basic human activity recognition based on sensor fusion in smartphones. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pages 1182–1185. IEEE, 2017.

[39] Valery Serov. *Fourier series, Fourier transform and their applications to mathematical physics*, volume 197. Springer, 2017.

[40] Muhammad Shoaib, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. A survey of online activity recognition using mobile phones. *Sensors*, 15(1):2059–2085, 2015.

[41] Muhammad Shoaib, Hans Scholten, and Paul JM Havinga. Towards physical activity recognition using smartphone sensors. In *2013 IEEE 10th international conference on ubiquitous intelligence and computing and 2013 IEEE 10th international conference on autonomic and trusted computing*, pages 80–87. IEEE, 2013.

[42] Krishna Kant Singh, Mohamed Elhoseny, Akansha Singh, and Ahmed A Elngar. *Machine Learning and the Internet of Medical Things in Healthcare*. Academic Press, 2021.

[43] Sameer Singh, Marco Tulio Ribeiro, and Carlos Guestrin. Programs as black-box explanations. *arXiv preprint arXiv:1611.07579*, 2016.

[44] Vijay Srinivasan and Thomas Phan. An accurate two-tier classifier for efficient duty-cycling of smartphone activity recognition systems. In *Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones*, pages 1–5, 2012.

[45] Paolo Tamagnini, Josua Krause, Aritra Dasgupta, and Enrico Bertini. Interpreting black-box classifiers using instance-level visual explanations. In *Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics*, pages 1–6, 2017.

[46] James Theiler. Matched-pair machine learning. *Technometrics*, 55(4):536–547, 2013.

[47] Bipan Tudu, Arun Jana, Animesh Metla, Devdulal Ghosh, Nabarun Bhattacharyya, and Rajib Bandyopadhyay. Electronic nose for black tea quality evaluation by an incremental rbf network. *Sensors and Actuators B: Chemical*, 138(1):90–95, 2009.

[48] Yonatan Vaizman, Katherine Ellis, Gert Lanckriet, and Nadir Weibel. Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[49] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11, 2019.

[50] Jindong Wang, Yiqiang Chen, Lisha Hu, Xiaohui Peng, and S Yu Philip. Stratified transfer learning for cross-domain activity recognition. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10. IEEE, 2018.

[51] Jianyu Xiao, Linlin Chen, Haipeng Chen, and Xuemin Hong. Baseline model training in sensor-based human activity recognition: An incremental learning approach. *IEEE Access*, 9:70261–70272, 2021.

[52] Ke Yan and David Zhang. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, 212:353–363, 2015.