Recognition and Classification of Aggressive Motion using Smartwatches

Franck Tchuente

A thesis submitted to the Faculty of Engineering

in partial fulfillment of the requirements for the degree of

MASTER OF APPLIED SCIENCE

in Mechanical Engineering

Ottawa-Carleton Institute for Mechanical and Aerospace Engineering

University of Ottawa

Abstract

Aggressive motion can occur in clinical and elderly care settings with people suffering from dementia, mental disorders, or other conditions that affect memory. Since identifying the nature of the event can be difficult with people who have memory and communication issues, other methods to identify and record aggressive motion would be useful for care providers to reduce re-occurrences of this activity. A wearable technology approach for human activity recognition was explored in this thesis to detect aggressive movements. This approach aims to provide a means to identify the person that initiated aggressive motion and to categorize the aggressive action.

The main objective of this thesis was to determine the effectiveness of smartwatch accelerometer and gyroscope sensor data for classifying aggressive and non-aggressive activities. 30 able-bodied participants donned two Microsoft Bands 2 smartwatches and performed an activity circuit of similar aggressive and non-aggressive movements. Statistical and physical features were extracted from the smartwatch sensors signals, and subsequently used by multiple classifiers to determine on a machine learning platform six performance metrics (accuracy, sensitivity, specificity, precision, F-score, Matthews correlation coefficient).

This thesis demonstrated: 1) the best features for a binary classification; 2) the best and most practical machine learning classifier and feature selector model; 3) the evaluation metrics differences between unilateral smartwatch and bilateral smartwatches; 4) the most suitable machine learning algorithm for a multinomial classification.

Table of Contents

List of Figures

List of Tables

Abbreviations and Definitions

Acc	Acceleration
BW	Both Wrists
C	Correlation
CM-FS	Classification Model and Feature Selector
Diff	Maximum difference, Var
DT	Decision Tree
DW	Dominant Wrist
FN	False Negative
FP	False Positive
FS	F-Score
Gyr	Gyroscope
HAR	Human Activity Recognition
IG	Infogain
kNN	k-Nearest Neighbours
MCC	Matthews correlation coefficient
Med	Median
MP	Multilayer Perceptron Neural Network
NB	Naïve Bayes
NDW	Non Dominant Wrist
Pcc	Pairwise Correlation Coefficient
Prec	Precision

ReF	ReliefF
RF	Random Forests
RGB	Red, Green, Blue
Sens	Sensitivity
Skew	Skewness
SMA	Signal magnitude area
Spec	Specificity
Std	Standard deviation
SVM	Support Vector Machines
TN	True Negative
TOHRC	The Ottawa Hospital Rehabilitation Centre
TP	True Positive

Acknowledgments

I would like to express my gratitude towards my supervisors Edward Lemaire and Natalie Baddour. Not only were they my supervisors, but also my mentors who taught me the soft and hard skills (critical thinking, time management) I needed to become a better researcher and a multivalent engineer. They guided me from the beginning and were always professional and friendly. They made me stretch by applying positive pressure on me.

I will also thank my colleagues at the University of Ottawa and The Ottawa Hospital Rehabilitation Center who helped me through their advice, debates, questions, opinions, and company. Finally, I would like to thank my girlfriend who was a great emotional support, and my family members (Marie-Claire, Bernard, Vanina, and Jordan) who assisted me morally and financially throughout this very interesting adventure.

Thank you to all the volunteers who participated in the experiments.

1. Introduction

Dementia is a mental disorder that affects more than 35 million people in the world, and is expected to double over the next 20 years [1]. In long-term residential care facilities, more than forty percent of the elderly (older than 65) are affected by this disorder. People suffering from dementia can sometimes become quickly agitated, verbally and even physically aggressive [2]. Kicking, hitting, or pushing are some aggressive motions frequently observed and listed in conventional scales such as the Cohen-Mansfield Agitation Inventory [3].

Direct observation of these activities is currently the main method used by caregivers to determine the events that took place [4]. However, such a method is very subjective, prone to diagnosis errors, and might increase the caregiver time load. The ability to quickly recognize aggressive situations could result in prompt intervention and a better understanding of the person's activity. Tailored care could then be adopted to better solve this problem and help both the caregivers and the patients.

Technology, by the means of Human Activity Recognition (HAR), could be used to address the problem of identification of aggressive motion. The use of smart technologies is increasingly applied in health-related applications and will be explored in this thesis.

1.1. Rationale

In the literature, aggressive motions and actions are predominantly monitored and analysed with intelligent vision systems. For instance, computer vision methods would detect when two people are fighting and resources can be put in place to de-escalate such situations. High surveillance areas such as prisons, airports, and healthcare settings might benefit from this technology to increase safety and security.

Nevertheless, computer vision is financially and computationally expensive. For example, for healthcare applications, cameras installation would be required in several locations of the care facility, and the corresponding software for analysis would also be required. Additionally, surveillance prompts the issue of privacy, where it could be inappropriate to record all patient activities in a hospital setting.

A unique way of determining aggressive movements could involve wearable devices. The need exists for a wearable smartwatch approach that is non-obtrusive, easy to implement, and can be worn by people who initiate aggressive motion (dementia, etc.) to identify aggressive movements or physical escalating situations in an institutional setting. Thereby, smartwatches would enhance understanding of events not observed by staff and possibly provide an alarm to alert staff when an event has started.

Utilizing smartwatches could enable people at the hospital to quickly intervene in some situations and help diagnose the occurrence of mental outbreaks if an alarm-based event system is adopted. Appropriate measures could be taken to help the elderly community, since care providers could better determine the best methods to reduce activity re-occurrence.

Smartwatch studies typically do not address aggressive motion but mostly record daily activities that include walking, running, and exercising. In this thesis, we propose and evaluate combining machine learning classifiers and smartwatches to determine aggressive activity events.

1.2. Objectives

This thesis examined the use of smartwatch sensors for aggressive motion detection. The aim of the research was to determine if smartwatch technologies could correctly classify aggressive and non-aggressive movements. This aim was divided into four thesis objectives:

1. Determine the best set of smartwatch sensor features to distinguish aggressive from non-aggressive movements.

   Research questions

   a) How will the selected features perform?
   b) Will different selection methods output similar features?

2. Determine the best machine learning classifier and feature selection model

   Research question

   a) What is the best classifier for recognizing aggressive from non-aggressive motion?

3. Determine the differences between bilateral smartwatches and unilateral smartwatches.

Research questions

a) Do bilateral watches yield better evaluation metrics than unilateral watches?
b) Does the dominant wrist configuration gives better results than non-dominant?

4. Determine the best machine-learning classifier for a multinomial aggressive classification

Research questions

a) Does multinomial classification perform worse than binary classification?
b) Which movements will be confused as false positives and negatives?

1.3. Thesis contributions

The thesis produced positive contributions to the HAR scientific community, including evidence that smartwatches can be effective for detecting aggressive movements. The primary contributions of the thesis are:

- An updated version of the TOHRC Data Logger Android app was built to connect the smart watches to an Android phone via Bluetooth. It is now possible to use smartwatches to record upper-body motion.
- It was shown that only one smartwatch on the non-dominant hand can be effective for a binary aggressive activity classification
- The best features were identified and coupled with appropriate machine learning classifiers to categorize aggressive and non-aggressive motion.
- This novel research classified aggressive movements using exclusively smartwatch accelerometer and gyroscope sensors, making this research transferrable to other wrist-worn wearable devices.

1.4 Thesis outline

The thesis follows a manuscript format and is composed of seven chapters.

Chapter 2 provides a literature review; including, equipment used for HAR, analysis methods, and areas of application.

Chapter 3 is a journal manuscript that addresses objective 1 and outputs the best set of features used for aggressive movement classification.

Chapter 4 is a journal manuscript and addresses objective 2, resulting in the best classification method and feature selection method used for binary aggressive/non-aggressive classification.

Chapter 5 comprises a conference paper that compares and evaluates bilateral watches to unilateral (dominant and non-dominant) performance.

Chapter 6 contains a journal manuscript that presents a multinomial aggressive classification evaluation of ten activities.

Chapter 7 summarizes the thesis, addresses the contributions, and proposed future work for HAR using smartwatches.

2. Literature Review

This chapter provides details on the concept of Human Activity Recognition (HAR), the evolution of HAR systems, and the main steps used in an activity monitoring analysis, ranging from data collection to the performance evaluation.

Human Activity Recognition involves visual or sensory pattern recognition that leads to interpreting and labeling a specific activity. The aim is to discover human physical activity patterns by analyzing movement data captured by multiple systems [5]. HAR is based on concepts such as mobility monitoring, movement tracking, and computer vision that determine the type of activity. As a growing field over the past decades, HAR helped to improve lives in different areas, including entertainment and healthcare.

2.1 HAR areas of application

In sports, several attempts have been made to identify and deconstruct distinct phases of a play. As an illustration, Anik [6] identified patterns to recognize badminton movements: serve, smash, return, or backhand. Data were collected with accelerometer and gyroscope sensors attached to the badminton bat. Additionally, a baseball movement analysis by Lapinski analyzed forces and torques from the players [7].

To improve participant golfing skills, Ghasemsadeh [8] developed a training system with motion sensors. The system determined the golfer’s skill level and expertise via signal progressing algorithms to calculate angular rotations of a person’s wrist during a golf swing. Feedback on movement quality was subsequently provided.

Entertainment systems have incorporated HAR to improve the gaming experience. The gaming industry, with technology improvements, has introduced visual and inertial sensors (thoroughly explained in Section 2.1.2) in console accessories. Nintendo Wii controllers incorporate accelerometers to interact with the system through gesture recognition [9]. The Xbox 360 Kinect also records and detects user motion to interact directly with the game, emulating an activity whether it is a specific dance move, a fight simulation, or a car driving experience.

Video surveillance can detect abnormal activity in some environments (police stations, force protection, antiterrorism settings, etc.). HAR technology would identify and recognize suspicious activities or unusual events by detecting anomalies [10], [11]. Understanding such unconventional situations might help to de-escalate potential dangerous situations.

HAR has been increasingly applied to the healthcare sector. Mobility monitoring, for example, can provide an objective portrait of patient's mobility outside a clinical setting. Capela [12] used smartphone technologies to identify walking differences between able-bodied, elderly, and people with disabilities. Activity recognition can also be a risk assessment tool to evaluate fall detection for elderly people and aid people with limited mobility [13].

HAR is applicable to other fields that include augmented and virtual reality, rehabilitation programs, motion disorder identification, and smart homes [14]. Two main types of HAR systems can be used for activity recognition monitoring: external sensors fixed on a known location or inertial sensors directly located on a person [15].

2.2 HAR external systems

External HAR systems have a longer history [16], with the sensor equipment attached to a fixed and predetermined location [17]. Activity recognition is via the interaction between people and sensors. The equipment can be vision-based (infrared cameras, depth cameras, RGB cameras, 3D motion capture sensors, Microsoft Kinect, Vicon cameras; Figure 2.1) or non-visual sensors (location sensors, light sensors, force plates, pressure mats) [15].

Activity recognition with optical sensors, also known as cameras, requires video and image software analysis, and segmentation. Video-based systems have recognized activities such as sign language and gestures [18]. Sardsehmukh used a 3D HAR video dataset captured with a Kinect sensor. The Kinect sensor provided RGB and depth information [19]. The dataset was useful to help distinguish activities that included handshake, punch, kick, push, lift bag, and throw object.

In clinical gait analysis, motion capture systems combined with force plates are used to extract parameters such as the cadence, step time, or velocity to assess gait impairment and walking phases [20]. These external systems can monitor entire human body movement. However, they tend to be computationally expensive (heavy software analysis) and financially costly (material installation and costs). Another disadvantage is video-based HAR system privacy issues [15], since people can be visually recorded and identified. Furthermore, during the analysis, parameters such as lighting conditions, body size, position, and observer angle [16] can affect system performance.

Figure 2.1: Vicon motion analysis camera (left) [21] and Microsoft Kinect sensor (right)

2.3 Inertial or wearable sensors

Wearable devices are smart electronic devices equipped with microcontrollers that are worn on the body as accessories or implants. Wearable sensors often use inertial measurement units or radio frequency identification tags to gather a person's behavioral information. The sensors could be stand-alone (accelerometer, GPS, personal thermometer) or integrated into devices such as cell phones, personal digital assistants, or laptops. Smart glasses, smartphones, smart bands, smart clothing, and data gloves (Figure 2.2) are some new and trendy wearable devices [18].

Figure 2.2: Smart devices for activity monitoring: activity tracker smart band [22] (left), smart cloth [23] (middle), and smart glove [24] (right).

Wearable sensors used to be cumbersome and obtrusive (Figure 2.3), wrapping around the body to determine the type of movement a person was performing. Technology advancements have had a massive impact on the size of inertial-based sensors, transforming them from bulky devices that would hinder user motion to integrated chips in wireless micro-devices.

Inertia sensors could be mounted on several body locations, depending on the type of study [25]. Common locations include the shoulder, neck, arm, forearm, thighs, and legs. Inertial systems could be located on one part of the body for a local analysis (e.g., band on a leg, smartphone in a pocket, smartwatch on a wrist). Alternatively, a more encompassing study could capture the entire body motion [26].

A balanced number of sensors is important for HAR. Even though more data sources can enrich the knowledge about the activity [27], too many sensors might be invasive, expensive, uncomfortable, and not suitable for activity recognition.

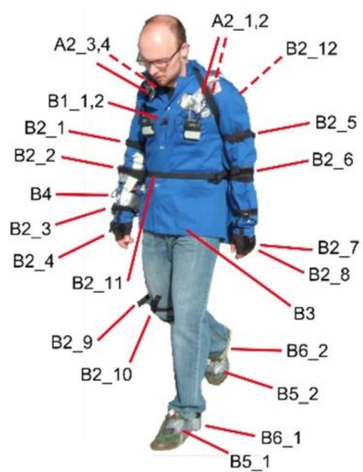


Figure 2.3: An experiment with sensors located all over the body [28]

2.4 Smartphones

Smartphones have become the most essential gadget in people's lives. People carry their phones all the time and bring them everywhere; therefore, they are great tools for HAR. Smartphones are unobtrusive and contain multiple integrated sensors such as accelerometer, gyroscope, GPS, altimeter, magnetic field sensor, or pedometer, which in the past were used individually. Smartphones also have excellent software development platforms (Android, iOS) for real-time HAR monitoring.

Several HAR studies used smartphones as monitoring devices to identify a variety of dynamic activities (Table 2.1). Akhavian [29] attached a smartphone on the upper arm to collect data and recognize construction worker's cutting, transporting, and installing lumber. The activities were sawing, loading, pushing, unloading, returning, hammering and turning the wrench.

Chen [5] evaluated five actions (descending stairs, ascending stairs, walking, jogging, jumping) on a MEXZU MX3 running on Android 4.4.x. The phone was positioned at multiple locations: right upper arm, right hand, right jacket pocket, right trousers pocket, and waist.

Table 2.1: HAR studies based on Smartphones

Author	**Smartphone**	**Placement**	**Activities**	**Accuracy**
Capela, et al. [12]	Blackberry Z10	Pelvis	Stand, sit, lie, walk, walk, stairs	Not specified
Chao, et al. [30]	MEXZU MX3	Waist	Downstairs, upstairs, walking, running, jumping	75.82%-90.65%
Akhavia, et al. [29]	Not Specified	Upper arm	Sawing, loading, pushing, unloading, returning, hammering and turning the wrench	78.57%-96.64%
Lee, et al. [31]	Huawei Nexus 6P	Held in hand, in pocket, in bag/knapsack	Run, walk, stay still	92.71%
Dernback, et al. [32]	Samsung Captivate	User choice	Simple: biking, stairs, driving, lying, running, sitting, standing walking Complex: cleaning, cooking, medication, sweeping, washing hands, watering plants	90% for simple activities and 50% for complex activities

The optimal location for smartphone activity recognition is ambiguous. Several positions have been adopted with preferences around the body center of gravity. Studies have claimed that the ideal position is the back within a holster [33]. Chest and stomach have also been considered.

Despite the smartphone capacity to integrate a multitude of sensors in one location, the main issues are with the phone positioning and phone orientation [34]. Acceleration and gyroscope data can be influenced by the phone's orientation and location, which vary during activity, leading to misclassifications and increased error rates. Some solutions to this problem include transformation matrices and using position/orientation-independent features [34]. Positioning and orientation might not be an issue for some wearables with consistent locations: smartwatches and smart bands.

2.5 Smartwatches

Smartbands and smartwatches are more natural to wear compared to smartphones, which require appropriate positions and an extra piece of equipment such as a holster. Wrist devices are more popular over the past few years, with the boom of the fitness and well-being industry. These devices use micro-sensors to track inertial data and provide daily physiological information for health monitoring (number of steps, calories, heart rate signal) [17].

Several studies have been conducted to classify activities; such as, writing, eating, sitting, and jumping using variants of smartwatches, with accuracies ranging from 80% to 90% (Table 2.2). Applications include a wrist-worn Actigraph to evaluate activity recognition and fall detection [35], a smartwatch system to identify gestures associated with writing the alphabet (accuracies between 94 and 99%) [36], and a Sony SWR50 smartwatch system that can detect stereotyped movements in children with a development disability [37]. Stereotyped movements consisted of clenching a fist, waving a hand, swinging an arm, raising an arm, lowering an arm, and throwing. Early detection of these movements could help to provide timely medical treatment to the children.

Combining external and inertial sensors can provide better classification results and improve HAR metrics. For example, Xi Liu combined accelerometers and RGB-D sensors to identify twenty movements [38]. Nurwanto used accelerometer sensors data from a smartphone attached to the upper arm and a smartwatch on the wrist of the same limb to differentiate light sport exercises such as push up, sit up, and squat jump with accuracies from 77% to 97% [39].

Table 2.2: Studies using smartwatches for activity recognition

Sources	Smartwatch types	Activities
Ardüser [36]	LG Watch R	Recognizing letters from the alphabet
Lee and Song [31]	SONY SWR50	Developmental disability activities: clenching a fist, waving a hand, swinging an arm, raising an arm, lowering an arm, throwing
Weiss [40]	LG G smartwatch Samsung Galaxy S4	Dribbling; catch; typing; handwriting; eat pasta, soup, sandwich, chips; drink
Mortazavi, et al. [41]	Samsung Galaxy Gear	Sit, stand, lie, transitions

2.6 Aggressive motion

Aggressive motion can be characterized by a quick, sudden, and high-intensity movement done by one individual when a situation physically escalates. Identifying these movements is important for injury prevention. Studies to determine people's aggressive or abnormal activity have typically been conducted using external sensors (video surveillance) or image datasets.

Ouanane recognized aggressive human activity using two computer vision methods, bag of features and skeleton graph [42]. The experiments were performed on an action dataset and delivered a recognition rate of 96% between six actions: boxing, hand clapping, hand waving, jogging, running, and walking. Koh [43] proposed a method for detecting driving aggressiveness using a Galaxy Note 2 smartphone lateral acceleration. Aggressive driving consisted of iterated high-intensity u-turns and smooth driving was conducted on a smooth rectangular course. More recently, a Microsoft Kinect and a wearable accelerometer sensor were used to classify aggressive and agitated activities: hitting, pushing, throwing, tearing, kicking, and wandering [4].

2.7 Steps for effective HAR monitoring

HAR is a complex task that requires different steps to make sure the activities are properly classified. The main phases are data collection, pre-processing, feature selection, and classification.

2.7.1 Data collection

To effectively collect smartwatch data, three units can be used: smartwatch, smartphone, computer. The smartwatch can be paired to a smartphone wirelessly (e.g., Bluetooth) and, by means of an application, information can be sent to a storage unit for further analysis. To recognize hand-based activities, Weiss used a custom designed app for a smartwatch (LG G) and smartphone (Samsung Galaxy S4). The app collected accelerometer and gyroscope data from the phone/watch and sent the data via email to a server [40] for further processing.

Smartwatches usually have low memory storage, and low capacity processing unit integrated, hence the need of a smartphone. However, some self-sufficient sophisticated smartwatch systems have been developed: the smartwatch features integrated chips and a processing unit that categorizes directly the activity and displays it on a screen. Mortazavi developed a tracking system to identify three posture states (sitting, standing and lying) using a Samsung Galaxy Gear smartwatch. An application was developed directly on the watch, which has an internal memory and a GUI providing directly the posture state to the user [41].

2.7.2 Pre-processing

Raw data may contain background noise such as outliers, errors, missing values or discrepancies. Pre-processing reduces the noise and diminishes the error rate after data collection. Raw signals can be pre-processed using methods such as filtering, Wavelet/Fourier transformation, or smoothing approximation [30]. Removing the timestamp, the orientation or transitions could also be considered as a preprocessing action [44].

Time series data are continuous. For easier activity recognition, cutting down the continuous signal into segments facilitates feature extraction. The process of dividing the time series data into a series of discrete segments is called segmentation. Segmentation can be divided into activity-based, event-based, and sliding window.

The activity-defined windowing procedure partitions the signal by activity changes [45]. Initial and end points of each activity are found during preprocessing. To identify the transition points, frequency changes of a signal can be tracked using wavelet decomposition or participants can provide feedback at the end of each activity by standing still for several seconds.

For activities conducted in a certain order or consisting of sporadic actions, specific events can be used for segmentation. For example, in gait analysis, heel strike and toe-off events are discrete occurrences for stride signal partitioning [46].

The sliding window technique consists of splitting the signal into windows of equal length with no gap between the windows. The windows may be overlapping or non-overlapping, depending on the analysis [46]. Adopting this approach is simpler to implement and does not require any pre-processing, contrary to the previous times-series segmentation methods. The sliding window approach is the most widely utilized for segmentation [45]. The sliding window is characterized by parameters such as the length and the step. The length could be fixed or dynamic [47]. The step determines if the window is overlapping. Sliding windows work well with periodic activities and are used frequently for medical applications.

2.7.3 Dimensionality and feature selection

Selecting high-quality features for pattern recognition system input will lead to better classification accuracy and decreased error rates. In HAR systems, features are selected based on a combination of intuition and empirical experience [48]. Zhang used statistical features such as the mean, variance, and mean crossing rate for handwriting recognition. Intuition based on the physical world and real-life situations is portrayed through the movement's physical features. For instance, eigenvalues of the dominant direction reflect the large vertical acceleration component that is expected when a person jumps [49]. Appendix A shows features that have been used in recent years for HAR.

Increasing the number of features may improve activity class recognition and reduce the probability of error. However, at some point, adding features might be computationally expensive and create overfitting problems, especially when the features are redundant or the training set is small. This issue commonly reflects the curse of dimensionality: the number of instances required to estimate an arbitrary function with a given level of accuracy grows exponentially with the number of input features (feature space dimensionality) of the function.

Two main techniques are used to decrease dimensionality: feature transformation and feature selection. Feature transformation creates new features based on combinations and transformations of the original extracted feature set [50]. Feature transformation includes unsupervised learning methods such as Principal Components Analysis, Factor Analysis [51], or non-negative matrix factorization [52].

Feature selection identifies a smaller subset of relevant features from the original set by removing irrelevant, redundant, or noisy features. Feature selection usually leads to better learning performance (i.e., higher accuracy, lower computational cost, better model interpretability) [53]. Feature selection methods include filter, wrapper, or embedded methods.

Filter methods select the features regardless of the classifier (learning algorithm). The least interesting features are suppressed and the model tends to be more robust to overfitting. Filter methods pick up the intrinsic properties (i.e. the relevance) of the features. Wrapper methods use a predictive model (classifier) to score feature subsets; the classifier performance metrics are used to evaluate and select the best features subset. Embedded methods incorporate feature selection in the classifier's training process.

Wrapper and Embedded methods are very classifier-dependent; therefore, the feature selection process is not generalizable across a variety of classifiers. Filter methods are preferred for HAR. Popular filter-based feature selection methods include ReliefF, Correlation, Chisquared, and Information Gain.

ReliefF is an instance-based evaluator that samples instances randomly and checks the instances nearby of the same and different classes. ReliefF has been heavily used in HAR studies [54]. Correlation evaluates the worth of a feature by measuring Pearson's correlations between that feature and the class, whereas ChiSquared evaluates features by computing the feature's Chi-square statistics with respect to the class [55]. Information Gain (InfoGain) is a single-feature evaluator that measures the feature's total entropy with respect to the class.

2.7.4. Choosing a classifier

Mitchell [56] defines a classifier as a mathematical function or an algorithm that maps input data (input features) to a category (class label). A classifier could also be constructed by a set of rules or methods. The algorithm learns the dataset patterns through training and testing, and leads to predicting performance metrics.

The learning method could be either supervised or unsupervised. In supervised learning, the dataset and the correct output (class labels) are provided. The algorithm predicts the output and compares it to the known correct classes. Unsupervised learning methods are applied in problems where there is no information about the output (data classes). In this case, the classification relies on methods such as clustering or association [51] to label a class. Capela claimed that HAR tends to adopt supervised learning [44].

The choice of a classifier is crucial to obtain satisfactory evaluation results. The "no free lunch" theorem states that one classifier cannot be considered as the best for a general HAR problem. In other words, there is no algorithm that is always superior to the others. For a specific application, the "best" classifier depends on parameters such as the problem hypotheses and the type of data analysed. The classifier choice also depends on data size (number of instances) and number of features (dimensionality of the feature vector).

Several studies have compared classifiers, with different accuracy results [57]. Six general classifiers are described in Table 2.3: Support Vector Machines (SVM), k-Nearest Neighbours (kNN), Decision Trees, Random Forests, Neural Networks, and Naïve Bayes.

SVM can be defined as a system using a hypothesis space of a linear function in a high dimensional feature space. This classifier relies on an optimal hyperplane to separate data classes. The optimal classifier is a probabilistic linear classifier computed to find the largest minimum distance between support vectors. SVMs perform well on data sets with a large number of features and can be used for linear or non-linear classification.

kNN is an instance-based learning technique where the function is approximated locally. kNN finds a group of k instances in the training set that are nearest to the desired class and allocates the class label based on the predominance of a class in the local neighbourhood. kNN is defined by three main parameters: a set of labelled features, the distance measure between two data points and the number of nearest neighbours (k). kNN algorithms can handle missing values, are robust to outlying data points, and ease difficult models' interpretation.

Decision Trees are machine learning models resembling tree structures. They start with a single node (root), which branches into possible outcomes. Decision nodes have two or more branches, and leaf nodes represent the final decision. Tree model target values are a discrete set of values (classification trees) or continuous values (regression trees).

Table 2.3: Description of the main classifiers

Classifier	Advantages	Drawbacks	Studies
Random Forest	• Easier to understand • Fast, robust, good accuracy	• Slower than trivial methods (Naïve Bayes, kNN) • Works best with equal classes (no imbalance)	[35]
SVM	• Adequate with problems that might not be linearly separable • Great accuracies • Fits high dimensional problems, large amount of data • Good for multiclass	• Can be inefficient to train • Memory intensive • Finding the kernel could be a challenge • Hard to interpret	[35], [37], [41],[58]
Decision Tree	• Easy to interpret and explain • Fast to train • Non-parametric (does not require linearly separable data, handles outliers) • No distribution requirement, good for a few categories of variables	• Easily overfit • Accuracy depends on data type	[35],[40], [58]
Naïve Bayes	• Simple to implement and converge fast • High Performance	• Variable independence assumption is constraining	[35],[40],[58]
Neural Network	• Less formal statistical training • Multiple training algorithms • Detect all possible interactions between the predictor variables • Great for complex problems, can approximate any function, regardless of its linearity	• Great computational burden	[37], [40]
kNN	• Very simple to implement • Works well on basic recognition systems	• Algorithm does not learn from training data • Slow	

Random Forests is an ensemble learning method for classification and regression, made of multiple decision trees. Random Forests mainly correct Decision Trees overfitting disadvantages.

Artificial neural networks have been inspired by biological neural networks present in the brain. They are based on a collection of connected units (nodes), organized in layers and communicating with each other. Data propagates from the first (input) layer to the last (output) layer through intermediate layers. Neural network structures include feed-forward propagation, back propagation, and perceptrons.

Naïve Bayes is a probabilistic classifier founded on Bayes' theorem. Features are assumed to be independent. The classifier is highly scalable and requires a large number of parameters.

2.7.5. Evaluation

Correctly and incorrectly classified instances can be used to evaluate a classifier's performance. For this thesis, an aggressive instance correctly identified by the classifier is considered a true positive (TP). An aggressive instance misclassified by the predictor is a false positive (FP). A non-aggressive instance correctly identified would be a true negative (TN). A misclassified non-aggressive instance is a false negative (FN). A confusion matrix (Table 2.4) maps these outcomes.

Table 2.4: Confusion Matrix

	Predicted Condition	
True Condition	True Positive (TP)	False Negative (FN)
	False Positive (FP)	True Negative (TN)

From the classified instances, three types of metrics can be extracted [59]: threshold metrics (accuracy, Lift, F-score), ordering/ranking metrics (ROC Area, Average Precision), probability metrics (RMS, Cross-entropy, Probability C).

Accuracy is a common performance metric but classifier performance cannot only be described by accuracy. When class imbalances occur, accuracy is insufficient. For example, a HAR algorithm fed with 900 instances of gentle movement and 100 instances of aggressive movements will achieve 90% accuracy even if **all** movements are classified non-aggressive (i.e., the algorithm can still claim to be very good despite all the aggressive movements being misclassified). To deal with such issues, other metrics (i.e. precision, recall, F-score, etc.) provide additional insight into classification performance (Table 2.5).

A summed ranking classifier selection method can be used to combine several metrics to rank classifier performance [60]. The summed ranking method ranks classification models in descending order (best results ranked as 1) according to each metric. The ranks for all six metrics are subsequently summed to provide an overall ranking for each model. Models are sorted in descending order, since the lowest rank value indicated the best model. The ranking method does not focus on one metric and involves results from all six parameters. This gives a better, wider, and more generalizable representation of model performance.

Table 2.5: Performance metrics

Metric	**Formula**
Accuracy	$\frac{\sum \mathrm{TP} + \sum \mathrm{TN}}{\mathrm{Total\ Population}}$
Sensitivity	$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$
Specificity	$\frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}$
Precision	$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$
F-score	$\frac{2.\mathrm{Precision}.\mathrm{Sensitivity}}{\mathrm{Precision} + \mathrm{Sensitivity}}$
Matthew Coefficient Correlation (MCC)	$\frac{\mathrm{TP}.\mathrm{TN} - \mathrm{FP}.\mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}}$

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

3. Feature Selection for Classification of Aggressive Movements using Smartwatches

This chapter addresses Objective 1 by determining the best machine learning features that can classify aggressive and non-aggressive movements. Feature selection methods are evaluated by a random forest classifier to choose the best twenty features. The study is a proof of concept to determine how effective smartwatches can be in determining aggressive activities. Six performance metrics are evaluated: accuracy, sensitivity, specificity, precision, F-score and Matthew Coefficient Correlation (MCC).

This chapter was submitted for publication to the Journal of Biomedical and Health Informatics:

Tchuente, F, Lemaire, ED, Baddour, N, "Feature Selection for Classification of Aggressive Movements using Smartwatches", Journal of Biomedical and Health Informatics, submitted July 2018.

3.1. Abstract

Aggressive activities can occur in clinical and elderly care settings with people suffering from dementia, mental disorders, or other conditions that affect memory. Since identifying the nature of the event can be difficult with people who have memory and communication issues, other methods to identify and record aggressive motion would be useful for care providers that need to determine the best methods to reduce reoccurrences of this activity. A wearable technology approach for human activity recognition was explored to detect aggressive movements.

Participants donned two Microsoft Bands 2 smartwatches and performed an activity circuit of similar aggressive and non-aggressive movements. Smartwatch accelerometer and gyroscope sensors captured data that was used to extract 136 features. Filter-based feature selection

methods (ReliefF, Information Gain, Chi-squared, Correlation, from the Waikato Environment for Knowledge Analysis, WEKA) were used to determine the best features for distinguishing between aggressive and non-aggressive movements. A Random Forest classifier with 5-fold cross-validation was used to evaluate performance metrics for each feature selection method.

Derivative, range, and standard deviation were the best overall features across the feature selection methods. The average results from the four methods were 99% accuracy, 94.6% sensitivity, 99.8% specificity, and 98.6% precision.

Accelerometer and gyroscope-based features (derivative, range, standard deviation) are relevant for classifying aggressive movements. The high-performance metrics support further investigation of a smartwatch approach for aggressive human activity recognition.

3.2. Introduction

The ability to identify and recognize human activities has improved considerably with the rapid pace of wearable technology advancements, such as miniaturized sensors embedded in phones, glasses, watches, and cameras. Smart devices have been applied for Human Activity Recognition (HAR) across the landscape of life: sports [61], entertainment [8], surveillance [10], healthcare [62]. Such tools may also be used to detect aggressive movements in clinical settings with patients suffering from dementia or mental illnesses and who exhibit periods of hostile behavior.

For HAR, smart devices output sensor signals that are filtered, pre-processed, and segmented can be used to extract relevant features for categorizing the target movements. The process can be summarized as pre-processing, segmentation, feature extraction and selection, and finally classification [63]. Feature selection and extraction can enhance the efficiency of the classifier, reduce the chance of overfitting, and increase classification accuracy by removing unneeded and redundant parameters [64].

Movement aggressiveness is a branch of HAR that is typically analyzed using computer vision. However, image and video processing are computationally intensive and raise privacy issues, especially in clinical establishments or nursing homes where residents do not want to be continuously video recorded. Ouanane [42] recognized aggressive human activity with two visual methods, bag of features and skeleton graph. The recognition rate was 96% for boxing, hand clapping, hand waving, jogging, running, and walking.

Few researchers have attempted to classify aggressiveness using smart technologies. Koh (9) used a Galaxy Note 2 smartphone's lateral acceleration to detect driving aggressiveness. Aggressive driving included high-intensity U-turn challenges and smooth driving was conducted on a smooth rectangular course. To date, published research is lacking on smartwatch applications for aggressive activity identification.

The goal of this paper is to determine a viable feature set from smartwatch acceleration and gyroscope output that can discriminate between aggressive and non-aggressive motions, using a machine learning classifier. Wrist mounted inertial measurement units have a great potential for broad application within healthcare and elderly-care facilities and provide a repeatable location for capturing upper-limb related aggressive activities. Identification of aggressive movements will improve service delivery for care providers by enabling alarm-based notification of event onset and also providing quantitative information on who initiated the aggressive event, which is often difficult to understand in elderly care environments where dementia is prevalent.

3.3. Methodology

3.3.1. Data Collection and Equipment

A convenience sample of 30 able-bodied adults(15 male, 15 female) were recruited from The Ottawa Hospital Rehabilitation Centre (TOHRC) staff, students, and volunteers, and the community. Characteristics included age (25.9±8.0), weight (70.2±11.9 kg), height (170.7± 8.6 cm), and right-handedness (20 out of 30). This large variability is important because it broadens the range of possible movements and thereby potentially broader applicability of the classified movements beyond the population sample. The study was approved by the Ottawa Health Science Network and the University of Ottawa Research Ethics Board. All participants read and signed an informed consent form.

Each participant wore a Microsoft Band 2 (MSB2) smartwatch on each wrist and performed a series of aggressive and non-aggressive movements. The MSB2 sensors used in this project were the tri-axial accelerometer and tri-axial gyroscope (Figure 3.1). Aggressive actions were performed on a Body Opponent Bag (BOB) (Figure 3.2). The TOHRC Data Logger [65] Android app was modified for signal acquisition from two MSB2, with Bluetooth communication between the MSB2 and a Nexus 5 smartphone. Videos of the movements were recorded with a separate smartphone camera to provide gold standard activity timing.

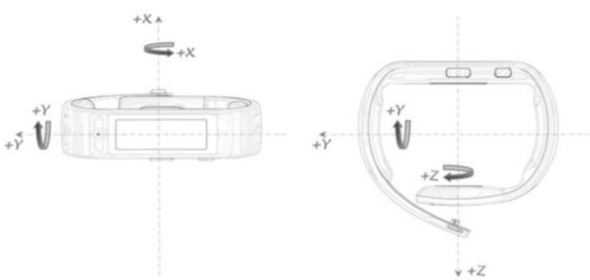


Figure 3.1: MSB2 accelerometer and gyroscope axes orientation

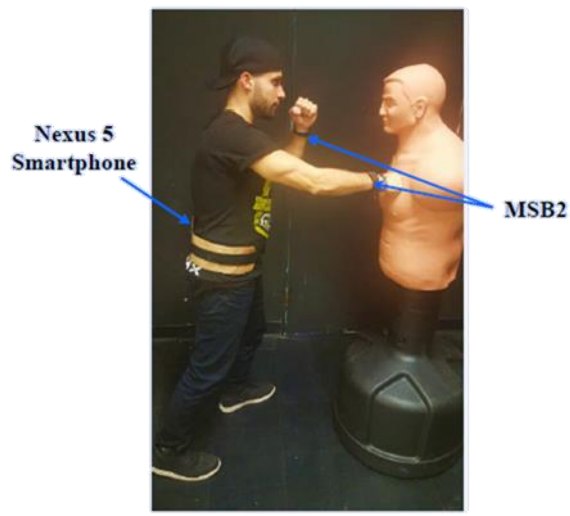


Figure 3.2: Participant punching the Body Opponent Bag

3.3.2. Circuit Activities

Participants performed an activity circuit that included non-aggressive and aggressive actions (Table 3.1). Similar activities, such as slapping and clapping, were chosen to present opportunities for misclassification.

3.3.3. Feature Selection and Evaluation

Raw data were extracted from two MSB2 accelerometers and gyroscopes sensors at a frequency of 50 Hz. The data were divided into 1-second sliding windows (50 data points) for feature extraction. One second represented enough time to capture each activity. Windows overlapped by 96% of the window length (window advanced by 2 data points). This resulted in 18,379 instances for aggressive activities and 103,928 instances for non-aggressive activities.

Table 3.1: Activities

Movement	Activity	Description
Aggressive movements	Punch	Participant punches BOB eight times, alternating hands
	Shove	Participant aggressively shoves BOB five times with both hands
	Slap	Participant aggressively slaps BOB ten times alternating hands
	Shake	Participant holds BOB's neck and shakes BOB back and forth five times
Non-Aggressive movements	Transitions	Set of movements between an aggressive action and non-aggressive action (i.e., sitting, standing, moving, still)
	Clap	Participant claps their hands ten times
	Wave	Participant waves with the preferred hand as if they are saying goodbye
	Handshake	Participant handshakes the project assistant
	Open/Close door	Participant opens and closes the door three times
	Type on a keyboard	Participant types the first verse of the Canadian National anthem

3.3.4.1 Feature Description

Features were extracted from the twelve sensor data streams (tri-axial accelerometer and gyroscope data per smartband). Time-domain features were preferred because they are less computationally expensive than frequency-domain for HAR [66]. Two types of features were considered: statistical features and physical features (detailed in Section 2.7.3). Statistical features included mean, variance, median, range, standard deviation, skewness, kurtosis, maximum difference and pairwise correlation coefficient. Each of the statistical feature had twelve components and were extracted over each window.

As an example, a window W made up of 50 x-acceleration data points $\{a_1, a_2, \dots, a_{50}\}$ would have its mean $\overline{x}$ defined as

$$\bar{x} = \sum_{i=1}^{n} \frac{a_x(i)}{n}$$

Physical features included area under the curve (4 components), signal magnitude area (SMA, acceleration magnitude summed over three axes within each window normalized by the window length), sum of all squares (acceleration magnitude squared and summed over three axes, normalized by window length, 4 feature components), mean movement intensity (mMI: Euclidean norm of the total acceleration vector, 4 feature components), and variant movement intensity (vMI; 4 feature components).

$$\mathrm{mMI} = \sum_{i=1}^{n} \frac{MI(i)}{n} \text{ where } \mathrm{MI(i)} = \sqrt{\mathrm{a}_\mathrm{x}^2(\mathrm{i}) + \mathrm{a}_\mathrm{y}^2(\mathrm{i}) + \mathrm{a}_\mathrm{z}^2(\mathrm{i})}$$

where $\mathrm{a_x}$ (i), $\mathrm{a_y}$ (i), and $\mathrm{a_z}$ (i) represent the acceleration sample of the *x*, *y*, and *z* axes and n is the number of data points. A total of 136 time-domain features were initially chosen to classify aggressive and non-aggressive movements. Appendix A provides more information on the features.

3.3.4.1 Feature Selection

Four filter-based WEKA feature selection methods were used to determine the best 20 features or attributes for aggressive movements: ReliefF, Correlation, Information Gain, and ChiSquared.

ReliefF is an instance-based evaluator that samples instances randomly and checks the instances nearby of the same and different classes. ReliefF has been heavily used in HAR studies [54].

Correlation evaluates the worth of an attribute by measuring Pearson's correlations between that attribute and the class; whereas, ChiSquared evaluates attributes by computing the feature’s Chi-square statistic with respect to the class [55].

Information Gain (InfoGain) is another single-attribute evaluator that measures the attribute’s information gain with respect to the class. InfoGain measures the total entropy for an attribute if, for each of the attribute’s values, a unique classification can be made for the result attribute [55].

All these methods employ a ranker search that provides a specific rank to the features during the evaluation process, corresponding to a weight. The more the weight is applied to a feature, the more important the feature is for the selection method. For example, a feature ranked 3 is the third best feature.

3.3.4.1 Evaluation

To evaluate the selected features and assure that the feature set achieves appropriate performance, a Random Forest machine learning classifier with 5-fold cross validation was applied to the dataset. 5-fold cross-validation is consistent with the number of participants; each fold would contain enough data instances to be trained and tested effectively.

Random Forest classification is fast and robust, thereby providing a viable approach for real-time aggressive activity identification. Random Forest has been an effective and powerful algorithm used in real-time HAR classification; for example, Sangjun et al. [67] used Random Forest to instantly detect hand gestures. Outcome measures were derived from a binary confusion table matrix and included accuracy, sensitivity, specificity, F-score, and MCC.

3.4. Results

3.4.1. Feature Selection

Table 3.2 shows the feature sets from the four feature selection methods. Interestingly, InfoGain and Chi-Squared resulted in the same features. “Maximum difference” was the predominant feature, across all sensor axes. For ReliefF, “correlation between the axes” feature occurred most often, followed by acceleration features such as mean, area under the curve, and skewness. The two watches were equally represented with ReliefF (10 features from band 1 and 10 features from band 2). The Correlation method had three predominant features: derivative, range, and standard deviation. These features were equally distributed between the two watches (10 features from the band 1 and 10 from the band 2). However, features based on the acceleration were more prominent (17 out of 20 features from the acceleration). All axes appeared in the features.

Table 3.2: Twenty best features selected for each method, in order of importance

ReliefF	Correlation	Chi Squared	InfoGain
Pcc_Gyr_yz_1	Std_Acc_x_1	Diff_Gyr_z_2	Diff_Gyr_z_2
Pcc_ Gyr xy_1	Range_Acc_x_1	Diff_Acc_x_2	Diff_Acc_x_2
Pcc_ Gyr_ yz_2	Std_Acc_x_2	Diff_Acc_z_1	Diff_Gyr_y_1
Pcc_ Acc_yz_1	Range_Acc_x_2	Diff_Gyr_y_1	Diff_Acc_y_1
Pcc_ Acc_ yz_2	Std_Acc_y_2	Diff_Acc_y_1	Diff_Acc_z_1
Pcc_ Gyr_ xz_1	Range_Acc_y_2	Diff_Acc_y_2	Diff_Acc_y_2
Pcc_ Acc_ xy_2	SMA_Acc_2	Diff_Gyr_z_1	Diff_Gyr_y_2
Pcc_Acc_xy_1	Range_Gyr_z_2	Diff_Gyr_y_2	Diff_Gyr_x_1
Pcc_ Gyr_ xz_2	AI_Acc_2	Diff_Gyr_x_1	Diff_Gyr_z_1
Pcc_ Gyr_ xy_2	AI_Acc_1	Diff_Acc_z_2	Diff_Acc_z_2
Pcc_ Acc_ xz_1	Range_Acc_y_1	Diff_Acc_x_1	Diff_Acc_x_1
Pcc_ Acc_ xz_2	SMA_Acc_1	Range_Acc_x_2	Range_Acc_x_2
Mean_Acc_y_2	Std_Acc_y_1	Range_Acc_x_1	Range_Acc_x_1
Area_Acc_y_2	Range_Gyr_z_1	Range_Acc_y_2	Range_Acc_y_2
Med_Acc_y_2	Var_Acc_x_1	Range_Acc_z_1	Range_Acc_z_1
Area_Acc_z_1	Diff_Acc_x_1	Diff_Gyr_x_2	Diff_Gyr_x_2
Mean_Acc_z_1	Range_Gyr_y_2	Range_Acc_y_1	Range_Acc_y_1
Med_Acc_z_1	Range_Acc_z_2	Var_Acc_x_1	Var_Acc_x_1
Skew_Acc_y_2	Diff_Acc_x_2	Std_Acc_x_2	Std_Acc_x_1
Area_Acc_y_1	Diff_Acc_y_1	Std_Acc_x_1	Std_Acc_x_2

Pcc=Pairwise Correlation Coefficient, Gyr=gyroscope, Acc=acceleration, Med=Median, Skew=skewness, SMA=signal magnitude area, Diff=maximum difference, Var=variance, Std=standard deviation, 1=left wrist, 2=right wrist

3.4.2. Feature evaluation

Table 3.3 displays the Random Forest confusion matrix from the full feature set. With aggressive actions as the positive, 154 false positives and 557 false negatives were reported.

Evaluation of the four methods (Table 3.4) resulted in an average accuracy of 99%, sensitivity of 94.6%, specificity of 99.8%, and precision of 98.6%. The feature set determined by ChiSquared and InfoGain was the best when coupled with Random Forest. While still scoring above 0.9, ReliefF had lower outcome scores than the other feature selection sets, especially for sensitivity.

Table 3.3: Confusion matrix of the full data set

		True Condition	
		Aggressive	Non-Aggressive
Predicted Condition	Aggressive	17822	154
	Non- Aggressive	557	103774

Table 3.4: Performance metrics

	Accuracy	Sensitivity	Specificity	Precision	F-score	MCC
Full Set	0.9941	0.970	0.999	0.991	0.980	0.977
ReliefF	0.9822	0.893	0.998	0.987	0.938	0.929
InfoGain	0.9918	0.960	0.997	0.985	0.972	0.967
Chisquared	0.9918	0.960	0.997	0.985	0.972	0.967
Correlation	0.9898	0.947	0.997	0.984	0.965	0.960

3.5. Discussion

This study demonstrated the viability of wrist-worn inertial sensors for classification of aggressive and non-aggressive movements. An appropriate twenty feature set achieved classification results above 0.96, and specificity near 1, which would indicate that this approach could be viable as an alarm or indicator within elderly or healthcare facilities.

InfoGain and ChiSquared methods resulted in the same features, although not in the same ranking order, even though these two methods worked differently (information gain versus statistical analysis). The feature set's strength is therefore highlighted. The main feature type for this set was maximum difference, with all twelve features represented. Range, which was the second discriminatory feature, was mainly based on three-dimensional accelerations. Since these features were based on the difference between the maximum and minimum gyroscope or acceleration values, aggressive movements could tend to have high differences in amplitude over short periods, compared to non-aggressive movements that tend to be more stable.

The top ReliefF feature rankings were for pairwise correlations between axes for the accelerometer and gyroscope signals, followed by the mean, area under a curve, and skewness of accelerometer signals. ReliefF features were substantially different from the other methods in terms of selected features and presented no similarities with the other methods. The three main feature types were maximum difference, range, and standard deviation. The maximum difference and the range were also selected by InfoGain and ChiSquared.

All these methods use signals from both smartwatches to obtain the features. The acceleration features occurred more frequently than the gyroscope features. Relief F, ChiSquared, and InfoGain used 14 accelerometer features out of 20, whereas Correlation feature selection only used 3 gyroscope features out of 20. Since accelerometer-only wrist-worn devices could be less expensive and less power consuming, future research could explore accelerometer-only applications for aggressive activity classification.

Few physical features were found in the reduced feature sets, with only integral and signal magnitude area (7^{th}, 9^{th},10^{th} and 12^{th} features) using the Correlation method. This raises questions about the usefulness of such features for binary aggressive/non-aggressive classification. In future research, these physical features could be evaluated using a subset attribute evaluator such as Correlation-based Feature Selection or other methods such as the minimum-redundancy maximum-relevance (mRMR) to determine if the resulting sub-sets include more physical features. Features were considered in three dimensions, but no preference in one movement plane helped differentiate between classes.

3.5.1. Random Forest Evaluation

The Random Forest classifier performance metrics had approximately the same results with 20 features compared to 136.

The high evaluation results, greater than 0.96, may reflect that this type of binary classification is easily differentiable because of the nature of the movements. Accuracy should not be the main metric to consider because of the large, but normal, class imbalance between the non-aggressive and aggressive movements. Class imbalance is a common phenomenon that is portrayed in many areas. In this case, real situational aggressive movements occur far less frequently than daily non-aggressive movements. The imbalance therefore reflects the real-life situation. Class imbalance could be addressed with a ranking method that combines all the performance metrics, as suggested by Drover [60].

Sensitivity evaluates the ratio between the true positives (predicted aggressive actions that are really aggressive) and the actual positive conditions (instances of aggressive actions). The value of 94.6% indicates that more than five percent of the aggressive actions were not detected by the smartwatch. On the other hand, specificity was very high at 99.8% (i.e., non-aggressive movements were detected by the algorithm almost all the time).

The precision of 98.6% evaluated the “false alarms” or the ratio between the predicted aggressive movements that were really aggressive (true positives) compared to the aggressive predictions of the classifier. Only 1.4% of the movements classified by Random Forest as aggressive were actually non-aggressive. Our focus is to detect aggressive movements, allowing care staff to intervene and/or the event to be recorded, but avoid false alarms for aggressive situations that would consume human resources when reacting to the alarm. For an aggressive motion alarm application, such as minimally staffed night shifts on a dementia ward, high sensitivity is essential to avoid responses to false positive events. Therefore, the excellent sensitivity and precision results address these requirements.

Overall, ReliefF took the longest time to compute and Random Forest methods had poorer results. The ReliefF feature set could be investigated with other classifiers to consider if it is viable for classification between aggressive and non-aggressive movements.

ChiSquared and InfoGain methods, which presented the same feature set, had overall the best scores, but still slightly less than the full feature set. The feature set extracted from the two methods was therefore recommended for use in aggressive movement classification applications.

From the high-performance metric scores, overfitting may have occurred. Overfitting may happen in HAR problems [68] and it is important to consider when generalizing results to other sets, especially in binary classes where two results might be easily differentiable. In this study, 5-fold cross-validation was used to reduce the chance of overfitting.

3.6. Conclusions and Future work

This research identified a 20-feature set that can differentiate between aggressive and non-aggressive movements using wrist-worn smartwatch acceleration and gyroscope signals. Out of the four attribute selection methods (ReliefF, Chi-Squared, Correlation, InfoGain), the best features set was obtained using ChiSquared and InfoGain methods, in terms of accuracy; fewer false positives, and fewer false negatives were obtained. These two methods had the same features, even though they did not follow the same order. Future research could test other feature selection methods such as Correlation-based Feature Selection and mRMR to determine any similarities with the selected features in this research. Additional classifiers such as Neural

Networks or Support Vector machines can be tested to compare performance metrics. Future work should also apply the chosen features and classifier on data from elderly people to observe how the selected features perform with a different population.

4. Classification of Aggressive Movements using Smartwatches

This chapter addresses Objective 2 by evaluating combinations of machine learning classifiers and feature selectors that will result in the best performance metrics to distinguish between aggressive and non-aggressive activities. Six classifiers (k-Nearest Neighbours, Random Forests, Neural networks, Support Vector Machines, Decision Tree, and Naïve Bayes) are associated with three feature selection methods (ReliefF, InfoGain and Correlation). The summed ranking is the method used to rank the eighteen possible models based on their evaluation metrics.

4.1. Abstract

Background

Recognizing aggressive motion is a challenging task in human activity recognition. Wearable smartwatch technology in combination with a machine learning classifier may be a viable approach for human aggressive motion classification. The objective of this paper was to determine the best Classification Model and Feature Selector (CM-FS) combination for separating aggressive from non-aggressive movements from data collected from the smartwatch.

Methods

A ranking method was used to select relevant CM-FS models across accuracy, sensitivity, specificity, precision, F-score and Matthews Correlation Coefficient (MCC). The Waikato Environment for Knowledge Analysis (WEKA) was used. Six WEKA machine learning classifiers (k-Nearest Neighbours (kNN), Random Forest, Support Vector Machine, Decision Tree, Naïve Bayes) coupled with three machine learning feature selectors (ReliefF, InfoGain, Correlation) formed the models. The activity circuit included punch, shove, slap, and shake (aggressive) and clapping hands, waving, handshaking, opening/closing a door, typing on a computer keyboard (non-aggressive).

Results

A combination of kNN and ReliefF was found to be the best CM-FS model for separating aggressive actions from non-aggressive actions, with 99.6% accuracy, 98.4% sensitivity, 99.8% specificity, 98.9% precision, 0.987 F-score, and 0.984 MCC. The kNN and Random Forest classifiers, combined with any of the three feature selectors, generated the top models. On the other hand, models with Naïve Bayes or Support Vector Machine classifiers had poor performance for sensitivity, F-score, and MCC.

Conclusion

The kNN and ReliefF combination results demonstrate that this smartwatch-based approach is a viable solution for identifying aggressive movements. This wrist-based wearable sensor approach could be used by care providers in settings where people suffer from dementia or mental health disorders, where random aggressive motions often occur.

4.2. Background

Human Activity Recognition (HAR) is a growing field that benefits different industries including sports and entertainment [6], [7], gaming, video surveillance [19], and healthcare [69]. In healthcare, HAR is used for a range of applications, from fall detection [35] to gait analysis [70] to rehabilitation [71]. Some clinical settings host people with dementia or mental illnesses, who may become aggressive and violent during their interactions with others in the living space. Identifying aggressive movements is therefore important for monitoring and preventing escalating situations.

The best method to detect aggressive motions has not been determined. Computer vision has been applied to this task, with research on fight prevention and de-escalating aggressive motion situations [7]. However, computer vision methods can be very expensive financially and computationally (i.e., to analyze and study huge amounts of digital video). Privacy is also a concern, since people do not necessarily want to be filmed continuously. Smartwatches may be a viable alternative to computer vision based approaches to determine aggressive occurrences.

Smartwatch HAR has been applied to several sectors: Gjoreski [35] used wrist-worn Actigraph to evaluate activity recognition and fall detection. In 2016, Arduser attempted to recognize text using smartwatch motion data. A gesture recognition system was built to identify gestures associated to writing 26 letters of the Roman alphabet. Data were collected using a commercial, off the shelf smartwatch with accuracies ranging from 94% to 99% for letter recognition [36].

On the other hand, smartwatches could be helpful to detect stereotyped movements in children with developmental disability [37]. Lee and Song used a SONY SWR50 smartwatch to train, recognize, and test stereotyped movements (swinging an arm, raising an arm, throwing) on teenagers diagnosed with developmental disability. However, literature is lacking for smartwatch applications in aggressive motion detection.

Appropriate machine learning is required for effective smartwatch HAR. A variety of classifiers have been used for wrist-worn sensor HAR: Random Forest, k-Nearest Neighbour, neural network, or decision trees [31], [64], [72]. To perform well, classifiers might use features from selection methods that include ReliefF and Correlation.

The aim of this preliminary research was to determine the best combined "machine learning classifier and feature selection" model for classifying aggressive and non-aggressive activities using smartwatch inertial sensor data. A wrist-worn wearable sensor approach for detecting aggressive motion events could be an important tool to help manage care in the resource-limited area of elderly care.

4.3. Methods

30 able-bodied and healthy participants (15 male, 15 female) were recruited from The Ottawa Hospital Rehabilitation Center (TOHRC) and the University of Ottawa. Characteristics included age (25.9±8.0), weight (70.2±11.9 kg), height (170.7± 8.6 cm), and righthandedness (20 out of 30). The study was approved by the Ottawa Health Science Network and the University of Ottawa Research Ethics Board. All participants read and signed an informed consent form.

The protocol consisted of aggressive and non-aggressive activities. Aggressive activities were punch, shove, slap, and shake a Body Opponent Bag (Figure 4.1). Non-aggressive activities were clapping hands, waving, handshaking, opening/closing a door, and typing on a computer keyboard. Similar activities (e.g., slap, clap; shove, push a door open) were chosen to challenge the classifiers and evaluate potential false positive events.

Figure 4.1: The Body Opponent Bag (BOB)

Participants wore one Microsoft Band 2 (MSB2) per wrist and donned a holster on their pelvis that carried a Nexus 5 smartphone. The MSB2 recorded upper-limb motion at 50 Hz via their integrated tri-axial accelerometer and gyroscope. The Nexus 5 smartphone was connected via Bluetooth to the smartwatches using the TOHRC Data Logger Android app [65], updated for signal acquisition from two MSB2. A second smartphone was used to video record the participant's movements and to serve as a gold standard comparator. The gold standard time was synchronized with smartwatch sensor output by shaking the hands at the beginning and end of the trial, providing a recognizable accelerometer signal and video event.

4.3.1. Machine learning classifiers and feature selector

Six machine learning classifiers that have been used extensively across HAR areas [73] were evaluated: Random Forests (RF), k-Nearest Neighbours (kNN), Multilayer Perceptron Neural Network (MP), Support Vector Machines (SVM), Naïve Bayes (NB), and Decision Tree (DT). Waikato Environment for Knowledge Analysis (WEKA) [74] was used for all classifications. These machine learning classifiers were fed by feature sets from three popular feature selection methods: ReliefF (ReF), Infogain (IG), and Correlation (C). These methods were used to select the 20 best features out of 136 from accelerometer and gyroscope sensor data (Tables 4.1 and 4.2) (more details on Chapter 3).

Table 4.1: Description of features from accelerometer and gyroscope signals, from the two smartwatches

Features	Description	Number of features
Statistical features		
Mean	Average for 3 axes	12
Variance	Variance for 3 axes	12
Median	Median for 3 axes	12
Range	Range for 3 axes	12
Standard Deviation	Standard deviation for 3 axes	12
Skewness	Degree of asymmetry for 3 axes	12
Kurtosis	Degree of peakedness for 3 axes	12
Pairwise Correlation Coefficient	Correlations between sensor axes (x,y; x,z; y,z)	12
Integral	Area under the curve for 3 axes	12
Maximum Difference or derivative	Difference between the highest and the lowest value	12
Physical Features		
Movement intensity	Average Movement Intensity (MI): Euclidean norm of the total acceleration vector after removing the static gravitational acceleration	8
Signal Magnitude Area (SMA)	Acceleration magnitude summed over three axes, normalized by window length	4
Sum of All Squares	Acceleration magnitude squared and summed over three axes, normalized by window length	4

Performance for each combination of Classification Method and Feature Selector (CM-FS) for differentiating aggressive and non-aggressive movements was calculated using accuracy, sensitivity, sensibility, precision, F-score, and Matthews Correlation Coefficient (MCC). For example, RF-ReF refers to the Random Forest classifier fed by the Relief-F feature set.

Table 4.2: Twenty best features selected for each method

ReliefF	Correlation Attribute	InfoGain
Pcc_Gyr_yz_1	Std_Acc_x_1	Diff_Gyr_z_2
Pcc_ Gyr xy_1	Range_Acc_x_1	Diff_Acc_x_2
Pcc_ Gyr_ yz_2	Std_Acc_x_2	Diff_Gyr_y_1
Pcc_ Acc_yz_1	Range_Acc_x_2	Diff_Acc_y_1
Pcc_ Acc_ yz_2	Std_Acc_y_2	Diff_Acc_z_1
Pcc_ Gyr_ xz_1	Range_Acc_y_2	Diff_Acc_y_2
Pcc_ Acc_ xy_2	SMA_Acc_2	Diff_Gyr_y_2
Pcc_Acc_xy_1	Range_Gyr_z_2	Diff_Gyr_x_1
Pcc_ Gyr_ xz_2	AI_Acc_2	Diff_Gyr_z_1
Pcc_ Gyr_ xy_2	AI_Acc_1	Diff_Acc_z_2
Pcc_ Acc_ xz_1	Range_Acc_y_1	Diff_Acc_x_1
Pcc_ Acc_ xz_2	SMA_Acc_1	Range_Acc_x_2
Mean_Acc_y_2	Std_Acc_y_1	Range_Acc_x_1
Area_Acc_y_2	Range_Gyr_z_1	Range_Accy_2
Med_Accy_2	Var_Acc_x_1	Range_Acc_z_1
Area_Acc_z_1	Diff_Acc_x_1	Diff_Gyr_x_2
Mean_Acc_z_1	Range_Gyr_y_2	Range_Acc_y_1
Med_Acc_z_1	Range_Accz_2	Var_Acc_x_1
Skew_Acc_y_2	Diff_Acc_x_2	Std_Acc_x_1
Area_Acc_y_1	Diff_Acc_y_1	Std_Acc_x_2

Pcc=Pairwise Correlation Coefficient, Gyr=gyroscope, Acc=acceleration, Med=Median, Skew=skewness, SMA=signal magnitude area, Diff=maximum difference, Var=variance, Std=standard deviation, 1=left wrist, 2=right wrist

The 18 CM-FS models were evaluated using a summed ranking method [60], where each performance metric was ranked for each CM-FS (best result ranked as 1) and then the ranks for all six metrics were summed to provide an overall ranking for each model. CM-FS models were sorted in descending order, since the lowest rank value indicated the best model. In the literature, accuracy tends to be the most used metric for evaluating classification results. However, this metric loses some of its power in cases of class imbalances, which occurs often in HAR. The ranking method does not focus on one metric, involving results from all six parameters. This gives a better, wider, and more generalizable representation of model performance.

4.4. Results

kNN-ReF was the best model combination, with 99.6% accuracy, 98.4% sensitivity, 99.8% specificity, 98.9% precision, 0.987 F-score, and 0.984 MCC (Table 4.3). kNN-ReF, RF-IG, kNN-C, kNN-IG, RF-C and RF-ReF were the top 6 models, with average metrics of 99.16% accuracy, 95.75% sensitivity, 99.77% specificity, 98.82% precision, 0.9715 F-score, and 0.9670 MCC.

Table 4.3: Classification method and feature selection combination sorted by summed rank (best to worst).

	Score						Rank					
	Acc	**Sens**	**Spec**	**Prec**	**FS**	**MCC**	**Acc**	**Sens**	**Spec**	**Prec**	**F**	**MCC**
kNN-ReF	0.996	0.984	0.998	0.989	0.987	0.984	1	1	1	2	1	1
RF-IG	0.992	0.962	0.998	0.998	0.974	0.97	4	4	1	1	4	4
kNN-C	0.995	0.979	0.997	0.985	0.982	0.979	2	2	6	5	2	2
kNN-IG	0.994	0.974	0.997	0.983	0.979	0.975	3	3	6	6	3	3
RF-C	0.990	0.949	0.998	0.986	0.967	0.962	5	5	1	4	5	5
RF-ReF	0.983	0.897	0.998	0.988	0.94	0.932	8	8	1	3	8	8
DT-IG	0.985	0.941	0.993	0.959	0.95	0.941	6	6	8	7	6	6
DT-C	0.983	0.932	0.992	0.956	0.944	0.934	7	7	9	8	7	7
MP-C	0.966	0.837	0.989	0.93	0.881	0.862	9	11	10	10	9	9
MP-IG	0.965	0.841	0.988	0.924	0.881	0.862	10	10	11	11	9	9
DT-ReF	0.959	0.844	0.98	0.883	0.863	0.839	11	9	12	12	11	11
SVM-C	0.949	0.774	0.98	0.876	0.822	0.794	12	14	12	13	12	12
SVM-ReF	0.870	0.152	0.998	0.931	0.261	0.346	17	18	1	9	18	18
SVM-IG	0.945	0.756	0.979	0.865	0.807	0.777	13	15	14	14	13	13
NB-C	0.935	0.822	0.955	0.767	0.793	0.756	14	12	16	16	14	14
MP-ReF	0.933	0.722	0.97	0.812	0.764	0.727	15	16	15	15	16	16
NB-IG	0.929	0.813	0.949	0.742	0.776	0.735	16	13	17	17	15	15
NB-ReF	0.855	0.54	0.911	0.52	0.53	0.444	18	17	18	18	17	17

Acc=Accuracy, Sens=Sensitivity, Spec=Specificity, Prec=Precision, FS=F-Score, MCC= Matthews correlation coefficient

The worst model used Naïve Bayes and ReliefF, with consistently lowest ranking across outcome measures. Naïve Bayes scored poorly regardless of the feature selector: NB-Ref was 18th, NB-IG was 16th (tie), and NB-C was 15th. Figure 4.2 compares performance of kNN-ReF (best model), RF-ReF (good), MP-IG (average), and NB-ReF (worst).

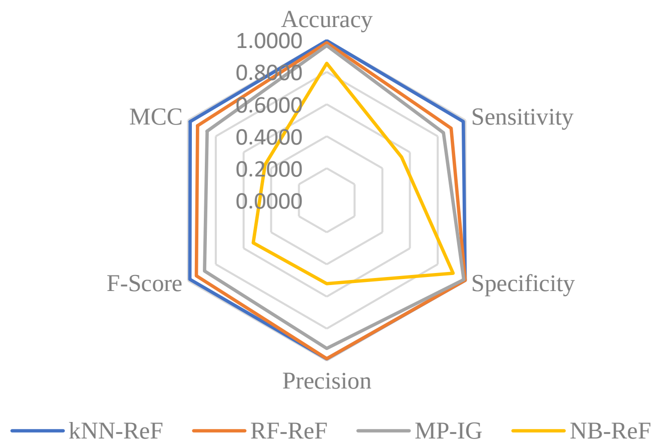


Figure 4.2: Performance of four CM-FS models

4.5. Discussion

This research demonstrated that it is possible to classify aggressive and non-aggressive motions using accelerometer and gyroscope data from smartwatches. kNN-ReF was the best combination for this binary classification. These results were comparative or better than computer vision approaches that scored between 91% and 96% [42], [75]. Therefore, the proposed smartwatch method represents a viable way of identifying aggressive movements, possibly leading to a wearable system for alerting care providers of an aggressive event and logging information to better understand the aggressive situation.

From all the evaluated combinations, the best model was kNN-ReF, which ranked 1st across all the performance metrics except precision, where it was 2nd. All kNN-ReF performance metrics were above 98.4%. Even though ReliefF was the worst ranked feature selector in general, it worked well with kNN. One explanation is that ReliefF uses inherently nearest neighbours to estimate attribute relevance, features that are selected, therefore ReliefF would be compatible with the kNN machine learning classifier. Villacampa [76] also noticed that combining ReliefF and kNN improved results in a binary classification. kNN-ReF was followed by RF-IG, kNN-C, kNN-IG, and RF-C in the group of five best models. Performance measures for these combinations were consistently in the top 5.

The worst model was NB-ReF, which had acceptable accuracy and specificity (85.46% and 91.10%, respectively) but low sensitivity (54%), precision (52%), F-score (0.53), and MCC (0.44). This would result in a high number of false positives (aggressive actions that are incorrectly classified as non-aggressive) and false negatives (non-aggressive actions that are incorrectly classified as aggressive). The five lowest ranking models were NB-ReF, NB-IG, MP-ReF, NB-C, and SVM-IG. SVM-ReF had especially low sensitivity (15.20%), F-score (0.2610) and MCC (0.346). Given these results, the Naïve Bayes and SVM classifiers in this research should not be used as machine learning tools to recognize aggressive movement.

For individual performance measures, sensitivity was high for kNN-ReF, kNN-C, and kNN-IG. High sensitivity indicates few false negatives (aggressive actions that are not detected by the classifier), so these models are ideal if the priority is to identify all aggressive movements. Precision represents false alarms (actions classified as non-aggressive that, in reality, are aggressive). SVM-ReF, NB-ReF, and MP-ReF were the best-ranked precision models. Therefore, these models are suitable if the main criterion is to minimize false positives. F-score combines sensitivity and precision, but does not consider the correctly identified non-aggressive actions. MCC is a balanced measure that takes into account all the four confusion matrix components and is very useful when there is a class imbalance. F-score and MCC ranking results were the same and displayed similarities with the general summed ranking results. kNN-ReF, kNN-C, and kNN-IG were the top 3 F-score and MCC models. These 3 classifiers are part of the summed ranking best 5. In a situation where the summed ranking cannot be applied because of limited resources on the number of performance metrics, MCC or F-score can be directly selected for model ranking.

A limitation of the approach in this research is that modelling and machine learning analysis were developed offline, meaning that the results were not obtained from a real-time system (i.e., a device that instantly notifies staff when an aggressive event occurs). Therefore, applying the selected models to a real-time platform might yield different performance metrics [77].

4.6. Conclusions

In this paper, a smart-watch based approach for identifying aggressive activity was investigated and the objective was to determine the best Classification Model and Feature Selector (CM-FS) combination for separating aggressive from non-aggressive movements. The kNN classifier and ReliefF feature selection combination provided excellent aggressive movement classification results, with all performance metrics above 98%. Using this model, alarm-based notification of aggressive movements would lead to a miss rate of only 0.2% (incorrectly classifying an aggressive action as non-aggressive) and a false alarm rate of less than 0.5% (incorrectly classifying a non-aggressive activity as aggressive). The metrics showed that this model could be used in a clinical setting to identify aggressive movements by means of smartwatches or other wrist-worn devices. Other powerful models such as RF-IG or kNN-C are suggested if the focus is on minimizing false positives or false negatives. Future research in this area should include model evaluation within a real-time system, testing with an elderly population that reflects people with dementia in healthcare settings, and a larger sample size that could provide more generalizable results.

5. Classification of Aggressive Movements with Unilateral or Bilateral Smartwatches

This chapter addresses Objective 3 by comparing bilateral watches to unilateral watches for binary classification of aggressive movements. The two previous chapters based their methodologies on two smartwatches (one smartwatch per wrist). This chapter seeks to determine if one smartwatch (placed either on the dominant wrist, or the non-dominant wrist) can be as effective as two smartwatches when classifying aggressive and non-aggressive activities.

This research was presented as a poster at the 40th IEEE Engineering in Medicine and Biology Society (EMBS).

Tchuente, F, Lemaire, ED, Baddour, N, "Classification of Aggressive Movements with Unilateral or Bilateral Smartwatches", 40th International Conference IEEE EMBS.

5.1. Abstract

Recognizing aggressive motion is a human activity recognition task that could be implemented using wearable technology, such as smartwatches. Wrist-worn wearable sensors could be on the dominant, non-dominant, or both wrists. This research explored whether unilateral or bilateral smartwatches are best for classifying aggressive motion. Participants donned two Microsoft Band 2 smartwatches and performed an activity circuit of similar aggressive and non-aggressive movements. Smartwatch accelerometer and gyroscope sensors captured data that were used to extract features. Three situations were evaluated: two smartwatches (one per wrist), dominant wrist smartwatch, and non-dominant wrist smartwatch. A Random Forest machine learning classifier coupled with three machine learning feature selectors (ReliefF, InfoGain, Correlation) was used to evaluate performance metrics from each situation. Bilateral smartwatches performed the best with 99.2% accuracy, 96% sensitivity, 99.7% sensibility, 98.5% precision, 0.972 F-score, and 0.967 Matthews Correlation Coefficient

(MCC), when Infogain feature selection was used. When only a single watch was available, the non-dominant wrist had better performance metrics than the dominant wrist. Results confirmed that wearing sensors on both wrists achieved the best classification results for aggressive and non-aggressive movements. This could be used to identify aggressive movements in healthcare facilities that host people with dementia or mental illnesses.

5.2. Introduction

Human Activity Recognition (HAR) benefits from wearable devices, which can be worn directly on the body. Examples of wearable devices include smartphones, smartglasses, smartclothes, and smartwatches [18]. To obtain the best results from wearable devices, sensor positioning and the number of sensors must be considered. Sensors can be put on different body locations depending on the type of study to be conducted [25]. Common locations include the pelvis [12], arm [78], waist [30] and wrist [39] .

It is also important to ensure a balanced number of sensors for HAR. Too many sensors are invasive, expensive, uncomfortable, and not suitable for activity recognition. Atallah et al. [79] stated that classification results tended to improve when more accelerometers are worn whereas Cleland et al. [78] did not find any significant improvement when combining more than two sensors. Therefore, using more sensors does not necessarily result in better classification results. Wang et al. [81] performed activity recognition using fewer sensors (2 sensors compared to 13 sensors), achieving comparable accuracy. Likewise, Mortazavi [82] proved that one sensor was able to produce similar accuracy metrics to multiple wearable sensors for posture tracking. The ideal situation would therefore be to minimize the number of sensors while maintaining or improving the performance metrics.

In previous work (Chapters 3 and 4), two smartwatches were used to record upper limb movements with the assumption that bilateral data would enable better classification results because two smartwatches would be more suitable to record data from any upper-limb. Classification models using the two smartwatches distinguished between aggressive and non-aggressive movements.

The goal of this chapter is to determine if a single watch (one wrist) for aggressive movement classification leads to different results from a two-smartwatch (two wrists) model. In other words, can one smartwatch yield better or comparable performance metrics than two smartwatches? The question was evaluated by comparing features extracted from bilateral smartwatches, dominant wrist smartwatch, and non-dominant wrist smartwatch. A successful classification model with only one wrist-worn wearable sensor would make this aggressive motion identification tool easier to implement in practice.

5.3. Methods

30 able-bodied and healthy participants (15 male, 15 female) were recruited from The Ottawa Hospital Rehabilitation Center (TOHRC) and the University of Ottawa. Characteristics included age (25.9±8.0), weight (70.2±11.9 kg), height (170.7± 8.6 cm), and right-handedness (20 out of 30). The study was approved by the Ottawa Health Science Network and the University of Ottawa Research Ethics Board. All participants read and signed an informed consent form.

The protocol comprised both aggressive and non-aggressive activities. Aggressive activities were punch, shove, slap, and shaking a Body Opponent Bag (Figure 5.1). Non-aggressive activities consisted of clapping hands, waving, handshaking, opening/closing a door, and typing on a computer keyboard. Similar activities (e.g., slap/clap; shove/push a door open) were chosen to challenge the classifiers and evaluate potential false positive events.

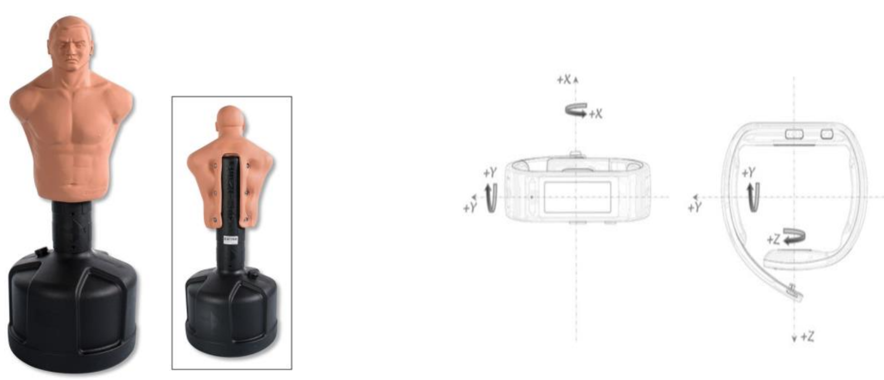


Figure 5.1: Body Opponent Bag (BOB) and Microsoft Band orientation

Participants wore one Microsoft Band 2 (MSB2) per wrist and donned a holster on their pelvis that carried a Nexus 5 smartphone. The MSB2 (Figure 5.1) recorded upper-limb motion at a frequency of 50 Hz via their integrated tri-axial accelerometer and gyroscope sensors. The Nexus 5 smartphone was connected via Bluetooth to the smartwatches using the TOHRC Data Logger [65] Android app, updated for signal acquisition from two MSB2. A second smartphone video recorded the participant's movements and served as a gold standard comparator. The gold standard time was synchronized with the smartwatch sensor output by shaking the hands at the beginning and end of the trial, providing a recognizable accelerometer signal and video event.

Three conditions were investigated: both wrists (BW), dominant wrist (DW), non-dominant wrist (NDW). Features from these three cases were extracted and three feature selection methods (ReliefF, InfoGain, and Correlation) were used to generate feature sets for classification (Table 5.1). BW was analyzed using the best 20 features (out of 136, for both wrists) (cf. Chapter 3). To maintain the same level of proportionality, DW and NDW used the 10 best features (out of 68). Features are described in Table 5.2 according to the following structure: FeatureAbbreviation_Sensor_AxisDirection_WatchNumber. The machine learning Random Forest classifier was used on the WEKA platform [12] with 5-fold cross-validation and then performance metrics were extracted.

Table 5.1: Feature descriptions per smartwatch

Features	Description	# features
Statistical	**Used to differentiate between two or more movements**	
Mean	Average of the signal	6
Variance	Variance of the signal	6
Median	Median of the signal	6
Range	Range of the signal	6
Standard Deviation	Measure of the spreadness of the signal	6
Skewness	The degree of asymmetry of the sensor signal distribution	6
Kurtosis	The degree of peakedness of the sensor signal distribution	6
Pairwise Correlation Coefficient	Correlation between two sensor axes, and between accelerometer and gyroscope sensors	6
Maximum Difference	Difference between the highest and the lowest value of over the window	6
Physical	**From the physical interpretation of the human motion**	
Mean of the Movement Intensity (mMI) Variance of the Movement intensity (vMI)	Movement Intensity (MI): Euclidean norm of the acceleration vector (accelerometer data) and velocity vector (gyroscope data) For example, considering the acceleration $MI(i) = \sqrt{a_x^2(i) + a_y^2(i) + a_z^2(i)}$ where $a_x\,(i), a_y\,(i), and\ a_z\,(i)$ represent the i_{th} acceleration data point of the x, y, and z axis in each window, respectively. Mean and variance over the window	4
Signal Magnitude Area (SMA)	The acceleration magnitude summed over three axes within each window normalized by the window length	2
Sum of All Squares	Acceleration magnitude squared and summed over three axes, normalized by window length	2

Table 5.2: Best features selected for each method

	ReliefF	Correlation	InfoGain
BW	Pcc_Gyr_yz_1	Std_Acc_x_1	Diff_Gyr_z_2
	Pcc_ Gyr xy_1	Range_Acc_x_1	Diff_Acc_x_2
	Pcc_ Gyr_ yz_2	Std_Acc_x_2	Diff_Gyr_y_1
	Pcc_ Acc_yz_1	Range_Acc_x_2	Diff_Acc_y_1
	Pcc_ Acc_ yz_2	Std_Acc_y_2	Diff_Acc_z_1
	Pcc_ Gyr_ xz_1	Range_Acc_y_2	Diff_Acc_y_2
	Pcc_ Acc_ xy_2	SMA_Acc_2	Diff_Gyr_y_2
	Pcc_Acc_xy_1	Range_Gyr_z_2	Diff_Gyr_x_1
	Pcc_ Gyr_ xz_2	AI_Acc_2	Diff_Gyr_z_1
	Pcc_ Gyr_ xy_2	AI_Acc_1	Diff_Acc_z_2
	Pcc_ Acc_ xz_1	Range_Acc_y_1	Diff_Acc_x_1
	Pcc_ Acc_ xz_2	SMA_Acc_1	Range_Acc_x_2
	Mean_Acc_y_2	Std_Acc_y_1	Range_Acc_x_1
	Area_Acc_y_2	Range_Gyr_z_1	Range_Accy_2
	Med_Accy_2	Var_Acc_x_1	Range_Acc_z_1
	Area_Acc_z_1	Diff_Acc_x_1	Diff_Gyr_x_2
	Mean_Acc_z_1	Range_Gyr_y_2	Range_Acc_y_1
	Med_Acc_z_1	Range_Accz_2	Var_Acc_x_1
	Skew_Acc_y_2	Diff_Acc_x_2	Std_Acc_x_1
	Area_Acc_y_1	Diff_Acc_y_1	Std_Acc_x_2
DW	Pcc_yz_Acc	Std_Ac_cx	Diff_Acc_y
	Pcc_xz_Gyr	Range_Acc_x	Diff_Gyr_y
	Pcc_xy_Gyr	AI_Acc	Diff_Gyr_z
	Pcc_xz_Acc	SMA_Acc	Diff_Acc_x_
	Pcc_yz_Gyr	RangeAccy	Diff_Gyr_x
	Pcc_xy_Acc	Std_Accy	Diff_Acc_z
	Area_Acc_y	Range_Gyr_z	Range_Acc_x
	Mean_Acc_y	Var_Acc_x	Range_Acc_y
	Med_Acc_y	Diff_Acc_x	Var_Acc_x
	Med_Acc_z	Range_Acc_z	Std_Acc_x
NDW	Pcc_yz_Acc	Std_Acc_x	Diff_Acc_y
	Pcc_xy_Acc	Range_Acc_x	Diff_Gyr_y
	Pcc_yz_Gyr	Std_Acc_y	Diff_Acc_z
	Pcc_xz_Acc	Range_Acc_y	Diff_Acc_x
	Pcc_xy_Gyr	SMA_Acc	Diff_Gyr_z
	Pcc_xz_Gyr	Range_Gyr_z	Range_Acc_x
	Mean_Acc_y	AI_Acc	Diff_Gyr_x
	Area_Acc_y	Range_Gyr_y	Range_Acc_y
	Med_Acc_y	Range_Acc_z	Range_Acc_z
	Skew_Acc_y	Diff_Gyr_z	Var_Acc_x

Pcc=Pairwise Correlation Coefficient, Gyr=gyroscope, Acc=acceleration, Med=Median, Skew=skewness, SMA=signal magnitude area, Diff=maximum difference, Var=variance, Std=standard deviation, 1=left wrist, 2=right wrist

5.4. Results

Evaluation metrics for the three feature selection approaches and smartwatch conditions are shown in Table 5.3. BW had the best results regardless of which feature selection method was used. The average results across the features selectors were 98.8% accuracy, 93.3% sensitivity, 99.7% specificity, 98.5% precision, 0.958 F-score, and 0.952 MCC. DW and NDW had comparable results. If the feature selector ReliefF was chosen, DW has a slight advantage on all the metrics, except for DW MCC (0.836) and NDW MCC (0.841). On the other hand, NDW had better statistics than the DW across the six performance metrics when InfoGain and Correlation methods were considered. As shown in Table 5.3, InfoGain was the best feature selector overall, with results in descending order of BW, NDW, and DW. ReliefF generally had worse results than the other feature selection techniques when using the Random Forest classifier. The differences between all methods were small.

Table 5.3: Performance metrics using ReliefF, Infogain, and Correlation methods

		Acc	**Sens**	**Spec**	**Prec**	**F**	**MCC**
ReliefF	**BW**	0.982	0.893	0.998	0.987	0.938	0.929
	DW	0.972	0.836	0.996	0.973	0.900	0.836
	NDW	0.961	0.766	0.996	0.968	0.855	0.841
Infogain	**BW**	0.992	0.960	0.997	0.985	0.972	0.967
	DW	0.986	0.925	0.996	0.978	0.951	0.943
	NDW	0.988	0.940	0.996	0.978	0.959	0.952
Correlation	**BW**	0.990	0.947	0.997	0.984	0.965	0.960
	DW	0.981	0.895	0.996	0.975	0.933	0.923
	NDW	0.984	0.911	0.997	0.980	0.944	0.936

BW=Both wrists, DW=Dominant wrist, NDW=Non-dominant wrist, Acc= Accuracy, Sens= Sensitivity, Spec= Specificity, Pres= Precision, F= F-score

5.5. Discussion

The smartwatch approach was successful at classifying aggressive and non-aggressive movements. Both unilateral and bilateral smartwatch approaches were effective, with excellent results. All the results were above 77%, regardless of selected wrist. Therefore, any approach could be used in a clinical care setting and achieve satisfactory results. However, the best results occurred with bilateral smartwatches.

BW had better results than DW and NDW across the six metrics. Differences were substantial using ReliefF feature selection (up to 13% of difference in terms of sensitivity, 8% in terms of F-score and 9% in terms of MCC between BW and NDW situations). When comparing BW and DW, the main differences appear in sensitivity (6%) and MCC (9%). The use of InfoGain and Correlation methods alleviates these differences to less than 5% across the performance metrics.

A comparison between DW and NDW showed similarities. For ReliefF, the maximum difference in sensitivity was 6.7%. InfoGain and Correlation methods showed differences of less than 2% in favor of NDW. The main difference was in sensitivity (false positives, where aggressive situations might not be detected by the algorithm). This is not a desirable trait, therefore 6.7% is quite high. A suggestion would be to replace ReliefF (displaying high discrepancies) with the InfoGain feature set for unilateral classification.

If achieving the highest performance metrics is a priority, smartwatches should be worn on both wrists. However, watches are generally not worn on both wrists, thus using two watches might be cumbersome. Two watches would collect more sensor data, hence requiring more resources for feature analysis and thus increasing processing time. Finally, it is also more expensive to equip all patients with two watches.

The minor differences between the use of two watches (BW) and one watch (NDW or DW) (with Infogain) would lead to the adoption of the unilateral condition. As mentioned in the results, NDW had a slight advantage over DW in terms of outcome measures, therefore NDW is the preferred configuration. These results are convenient because the non-dominant hand is the most commonly worn way for watches. With the objective to minimize false positives, NDW leads to 98.8% accuracy, 94% sensitivity, 99.6% specificity, 97.8% precision, 0.959 F-score and 0.952 MCC, which are very reasonable for a binary classification. There is nevertheless a slight trade-off between performance and convenience for the user and the installation team.

To maintain proportionality between features, 10 features were considered for conditions where one smartwatch was considered (DW or NDW), whereas 20 features were used in the BW condition. Increasing the number of features (DW or NDW) would break the proportionality implemented in this study but could lead to an improvement in unilateral smartwatch performance metrics.

5.6. Conclusion

Using one smartwatch per wrist provided excellent classification results, with performance metrics exceeding 77% when adopting a Random Forest classifier and the InfoGain feature selection method. This choice minimized false positives and false negatives that can easily occur in machine learning classification. Despite the impressive performance analytics, adopting two smartwatches requires more resources and processing time because the input features are doubled. To reduce these financial, computational, and practical issues, one smartwatch on the non-dominant wrist can be considered when implementing an aggressive movement identification application.

6. Multinomial Classification of Aggressive Movements using Smartwatches

This chapter addresses Objective 4 by determining the best machine learning classifier for multinomial activities. Previous chapters focused on the binary nature of activities (aggressive or non-aggressive). More insight is provided here on the activities, which are classified in ten different classes: punch, shove, slap, shake, transitions, open/close a door, clap, wave, handshake, and type on a keyboard.

6.1. Abstract

Human Activity Recognition (HAR) involves visual or sensory pattern recognition that leads to interpreting and labeling a specific activity. Wearable devices such as smart bands and smartwatches facilitate activity recognition by providing multiple motion and biological signals as input for movement models. In this research, participants donned two Microsoft Band 2 (MSB2) smartwatches and performed an activity circuit of ten similar aggressive and non-aggressive movements, to investigate aggressive movement recognition from machine learning models. A multinomial classification (resulting in three or more movement categories) will be the main approach to determine the correct movement nature.

136 time-domain features were extracted from MSB2 accelerometer and gyroscope data. These features fed five classifiers (kNN, Random Forest, Decisions Tree, Neural Network, Naïve Bayes) that were compared using a summed ranking method to determine the best algorithm for multinomial classification of punching, shoving, slapping, shake, transitions, opening/closing a door, clap, wave, handshake, and typing on a keyboard.

kNN performed the best, with 99.5% accuracy, 99.5% sensitivity, 99.9% specificity, 99.5% precision, 0.995 F-score, and 0.994 MCC. Random Forest and Decision Tree, with slightly lower metrics, may also constitute viable options for aggressive movement recognition.

The results support this machine learning approach for aggressive movement classification that could be used in high surveillance areas, prisons, mental illness facilities, elderly care centres for people with dementia, or other accommodations that host people with potentially aggressive motion.

6.2. Introduction

Clinical settings and healthcare facilities can host elderly people with bipolar disorder, schizophrenia, or dementia. Such mental illnesses may increase the likelihood of violent incidents [83]. A tool that identifies the onset of an aggressive incident, such as automatic activity recognition, could enable healthcare providers to appropriately intervene. As well, avoiding unnecessary interventions when an individual's movements are not aggressive would make better use of time and resources.

With progress in machine learning technologies, human activity recognition has been successfully applied in a broad range of sectors. For example, smartwatches and smartphones are increasingly used to classify simple and complex activities, from trivial daily movements (walking, standing, driving) to more sophisticated movements (cooking, watering plants, washing hands) with accuracies exceeding 90% [32]. Activities such as sitting, walking upstairs, and doing dishes have been helpful for evaluating elderly mobility routines [84]. Other applications have also been evaluated. For example, Lee and Song [37] used a smartwatch for activity classification for children with developmental disabilities, with average accuracies up to 95% for throwing objects, clenching a fist, or raising the arm.

Activity classification can be binary (two movement categories) or multinomial (three or more classes) [39],[64],[78]. In previous research, we used smartwatch accelerometer and gyroscope features and six machine learning classifiers for binary classification of aggressive or non-aggressive movements (Chapter 4). The binary classification provided high-level insight towards the aggressive (or non-aggressive) nature of the activities. However, binomial classification does not provide the specific activity that was labelled as aggressive. A multinomial approach would provide more information and specifics on the classified activities, which is beneficial for helping care providers determine an appropriate intervention plan or for researchers to better understand classification errors between similar aggressive movement and non-aggressive movements (e.g., slapping or clapping).

This research investigated the effectiveness of smartwatch accelerometer and gyroscope signal features for identifying aggressive movements that might occur in a healthcare or eldercare institution. The aim was to determine the most appropriate machine learning model for correctly classifying an aggressive movement, while avoiding misclassification from similar non-aggressive movements. A successful model would provide better information to care providers on the type of aggressive action, beyond the information provided by a simple binary alarm system that only identifies aggressive versus non-aggressive status.

6.3. Methods

6.3.1. Data Collection and Extraction

Thirty able-bodied and healthy participants (15 male, 15 female) were recruited from The Ottawa Hospital Rehabilitation Center (TOHRC) and the University of Ottawa. Characteristics included age (25.9±8.0), weight (70.2±11.9 kg), height (170.7± 8.6 cm), and right-handedness (20 out of 30). The study was approved by the Ottawa Health Science Network and the University of Ottawa Research Ethics Board. All participants provided informed consent.

The protocol consisted of ten aggressive and non-aggressive activities (Table 6.1). Aggressive actions were performed on a Body Opponent Bag (BOB, Figure 6.1). Similar activities (e.g., slap, clap; shove, push a door open) were chosen to challenge the classifiers and evaluate potential false positive events (Table 6.1).

Figure 6.1: Body Opponent Bag (BOB)

Table 6.1: Activities

Movement	Activity	Description
Aggressive	Punch	Participant punches BOB eight times, alternating hands
	Shove	Participant aggressively shoves BOB five times with both hands
	Slap	Participant aggressively slaps BOB ten times with both hands
	Shake	Participant holds BOB's neck and shakes BOB back and forth five times
Non-Aggressive	Transitions	Set of movements between an aggressive action and non-aggressive action (i.e., sitting, standing, moving, still)
	Open/Close door	Participant opens and closes the door three times
	Clap	Participant claps their hands ten times
	Wave	Participant waves with the preferred hand as if they are saying goodbye
	Handshake	Participant handshakes the project assistant
	Type on a keyboard	Participant types the first verse of the Canadian National anthem

Participants wore one Microsoft Band 2 (MSB2) per wrist and donned a holster on their pelvis that carried a Nexus 5 smartphone. The MSB2 recorded upper-limb motion using the integrated tri-axial accelerometer and gyroscope. A Nexus 5 smartphone was connected via Bluetooth to the smartwatches using a modified TOHRC Data Logger Android app [65], updated for signal acquisition from two MSB2. A second smartphone video recorded the participant's movements and serve as a gold standard comparator. The video time was synchronized with smartwatch sensor output by shaking the hands at the beginning and end of the trial, providing a recognizable accelerometer signal and video event.

Raw data were extracted from the two MSB2 sensors at 50 Hz. The continuous raw data were divided into 1-second sliding windows (50 data points) for feature extraction. Windows overlapped by 96% of the window length (window advanced by 2 data points). This resulted in 122,307 instances of ten multinomial activities (Figure 6.2).

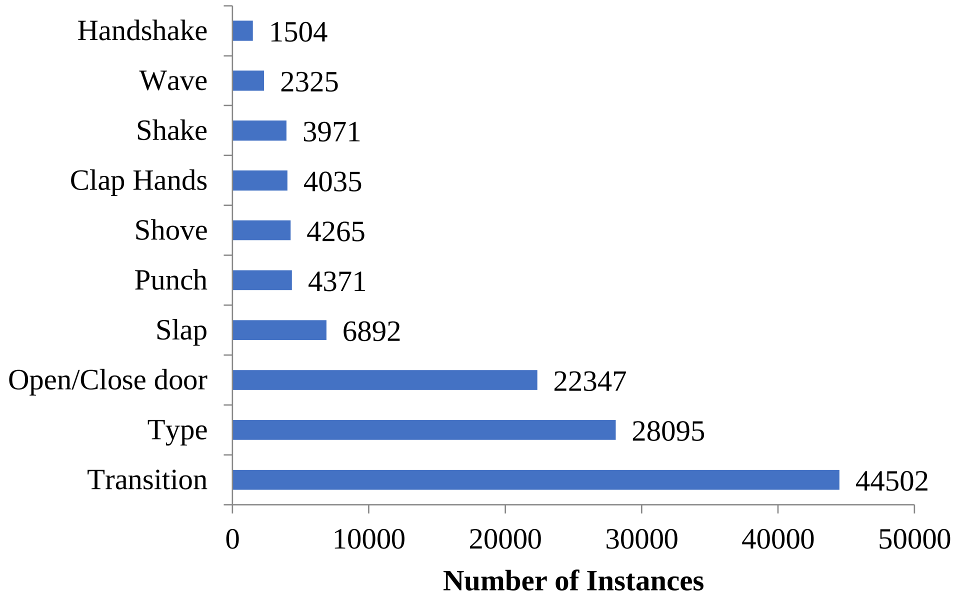


Figure 6.2: Activity instances

6.3.2. Classification and Evaluation

136 time-domain features from the accelerometer and gyroscope smartwatch sensors were evaluated to determine the best classification results (Table 6.2). Five common machine learning classifiers were implemented using the Waikato Environment for Knowledge Analysis (WEKA) platform [74] and evaluated to determine the best model for distinguishing the different movements. Classifiers included k Nearest Neighbours (kNN), Multilayer Perceptron neural network (MP), Naïve Bayes (NB), J48 Decision Tree (DT), and Random Forest (RF).

Evaluation metrics included accuracy, sensitivity, specificity, precision, F-score, and Matthews Correlation Coefficient (MCC) [85], [86]. The classifiers were evaluated using a summed ranking method [60], where each performance metric was ranked for each classifier and then the ranks for all six metrics were summed to provide an overall ranking for each model. Accuracy tends to be the most common metric for evaluating classification results but is not completely reliable when class imbalances occur, which happens frequently in Human Activity Recognition applications. The ranking method does not focus on only one metric, but involves results from all six parameters. A better, wider, and more generalizable representation of model performance is therefore provided.

Table 6.2: Description of features from accelerometer and gyroscope signals, from the two smartwatches

Features	**Description**	**Number of features**
Statistical features		
Mean	Average for each of 3 axes	12
Variance	Variance for each of 3 axes	12
Median	Median for each of 3 axes	12
Range	Range for each of 3 axes	12
Standard Deviation	Standard deviation each of for 3 axes	12
Skewness	Degree of asymmetry for each of 3 axes	12
Kurtosis	Degree of peakedness for each of 3 axes	12
Pairwise Correlation Coefficient	Correlations between sensor axes (x,y; x,z; y,z)	12
Integral	Area under the curve for 3 axes	12
Maximum Difference or derivative	Difference between the highest and the lowest value of the acceleration	12
Physical Features		
Mean of the Movement Intensity (mMI) Variance of the Movement intensity (vMI)	Movement Intensity (MI): Euclidean norm of the acceleration vector (accelerometer data) and velocity vector (gyroscope data) For example, considering the acceleration $MI(i) = \sqrt{a_x^2(i) + a_y^2(i) + a_z^2(i)}$ where $a_x\ (i), a_y\ (i), and\ a_z\ (i)$ represent the i_{th} acceleration data point of the x, y, and z axis in each window, respectively. The mean and variance over the window are extracted from MI	8
Signal Magnitude Area (SMA)	Acceleration magnitude summed over three axes, normalized by window length $SMA = \sum \lvert a_x(i) \rvert + \sum \lvert a_y(i) \rvert + \sum \lvert a_z(i) \rvert$	4
Sum of All squares	Acceleration magnitude squared and summed over three axes, normalized by window length $SaS = \sum a_x^2(i) + \sum a_y^2(i) + \sum a_z^2(i)$	4

6.4. Results

kNN was the best classifier, with 99.5% accuracy, 99.5% sensitivity, 99.8% specificity, 99.5% precision, 0.995 F-score, and 0.994 MCC (Table 6.3). Compared to the other algorithms, kNN was ranked first on every evaluation metric and followed by Random Forest, which was also a viable classifier for this application with results all above 99%. The Naive Bayes classifier did not perform well and therefore was not appropriate for this type of classification.

Table 6.3: Classifier evaluation metrics with classifiers sorted by summed rank (best to worse)

	kNN	**RF**	**DT**	**MP**	**NB**
Accuracy	0.995	0.992	0.961	0.883	0.554
Sensitivity	0.995	0.992	0.961	0.883	0.554
Specificity	0.998	0.997	0.989	0.962	0.909
Precision	0.995	0.992	0.961	0.883	0.685
F-score	0.995	0.992	0.961	0.882	0.501
MCC	0.994	0.990	0.950	0.82	0.477

The kNN confusion matrix (Table 6.4) shows the classified activity classes compared to the true classes. False positives and negatives (off-diagonal numbers) were very small compared to the total number of instances.

Table 6.4: kNN confusion matrix. 0: Transition, 1: Punch 2: Shove, 3: Slap, 4: Shake, 5: Open/close door, 6: Clap hands, 7: Wave, 8: Handshake, 9: Type

		Classified as									
	Class	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
True Class	**0**	**44230**	30	32	33	28	30	31	22	31	35
	1	44	**4326**	1	0	0	0	0	0	0	0
	2	26	1	**4231**	0	5	0	1	0	0	1
	3	35	0	0	**6854**	1	0	0	0	0	2
	4	33	1	4	0	**3932**	0	0	0	0	1
	5	36	0	0	0	0	**22311**	0	0	0	0
	6	32	0	0	0	1	0	**3998**	0	0	4
	7	33	0	0	0	0	0	0	**2292**	0	0
	8	27	0	0	0	0	0	0	0	**1477**	0
	9	34	0	0	7	0	0	0	0	0	**28054**

Performance metrics relative to each activity are shown in Table 6.5. Open/Close door and Type were ranked the best (i.e., easiest activities to classify), whereas wave and handshake had lower numbers, especially in terms of sensitivity and F-score.

Table 6.5: Multinomial performance metrics for kNN

	Sensitivity	**Specificity**	**Precision**	**F-Score**	**MCC**
Transition	0.994	0.996	0.993	0.994	0.99
Punch	0.990	0.999	0.993	0.991	0.991
Shove	0.992	0.999	0.991	0.992	0.991
Slap	0.994	0.999	0.994	0.994	0.994
Shake	0.990	0.998	0.991	0.991	0.990
Open/Close door	0.998	0.999	0.999	0.999	0.998
Clap	0.991	0.999	0.992	0.991	0.991
Wave	0.986	0.999	0.990	0.988	0.988
Handshake	0.982	0.999	0.979	0.981	0.981
Type	0.999	0.999	0.998	0.999	0.998
Weighted Average	0.995	0.998	0.995	0.995	0.994

6.5. Discussion

The activity classification results show that smartwatches constitute a viable and effective approach for multinomial aggressive and non-aggressive movement classification. The kNN algorithm successfully classified multiple aggressive and non-aggressive categories, potentially leading to a wearable system that would alert care providers of a specific aggressive event and log information to better understand the aggressive situation. This logged information would provide insight on the person's predominant aggressive motion. Knowing the type of aggressive activity may lead to a better care plan tailored to the person's activity.

kNN had the best performance metrics of all the classifiers, with all metrics exceeding 99.4%. A deeper analysis of the best classifier suggests that all the movements, whether aggressive or non-aggressive, were well classified (Table 6.5). From punching to typing on a computer, few false positive and false negatives occurred among the movements.

The maximum number of confused instances was 7, for shaking and typing. This result was surprising given the different nature and lack of similarities between the two movements. On the other hand, 5 false occurrences occurred between shaking and shoving. Both movements were considered aggressive and may have been confounded by participants during the experiment.

More activities were anticipated to be mismatched by the classifier, since the experiment was designed with similar aggressive and non-aggressive activities to test the machine learning classifier strength. Thus, it was expected that movements such as clap and slap, shove and open/close door would have yielded a higher number of false positives and negatives. However, kNN was excellent at correctly identifying and effectively differentiating these activities.

Despite the excellent kNN performance for target movements, transition movements represented an exception. The transition category was confused with all other activity classes of activities at least 22 times. Reciprocally, all the classes of activities were misidentified as transitions up to 44 times. Transitions consist of a variety of non-classified movements such as walking around, sitting, staying still, and random movements (e.g., scratching head, adjusting sleeves) that happened during the experiment and may be confused with other activities.

Several studies in the literature showed that some activities are difficult to distinguish and are typically misclassified. For example, walking and climbing stairs (participants often classified as climbing stairs when walking) [12], movements ranging from walking to cycling (maximum accuracy of 77%) [35], and complex activities (cooking, sweeping, taking medication; 50% accuracy) [32]. Despite the misclassification occurrences, misclassified transition actions represented a tiny percentage (less than 0.1%) of the total activities instances. Therefore, they were not sufficient to affect the kNN overall performance. Performance indicators were consistently high across the different activities.

Comparable results were obtained in a previous study with a binary classification between aggressive and non-aggressive movements was conducted (Chapter 4), where kNN and a set of 20 features yielded 99.6% accuracy, 98.4% sensitivity, 99.8% specificity, 98.9% precision, 0.987 F-score and a 0.984 MCC. Besides this study, kNN was also the preferred classifier for Amancio [87] who compared supervised learning classifiers in HAR with Logistic Regression, SVM, CART, and Neural Networks. Despite the excellent performance metrics with offline analysis, research is needed to prove that the same results will occur when implemented as a real-time system that provides instant notification. Gillian argued that kNN could sometimes be slow for gesture recognition in real-time if there were a large number of training examples [88]. In this case, kNN computing time would be a trade-off to its performance indicators.

Random Forest and Decision Tree classifiers could be considered when classifying aggressive and non-aggressive activities, if kNN is not adaptable for a real-time system. Their performance metrics were between 95% and 99%, which are acceptable in the Human Activity Recognition field. Random Forests have already been used in various real-time systems, have been proven to be more efficient in real-time applications, and might have an advantage for online classification [89],[67].

6.5.1 Limitations

This study had limitations for video-segmentation and the population cohort. For manual video-segmentation of the gold-standard comparator, identifying one frame to represent the start or end of a transition state was typically difficult. Interpretation differences of several frames would be identified as errors, but these false positives or false negatives would be video interpretation variability instead of poor classifier performance. In a real-time system, a threshold could be implemented to handle this transition variability.

Participants were young adults. Since an elderly population may perform the aggressive movements differently (slower and less intensity), future research is needed to test the model with older adults.

6.6. Conclusion

The best machine learning classifier for multinomial aggressive classification was kNN, which outperformed all others for accuracy, sensitivity, specificity, precision, F-score and MCC. kNN represents an effective classifier, which only misclassified very few instances of the ten activities, whether they were aggressive or not. Results show less than 0.1% of false positives and false negatives for each activity; it is therefore thus a great choice for a multinomial classification.

Smartwatch accelerometer and gyroscope signals were effective for correctly identifying the target activities with excellent performance metrics. Random Forest and Decision Tree machine learning alternatives can be successfully applied to a real-time classification that would provide instant notification at the onset of an aggressive situation. Future research should address how kNN works in a real-time aggressive activity environment and evaluation with a sufficiently large sample of elderly participants.

7. Thesis Conclusions and Future work

This thesis provided evidence that aggressive movements can be distinguished from non-aggressive movements using accelerometer and gyroscope smartwatch sensor signals. The best feature sets were identified, as well as the best machine learning classifier. The thesis objectives and the corresponding questions are discussed below:

7.1. Objective 1: Determine the best set of smartwatch sensor features to distinguish aggressive from non-aggressive motion

Question 1: How will the selected features perform?

The performance metric target to prove that smartwatch sensors can effectively distinguish aggressive from non-aggressive movements was 75%, which was satisfactory for this proof of concept study. Four feature selection sets (ReliefF, InfoGain, Chi-Square, and Correlation) were evaluated with a Random Forest classifier and resulted in an average 99% accuracy, 94.6% sensitivity, 99.8% specificity, and 98.6% precision, 0.965 F-score and 0.96 MCC. All the results were substantially higher than the 75% threshold. Smartwatch sensors are therefore viable for HAR aggressive classification.

Question 2: Will different selection methods output similar features?

Different feature selection methods produced different feature sets. Four feature selection methods (ReliefF, InfoGain, Chi-Squared, Correlation) were considered to determine the twenty best high-performance features. ReliefF prioritized features such as the pairwise correlations between the three axes, mean accelerations in the three directions, and area under the curve attribute. However, the ReliefF feature set was the worst set when tested with a Random Forest classifier.

The correlation method had three predominant features based on accelerometer and gyroscope signals: derivative, range, and standard deviation. Correlation also incorporated physical features such as the Signal Magnitude Area and Movement Intensity, features that were not selected by the other methods.

InfoGain and Chisquared provided the same feature set; the maximum difference and the range of the acceleration were ranked at the top. They produced the best feature set out of the four feature selection methods.

7.2. Objective 2: Determine the best machine learning classifier and feature selection model

Question: What is the best classification model for recognizing aggressive from non-aggressive motion?

Random Forest, which was selected in Chapter 3, was not the overall best classifier to distinguish aggressive from non-aggressive activities. kNN-ReF (kNN and ReliefF) was the best model combination and ranked first across almost all the performance metrics with 99.6% accuracy, 98.4% sensitivity, 99.8% specificity, 98.9% precision, 0.987 F-score, and 0.984 MCC. The top four models of this study are dominated by kNN, which would be considered the best classifier.

Nevertheless, the Random Forest classifier was second to best and still produced excellent performance metrics. Random Forest combined with Infogain had 99.2% accuracy, 96.2% sensitivity, 99.8% specificity, 99.8% precision, 0.974 F-score, and 0.970 MCC.

7.3. Objective 3: Determine the differences between bilateral smartwatches and unilateral smartwatches

Question 1: Do bilateral watches yield better evaluation metrics than unilateral watches?

Bilateral watches outperformed unilateral watches. It was important to compare two wrists (BW) with one wrist (DW or NDW) wearable devices to determine if less resources can be attributed in practice. All performance metrics were above 77%, regardless of the selected wrist; however, BW had better results than DW and NDW across the six metrics, regardless of the feature selection method. The average results across the features selectors were 98.8% accuracy, 93.3% sensitivity, 99.7% specificity, 98.5% precision, 0.958 F-score, and 0.952 MCC.

Differences were substantial using ReliefF feature selection (up to 13% difference between BW and NDW for sensitivity, 8% for F-score, 9% for MCC). When comparing BW and DW, the main differences appear in sensitivity (6%) and MCC (9%). The use of InfoGain and Correlation methods alleviated these differences to less than 5% across the performance metrics.

Question 2: Does the dominant wrist configuration gives better results than non-dominant?

NDW and DW had very similar results; however, NDW had a slight advantage. NDW had better results than DW across the six performance metrics when the best feature selection methods (InfoGain or Chi-Squared) were considered. InfoGain and Correlation methods had differences of less than 2% in favor NDW.

NDW is the preferred configuration, which is convenient because the non-dominant hand is the most natural location when wearing a watch. With the objective to minimize false positives, NDH lead to 98.8% accuracy, 94% sensitivity, 99.6% specificity, 97.8% precision, 0.959 F-score and 0.952 MCC, which are very reasonable for a binary classification.

7.4 Objective 4: Determine the machine-learning classifier for a multinomial aggressive classification

Question 1: Does multinomial classification perform worse than binary classification?

Chapter 4 compared classifiers across the six performance metrics on a binary classification, whereas Chapter 6 provided insight on a multinomial classification. Inconsistent with the hypothesis, generally, the multinomial classification performed slightly better than the binary classifier when using their top classifier, kNN. However, binary classification relied on 20 features for classification whereas the multinomial classification considered the whole set of 136 features.

Question 2: Which movements will be confused as false positives and negatives?

The maximum number of confused instances was 7, for shaking and typing. This result was surprising given the different nature and lack of similarities between the two movements. On the other hand, 5 false occurrences happened between shaking and shoving. Both movements were considered aggressive and may have been confounded by participants during the experimentation phase. However, more activities were expected to be mismatched by the classifier than was actually found, since the experiment was designed with similar aggressive and non-aggressive activities to test the machine learning classifier strength. Thus, movements such as clap and slap, shove and open/close door could have yielded a higher number of false positives and negatives. This was not the case since less than 0.1% of the movements were confused, with only a few instances between shoving and shaking. The main confused instances happened with the most dominant class, transitions.

7.5. Future Work

This research provided evidence for aggressive movement classification using smartwatch sensors. Improvements could be made on certain aspects.

1. The study only considered able-bodied participants for the proof of concept. However, a target market is the elderly, who may be affected by mental disorders such as dementia. Elderly aggressive movements could be slower and less intense. Classifier performance should be determined for this population. Elderly people with dementia who adopt perseverative activity and rip off their bandages, might do the same when wearing watches.
2. Only 30 participants were recruited for the experiments. This represents a sufficient number but a larger sample combined with techniques different than cross-validation (i.e., Leave-one-out cross-validation or hold-out) might be considered. The results should be more generalizable.
3. The thesis only relied on the smartwatch accelerometer and gyroscope sensors. Other sensors such as the galvanic skin response or heart rate might provide additional input to the aggressive nature of a movement. Additionally, smartphone accelerometer and

gyroscope sensors could also be considered for the classification. Future research can address if a combination of smartwatches and smartphones sensors provide better results than only smartwatches.

4. The thesis is only based on four upper-limb aggressive movements. More movements can be evaluated. For example, it is unknown if smartwatches can detect a lower-limb aggressive movement such as a kick.

5. A smartwatch real-time system that alerts the staff in case of an aggressive motion can be developed (at the onset of an aggressive movement). Research is needed to determine if the recommended classifiers in this thesis will perform well in a real-time system.

References

[1] S. C. Government of Canada, "Alzheimer's disease and other dementias in Canada," 18-May-2016.[Online].Available:https://www.statcan.gc.ca/pub/82-003-x/2016005/article/14613-eng.htm. [Accessed: 02-Apr-2018].

[2] T. Majić, J. P. Pluta, T. Mell, Y. Treusch, H. Gutzmann, and M. A. Rapp, "Correlates of agitation and depression in nursing home residents with dementia," *Int. Psychogeriatr.*, vol. 24, no. 11, pp. 1779–1789, Nov. 2012.

[3] R. J. Miller, J. Snowdon, and R. Vaughan, "The Use of the Cohen-Mansfield Agitation Inventory in the Assessment of Behavioral Disorders in Nursing Homes," *J. Am. Geriatr. Soc.*, vol. 43, no. 5, pp. 546–549, 1995.

[4] B. Chikhaoui, B. Ye, and A. Mihailidis, "Feature-level combination of skeleton joints and body parts for accurate aggressive and agitated behavior recognition," *J. Ambient Intell. Humaniz. Comput.*, vol. 8, no. 6, pp. 957–976, Nov. 2017.

[5] Y. Chen and C. Shen, "Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition," *IEEE Access*, vol. 5, pp. 3095–3110, 2017.

[6] M. A. I. Anik, M. Hassan, H. Mahmud, and M. K. Hasan, "Activity recognition of a badminton game through accelerometer and gyroscope," in *2016 19th International Conference on Computer and Information Technology (ICCIT)*, 2016, pp. 213–217.

[7] M. Lapinski, E. Berkson, T. Gill, M. Reinold, and J. A. Paradiso, "A Distributed Wearable, Wireless Sensor System for Evaluating Professional Baseball Pitchers and Batters," in *2009 International Symposium on Wearable Computers*, 2009, pp. 131–138.

[8] H. Ghasemzadeh, V. Loseu, & R. Jafari. Wearable coach for sport training: A quantitative model to evaluate wrist-rotation in golf. *J. Amb. Intel. and Smart En.*, 1(2), 2009, 173-184.

[9] T. Schlömer, B. Poppinga, N. Henze, & S. Boll. Gesture recognition with a Wii controller. In Proceedings of the 2nd international conference on Tangible and embedded interaction (pp. 11-14). ACM, 2008.

[10] O. P. Popoola and K. Wang, "Video-Based Abnormal Human Behavior Recognition; A Review," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 6, pp. 865–878, Nov. 2012.

[11] X. Li and Z. Cai, “Anomaly detection techniques in surveillance videos,” in *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2016, pp. 54–59.

[12] N. A. Capela, “Improving a Smartphone Wearable Mobility Monitoring System with Feature Selection and Transition Recognition,” thesis, Université d’Ottawa / University of Ottawa, 2015.

[13] J. Howcroft, J. Kofman, and E. Lemaire, “Prospective Fall-Risk Prediction Models for Older Adults based on Wearable Sensors,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. PP, no. 99, pp. 1–1, 2017.

[14] S. Ranasinghe, F. Al Machot, and H. C. Mayr, “A review on applications of activity recognition systems with regard to performance and evaluation,” *Int. J. Distrib. Sens. Netw.*, vol. 12, no. 8, p. 1550147716665520, Aug. 2016.

[15] B. Florentino-Liaño, N. O’Mahony, and A. Artés-Rodríguez, “Human activity recognition using inertial sensors with invariance to sensor orientation,” in *2012 3rd International Workshop on Cognitive Information Processing (CIP)*, 2012, pp. 1–6.

[16] Y. Chen and C. Shen, “Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition,” *IEEE Access*, vol. 5, pp. 3095–3110, 2017.

[17] M. D. Tundo, “Development of a Human Activity Recognition System Using Inertial Measurement Unit Sensors on a Smartphone,” Thesis, Université d’Ottawa / University of Ottawa, 2014.

[18] B. Hartmann, *Human Worker Activity Recognition in Industrial Environments*. KIT Scientific Publishing, 2011.

[19] M. M. Sardsehmukh, M. T. Kolte, P. N. Chatur, and D. S. Chaudhari, “3-D dataset for Human Activity Recognition in video surveillance,” in *2014 IEEE Global Conference on Wireless Computing Networking (GCWCN)*, 2014, pp. 75–78.

[20] D. Meldrum, C. Shouldice, K. Jones, and M. Forward, “A study to investigate the reliability of computerised analysis of human gait using VICON motion analysis system,” *Gait Posture*, vol. 36, Supplement 1, p. S80, Jun. 2012.

[21] "UCI Machine Learning Repository: Vicon Physical Action Data Set Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Vicon+Physical+Action+Data+Set. [Accessed: 23-May-2017].

[22] "IoT Research – Smartbands," *Securelist - Kaspersky Lab's cyberthreat research and reports*.

[23] "Moving away from the wrist - smart clothing you'll actually want to wear." [Online]. Available: https://gadgetsandwearables.com/2018/04/14/best-smart-clothing/. [Accessed: 28-Aug-2018].

[24] "MUSICAL GLOVE," *SUNNIE SANG*. [Online]. Available: http://www.sunnie-sang.com/musical-glove/. [Accessed: 28-Aug-2018].

[25] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *2016 IEEE Sensors*, 2016, pp. 1–3.

[26] Y. Lee, M. Song. Using a Smartwatch to Detect Stereotyped Movements in Children With Developmental Disabilities. *IEEE Access*. 2017;5:5506-14.

[27] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Commun. Surv. Tutor.*, vol. 15, no. 3, pp. 1192–1209, Third 2013.

[28] Daniel Roggen, "Wearable Computing - Part II: Sensors," 06:21:38 UTC.

[29] R. Akhavian and A. Behzadan, "Wearable sensor-based activity recognition for data-driven simulation of construction workers' activities," in *2015 Winter Simulation Conference (WSC)*, 2015, pp. 3333–3344.

[30] C. Shen, Y. Chen, and G. Yang, "On motion-sensor behavior analysis for human-activity recognition via smartphones," in *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2016, pp. 1–6.

[31] S.-M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using Convolutional Neural Network," in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017, pp. 131–134.

[32] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, “Simple and Complex Activity Recognition through Smart Phones,” in *2012 Eighth International Conference on Intelligent Environments*, 2012, pp. 214–221.

[33] L. Atallah, B. Lo, R. King, and G. Z. Yang, “Sensor Placement for Activity Detection Using Wearable Accelerometers,” in *2010 International Conference on Body Sensor Networks*, 2010, pp. 24–29.

[34] S.-M. Lee, S. M. Yoon, and H. Cho, “Human activity recognition from accelerometer data using Convolutional Neural Network,” in *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2017, pp. 131–134.

[35] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, “How Accurately Can Your Wrist Device Recognize Daily Activities and Detect Falls?,” *Sensors*, vol. 16, no. 6, p. 800, Jun. 2016.

[36] L. Ardüser, P. Bissig, P. Brandes, and R. Wattenhofer, “Recognizing text using motion data from a smartwatch,” in *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, 2016, pp. 1–6.

[37] Y. Lee and M. Song, “Using a Smartwatch to Detect Stereotyped Movements in Children with Developmental Disabilities,” *IEEE Access*, vol. 5, no. 99, pp. 1–1, 2017.

[38] X. Liu, L. Liu, S. J. Simske, and J. Liu, “Human Daily Activity Recognition for Healthcare Using Wearable and Visual Sensing Data,” in *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, 2016, pp. 24–31.

[39] F. Nurwanto, I. Ardiyanto, and S. Wibirama, “Light sport exercise detection based on smartwatch and smartphone using k-Nearest Neighbor and Dynamic Time Warping algorithm,” in *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2016, pp. 1–5.

[40] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber, “Smartwatch-based activity recognition: A machine learning approach,” in *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2016, pp. 426–429.

[41] B. Mortazavi *et al.*, “Can Smartwatches Replace Smartphones for Posture Tracking?,” *Sensors*, vol. 15, no. 10, pp. 26783–26800, Oct. 2015.

[42] A. Ouanane and A. Serir, “New paradigm for recognition of aggressive human behavior based on bag-of-features and skeleton graph,” in *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*, 2013, pp. 133–138.

[43] D. W. Koh and H. B. Kang, “Smartphone-based modeling and detection of aggressiveness reactions in senior drivers,” in *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 12–17.

[44] Y. Saez, A. Baldominos, and P. Isasi, “A Comparison Study of Classifier Algorithms for Cross-Person Physical Activity Recognition,” MDPI, 30-Dec-2016. [Online]. Available: http://www.mdpi.com/1424-8220/17/1/66.

[45] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, “Window Size Impact in Human Activity Recognition,” *Sensors*, vol. 14, no. 4, pp. 6474–6499, Apr. 2014.

[46] Wilde, G. Adriana, "An overview of human activity detection technologies for pervasive systems." Department of Informatics University of Fribourg, Switzerland 212 (2010): 72-112.

[47] J. O. Laguna, A. G. Olaya, and D. Borrajo, “A Dynamic Sliding Window Approach for Activity Recognition,” SpringerLink, 11-Jul-2011. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-22362-4_19.

[48] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, “Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1553–1567, Oct. 2006.

[49] “A feature selection-based framework for human activity recognition using wearable multimodal sensors.” [Online]. Available: http://dl.acm.org/citation.cfm?id=2318798. [Accessed: 13-Jul-2017].

[50] M. Zhang and A. A. Sawchuk, “A Feature Selection-based Framework for Human Activity Recognition Using Wearable Multimodal Sensors,” in *Proceedings of the 6th International Conference on Body Area Networks*, ICST, Brussels, Belgium, Belgium, 2011, pp. 92–98.

[51] Duda, O. Richard, E. Hart, and D.G. Stork. Pattern classification. John Wiley & Sons, 2012.

[52] “Nonnegative Matrix Factorization - MATLAB & Simulink.” [Online]. Available: https://www.mathworks.com/help/stats/nonnegative-matrix-factorization.html. [Accessed: 23-May-2017].

[53] C. Sammut and G. I. Webb, Eds., *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US, 2016.

[54] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, H. Liu, Advancing feature selection research. ASU feature selection repository. 2010:1-28.Z. Zhao and F. Morstatter, “Advancing Feature Selection Research,” p. 28.

[55] Z. M. Hira and D. F. Gillies, “A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data,” *Advances in Bioinformatics*, 2015. [Online]. Available: https://www.hindawi.com/journals/abi/2015/198363/. [Accessed: 13-Jul-2017].

[56] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2016.

[57] Edmond and Mitchell, “A machine learning framework for automatic human activity classification from wearable sensors,” *DORAS - DCU*, 01-Mar-2015. [Online]. Available: http://doras.dcu.ie/20336/.

[58] Fernández-Delgado, Manuel, et al. "Do we need hundreds of classifiers to solve real world classification problems." *J. Mach. Learn*. Res 15.1 (2014): 3133-3181.

[59] “Which Supervised Learning Method Works Best for What? An Empirical Comparison of Learning Methods and Metrics.” [Online]. Available: http://videolectures.net/solomon_caruana_wslmw/. [Accessed: 13-Jul-2017].

[60] D. John, “Evaluation of Accelerometer-Based Walking-Turn Features for Fall-Risk Assessment in Older Adults,” UWSpace - Waterloo's institutional repository, 17-May-2017. [Online]. Available: https://uwspace.uwaterloo.ca/handle/10012/11914.

[61] M. Lapinski, E. Berkson, T. Gill, M. Reinold, and J. A. Paradiso, “A Distributed Wearable, Wireless Sensor System for Evaluating Professional Baseball Pitchers and Batters,” in *2009 International Symposium on Wearable Computers*, 2009, pp. 131–138.

[62] J. Howcroft, J. Kofman, and E. D. Lemaire, “Prospective Fall-Risk Prediction Models for Older Adults Based on Wearable Sensors,” *IEEE T. Neur. Sys. Rehab. Eng.*, vol. 25, no. 10, pp. 1812–1820, 2017

[63] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, “Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey,” in *23th International Conference on Architecture of Computing Systems 2010*, 2010, pp. 1–10.

[64] H.O. Hessen, A. Tessem. “Human Activity Recognition with Two Body-Worn Accelerometer Sensors” (Master's thesis, NTNU), 2016.

[65] “TOHRC Data Logger. Google Play” [Online]. Available: https://play.google.com/store/apps/details?id=ca.irrd.datalogger. [Accessed: 03-Aug-2017].

[66] N. Capela, E.D. Lemaire, and N. Baddour, “Feature selection for wearable smartphone-based human activity recognition with able bodied, elderly, and stroke patients,” *PLoS ONE*, vol. 10, no. 4, p. e0124414.

[67] S. O., R. Mallipeddi, and M. Lee, “Real Time Hand Gesture Recognition Using Random Forest and Linear Discriminant Analysis,” in *Proceedings of the 3rd International Conference on Human-Agent Interaction*, New York, NY, USA, 2015, pp. 279–282.

[68] H. Mazaar, E. Emary, and H. Onsi. "Regression-Based Feature Selection on Large Scale Human Activity Recognition." *Int. J. Adv. Comput. Sci Applic.* Vol. 7, no. 2, pp. 668-674, 2016.

[69] J. P. Gupta, N. Singh, P. Dixit, V. B. Semwal, and S. R. Dubey, “Human Activity Recognition Using Gait Pattern,” *Int. J. Comput. Vis. Image Process. (IJCVIP)*, 01-Jul-2013. [Online]. Available: https://www.igi-global.com/article/human-activity-recognition-using-gait-pattern/95969.

[70] A. M. Sabatini, C. Martelloni, S. Scapellato, and F. Cavallo, “Assessment of walking features from foot inertial sensing,” *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 486–494, Mar. 2005.

[71] D. Leightley, J. Darby, B. Li, J. S. McPhee, and M. H. Yap, “Human Activity Recognition for Physical Rehabilitation,” in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 261–266.

[72] J. Lockman, R. S. Fisher, and D. M. Olson, “Detection of seizure-like movements using a wrist accelerometer,” *Epilepsy Behav.*, vol. 20, no. 4, pp. 638–641, Apr. 2011.

[73] “A machine learning framework for automatic human activity classification from wearable sensors - Aficionado.” [Online]. Available: http://ei.aficionado.ie/publications/36710. [Accessed: 13-Feb-2018].

[74] “Weka 3 - Data Mining with Open Source Machine Learning Software in Java.” [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/. [Accessed: 07-Feb-2018].

[75] E.B. Nievas, O.D. Suarez, G.B. García, R. Sukthankar. “Violence detection in video using computer vision techniques”. International conference on Computer analysis of images and patterns 2011 Aug 29 (pp. 332-339). Springer, Berlin, Heidelberg.

[76] O. Villacampa, “Feature Selection and Classification Methods for Decision Making: A Comparative Analysis,” *CEC Theses Diss.*, Jan. 2015.

[77] J. Suto, S. Oniga, C. Lung, and I. Orha, “Recognition rate difference between real-time and offline human activity recognition,” in *2017 International Conference on Internet of Things for the Global Community (IoTGC)*, 2017, pp. 1–6.

[78] R. Akhavian and A. Behzadan, “Wearable sensor-based activity recognition for data-driven simulation of construction workers’ activities,” in *2015 Winter Simulation Conference (WSC)*, 2015, pp. 3333–3344.

[79] L. Atallah, B. Lo, R. King, and G. Z. Yang, “Sensor Positioning for Activity Recognition Using Wearable Accelerometers,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 4, pp. 320–329, Aug. 2011.

[80] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean, D. Finlay. Optimal placement of accelerometers for the detection of everyday activities. *Sensors*. vol. 13, no. 7, pp. 9183-9200, Jul 2013.

[81] Robust human activity recognition using lesser number of wearable sensors - IEEE Conference Publication. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8304292/.

[82] B. Mortazavi, E. Nemati, K. VanderWall, H.G. Flores-Rodriguez, J.Y. Cai, J. Lucier, A. Naeim, M. Sarrafzadeh. Can smartwatches replace smartphones for posture tracking? *Sensors*. vol. 15, no. 10, pp. 26783-26800, Oct 2015.

[83] M. E. Rueve and R. S. Welton, “Violence and Mental Illness,” *Psychiatry Edgmont*, vol. 5, no. 5, pp. 34–48, May 2008.

[84] N. A. Capela, E. D. Lemaire, and N. Baddour, “Improving classification of sit, stand, and lie in a smartphone human activity recognition system,” in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, 2015, pp. 473–478.

[85] M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation,” in *AI 2006: Advances in Artificial Intelligence*, 2006, pp. 1015–1021.

[86] S. Boughorbel, F. Jarray, and M. El-Anbari, “Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric,” *PLOS ONE*, vol. 12, no. 6, p. e0177678, Jun. 2017.

[87] D. R. Amancio *et al.*, “A Systematic Comparison of Supervised Classifiers,” *PLOS ONE*, vol. 9, no. 4, p. e94137, Apr. 2014.

[88] “KNN — NickGillianWiki.” [Online]. Available: http://www.nickgillian.com/wiki/pmwiki.php/GRT/KNN. [Accessed: 31-Mar-2018].

[89] X. Zhao, Z. Song, J. Guo, Y. Zhao, and F. Zheng, “Real-Time Hand Gesture Detection and Recognition by Random Forest,” in *Communications and Information Processing*, Springer, Berlin, Heidelberg, 2012, pp. 747–755.

[90] A. Z. Lein, “Microsoft Band 2 Review: a massive hardware improvement,” *Pocketnow*, 29-Oct-2015.

Appendix A: List of possible features

Time domain Features

Features	Description
Statistical features	Used to differentiate between two or more movements
Mean	Average of the signal
Median	Median of the signal
Maximum	Maximum of the signal
Minimum	Minimum of the signal
Range	Range of the signal
Standard Deviation	Measure how spread the signal is
Root Mean Square	The quadratic mean value of the signal
Averaged Derivatives	The mean value of the first order derivatives of the signal (in this case, this corresponds to jerk)
Skewness	The degree of asymmetry of the sensor signal distribution
Kurtosis	The degree of "peakedness" of the sensor signal distribution
Interquartile Range	Measure of the statistical dispersion, being equal to the difference between the 75th and the 25th percentiles of the signal
Mean Crossing Rate	The total number of times the signal changes from below average to above average or vice versa normalized by the window length
Pearson Correlation	Correlation between two sensor axes, and between accelerometer and gyroscope sensors
Physical Features	From the physical interpretation of the human motion
Mean of sum of square (all 3 acceleration axes)	Average movement Intensity (MI): The Euclidean norm of the total acceleration vector after removing the static gravitational acceleration, where $a_x\ (i), a_y\ (i), and\ a_z\ (i)$ represent the t_{th} acceleration sample of the x, y, and z axis in each window, respectively. $$MI(t) = \sqrt{a_x^2(i) + a_y^2(i) + a_z^2(i)}$$
Stdev of sum of squares (all 3 acceleration axes)	Standard deviation of $(a_x^2(i) + a_y^2(i) + a_z^2(i))$
Signal Magnitude Area (SMA)	The acceleration magnitude summed over three axes within each window normalized by the window length
Eigenvalues of Dominant Directions (EVA)	The eigenvectors of the covariance matrix correspond to the dominant directions along which intensive human motion occurs The eigenvalues measure the corresponding relative motion magnitude along the directions.

Impulse	Sum of absolute values x delta time
Number of Peaks	Number of maxima peaks
Sum of ranges (SoR)	$SoR = range_x + range_y + range_z$
Sum of standard deviation (SStD)	$SStD = SD_x + SD_y + SD_z$
Simple moving average	Sum of maximum accelerations for current windows, before, and after
Gravity range Difference	Sum of range of X and Z gravity components
Averaged Rotation Angles related to Gravity Direction	It calculates the cumulative rotation angles around gravity direction. The cumulative sum is then divided by the window length. This feature captures the rotation movement of the upper limb around gravity direction.

Frequency Domain Features

Fast Fourier Transform (FFT) Quartile: Percentage of acceleration frequencies in the first quartile (i.e., frequencies ≤ 12.5 Hz) of an FFT frequency plot for vertical, AP, and ML axes.
Averaged acceleration energy
Spectral entropy measure
The frequency range power
Abscise of the first frequency peak
The dominant frequency over the window
Median frequency
Averaged velocity along heading direction
Correlation between acceleration along gravity and heading directions
Averaged velocity along gravity direction
Averaged rotation angles related to gravity direction
Averaged acceleration energy
Averaged rotation energy
Dominant frequency

Appendix B: Experimental Protocol

A project assistant will review the protocol with each participant at the start of the test session and then ask the participant to complete a consent form. The project assistant will record the person's age, sex, weight, height, and handedness.

Before testing, the person will don a Microsoft Band smart watch on each wrist. The watch screen will be positioned on the front of the wrist (the side of the wrist that is down when the palm of the hand is down).

The project assistant will start the Data Logger application on a smartphone and ensure that the two watches are connected to the phone via Bluetooth (i.e., two green checkmarks appear on the application screen). The project assistant will slide the button on the Data Logger screen to the right to start recording and then place the smartphone in a belt-holster at the posterior pelvis. The participant will stand still and then shake their wrists two times, to provide a synchronization movement between video and sensor data. At the end of the trial, the project assistant will swipe the Data Logger button to the right to stop data collection and save the data on the smartphone.

For every trial, the person will be videoed using a separate smartphone. The smartphone video will be synchronized with the data recorded on the phone by the first accurately detected "shake". Digital video will be used to validate the detected movements and provide contextual information.

After the wrist shaking activity, each person will be instructed to complete a circuit of activities:

1. From a standing position, shake both hands to indicate the start of the trial.
2. Continue standing for at least 10 seconds. This standing phase can be used for phone orientation calibration.

3. Walk towards the Body Opponent Bag, stand still for 2-3 seconds, then aggressively punch the bag four times, alternating hands for each punch.
4. Walk to a chair and sit down, Clap your hands ten times as if you were cheering at a game
5. Stand and then wave one hand, as saying goodbye. Person can use their preferred hand (i.e., no instruction as to which hand to wave)
6. Stand up, walk to the BOB, stand in front of the BOB for 2-3 seconds, and aggressively shove the BOB with both hands. Repeat five times.
7. Handshake with the project assistant
8. Walk to the laboratory door, open, and walk through. Project assistant closes the door without latching. Push the door open and walk through. Repeat 5 times.
9. Walk to the BOB. Instruct the person to "Aggressively slap the head 5 times with each hand, alternating hands with each slap". A total of 10 slaps will be performed by the participant.
10. Walk towards a computer, sit down, and type the first verse of the Canadian national anthem
11. Walk towards the BOB, stand still for 2-3 seconds, shake the BOB violently back and forth
12. Standing still for 10 seconds.
13. Shake both hands to indicate the end of the trial.

Appendix C: Methods and Equipment

Data logger and Microsoft Band 2 (MSB2)

The TOHRC data logger is an app used to capture sensors data from a phone. The app was originally meant for Blackberry platforms [65]. However, the data logger has been adapted to the Android platform. The original code was updated to incorporate two MSB2 smartwatches signal sensors. Sensors from the smart bands (such as the X, Y, Z accelerometers and the gyroscopes or the galvanic skin response) were added. A Bluetooth connection was established between an Android Nexus 5 phone and the smart watches to facilitate the data transfer.

Figure C.1 shows the differences between the two data logger versions (Picture of the Data Logger with Band 1 connected for example).

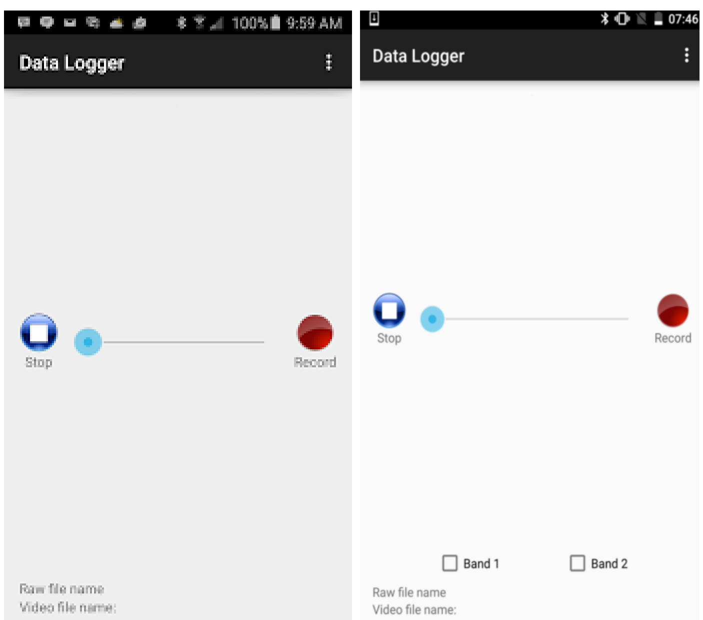


Figure C.1: Data Logger Version 1 and Version 2

The Microsoft Band 2 (Microsoft second-generation smartwatch) is the wearable device used in this thesis. The health and fitness band was announced in October 2015. MSB2 is compatible to Android, Apple and Windows phone. The band incorporates multiple sensors (Figure C.2) that gather user physical data.

The body of the band is made up of metal and surround a curved gorilla glass screen. The wrist band is made up of a thick strip of rubber that will be around the wrist. The device is made up of silicon and curved glass as well as a variety of sensors. The Band 2 has a 32 mm x 12.8 mm, 320 x 128 pixel curved AMOLED screen with a lithium-polymer battery that could last about 48 hours [90].

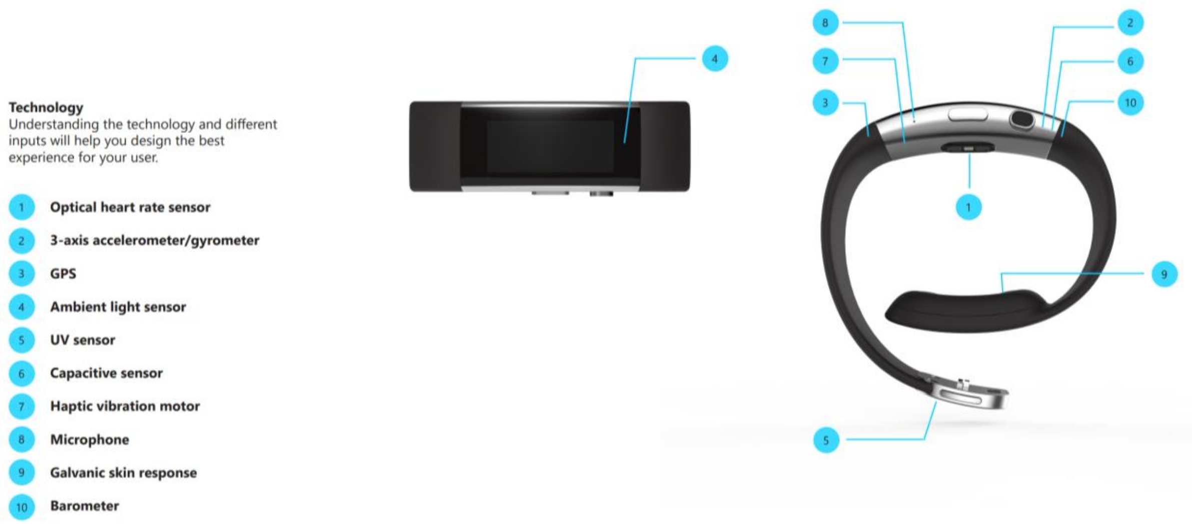


Figure C.2: Microsoft Band 2 sensors

The experiments involved the use of two MSB2 (one for each wrist). This thesis focuses on the accelerometer and gyroscope sensors. Figures C.2 and C.3 display the accelerometer and gyroscope axes.

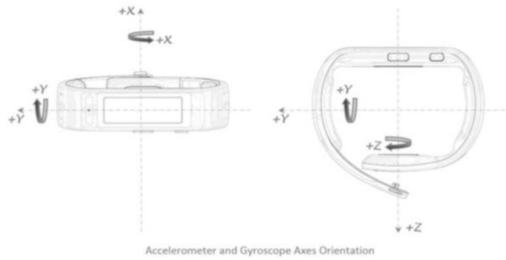


Figure C.3: X, Y and Z axes for linear and angular accelerations

The Body Opponent Bag (BOB)

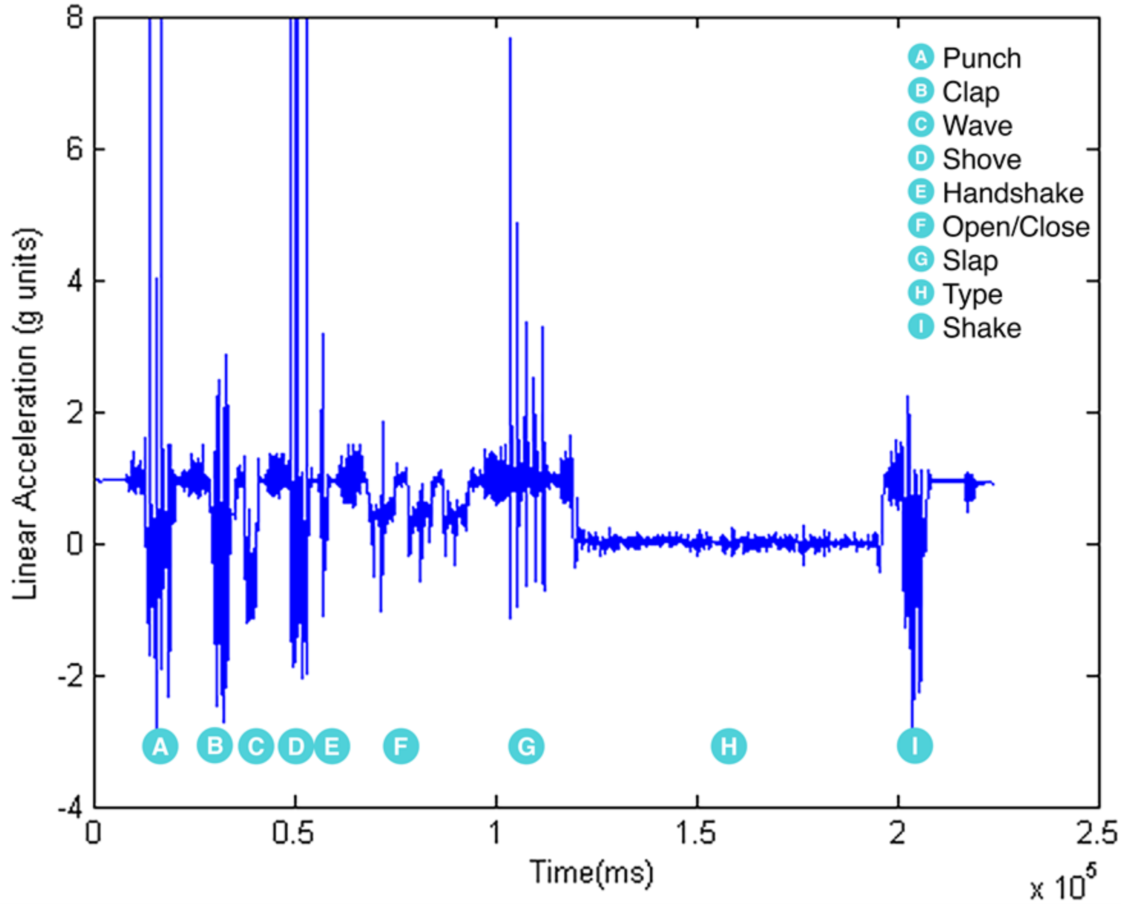
BOB is realistic, humanlike, and height adjustable equipment on which aggressive movements were performed. The covering is high strength Plastisol and an inner cavity is filled with tough urethane foam. The relatively soft covering avoids participants to get hurt when acting aggressive and the fact that BOB moves around its base has a positive output:

- All the energy from the aggressive motion is absorbed by the equipment and not released on the participants hand (as it would happen with a wall made of concrete)
- BOB mimics more the human being activity: a person who is hit will not stay still but move around

The base was filled with water.

Figure C.4: Body Opponent Bag (BOB)

Accelerometer linear acceleration over time



Appendix D: Ottawa Health Science Network Research Ethics Board Approval

Ottawa Hospital Research Institute
Institut de recherche de l'Hôpital d'Ottawa

uOttawa

UNIVERSITY OF OTTAWA HEART INSTITUTE
INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA

Ottawa Health Science Network Research Ethics Board/ Conseil d'éthique de la recherche du Réseau de science de la santé d'Ottawa

Civic Box 411 725 Parkdale Avenue, Ottawa, Ontario K1Y 4E9 613-798-5555 ext. 14902 Fax : 613-761-4311
http://www.ohri.ca/ohsn-reb

August 10, 2017

Dr. Edward Lemaire
[redacted]

Dear Dr. Lemaire:

Re: Protocol # 20170573-01H **Recognition and Classification of Aggressive Movements using a Smar Watch**

Protocol approval valid until - October 10, 2017

Thank you for the e-mail from Franck Fabrice Tchuente Kemdjo, received August 04, 2017 confirming that French translation has been requested for the English Recruitment Email Scrip and the English Participant Informed Consent Form.

This protocol underwent delegated review by the Ottawa Health Science Network Research Ethics Board (OHSN REB). You have met the requirements of the OHSN-REB and your protocol has been granted approval by the OHSN-REB for two months to begin recruitment of English speaking participants. No changes, amendments or addenda may be made to the protocol or the consent form without the OHSN-REB's review and approval.

Please note that the OHSN-REB approval is conditional upon the University of Ottawa REB approval.

Approval is for the following:
- Study Protocol, version dated March 21, 2017
- Participant Data Sheet, no version date, uploaded July 27, 2017
- English Recruitment Email Script, no version date, uploaded July 27, 2017
- English Participant Informed Consent Form, version dated July 27, 2017

Upon receipt and review of the French version of the Recruitment Email and the Informed Consent Form, ethical approval will be extended to August 09, 2018 and the recruitment of French speaking patient participants may begin.

OHSN-REB complies with the membership requirements and operates in compliance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans; the International Conference on Harmonization - Good Clinical Practice: Consolidated Guideline; and the provisions of the Personal Health Information Protection Act 2004.

Yours sincerely,

[redacted]

Raphael Saginur, M.D.
Chairperson
Ottawa Health Science Network Research Ethics Board
/hl

Appendix E: The University of Ottawa Health Sciences and Science Research Ethics Board Approval

Université d'Ottawa University of Ottawa

Bureau d'éthique et d'intégrité de la recherche Office of Research Ethics and Integrity

LETTRE D'APPROBATION ADMINISTRATIVE | LETTER OF ADMINISTRATIVE APPROVAL

Numéro de dossier / Ethics File Number	A08-17-03
Titre du projet / Project Title	Recognition and Classification of Aggressive Movements using a Smart Watch
Type de projet / Project Type	Professor's research
CÉR primaire / Primary REB	OHSN-REB
Statut du projet / Project Status	Approbation administrative / Administrative Approval
Date d'approbation (jj/mm/aaaa) / **Approval Date** (dd/mm/yyyy)	29/08/2017
Date d'expiration (jj/mm/aaaa) / **Expiry Date** (dd/mm/yyyy)	10/10/2017

Équipe de recherche / Research Team

Chercheur / Researcher	***Affiliation***	***Role***
Edward Lemaire	École de médecine / School of Medicine	Principal Investigator
Franck Fabrice Tchuente Kemdjo	École de médecine / School of Medicine	Co-Investigator

Conditions spéciales ou commentaires / Special conditions or comments:

L'Université d'Ottawa a signé une Entente, conforme aux exigences de la plus récente version de l'EPTC et tout autre règlement ou législation applicable, permettant au CÉR ci-haut nommé d'être désigné comme CÉR primaire pour les projets de recherche où

1) les activités principales de recherche sont menées sous l'autorité ou sous les auspices de l'établissement lié au CÉR primaire et
2) Une partie du projet est également réalisé sous l'autorité ou sous les auspices de l'Université d'Ottawa.

Cette lettre confirme que l'Université d'Ottawa a autorisé que le CÉR primaire soit le CÉR officiel pour l'évaluation et la supervision de ce projet de recherche. Ceci n'est pas une approbation éthique.

Afin de nous aider à garder votre dossier à jour, veuillez soumettre une copie de toutes demandes de modification, renouvellement d'approbation éthique etc. soumis à et approuvé par le CÉR primaire dès qu'elles sont disponibles.

Cette approbation administrative est valide pour la durée indiquée ci-haut et est sujette aux conditions énumérées dans la section intitulée « Conditions spéciales ou commentaires ».

The University of Ottawa has signed an Agreement, compliant with current TCPS guidelines and any other applicable guidelines or legislation regarding multisite review, allowing the REB named above to serve as Board of Record (BoR) for research projects where

1) the main research activities are conducted within the auspices or jurisdiction of the BoR's institution and
2) parts of the project are also conducted under the jurisdiction or auspices of the University of Ottawa.

This letter confirms that the University of Ottawa has authorized the REB named above to serve as Board of Record for the review and oversight of this research project. This is not an REB approval.

In order to help us keep your file up to date, please submit a copy of all amendment requests, project renewals or any other changes submitted to and approved by the BoR, as they become available.

Administrative approval is valid for the period indicated above and is subject to the conditions listed in the section entitled "Special conditions or comments".

[redacted]

Catherine Paquet
Directrice/Director

Appendix F: Recruitment Notice

The Ottawa Hospital | L'Hôpital d'Ottawa

Recognition and Classification of Aggressive Movements Using a Smart Watch

Dear Ms./Sir,

I would like to make you aware of a new research project that is being done at The Ottawa Hospital Rehabilitation Centre (TOHRC). It involves developing a smartwatch app that can recognize aggressive movements. Our project team is looking for participants to help with this study.

As a participant in this study, you would perform an activity circuit that includes gentle and aggressive actions (e.g. shake hands, punch a boxing bag, clap), with a Microsoft Band smartwatch on both wrists and a smartphone on a belt. These devices will record sensor data while you complete the circuit. During the test, a member of the research team will use a second smartphone to record video of you. This video will be used by the project team to relate your actions to the smartwatch sensor data. Testing will take approximately 25 minutes.

This research has been approved by both the Ottawa Health Science Network and University of Ottawa Research Ethics Boards. If you decide to participate, you will be free to withdraw from the study at any time without any negative consequences and without it affecting any of your present or future relationships with your health care providers or employers at The Ottawa Hospital or TOHRC.

For any questions, more information, or to arrange for an appointment please contact Franck Tchuente [redacted] or me at [redacted], at your convenience.

Thank you.

Edward Lemaire, PhD

Appendix G: Consent form

PARTICIPANT INFORMED CONSENT FORM

Title of Study: Recognition and Classification of Aggressive Movements using a Smartwatch

Principal Investigator (PI): Dr. Edward Lemaire [redacted]

Funding Agency: Natural Science and Engineering Research Council of Canada (NSERC)

Participation in this study is voluntary. Please read this Participant Informed Consent Form carefully before you decide if you would like to participate. Ask the principal investigator and study team as many questions as you like. We encourage you to discuss your options with family, friends or your healthcare team.

Why am I being given this form?

You are being asked to participate in this research study to help us develop a method for identifying aggressive movements from smartwatch sensor data.

Why is this study being done?

Aggressive behaviour can occur in clinical and elderly care settings with people suffering from dementia, mental disorders, or other conditions that affect behaviour. Since identifying the nature of the event can be difficult with people who have memory and communication issues, other methods to identify and record aggressive behaviour would be useful for care providers that need to determine the best methods to reduce or contain aggressive events.

Smartwatches are wearable devices that can be used to record arm movements and therefore could be used to determine if a particular motion is intended to be violent. This study will create and test a smartwatch app for human activity recording and aggressive movement detection. This approach could be used to identify if the person wearing the watch initiated the aggressive behaviour.

What is expected of me?

After having any questions answered by the project team and signing this consent form, a project assistant will write down your age, sex, weight, height, and handedness. You will be asked to wear a Microsoft Band smartwatch on both wrists. You will also be asked to wear a smartphone on a belt at your lower back.

You will follow a circuit where you shake hands, punch a boxing bag, clap, wave goodbye, shove the boxing bag, push and pull a door, slap the boxing bag, type on a keyboard and shake the boxing bag.

Smartwatch sensor data will be recorded for the entire circuit. During the test, a member of the research team will record video of you with a second smartphone. This video will be used by the project team to relate your actions to the smartwatch sensor data.

Will my samples or research data be used in future research?

Sensor data from the smartwatch and smartphone may be used by other engineering researchers at TOHRC and other students for research on recognizing movements with wearable technology. This sensor data does not include personal information that would identify you. Video will not be used for future research since your actions will have already been recorded in the current study.

How long will I be involved in the study?

Your test session will be approximately 25 minutes, including the time to complete the consent form, wear the watch, and complete the circuit.

What are the potential risks I may experience?

You may experience discomfort in your hand or wrist after punching or slapping the boxing bag. The boxing bag has a foam cover and moves when hit, that reduces injury risk from hitting a hard surface.

Can I expect to benefit from participating in this research study?

You may not receive any direct benefit from participating in this study.

Do I have to participate?

Your participation in this study is voluntary. You may decide not to be in this study, or to be in the study now, and then change your mind later without affecting the medical care, education, or other services to which you are entitled or are presently receiving at this institution, or your relationship with employers at The Ottawa Hospital or at the University of Ottawa.

If I agree now, can I change my mind and withdraw later?

You may withdraw from the study at any time. Any information collected specifically for the research study will be destroyed in the event that you withdraw from the study.

What compensation will I receive if I am injured or become ill in this study?

In the event of a study-related injury or illness, you will be provided with appropriate medical treatment and care. You are not waiving any of your legal rights by agreeing to participate in this study. The Ottawa Hospital Rehabilitation Centre still has their legal and professional responsibilities.

How is my personal information being protected?

- All personal health information (PHI) and your personal identifying information (PII), such as your name and age will be kept confidential.
- Release of your PHI/PII information will only be allowed if it is legally required.
- As a participant, you will be assigned a coded study number that will be used throughout the study on all your study records.
- A Master List provides the link between your identifying information and the coded study number. This list will only be available to Dr. Lemaire and his staff and will not leave this site.
- The Master List and coded study records will be stored securely in a password protected computer and locked file cabinet.
- For audit purposes only, your research records may be reviewed under the supervision of Dr. Lemaire staff by representatives from:
  - the Ottawa Health Science Network Research Ethics Board (OHSN-REB) and
  - the Ottawa Hospital Research Institute
  - The University of Ottawa
- You will not be identified in any publications or presentations resulting from this study.
- Research records will be kept for 10 years, as required by the OHSN-REB.
- At the end of the storage time, all paper records will be shredded and all electronic records will be securely deleted.

Do the investigators have any conflicts of interest?

There are no conflicts of interest to declare related to this study. The Principal Investigator is receiving financial payment from the funding source to cover the cost of conducting this study.

What are my responsibilities as a study participant?

It is important to remember the following things during this study:

- Ask the project assistant if you have any questions or concerns.

Who do I contact if I have any further questions?

If you have any questions about this study, or if you feel that you have experienced a study-related injury or illness, please contact Dr. Edward Lemaire at 613-737-7350 x75592 or elemaire@ohri.ca.

The Ottawa Health Science Network Research Ethics Board (OHSN-REB) has reviewed this protocol. The Board considers the ethical aspects of all research studies involving human participants at the Ottawa Hospital Rehabilitation Centre. If you have any questions about your rights as a study participant, you may contact the Chairperson at 613-798-5555, extension 16719.

The Ottawa Hospital | L'Hôpital d'Ottawa
RESEARCH INSTITUTE | INSTITUT DE RECHERCHE

uOttawa

UNIVERSITY OF OTTAWA HEART INSTITUTE
INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA

Recognition and Classification of Aggressive Movements using a Smartwatch

Consent to Participate in Research

- I understand that I am being asked to participate in a research study about smartwatch detection of aggressive movements.
- This study was explained to me by a project assistant.
- I have read, or someone has read to me, each page of this Participant Informed Consent Form.
- All of my questions have been answered to my satisfaction.
- If I decide later that I would like to withdraw my participation and/or consent from the study, I can do so at any time.
- I voluntarily agree to participate in this study.
- I will be given a copy of this signed Participant Informed Consent Form.

Participant's Printed Name	Participant's Signature	Date

Investigator or Delegate Statement

I have carefully explained the study to the study participant. To the best of my knowledge, the participant understands the nature, demands, risks and benefits involved in taking part in this study.

Investigator/Delegate's Printed Name	Investigator/Delegate's Signature	Date