Using Machine Learning to Infer Affect from Touch in the context of Fabric Handling via Wearable Sensors

Candidate Number: QMKR4 [1]

Programme: MSc Data Science and Machine Learning

Supervisors: Nadia Berthouze and Temitayo Olugbade

Submission date: September 2022

[1]**Disclaimer:** This report is submitted as part requirement for the MSc Data Science and Machine Learning at UCL. It is substantially the result of my own work except where explicitly indicated in the text. The report may be freely copied and distributed provided the source is explicitly acknowledged.

The code for this project is included in the following public repository: `https://github.com/nswarawita/Final-Project`

Abstract

The fashion industry is one of the most significant contributors to the current climate crisis and is responsible for approximately 10% of greenhouse gas emissions and roughly 20% of all wastewater. The rise of fast fashion has led to consumers impulsively purchasing and ultimately discarding clothes at an alarming rate. This is primarily due to consumers impulsively purchasing clothes without reflecting on if they are likely to enjoy the material and design. This project proposes a novel machine learning method to estimate people's sensations and liking of the textile being touched. This project uses machine learning to infer affect from touch in the context of fabric handling using Electromyography (EMG), accelerometer and quaternion data collected via wearable sensors. Previously conducted studies that infer tactile perception using wearable sensors in the context of fabric handling have only used statistical techniques for feature selection and model building. This project aims to build a two-part novel machine learning model, the Property Rating Hierarchical Model (PRHM), to predict the assessed property and the corresponding property rating. The best PRHM yielded a macro F1 score of 0.41 and overall classification accuracy of 45% when predicting properties and had a weighted F1 score of 0.73 and overall classification accuracy of 78.6% when predicting ratings.

Contents

Chapter 1

Introduction

1.1 Motivation

The fashion industry is one the most significant contributors to the current climate crisis. Fashion is responsible for approximately 10% of greenhouse gas emissions and roughly 20% of all wastewater worldwide [1, 2, 3]. In addition to its disastrous environmental impact, the fast fashion industry has substantial implications on an individual's overall well-being and finances. The fashion industry has recently leveraged digital technology to create a more sustainable and circular economy. Instances such as virtual clothing try-ons and digital clothing for virtual presence aim to lessen a consumer's one-time use.

This project's findings significantly contribute to the end goal of building a smart chatbot that learns how people touch and engage with clothes. The chatbot aims to transform clothes shopping into a multi-sensory, reflective and recognition-based experience for the consumer, which will drive them to optimise their purchases and thereby reduce wasteful purchasing, resulting in a more circular and sustainable fashion industry.

This project supports the digital initiative of developing digital technology to create a more sustainable and circular economy by building a novel machine learning model to infer affect [1] from touch in the context of fabric handling. The data input to the model (Electromyography (EMG), accelerometer and quaternion) was collected from both hands (left and right) via wearable sensors. The model was built to achieve two research goals. The model should first detect the property an individual is assessing and then accurately

[1] The terms affect and emotion are widely used synonymously in the context of affective computing [4]

predict the corresponding property rating. The novel machine learning method proposed in this project is the Property Rating Hierarchical Model (PRHM), a two-part model that predicts the assessed property and the corresponding property rating.

1.1.1 Fast Fashion and its Environmental Impact

'Fast fashion' refers to speedily and mass-produced, low-quality clothing that is quickly circulated through high street stores to satisfy the latest trends and maximise consumer demand [5]. Fast fashion garments are cheaply produced and priced and replicate the latest celebrity or catwalk styles [1, 6]. Fast fashion involves the rapid design, production, distribution, and marketing [1]. This allows retailers to obtain a larger quantity of assorted designs, and in turn, consumers are presented with a more extensive choice of inexpensive clothing [1].

As fast fashion relies on cheap and quick production, it promotes overproduction [6]. Fast fashion also encourages overconsumption because consumers are attracted to cheap and trendy clothing that copies current trends compared to relatively expensive, long-lasting items that fall out of style shortly [5, 6]. This toxic system of constantly buying clothing and almost immediately discarding them due to its low quality is the most significant pitfall of fast fashion [5, 6]. As a result, fast fashion massively harms the environment [6].

The environmental impact of fast fashion includes large-scale emissions of greenhouse gases, the use of massive amounts of water and energy, and the depletion of non-renewable resources. Fast fashion is also one of the world's largest polluters.

According to the United Nations Environment Programme (UNEP), textile production accounts for up to 10% of total global carbon emissions (1.2 billion tonnes) [1, 2]. This figure is larger than the emissions from all international flights and maritime shipping combined. Further, according to the United Nations (UN) Framework Convention on Climate Change, global carbon emissions are estimated to skyrocket by more than 60% by 2030 [7].

The fashion industry is also the second largest consumer of the world's water supply [1, 8]. Approximately 700 gallons (3182.26 litres) of water is required to produce a single cotton shirt; this is enough water for an individual to drink at least eight cups per day for three-and-a-half years [1, 8]. Roughly 2 000 gallons (9092.18 litres) of water is used to produce a pair of jeans; this is more than enough for one person to drink eight cups per day for ten years. Further, the United Nations Environment Programme (UNEP) discovered that the fashion industry produces 20% of the world's wastewater and that fabric dyeing is the second largest water polluter because the water leftover from the dyeing process is dumped into rivers, streams, and other water bodies [3]. This severely affects marine life and the aquatic ecosystem.

Synthetic materials such as nylon, polyester and acrylic are created from fossil fuels and currently comprise over two-thirds of the materials used in the apparel industry [1, 9, 10]. Such materials take over hundreds of years to biodegrade [1]. A 2017 report from the International Union for the Conservation of Nature (IUCN) estimated that 35% of all microplastics in the ocean came from laundering synthetic textiles like polyester [1, 8, 10]. It is also estimated that microplastics cause up to 31% of plastic pollution in the ocean [8, 10].

There are also massive amounts of monetary and resource wastage resulting from fast fashion. According to Business Insider, 85% of textiles of all textiles go to the landfills each year – this is enough to fill the Sydney harbour annually [8, 11, 12]. Further, the equivalent of one garbage truck full of clothes is dumped in a landfill or burned every second [8, 13].

Consumers are drawn to fast fashion due to its affordability, accessibility and variety. Consumers impulsively buy fast fashion clothing and do not spend time engaging with and reflecting on the garment they are purchasing. The Property Rating Hierarchical Model (PRHM) aims to contribute to the end goal of building a smart chatbot that learns how people touch and engage with clothes. The chatbot aims to create a reflective clothes shopping experience for the consumer by making consumers touch, reflect and critically think about the clothes they are about to purchase. This innovative tactile interaction with clothes may compel consumers to only purchase garments they will wear more than once and genuinely love, thereby slowing down fast fashion and its adverse impacts.

1.1.2 The Importance of Touch when Purchasing Clothes

Multiple studies have discovered that consumers heavily rely on affective touch and tactile experience when purchasing clothes. Morrison defined affective touch as tactile processing with a hedonic or emotional component, i.e. the emotional aspect of touch [14].

Previous studies have discovered that direct sensory contact with fabrics and garments provides valuable product information to consumers to make an informed choice [15, 16]. Other studies have also observed that tactile input is preferred over macro-spatial characteristics (shape and size) when assessing an item's physical properties (such as softness, smoothness, flexibility) [17, 18]. According to a study led by Holbrook [19], tactile cues were more important than visual cues in consumer perception and assessment of sweaters. Moreover, the principal effect of tactile cues may differ from one item to another [20]. For example, consumers will rely more on tactile inputs when assessing some objects (a coat with various properties such as weight, thickness and texture) than others (a standard AAA battery). Therefore, touch is a crucial criterion when assessing items that differ in their textual properties [21]. Consequently, it is plausible that conscious or unconscious tactile emotions (affective touch) play a leading role in consumer perception of clothing.

Although consumers can readily assess garments by touching them when shopping in stores, the role of touch has inevitably diminished when shopping online. With the expansion of fast fashion and the development of the internet, online clothes shopping has become increasingly popular. However, online shopping comes with the caveat that consumers cannot physically touch and engage with the clothes they purchase. This is because online shopping mainly uses audio and visual channels to communicate product information with consumers. Multiple studies on internet retail [22, 23] have confirmed that the main drawback of online shopping is the inability of the consumer to touch the products. Therefore, consumers find it challenging to develop a comprehensive evaluation of the product they are purchasing purely through online shopping. Studies have also discovered that some consumers feel frustrated or disappointed if they do not have the opportunity to touch and examine the products in real life. This is particularly true for consumers with a higher need for touch (NFT) [15, 22].

The Property Ranking Hierarchical Model (PRHM) aims to infer the rating of several properties for each fabric. These ratings will inform the consumers roughly how smooth, thick, warm, flexible and soft the garment will be. Therefore, consumers can make more informed decisions when shopping online. Although the PRHM is not an identical substitute for physically touching a new item of clothing, it will give consumers a better understanding of the garment's textual and physical properties.

1.1.3 Crowdsourcing Tactile Perception

When shopping online, consumers can only gauge if the tactile sensation of the garment suits their needs based on the textual and visual (sometimes even video) description provided. Therefore, consumers regularly return online clothes that do not meet their tactile expectations. This is another major pitfall associated with fast fashion. The Financial Times reports that thousands of packages containing clothes are returned daily and that most returned clothes are either sold at a discount or binned because they are out of season [24].

As tactile information cannot be substituted by haptic feedback, recording and displaying the tactile information of an item of clothing may help the consumer get a better idea of its tactile experience [25]. This may help decrease the wastage of money and energy resources caused by returning fast fashion clothes. Further, crowdsourcing how individuals touch a product and their tactile perception of the properties may be a close substitute to someone physically touching the product before buying it.

Crowdsourcing is levered on as a feasible and low-cost method to present tactile experiences online [26]. Online product reviews provide information regarding an item's quality and how it feels to the touch. Such reviews are very common and help indecisive consumers decide whether to purchase the reviewed product [27]. Therefore, crowdsourcing tactile perception of an item of clothing may work as an alternative to physically touching the garment before buying it. Previous studies regarding crowdsourcing in the fashion/textile industry have mainly been in the form of text [26, 28]. The data used in this project was collected via crowdsourcing Electromyography (EMG) and kinematics data while participants assessed predetermined properties by touching items of clothing. Developing machine learning models that crowdsource tactile experiences of clothes via sensors that track motion data from the user's hand is an exciting and innovative concept explored in this project.

1.2 Objective

Past studies have discovered that consumers rely heavily on affective touch and tactile experience when purchasing clothes. Therefore, developing a method to comprehensively understand what individuals experience when touching a textile, how they feel after touching it, and how it influences their choice may help to reduce the impulsive buying habits of consumers. This study proposes a novel machine learning approach to estimate people's sensations and liking of the textile being touched.

The main objective of this project is to build a novel machine learning model to infer tactile affect in the context of fabric handling. The Electromyography (EMG) and Inertial Measurement Unit (IMU) data input to the model were collected from both hands (left and right) via wearable sensors. The two research questions that this study aimed to answer were:

1. Can the model predict what property an individual is assessing based on the muscle and hand movement collected via the wearable sensors?
2. Can the model predict the property rating an individual gives each of the properties assessed for a specific garment, given the property?

The long-term aim of this project is to build a chatbot that helps reduce fast fashion's impact. The chatbot aims to transform clothes shopping into a multi-sensory, reflective and recognition-based experience. Firstly, the chatbot will ask individuals to touch new clothes, engage with the fabric, **reflect** on whether they like the garment and if they will wear it. Secondly, the chatbot will have previously collected individual-specific data (e.g. what clothes an individual has in their cupboards, what their favourite and most worn clothes are and what type of clothes they like) and generic crowdsourced data regarding the tactile perceptions of different clothes. Based on this data, the chatbot will look at the cloth the individual is looking at and **recognise** if the individual likes it and will wear it.

Chapter 2

Background and Related Work

In order to build a model that can recognise the affective and experiential meaning of a tactile gesture, we first review the literature to understand if a language of affective touch exists. We then review the existing literature on automatic recognition of affective and experiential touch to identify the current state-of-the-art methods and gaps in the existing literature to be addressed.

2.1 Recognising Affect through Interpersonal Touch

As humans, we use touch in our daily life for various reasons. We rely on touch to perform actions such as attracting attention (waving hands), offering congratulations (handshake or a pat on the back), flirting (gently stroking face, hair, or arm), and when thanking others (hugging and gently squeezing). We also express emotions such as love (embracing, hugging, and stroking), sympathy (embracing, stroking back), fear (squeezing hand) and anger (gently slapping) through touch.

Although touch plays an essential role in human life, it has received less attention in affective science than facial and vocal displays of emotion [29]. Initially, studies regarding touch as an affective modality claimed that it was mainly used to communicate the hedonic tone of emotions (positive and negative) [30, 31, 32, 33, 34] and increase the intensity of emotion-related communication [30, 34]. However, two consecutive studies by Hertenstein in 2006 and 2009 have argued that touch plays a much larger role in emotional communication.

The first study discovered that participants could identify several distinct emotions when they were touched on their forearm by their partner (a stranger), even though they could not communicate visually or verbally [30, 35]. This study observed that the tactile modality could distinguish between the emotions of love, gratitude, sympathy, fear, disgust and anger with accuracy rates between 48% and 83% [30]. The second study confirmed these results and observed that happiness and sadness could also be recognised with accuracy rates higher than chance [29, 35]. Therefore, the recognition performance attained with touch alone was comparable to those observed in studies of facial displays, and vocal communication [30, 35, 36, 37].

The second study investigated 23 types of tactile behaviour (ex: squeezing, stroking, pushing) and revealed systematic differences in how touch was used to communicate different emotions [29, 35]. For example, love was associated with stroking; gratitude was associated with shaking of the hand; sympathy was associated with stroking and patting; disgust was associated with a pushing motion; fear was associated with trembling, and anger was associated with hitting and squeezing [29]. However, as some types of tactile behaviour were used to communicate multiple emotions (for example, stroking was used to express both love and sympathy), the study concluded that tactile behaviour alone is insufficient to differentiate between different affective emotions [29, 35].

The second study by Hertenstein also discovered that emotions could be categorised according to differences in intensity and duration [29]. For example, love and sympathy were characterised by a moderate-intensity touch for a longer duration, whereas anger was characterised by a vigorous intensity of touch for a moderate duration [29].

2.2 Emotion Recognition based on Human-Computer Touch

Several studies have investigated the possibility of recognising affect based on how humans touch mobile devices. Gao et al. conducted a study to decode affective touch in the context of video games [35]. This study aimed to capture the emotional state of players in a real-world touch screen gameplay setting by using finger-stroking motions detected by an iPad [35]. Samurai Fruit is a modified version of Fruit Ninja (an iPhone game), and this game was used to evoke different emotions within participants and detect and store features related to their finger strokes, such as coordinates corresponding to the stroke and the duration of a stroke). Participants were asked to play twenty levels of the modified game and fill out a questionnaire regarding their emotional state (bored, frustrated, excited and relaxed) at the end of each level. The results revealed that the length and pressure of the strokes were important factors when classifying the hedonic tone of emotions (positive and negative). The speed and the direction of the stroke provided information regarding the player's level of arousal. This study discovered that tactile behaviour contains a wealth of information regarding users' affective states in touch screen gameplay. Another study conducted by Shah et al. used finger strokes to predict an individual's affective states in a more general touch screen usage setting [38]. Seven features related to finger strokes were used to build a regression model, and this model predicted the user's positive, negative and neutral affective states with an accuracy of 90.5%.

Multiple industries aim to leverage automatic emotion recognition via Human-Computer touch for various reasons. The entertainment industry uses technology that automatically retrieves media (such as images and music) based on what an individual feels or wants to feel [39, 40, 41, 42]. The communication industry is attempting to use technology that automatically integrates text messages with the sender's emotions. The technology introduced in [43] allows the sender to add emotions to a text message by shaking their mobile phone differently. The new SAMSUNG smartphone is creating a smartphone that can detect the emotional state of an individual based on how they tweet (number of spelling and grammatical errors, speed of typing, use of special symbols and emoticons) [44]. Further, [45] investigated how affective technology could help students better their emotional state to facilitate cognitive processing and stimulate motivation whilst e-learning [35].

2.3 Automatic Affect Recognition Using Machine Learning

Several previous studies have used Machine Learning methods for automatic affect recognition. Machine Learning algorithms used for classification tasks such as k-Nearest Neighbors (kNN), Support Vector Machines (SVM), Multi-Layered Perceptron (MLP), Linear Discriminant Analysis (LDA), and Random Forests (RF) have previously been used for affect recognition [46]. However, these methods require human intervention as the input features must be extracted and carefully engineered from the raw data [47].

Several examples in the literature typically extract too many features from Electromyography (EMG) and Inertial Measurement Unit (IMU) data, which can often be time and resource-consuming [48]. After carefully analysing the literature, low-level features such as signal statistics and waveform traits were used as input features in this project. The maximum, mean and standard deviation were calculated for each sampled slice. Other commonly used techniques for feature extraction include various statistical approaches [49], time-frequency transformations [50], and symbolic representations [51, 47].

However, when considering computational and statistical efficiency and human intervention, these machine learning methods may not be the most efficient way to approach challenging learning problems such as affect recognition [52, 53].

Deep Learning (DL) is a subfield of Machine Learning that has revolutionised many areas, including computer vision, natural language processing, and speech recognition [54]. The main advantages of deep architectures are derived from their depth [53]. Increasing the depth of a neural network allows the learned features to be reused [53, 55]. Further, the deep hierarchy of feature representations allows learning at different levels of abstraction[53, 55].

Deep Learning methods differ from traditional ML techniques as feature extraction is part of the model process and eliminates the need for handcrafted features [46]. Therefore, neural networks have been used to extract temporal features. DL models can also learn descriptive features from complex data using advanced computing resources such as GPUs [46]. This exceptional learning ability enables deep learning models to thoroughly analyse multimodal sensory data for accurate recognition when performing affect recognition tasks [46].

Various deep learning model architectures have encoded features from multiple perspectives in affect recognition tasks. For example, convolutional neural networks (CNNs) can capture the local connections between multimodal sensory data [46]. The translational invariance introduced by locality leads to accurate recognition [46, 56]. Recurrent neural networks (RNNs) can extract temporal relationships and incrementally learn information across time intervals [46]. Therefore, RNNs are appropriate for streaming sensory data in affect recognition. However, traditional RNN cells suffer from vanishing/exploding gradients problems. The Long Short-Term Memory (LSTM) units are usually used to overcome this issue [46, 57]. LSTM RNNs successfully learn long-term dependencies by providing gate mechanisms to selectively add and forget information [58]. Gated recurrent units (GRUs) are a recently proposed gating mechanism used in sequence-to-sequence processing [59].

2.4 Using Sensors to Recognise Affective Touch

Different sensors are currently being used to recognise affective touch and human activity. Further, research has demonstrated that combining diverse sensing modalities can obtain better results than only one particular sensor [47, 60].

2.4.1 Electromyography (EMG)

Electromyography (EMG) measures the muscle response or electric potential generated by the muscle [61, 62]. EMG is preferred over other sensor modalities used to capture hand gesture information because EMGs capture and evaluate the muscles' electrical activity; the physical phenomenon that results in hand gestures [46]. EMG data is recorded via both invasive and non-invasive methods [46]. Surface electromyography (sEMG) is a non-invasive technique that measures a muscle's action potential from the skin's surface [46]. This technique is preferred over other invasive methods that penetrate the skin to reach the muscle [46]. The acquisition of sEMG signals involves one or more sensors attached around the target muscle group [46].

Saponas et al. explored whether muscle-computer interfaces (muCIs) could be built using EMG technology [63, 62]. muCIs are an interaction methodology that directly senses and decodes human muscular activity rather than relying on physical devices or user actions. In the study, participants wore an armband containing 10 EMG sensors on their upper forearm and performed four distinct sets of finger gestures [63]. The study aimed to determine if the researchers could distinguish between the four sets of finger gestures using EMG data [63]. Saponas et al. wanted to test if they could classify position and pressure of finger presses and differentiate among all fingers during tapping and lifting. They discovered they could easily do so. The classification accuracy was 78% for position, 84% for pressure, 78% for tap, and 95% for lift. Further, Cary identified that lifting and pressing (related to pressure) were two commonly used gestures when exploring fabrics [64, 62].

The high accuracy associated with classifying these gestures proves that EMG can be used to recognise affective touch in the context of fabric handling.

2.4.2 Accelerometer

A 3D accelerometer is used to measure kinematic information such as acceleration and velocity in a three-dimensional space [62].

Humayoun et al. conducted a study to define a set of gestures based on accelerometer data as accelerometers are embedded in most modern mobile devices [65]. Nine commonly used gestures related to mobile devices (tap, double-tap, left flick, right flick, press, drag, rotate, zoom-in, zoom-out) were transformed into appropriate accelerometer-based gestures [65]. Most participants found it easy to perform tasks using these 3D accelerometer-based gestures, thereby indicating the immense potential of accelerometers in facilitating gesture-controlled interaction [62, 65]. Further, as the accelerometer-based gestures are calculated according to the change in angle along the x, y and z axes, it can capture aspects of textile handling that EMG cannot [62].

2.5 Recognising Affective Touch using Sensors in the context of Fabric Handling

Multiple studies have recently explored how muscle activity and hand movement data can be collected via wearable sensors to assess properties and property ratings of textiles in the context of fabric handling [62, 66, 25].

Wang explores the feasibility of incorporating EMG and accelerometer data to assess the tactile experience of fabrics [62]. In the first part of the study, participants were asked to touch four fabrics and then rate each fabric's softness, warmth, thickness, and smoothness [62]. The results from the discriminant analysis revealed that the EMG and accelerometer features could distinguish between the warmth, thickness and smoothness properties with a percentage accuracy of 55.1%. In the second part of the study, participants were asked to assess the enjoyment gained by caressing, scratching, squeezing, and rubbing the provided garment using varying speeds and strength. This part aimed to identify if changes in speed and strength would influence the enjoyment a participant feels during fabric handling. The discriminant analysis results revealed that classification accuracy was 75.4% when EMG, acceleration and speed features were used to discriminate between the four gestures used in textile handling. Therefore, the results from this study imply that sensor technology can be used to identify what individuals feel when touching textiles.

Gao explored the possibility of using motion sensors to investigate the automatic recognition of properties assessed in tactile interaction with textiles [66]. Participants were instructed to evaluate the textiles' smoothness, softness, warmth, thickness, and durability in both a lab and physical shopping environment. A wearable sensor was used to record kinematic data (EMG and IMU data) corresponding to the hand movements and gestures used when participants explored the textiles. The discriminant analysis results revealed that the classification model could accurately distinguish between the five properties with an average accuracy of 36.5% in a lab setting. The classification accuracies for all five properties were above the chance classification accuracy (20%). However, the classification model failed to distinguish between the properties using the data collected in the field environment. Further, the performance of the classification model increased when the field data was added to the experimental data, thus suggesting the importance of a large dataset for the model's accuracy. Although this project was only partially successful, it demonstrated the possibility of automatic recognition of textile properties and the feasibility of using sensors to record hand movement and gesture data. The results of Gao also confirm that individuals tend to use particular gestures when evaluating specific features of textiles, which is in line with the findings of Wang [62, 66].

Lin conducted two experiments to analyse the feasibility of crowdsourcing quantitative fabric-hand perception via wearable sensors to stimulate a consumer's online shopping experience and improve textile communication within the fashion industry [25]. Lin explored whether the collected EMG and IMU could automatically detect what property an individual is assessing and the rating corresponding to the assessed property. This was done by using discriminant analyses as the classification algorithm. The success of Lin's models confirms the feasibility of inferring fabric-hand perception from wearable sensors.

This project builds on the findings of Wang, Gao, and Lin [62, 66, 25]. However, these studies have mainly used statistical techniques for feature selection and model building. This project uses machine learning techniques with appropriate cross-validation methods to build a novel model to infer affect from touch in the context of fabric handling using EMG, accelerometer and quaternion data.

Chapter 3

Extension of An Existing Textile Touch Dataset

The existing dataset contains EMG and IMU data collected via wearable sensors when exploring different properties and property ratings for various textiles. The dataset was collected in 2021 by Lin [25] as part of her MSc Final project. The author of this project and two other researchers collected data to extend this existing dataset. The researchers had different motivations for collecting data, including developing a visualisation system for consumers. The author of this project wanted to extend the existing dataset for two main reasons. Firstly, the author wanted to expand the current dataset so that the machine learning models would have more input data. Secondly, the author wanted to observe if a model built using the existing data could be generalised to a new type of garment that the model had never seen before.

The organisation of this chapter is as follows. Sections 3.1 and 3.2 provide an overview of the existing and new datasets. Section 3.3 explains the materials used in data collection, and section 3.4 explains the experimental procedure followed when collecting the two datasets highlighting differences between the original data collection and the one done for this project where necessary.

The study conducted to collect the existing dataset was approved by the UCL Research Ethics Committee (Project ID Numbers: 5095/00167) to continue the research within the UKRI Textile Circularity Centre (TCC) [25]. The study to collect the new dataset was approved by the UCL Research Ethics Committee (Project ID Number: 2021_018).

3.1 Existing Dataset

Nine participants (eight females and one male) were recruited for the study that collected the existing dataset. [25]. Eight were students of the Human Computer Interaction department at University College London (UCL), while one was from another university [25]. These participants were all Chinese and righthanded [25]. Participants received an Amazon voucher as a reward for their time [25].

Each participant was asked to select six different clothes from their wardrobe to explore their textile properties. The only limiting criterion was that they had to choose different types of clothes with different textile properties and tactile experiences [25]. For example, they could not choose six t-shirts as the t-shirts may have a similar tactile experience.

All nine participants in the study that collected the existing dataset recorded the data in their houses. Participants received two wearable armbands that each contained eight EMG sensors and one IMU sensor, a phone with an app which paired the armbands via BlueTooth, chargers for the phone and armbands and disinfecting wipes to carry out the experiment [25]. The sensors are discussed in detail in section 3.3.

Each participant was asked to assess five physical properties (smoothness, thickness, warmth, flexibility and softness) and enjoyment associated with all six items of clothing by touching them for a set period. For this project, an 'instance' is defined as when a participant explores a single property of a cloth. For example, if participant two explores the 'softness' of the tenth item of clothing, this is an instance.

The resulting dataset consisted of 324 (9 participants $\times$ 6 clothes $\times$ 6 properties) instances (20 sec of sensor data). However, the data for a single participant had to be discarded because they collected it incorrectly.

3.2 New Dataset

Six (three females and three males) participants participated in the study that collected the new dataset. Five participants were PhD students at University College London, and one was a visiting researcher at University College London Interaction Centre (UCLIC). Five out of the six participants were right-handed, whereas the other was left-handed. Participants received an Amazon voucher worth £15 as a token of appreciation for their participation. The same sensors and smartphone app used to collect the existing dataset were used to collect the new dataset. The only difference was that the new data was collected in the University College London Interaction Centre (UCLIC) lab.

Since the existing dataset did not contain any sock garment, all participants involved in the data collection of the new dataset were asked to touch the same six pairs of socks shown in Figure 1. Participants were asked to touch and evaluate the same six properties assessed for the existing dataset for all six pairs of socks in Figure 1.

Figure 1: Socks used in the subsequent study

The experiment carried out to collect the new dataset was conducted by the author and the two other researchers at UCLIC. The new dataset consisted of 216 (6 participants × 6 clothes × 6 properties) instances. It should be noted that the other researchers at UCLIC used the dataset for other reasons, including the visualisation of tactile patterns.

3.3 Material and Sensors

The same equipment and software were used to create the existing and the new datasets.

1. **Armband with sensors**: Two OYMotion gForcePro+ EMG Armbands (refer Figure 2) were used to collect raw EMG and motion data from the participants' left and right hands (the armbands were labelled to differentiate between the left and the right). These armbands are examples of smart wearable human interface devices used for gesture recognition [67]. Each armband contains eight EMG sensors with differential dry electrodes, nine-axis IMU[1] motion sensors and communicates using Bluetooth BLE 4.2.[70, 25]. The armbands recognise gestures according to the sEMG signals of human forearms and calculate orientation data in quaternions or Euler Angles using its built-in nine-axis IMU sensors [67]. This study used the armbands to collect raw EMG and IMU data. Quaternions were used (instead of Euler Angles) to represent the IMU data. Although quaternions are much less intuitive than Euler Angles, rotations defined by quaternions can be computed more efficiently and with more stability [71]. Quaternium representations are also less sensitive to gimbal lock[2] problems than Euler Angles.

2. **Phone**: A Motorolla moto g^9 power phone (refer Figure 3) with both the GForceTextileHand App [25] developed by Lin [25] and the gForceApp [70] developed by OYMotion installed were used to collect the data. This phone uses Android 10 with a Qualcomm® Snapdragon™ 662 processor and Bluetooth® 5.0 as its operating system [74, 25].

3. **Smartphone App**: The GforceTextileHand app was mainly used for data collection, whilst the gForceApp was also used. The user interface for the GforceTextileHand is given in Figure 4 [25]. The software used in this app employed SQLite to store experimental, EMG, Euler Angle and IMU (accelerometer, gyroscope, magnetometer, and quaternion) data [25]. SQLite is a C-language library that implements a fast and

[1] An Inertial Measurement Unit (IMU) is an electronic device that can measure and report an object's distinct force, angular rate, and sometimes its orientation [68]. An IMU typically consists of accelerometers and gyroscopes and sometimes contains magnetometers [69, 68]. The two commonly used methods of representing IMU data are Euler angles and Quaternions. Both Euler Angles and Quaternions are used to represent a rotation in 3D space.

[2] Gimbal lock occurs when the axes of two gimbals in a three-gimbal device are driven into a parallel configuration, thereby losing a single degree of freedom and "locking" the system into rotation in two-dimensional space [72, 73].

Figure 2: OYMotion gForcePro+ EMG Armband [67, 70]

Figure 3: Motorolla moto g^9 power phone [74]

high-performance SQL database engine [25, 75]. This project only uses the EMG, accelerometer and quaternion data.

3.4 Experimental Procedure

Identical experimental procedures were followed when collecting the existing and new datasets. As mentioned before, the only difference was the experiment's location. For

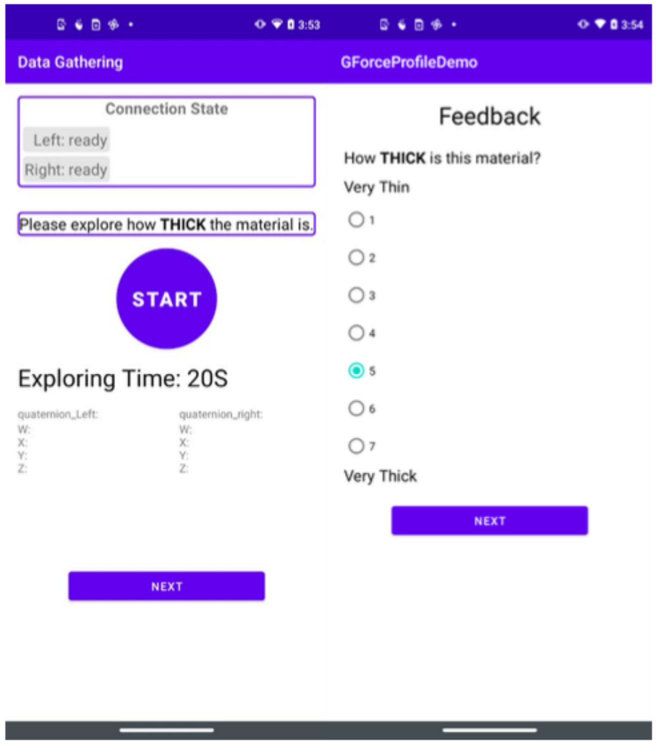


Figure 4: Sample User Interface of theGforceTextileHand App [25]

both experiments, the participants were sent an information sheet (included in Appendix A.1) by email and were asked to sign the consent form (included in Appendix A.2) via REDCap before the experiment started.

The data collection process consisted of two components. Firstly, we recorded the baseline readings. Then, participants were asked to assess six properties (five physical properties and enjoyment) for six items of clothing. The five physical properties evaluated were smoothness, thickness, warmth, flexibility and softness. Lin [25] selected these properties when creating the existing dataset based on the study done by Gao [66]. However, Lin [25] replaced durability with flexibility as Gao discovered that durability was challenging to identify using only touch [66]. After evaluating a property, participants were asked to rate the property on a scale from 1 to 7. Ratings for each property were obtained using Likert scale questions with seven choices (for example, for the flexibility property, 1 was not at all flexible, and 7 was very flexible).

The procedure followed to collect the EMG and IMU data is as follows.

1. Before assessing the six properties for all clothes, participants received 10 seconds to touch the cloth freely, engage with it, and determine how they wanted to touch it to best assess the properties.

2. The participants touched their first item of clothing for a pre-determined time period and assessed a particular physical property. In the existing dataset, the pre-determined period was 20 seconds, as determined by previous studies [66, 62]. When extending the existing dataset, the time was shortened to 15 seconds for fear that the participants would get bored and not fully cooperate.

3. Then, the participants rated the property from 1 to 7 on a Likert scale.

4. Participants repeated steps 2 and 3 for the four remaining physical properties (for the first selected item of clothing).

5. Afterwards, the participants touched the cloth for a pre-determined period and explored how enjoyable it was to touch the cloth using their preferred gestures.

6. The enjoyment gained from touching the cloth was then rated on a Likert scale from 1 to 7.

7. The participants were then asked to reflect on their feelings when touching the cloth and comment on it. This included anything they enjoyed or disliked and whether the item of clothing had any characteristic features.

8. Participants repeated steps 1-7 for the remaining five clothing items.

The EMG, accelerometer and quaternion data were obtained using the armbands as mentioned previously. The armbands were worn as indicated in Figures 5 and 6. As shown, armbands on the left and right should be worn differently (the armband on the left should have the USB port facing the elbow, and the one on the right should have its USB port facing the wrist). This is so that when recording the data, the EMG channels are the same, but the IMU data signs are swapped.

There were two main reasons for obtaining the baseline readings [25]. The first reason was to check if the participants were wearing the armbands correctly. The position of the armbands (how high or low they sit on the forearm) varied depending on variables such as participants' gender and forearm circumference. This difference could also affect the data collected because if the participant had a smaller forearm circumference, the sensors on the armband might not have continuously had contact with the participant's skin. Therefore, the baseline readings were also used to normalise the collected data to account for personal idiosyncrasies in touching behaviour and strength differences between participants.

Participants had to perform two tasks when recording the baseline readings. Firstly, the participants had to lay both arms, relaxed on the table, palms facing upwards (as shown in Figure 5). This was to record the minimal contraction of a participant's muscles. Secondly, the participants had to clench their hands into a fist as tightly as possible (refer Figure 6) to record the maximum contraction of muscles.

Figure 5: Baseline 1 - Relaxed

Figure 6: Baseline 2 - Fist

After freely touching the cloth, participants would touch the cloth for a pre-determined time period to assess each property. When collecting the existing dataset, the order of assessing properties was fixed for each garment, and each participant [25]. This was because the participant conducted the experiment, and Lin wanted to keep the experimental procedure simple and uncomplicated [25]. However, the order of properties assessed for each cloth and the order of clothes for a participant were randomised when extending the existing dataset. This order can be observed in Appendix A.3.

After touching the item of clothing for the pre-determined time to explore a property, participants were asked to rate the property on a Likert scale with values ranging from 1 to 7. For example, when assessing the property flexibility, 1 would signify that the cloth is not at all flexible (very stiff), and 7 represented a very flexible cloth.

It is important to note that the participants were not explicitly informed how to touch the cloth as we aimed to capture natural movements used in everyday life rather than prescribed gestures.

A nested Leave One Participant Out Cross-Validation process was used for all three PRHM model. The inner loop was used for hyperparameter tuning, and the outer loop was used to test the model performance.

Chapter 4

Automatic Detection of Textile Touch Behaviour and Experience: Methods

4.1 Property-Rating Hierarchical Model (PRHM)

The Property-Rating Hierarchical Model (PRHM) was built to predict the property assessed and the corresponding rating. The PRHM is a two-part model that first predicts the assessed property based on the input EMG, accelerometer and quaternion data and then predicts the corresponding property rating. The first part of the PRHM is the PRHM property classification section, whereas the second part of the PRHM is the PRHM property classification section.

Different machine learning techniques were used to create three PRHMs to predict the assessed property and corresponding rating. Two variations of each PRHM were created. In the first variation, the property variable was not an input for the model that predicted rating. In the second variation, the property variable was one hot encoded and entered as input for the rating model. This variation of the PRHM assumed that all the properties were perfectly predicted when predicting the corresponding ratings. As expected, the second variation provided better predictions. Therefore, the second variation will be discussed in this report.

4.1.1 Random Forest Property Rating Hierarchical Model (RF PRHM)

Bootstrapping is a sampling technique which randomly samples data with replacement from the main data set [76]. Bagging (or bootstrap aggregating) is an ensemble algorithm that fits multiple models on different subsets of a training dataset and then combines the predictions from all models. In bagging, m decision trees are trained on m independent bootstrapped samples/training datasets (m is a hyperparameter that can be tuned). The final predicted value is the average of the decision trees [76].

A single decision tree has a high variance and tends to overfit [76]. Bagging overcomes this problem by combining many weak learners to create a strong learner [77]. Random forest is an extension of bagging and was used to create the first PRHM.

4.1.2 Fully-connected Property Rating Hierarchical Model (FC PRHM)

Feedforward Deep Neural Networks are densely connected layers where inputs influence each successive layer which then influences the final output layer [78]. Information is only passed forward as these Neural Networks do not contain loops [79].

A fully-connected feedforward neural network consists of a series of fully-connected layers that connect every neuron in one layer to every neuron in the next layer [80]. The main advantage of fully-connected networks is that they are "structure agnostic", i.e., no special assumptions need to be made about the input [80].

A Deep Fully-Connected Feedforward Neural Network Architecture can be seen in Figure 7. The leftmost layer (coloured in green) is called the input layer. The neurons in this layer are called input neurons, and these neurons receive input and pass it on to the other layers of the network [79, 81]. The number of features in the dataset should match the number of neurons in the input layer [79]. The right-most layer (coloured in red) represents the predicted feature [79]. The middle layers are called hidden layers. Neurons in these layers (coloured in purple) transform the input before transferring it to the next layer [79]. The Fully Connected PRHM consisted of two parts. The first part predicted properties, and the second part predicted the corresponding property rating. Each part of the PRHM

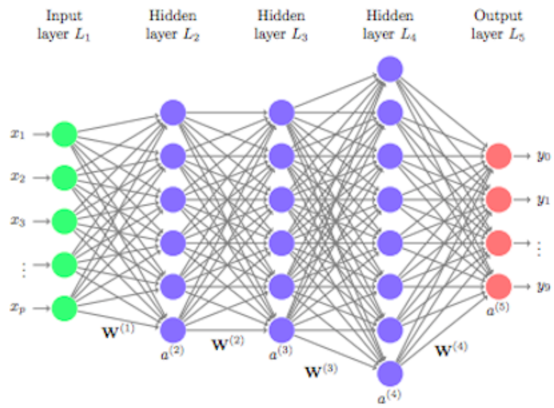


Figure 7: Deep Fully-Connected Feedforward Neural Network Architecture [78]

contained three linear hidden layers (see Figure 8). After both the first and second layers, a tanh activation function given in equation 1 was applied.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{1}$$

A softmax activation function (Equation 2) was applied after the third hidden layer.

$$f(x) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{2}$$

Figure 8 shows the architecture of a single part of the FC PRHM. Here, n* is the batch size, and this depends on what variation of the dataset was used. The labels depended on what part of the PRHM model was used. If the first part was used, the labels were 1-5; if the second part was used, the labels were 1-3 (refer section 4.4 for further details). The number of features depended on the feature set used and what part of the PRHM model was used. The number of features took four values.

- features = 180 when all the features were used to predict properties
- features = 185 when all the features + one hot encoded properties were used to predict the property ratings
- features = 48 (8 EMG channels × 3 statistics × 2 hands) when only the EMG features were used to predict properties
- features = 53 when only the EMG features + one hot encoded properties were used to predict the property ratings

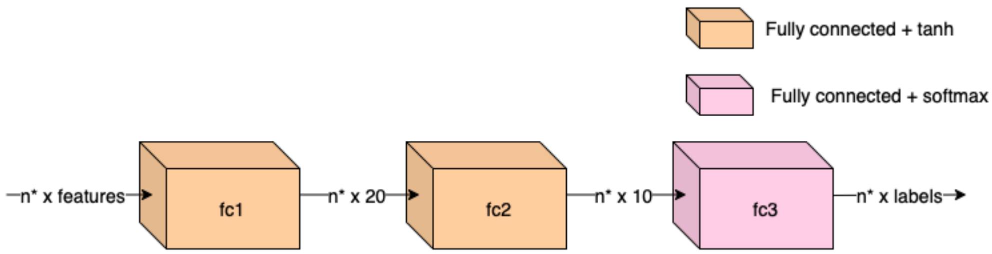


Figure 8: Architecture of a single part of the Fully-connected PRHM

4.1.3 Long Short-Term Memory Property Rating Hierarchical Model (LSTM PRHM)

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) capable of learning order dependence in sequence prediction problems [82]. RNNs differ from Feed Forward neural networks as RNNs are bi-directional and pass information forwards (from input to output) and backwards (from output to input). RNNs save the output of the processing nodes and feed the result back into the model [83].

Figure 9 shows the architecture of a single part of the LSTM PRHM. n* in this diagram is the batch size which depends on the variation of the dataset used. The labels are the same as those used in the FC PRHM. The number of features depended on the feature set used and what part of the PRHM model was used.

In each part of the LSTM PRHM, the left and right-hand data were separately passed through an LSTM followed by a tanh activation function. Then the datasets for both hands were combined. The combined data was then passed through three linear layers. After both the first and second layers, a tanh activation function given in equation 1 was applied. A softmax activation function (Equation 2) was applied after the third hidden layer.

4.2 Data Pre-processing and Feature extraction

The existing datasets for EMG, accelerometer and quaternion consisted of 324 (9 participants $\times$ 6 clothes $\times$ 6 properties) instances, and the newly collected datasets consisted of

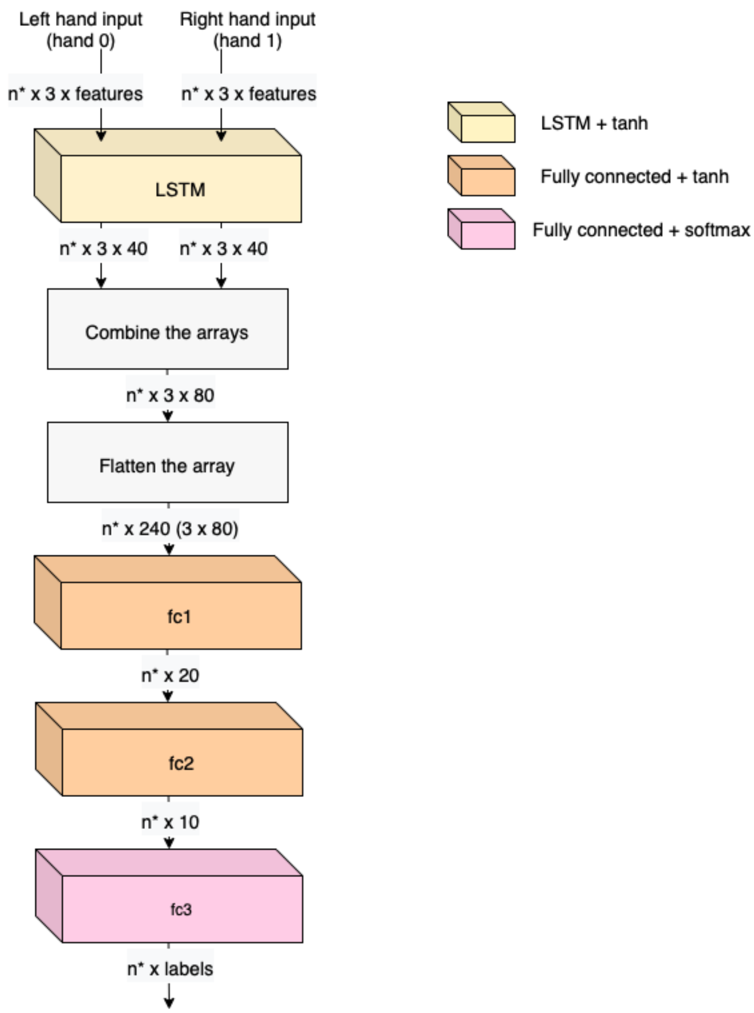


Figure 9: Architecture of a single part of the LSTM PRHM

216 (6 participants × 6 clothes × 6 properties) instances. The data for a single participant in the existing data set had to be discarded because they collected it incorrectly.

4.2.1 Sensor Data Preparation

As explained previously, this project uses the EMG and IMU data collected. The collected data needed to be pre-processed before it could be used in the machine learning models. To preserve uniformity, the same pre-processing procedure was followed for both the existing and newly collected data.

Firstly, the baseline readings and free exploration data were removed from all three (EMG, accelerometer and quaternion) subsets of features. Then, the data regarding rating was attached to each of the three subsets of features.

As previously mentioned, an instance lasted twenty seconds when collecting data for the existing dataset and fifteen seconds when collecting the new dataset. Lin removed the first five seconds of data from each instance in the current dataset to account for human errors and slow reaction time [25].

EMG Data Preparation

The EMG data was collected via the eight channels of the armband. The preparation of EMG data consisted of rectification and normalisation.

Raw EMG data consists of both positive and negative components. Rectification translates these components to a signal with a single polarity (usually positive) [84, 85]. This procedure is done so that the EMG signals do not average zero. The two commonly used rectification methods are half-wave rectification and full-wave rectification [84]. Full-wave rectification was used in this project because it does not get rid of any data, whereas half-wave rectification removes some signals [84, 86]. Full wave rectification translates data below the baseline to data above the baseline so that all data is positive [84]. If the baseline is zero, this is equivalent to taking the absolute value of the signal [87]. The data for the existing and the newly collected datasets were centred around 120. Therefore, the EMG data was centred on zero before obtaining its absolute value.

Normalisation was the second step of the EMG pre-processing process, and this refers to the conversion of the EMG signal to a relative scale by a reference value [88]. As EMG signals are inherently prone to variability, these signals require normalisation for physiologic interpretation and comparison between different participants [89]. The data was normalised by separating the data into left and right hands and then dividing the data in each channel by the maximum observed signal (in the channel).

Accelerometer Data Preparation

The accelerometer data included linear acceleration along the x, y and z axes. The linear velocity and jerkiness along the x, y and z directions were computed using the provided linear acceleration as follows.

$$a_t = \frac{v_t - v_{t-1}}{t_t - t_{t-1}} \iff v_t = v_{t-1} + a_t(t_t - t_{t-1}) \text{ (Note: } v_0 = 0, t_0 = 0) \tag{3}$$

$$j_t = \frac{a_t - a_{t-1}}{t_t - t_{t-1}} \tag{4}$$

Where $t \in \{0, ..., T\}$ (T=15 seconds), a_t is the linear acceleration at timestamp t, v_t is the linear velocity at timestamp t and j_t is the linear jerk at timestamp t.

Quaternion Data Preparation

The quaternion dataset contained data regarding the real and 3D imaginary components (w, x, y and z) of the raw quaternion[1]. The angular velocity for the real and 3D imaginary components of the quaternions were computed using the quaternion library [90]. Then the angular acceleration and jerk for the quaternion's real and 3D imaginary components were computed using equations 3 and 4.

4.2.2 Extracting Low Level Features

Each instance was split into subwindows, and each subwindow was further divided into slices to extract meaningful low-level features from the data. Each instance was split into subwindows to create a larger number of sub-instances with a smaller time duration. These sub-instances were further divided into slices to differentiate the timesteps within a subwindow. This process of creating subwindows and slices was repeated using three different numbers of subwindows and slices. The three variations are as follows.

1. **3 window segmentation (Variation 1)**: Each instance was divided into three subwindows (subwindow duration = five seconds). The subwindows were not further divided into slices. This was what Lin did in her project and was therefore included [25].

[1] Quaternions contain both real and imaginary components and can be written as $q = w + xi + yj + zk$ where $w, x, y, z \in \mathbb{R}$ and i,j and k represent the 3D imaginary component axes [71].

2. **15 window segmentation with 3 timesteps (Variation 2)**: Each instance was divided into fifteen subwindows (subwindow duration = one second), and each subwindow was divided into three slices (slice duration = a third of a second)

3. **15 window segmentation with 10 timesteps (Variation 3)**: Each instance was divided into fifteen subwindows (subwindow duration = one second), and each subwindow was divided into three slices (slice duration = 0.1 seconds)

The maximum, mean and standard deviation were calculated for each slice (of each subinstance) for all the independent subsets of variables in the EMG, accelerometer and quaternion datasets. This process was done for both the left and right hands.

- **Maximum** $= max\left\{x_1, x_2, ..., x_n\right\}$ where $x_1, ..., x_n$ are the observations in a slice and n is the total number of observations in the slice

- **Mean** $= \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ where x_i is an observation in a slice and n is the total number of observations in the slice.

- **Standard Deviation** $= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})$ where x_i is an observation in a slice, $\bar{x}$ is the mean of the observations in the slice and n is the total number of observations in the slice.

The EMG dataset contained eight independent variables corresponding to the data collected from the eight sensors of the armband. The accelerometer dataset contained nine independent variables (linear velocity, acceleration and jerk in the x, y and z directions - 3×3 variables). The quaternion dataset contained thirteen independent variables (the real and 3D imaginary components of the raw quaternion data (4 variables) and the angular velocity, angular acceleration, and angular jerk of the 3D imaginary components of the quaternions (3×3 variables)). Therefore, the EMG, accelerometer and quaternions contained 30 $(8 + 9 + 13)$ independent variables in total.

4.3 Dataset

The first dataset was created by merging the EMG, accelerometer and quaternion data that Lin collected [25]. This data was pre-processed using the methodology explained in sections 4.2.1 and 4.2.2. The second dataset was created by merging the pre-processed newly collected EMG, accelerometer and quaternion data. Both the first and the second dataset consisted of 180 independent features (the maximum, mean and standard deviation were calculated for each of the 30 independent variables for both the left and right hands - $30 \times 3 \times 2$)

The dataset for this project was then created by combining both these datasets. Three variations of the dataset were created based on how the subwindows and slices were created, and these variations are summarised in Table 1. The three variations of the dataset are given below.

1. **Dataset V1 - 3 window segmentation**
2. **Dataset V2 - 15 window segmentation with 3 timesteps**
3. **Dataset V3 - 15 window segmentation with 10 timesteps**

	Subwindows per instance	**Subwindow duration (seconds)**	**Slices for each subwindow**	**Slice duration (seconds)**	**Frames per slice - EMG**	**Frames per slice - Accelerometer and Quaternion**
Dataset V1	3	5	-	-	-	-
Dataset V2	15	1	3	0.33	10	16
Dataset V3	15	1	10	0.1	30	48

Table 1: Variations of the dataset

4.4 Predicted Labels

4.4.1 Textile Properties

Participants had to assess six properties (five physical properties and enjoyment) for each clothing item. However, only the five physical properties were considered for this project. This was because of the time constraints involved and because enjoyment was very subjective. The five physical properties were represented by positive integers in the datasets.

- Smoothness (represented as 1)
- Thickness (represented as 2)
- Warmth (represented as 3)
- Flexibility (represented as 4)
- Softness (represented as 5)

4.4.2 Textile Property Ratings

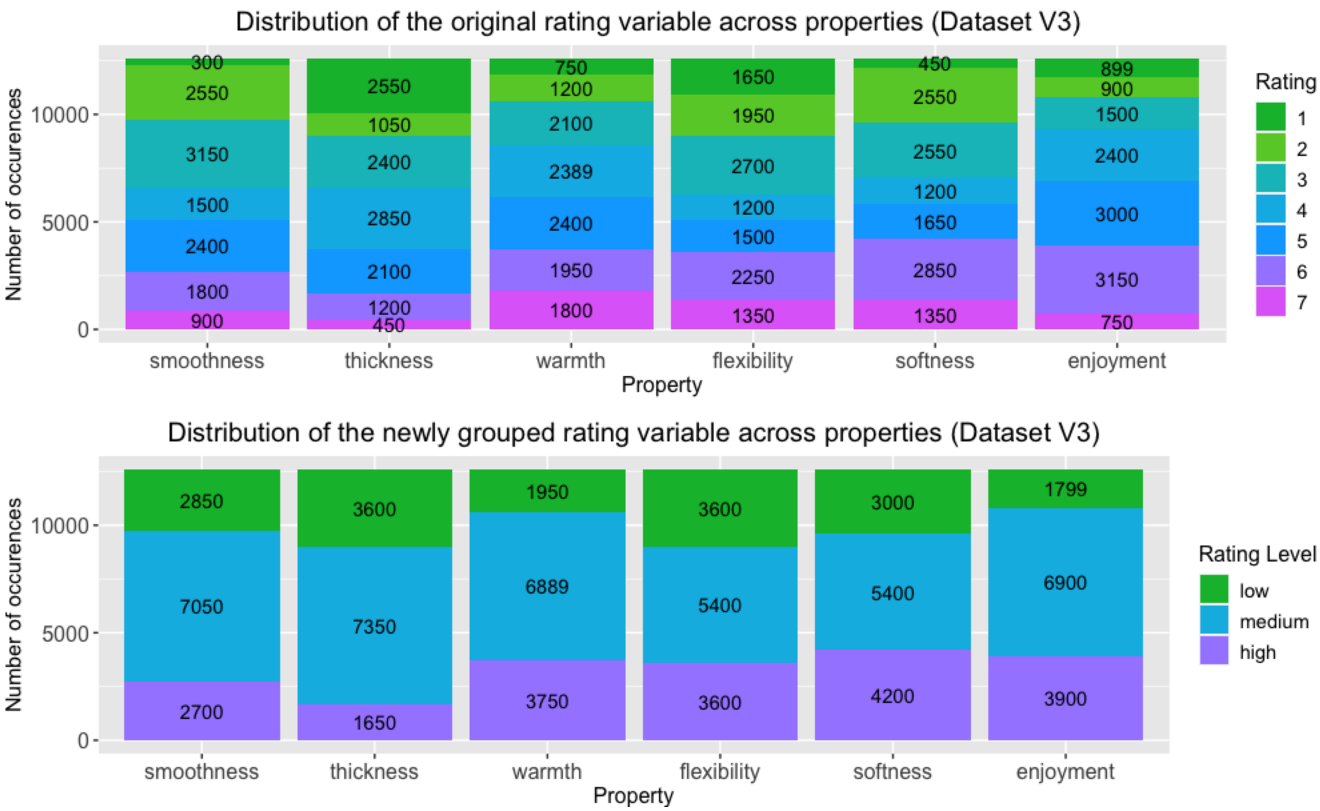


Figure 10: Distribution of the rating variable across properties for Dataset V3

The distribution of ratings by property for Dataset V3 (third variation of the dataset - 15 window segmentation with 10 timesteps) is shown in Figure 10 (top). This distribution is imbalanced, and the modal (most occurring) rating varies across the properties. Based on this distribution, the rating levels were grouped as follows: 1-2 into ‘low’, 3-5 into ‘medium’ and 6-7 into ‘high’. Positive integers were used in the datasets to represent the newly grouped rating variable.

- Low (represented as 1)
- Medium (represented as 2)
- High (represented as 3)

The rating variable was grouped the same for all properties to maintain uniformity. As observed in Figure 10 (bottom), this grouping works very well for some properties, such as flexibility and softness but not so well for others. The rating variable was grouped for two main reasons [25]. Firstly, the rating was very subjective and could greatly vary across participants. Secondly, the dataset was relatively small and thus would present problems if all seven levels were used in granular modelling.

The distribution of the rating variable across properties for the existing and the newly collected datasets are provided in Appendix B.3. The distribution of ratings across properties for Dataset V1 and V2 are very similar to the distribution of ratings across properties for Dataset V3 (See Appendices B.1 and B.2). The grouping criteria used to group the ratings for Dataset V3 was used to group the rating variable for Datasets V1 and V2. The distribution of the newly grouped variable across properties for Datasets V1 and V2 can be found in Appendices B.1 and B.2 respectively.

4.5 Evaluation methods

4.5.1 Cross-validation methods used

Usually, a machine learning model is trained (built) using the training dataset, and the model's performance is assessed using the test dataset [91]. A ratio of 80:20 is commonly used as a training-testing split (80% of the data is used as training data, and the remaining 20% is used as test data) [91]. However, splitting the dataset further into training and testing is not practical if the dataset is small. K-fold cross-validation is used to overcome this issue.

K-fold cross-validation is used to assess the performance of a machine learning model when predicting unseen data (data not used in the training of the model) [92]. K-fold cross-validation has a single hyperparameter (K), which determines the number of subsets a dataset is split into [92]. Once the dataset is split, each subset is used as the test set, whilst all the other subsets are combined and used as the training dataset [92].

Leave One Participant Out Cross-Validation (LOPOCV) was used in this project. LOPOCV is a configuration of K-fold cross-validation where K is set to be the number of participants [92]. As an example, assume that a dataset has three participants. If LOPOCV is used to validate the model, three folds (iterations) would be completed [91]. In each iteration, the model would be trained using two people, and its performance would be assessed using the other participant [91]. This process is shown in Figure 11.

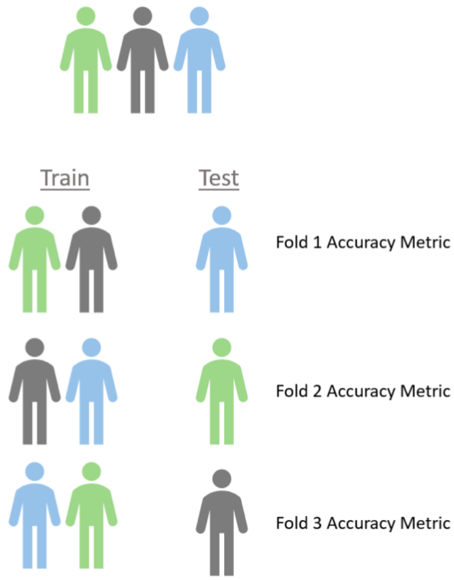


Figure 11: Leave One Participant Out Cross-Validation [91]

4.5.2 Hyperparameters explored

A grid search was conducted to find the optimum number of trees (n_estimators in sklearn) for the Random Forest PRHM. The options for the number of trees were 100, 500, 1000 or 1500. Building random forests with a larger number of trees will create a more robust aggregate model with a lower variance. However, building a model with a large number of trees will take a long time to train.

The learning rate and the number of epochs were hyperparameters for the three neural network models. A grid search was conducted over the values 0.0001, 0.001, 0.005, and 0.01 to determine the optimal learning rate. The number of epochs was selected by analysing the graph containing training loss and validation loss against the number of epochs. The first 150 epochs were observed, and the optimal number of epochs was manually selected to avoid overfitting.

4.5.3 Loss functions and evaluation metrics used

Loss functions used

The cross-entropy loss was used as the loss function in all the models. Cross entropy is the default loss function used in multi-class classification problems [93]. Cross-entropy calculates a score that summarises the average difference between the actual and predicted probability distributions for all classes [93]. The formula used to calculate the multi-class cross-entropy loss is given in Equation 5.

$$L(\hat{y}, y) = \sum_{k}^{K} y^{(k)} \log \hat{y}^{(k)} \tag{5}$$

Where

- K is the total number of classes
- log denotes the natural logarithm
- $y^{(k)} \in \{0, 1\}$. $y^{(k)} = 1$ if the observation is predicted the correct class and $y^{(k)} = 0$ if the observation is predicted the wrong class
- $\hat{y}^{(k)}$ is the probability that the observation belongs to class k

Evaluation metrics used

The confusion matrix, F1 score and accuracy were the three main evaluation metrics used to assess the performance of the machine learning models. All these metrics were computed using sklearn [94].

1. **Confusion matrix**: A confusion matrix is a cross table that summarises the prediction results of a classification problem [95, 96]. The number of correct and incorrect predictions are summarised with count values and are broken down according to each class as shown in figure 12 [95]. Let $C_{i,j}$ be the value of the cell corresponding to the i^{th} row and j^{th} column of a confusion matrix. $C_{i,j}$ is the number of observations expected to be in group i and predicted to be in group j [94]. For example, in Figure 12, $C_{a,c} = 1$. If a model perfectly classifies the outcome variable, the confusion matrix will be a diagonal matrix (i.e. $C_{i,j} = 0$ if $i \neq j$).

		PREDICTED classification				
	Classes	a	b	c	d	Total
ACTUAL classification	a	6	0	1	2	9
	b	3	9	1	1	14
	c	1	0	10	2	13
	d	1	2	1	12	16
	Total	11	11	13	17	**52**

Figure 12: An example confusion matrix [96]

The F1 score and accuracy are calculated from the confusion matrix. True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) are metrics that are calculated for each class based on the confusion matrix. Let $\mathrm{TP_i}$, $\mathrm{TN_i}$, $\mathrm{FP_i}$ and $\mathrm{FN_i}$ be the TP, TN, FP and FN for class i ($i \in a, b, c, d$) in Figure 12.

- $\mathrm{TP_i} = C_{i,i}$
- $\mathrm{TN_i} = \sum C_{j,k}$ where $i \neq j, k$ $(j, k \in a, b, c, d)$
- $\mathrm{FP_i} = \sum C_{j,i}$ where $i \neq j$ $(j \in a, b, c, d)$
- $\mathrm{FN_i} = \sum C_{i,j}$ where $i \neq j$ $(j \in a, b, c, d)$

2. **F1 score**: The F1 score is the harmonic mean of precision and recall and is calculated using Formula 6. The F1 score reaches its best value at 1, and worst score at 0 [96]. Precision measures the proportion of correct positive predictions. In contrast, recall measures the ratio of the number of correct positive predictions to the number of all the positive predictions that could have been made [97, 98]. Formulas 7 and 8 represent the precision and recall for a generic class, i.

$$\text{F1 score} = \frac{2}{\text{Precision}^{-1} + \text{Recall}^{-1}} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

$$\mathrm{Precision_i} = \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FP_i}} \tag{7}$$

$$\mathrm{Recall_i} = \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FN_i}} \tag{8}$$

Several types of F1 scores are used in multi-class classification problems based on how the data is averaged. This project uses the micro F1 score, macro F1 score and weighted F1 score. The micro F1 score calculates the metric globally by counting the total number of TP, FN and FP [94]. The macro F1 score calculates the F1 score for each class and finds their unweighted mean [94]. This metric is commonly used when there is a similar number of observations for each class and was used to assess how well the PRHM model predicted properties. The weighted F1 score calculates the F1 score for each class and finds their average weighted by support (the number of true instances for each class) [94]. This metric changes the macro F1 score to account for label imbalance and was used to evaluate how well the PRHM model predicted rating.

3. **Accuracy**: Accuracy is a popular metric used in multi-class classification problems and measures the number of predictions the model correctly predicted [94].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{9}$$

Chapter 5

Automatic Detection of Textile Touch Behaviour and Experience: Results and Discussion

This chapter consists of two main sections. The first section discusses how well the PRHM predicted the assessed property and the corresponding rating when it was trained using both the existing and the newly collected data. The second section of this chapter explores if the PRHM models could generalise to an unseen garment. In this section, two PRHM property classification models were trained using the existing dataset, and their performance was assessed using the newly collected data.

5.1 Training the PRHM using Both the Existing and Newly Collected Data

The PRHM is a two-part model that first predicts the assessed property and then predicts the corresponding property rating. The first part of the PRHM is the PRHM property classification section and the second part of the PRHM is the PRHM Rating Classification section. 18 (3 dataset variations $\times$ 3 model architectures $\times$ 2 sets of features) PRHM models were used to predict the assessed property and corresponding rating.

The classification performance of a machine learning model is often compared against the chance level. The chance level is a model's average expected classification accuracy if it predicted classes at random. For example, if the PRHM property classification model randomly predicted properties, it would, on average, predict properties correctly 20% of the time (as there are five properties). Therefore, the chance level for the PRHM property classification model is 20%. Similarly, the chance level for the PRHM Rating Classification model is 33% (as the property rating falls into one of three classes - low, medium and high).

5.1.1 PRHM Property Classification Results

Tables 2, 3 and 4 reveal the macro F1 score, percentage classification accuracy and F1 score for each property when the 18 PRHM property classification models were used. These tables assess how well each model predicted the five properties. On average, the models where all 180 features were used performed better than when only the EMG features were used. Further, the best-performing model depended on the variation of the dataset used.

	Dataset V1		**Dataset V2**		**Dataset V3**	
	All 180 Features	**Only EMG**	**All 180 Features**	**Only EMG**	**All 180 Features**	**Only EMG**
RF PRHM Property Classification	**0.32**	0.27	0.3	0.25	0.27	0.24
LSTM PRHM Property Classification	0.12	0.15	**0.39**	0.35	0.27	0.38
FC PRHM Property Classification	0.26	0.17	0.33	0.37	**0.41**	0.31

Table 2: Average macro F1 scores when predicting properties

	Dataset V1		**Dataset V2**		**Dataset V3**	
	All 180 Features	**Only EMG**	**All 180 Features**	**Only EMG**	**All 180 Features**	**Only EMG**
RF PRHM Property Classification	**37.2%**	32.2%	34.4%	29.8%	32.2%	27.9%
LSTM PRHM Property Classification	22.6%	23.3%	**42.4%**	38.8%	32.8%	41.4%
FC PRHM Property Classification	33.1%	25.3%	38.6%	39.5%	**45.0%**	34.8%

Table 3: Average percentage classification accuracy when predicting properties

For convenience, the best models for each variation of the dataset have been colour coded. These colour coded models will be discussed in depth. The best models were colour coded as follows.

- Purple - The Random Forest PRHM Property Classification Model using all 180 features was the best performing model for Dataset V1.
- Blue - The LSTM PRHM Property Classification using all 180 features was the best performing model when Dataset V2 was used.

- Pink - The Fully-Connected (FC) PRHM Property Classification using only the EMG features was the second best performing model when Dataset V2 was used.

- Green - The FC PRHM Property Classification using all 180 features was the overall best performing model.

- Yellow - The LSTM PRHM Property Classification using only the EMG features was the second best performing model when Dataset V3 was used.

			RF PRHM	**LSTM PRHM**	**FC PRHM**
Dataset V1	**All 180 Features**	**Smoothness**	0.46	0.29	0.39
		Thickness	0.41	0.12	0.34
		Warmth	0.28	0.22	0.21
		Flexibility	0.4	0.32	0.42
		Softness	0.25	0.06	0.24
	Only EMG	**Smoothness**	0.39	0.21	0.35
		Thickness	0.27	0.16	0.13
		Warmth	0.27	0.21	0.21
		Flexibility	0.42	0.33	0.34
		Softness	0.2	0.22	0.1
Dataset V2	**All 180 Features**	**Smoothness**	0.43	0.42	0.44
		Thickness	0.31	0.39	0.33
		Warmth	0.29	0.43	0.32
		Flexibility	0.41	0.5	0.49
		Softness	0.22	0.38	0.05
	Only EMG	**Smoothness**	0.36	0.41	0.48
		Thickness	0.11	0.27	0.3
		Warmth	0.29	0.41	0.35
		Flexibility	0.41	0.51	0.45
		Softness	0.21	0.29	0.37
Dataset V3	**All 180 Features**	**Smoothness**	0.4	0.4	0.49
		Thickness	0.25	0.28	0.44
		Warmth	0.29	0.2	0.33
		Flexibility	0.39	0.42	0.51
		Softness	0.22	0.3	0.44
	Only EMG	**Smoothness**	0.35	0.45	0.38
		Thickness	0.09	0.32	0.3
		Warmth	0.27	0.44	0.24
		Flexibility	0.38	0.48	0.45
		Softness	0.2	0.36	0.32

Table 4: F1 score for each property

The Fully-Connected (FC) PRHM Property Classification using all 180 features of Dataset

V3 (15 window segmentation with 10 timesteps) predicted properties the best. This model yielded a macro F1 score of 0.41 and an average percentage classification accuracy of 45% (from Tables 2 and 3). This model also predicted the smoothness, thickness, warmth, flexibility and softness properties with F1 scores of 0.49, 0.44, 0.44, 0.51 and 0.44, respectively (refer Table 4). Therefore, this model classified all five physical properties more than twice as well as the chance level.

Using Dataset V1 (3 window segmentation)

The Random Forest PRHM (RF PRHM) Property Classification model that used all 180 features provided the best classification results when Dataset V1 was used. This model had an average macro F1 score of 0.32 and an average percentage classification accuracy of 37.2% (Refer Tables 2 and 3). These values were almost double that of the chance level of 20%. Further, the RF PRHM Property Classification model that used all 180 features predicted all properties better than the chance level. This model could easily distinguish between the smoothness, thickness and flexibility properties (See Table 5). This model had trouble classifying the warmth and softness properties as they were only predicted marginally better than the chance level (correctly predicted 21% of the time). This may be because participants were found to have different interpretations of these two properties [25]. For example, when assessing warmth, some participants explored if the fabric's surface temperature was high or low, whereas others assessed whether the garment could keep them warm. The former interpretation is assessed by touching the surface of the cloth, and the latter is assessed by placing the hand inside the cloth. As these two interpretations are assessed differently, this will impact the model's ability to classify warmth.

The LSTM PRHM Property Classification model had a performance worse than the chance accuracy level. This may be because neural network typically performs better for larger datasets and Dataset V1 was very small. Small datasets don't allow neural networks to pick up complex patterns in the data [99, 100]

Using Dataset V2 (15 window segmentation with 3 timesteps)

The LSTM PRHM Property Classification model that used all 180 features provided the best results when Dataset V2 was used. This model had an average percentage classification accuracy of 42.4%, an average macro F1 score of 0.39 and predicted all five physical properties better than the chance level (Refer Tables 2, 3 and 4). The FC PRHM Property

		PREDICTED Classification					
		Smoothness	Thickness	Warmth	Flexibility	Softness	Total
ACTUAL Classification	Smoothness	132 (52%)	40 (16%)	26 (10%)	43 (17%)	11 (4%)	252 (100%)
	Thickness	36 (14%)	107 (42%)	16 (6%)	55 (22%)	38 (15%)	252 (100%)
	Warmth	62 (25%)	26 (10%)	54 (21%)	66 (26%)	44 (17%)	252 (100%)
	Flexibility	33 (13%)	41 (16%)	18 (7%)	122 (48%)	38 (15%)	252 (100%)
	Softness	56 (22%)	62 (25%)	14 (6%)	66 (26%)	54 (21%)	252 (100%)
	Total	319	276	128	352	185	1260

Table 5: Confusion matrix - RF PRHM Property Classification using all features of Dataset V1

Classification model that used only the EMG features classified properties second best when Dataset V2 was used.

Tables 6 and 7 reveal that both models classified properties well, as the majority of the predictions are concentrated in the main diagonals of the two confusion matrices. The main diagonal consists of the cells that lie on the diagonal that runs from the top left of the confusion matrix to the bottom right [101]. The elements on the main diagonal of a confusion matrix represent the number of data points correctly classified (the predicted label is equal to the true label) by the model. In contrast, the off-diagonal elements are points that have been mislabeled or misclassified [94]. Most data points for both tables 6 and 7 lie on the main diagonal. Thus, signalling that the models are capable of predicting the five physical properties.

		PREDICTED Classification					
		Smoothness	Thickness	Warmth	Flexibility	Softness	Total
ACTUAL Classification	Smoothness	4144 (53%)	896 (11%)	968 (12%)	600 (8%)	1200 (15%)	7808 (100%)
	Thickness	1464 (20%)	2048 (27%)	824 (11%)	1720 (23%)	1448 (19%)	7504 (100%)
	Warmth	1640 (22%)	1304 (18%)	2184 (30%)	1056 (14%)	1200 (16%)	7384 (100%)
	Flexibility	808 (11%)	928 (13%)	592 (8%)	3568 (48%)	1504 (20%)	7400 (100%)
	Softness	1472 (19%)	976 (13%)	704 (9%)	1520 (20%)	2976 (39%)	7648 (100%)
	Total	9528	6152	5272	8464	8328	37744

Table 6: Confusion matrix - FC PRHM Property Classification using EMG features of Dataset V2

The confusion matrix for the FC PRHM Property Classification using the EMG features of Dataset V2 (see Table 6) show that the model sometimes got confused when classifying thickness. The model confused the thickness property with the smoothness and flexibility properties 20% and 23% of the time. This model also misclassified warmth as smoothness 22% of the time. The confusion matrix for the LSTM PRHM Property Classification using all 180 features of Dataset V2 (see Table 7) shows that this model sometimes got confused when classifying softness. This model mislabeled softness as thickness and flexibility 22% and 21% of the time.

		PREDICTED Classification					
		Smoothness	Thickness	Warmth	Flexibility	Softness	Total
ACTUAL Classification	Smoothness	960 (38%)	408 (16%)	424 (17%)	392 (15%)	360 (14%)	2544 (100%)
	Thickness	264 (11%)	1008 (40%)	328 (13%)	384 (15%)	528 (21%)	2512 (100%)
	Warmth	320 (13%)	416 (17%)	1008 (41%)	496 (20%)	224 (9%)	2464 (100%)
	Flexibility	272 (11%)	264 (11%)	184 (8%)	1400 (58%)	312 (13%)	2432 (100%)
	Softness	232 (9%)	576 (22%)	296 (11%)	544 (21%)	944 (36%)	2592 (100%)
	Total	2048	2672	2240	3216	2368	12544

Table 7: Confusion matrix - LSTM PRHM Property Classification using all features of Dataset V2

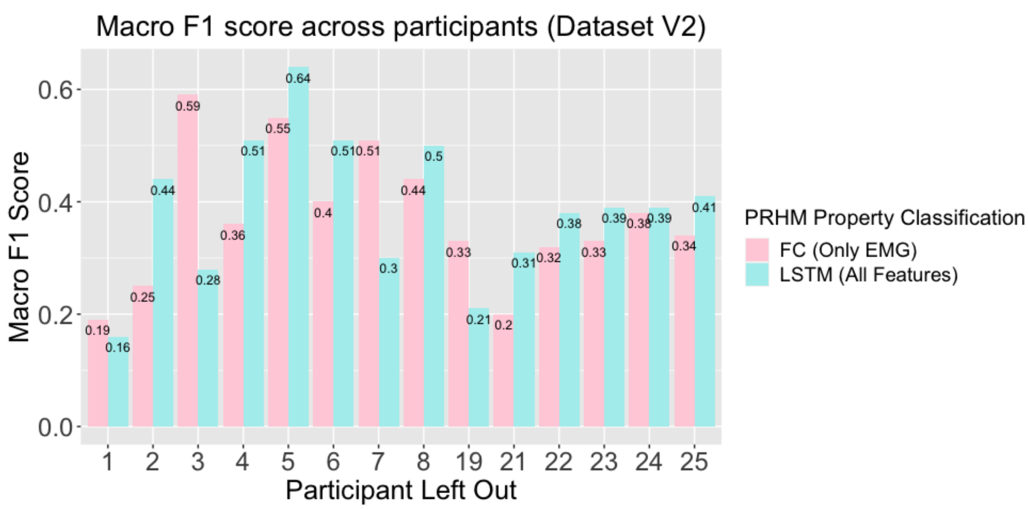


Figure 13: Macro F1 score across participants for the best performing models (Dataset V2)

Leave One Out Cross-Validation (LOOCV) was used in all 18 PRHM Property Classification Models. LOOCV trains the machine learning model using all but one participant ('leaves out' one participant) and assesses the model performance using the participant left out. This process is repeated for each participant in the dataset. Figure 13 shows the macro F1 score across all participants when predicting properties using the FC PRHM Property Classification model using only the EMG features of Dataset V2 and the LSTM PRHM Property Classification using all 180 features of Dataset V2. Based on Figure 13, both models could classify physical properties better for participants 1-8 than for participants 19-25. This means that both models classified properties better for the existing dataset than for the new dataset. However, the macro F1 scores when participants 19-25 were left out have less variation than when participants 1-8 were left out.

Using Dataset V3 (15 window segmentation with 10 timesteps)

The FC PRHM Property Classification model that used all 180 features provided the best results when Dataset V3 was used. This model was the overall best model when predicting properties and had an average percentage classification accuracy of 45% and an average macro F1 score of 0.41 (Refer Tables 2 and 3). The LSTM PRHM Property Classification model that used only the EMG features classified properties second best when Dataset V3 was used.

The FC PRHM Property Classification using all 180 features model predicted all five physical properties accurately much better than the chance level. The smoothness, thickness, flexibility and softness properties were predicted accurately 49%, 48%, 65% and 40% of the time (Refer to the diagonal of Table 8). Warmth was also predicted 24% of the time accurately. Approximately 50% of the time, warmth was misclassified as thickness or flexibility. However, The LSTM PRHM Property Classification using only the EMG features correctly predicted warmth 45% of the time (from Table 9).

		PREDICTED Classification					
		Smoothness	**Thickness**	**Warmth**	**Flexibility**	**Softness**	**Total**
ACTUAL Classification	**Smoothness**	12328 (49%)	4640 (18%)	2000 (8%)	3936 (16%)	2456 (10%)	25360 (100%)
	Thickness	3488 (14%)	11904 (48%)	1120 (5%)	5424 (22%)	2864 (12%)	24800 (100%)
	Warmth	3704 (14%)	5184 (20%)	6240 (24%)	7216 (28%)	3280 (13%)	25624 (100%)
	Flexibility	2776 (11%)	3040 (12%)	592 (2%)	15920 (65%)	2304 (9%)	24632 (100%)
	Softness	3112 (12%)	4416 (17%)	1944 (8%)	5824 (23%)	10264 (40%)	25560 (100%)
	Total	25408	29184	11896	38320	21168	125976

Table 8: Confusion matrix - FC PRHM Property Classification using all features of Dataset V3

		PREDICTED Classification					
		Smoothness	**Thickness**	**Warmth**	**Flexibility**	**Softness**	**Total**
ACTUAL Classification	**Smoothness**	1136 (47%)	384 (16%)	480 (20%)	168 (7%)	248 (10%)	2416 (100%)
	Thickness	392 (15%)	728 (28%)	592 (23%)	528 (21%)	320 (12%)	2560 (100%)
	Warmth	600 (22%)	296 (11%)	1256 (45%)	384 (14%)	240 (9%)	2776 (100%)
	Flexibility	176 (8%)	192 (8%)	280 (12%)	1256 (54%)	416 (18%)	2320 (100%)
	Softness	344 (14%)	376 (15%)	352 (14%)	584 (24%)	816 (33%)	2472 (100%)
	Total	2648	1976	2960	2920	2040	12544

Table 9: Confusion matrix - LSTM PRHM Property Classification using EMG features of Dataset V3

The FC PRHM Property Classification model that used all the features predicted well for both the existing and newly collected data. However, the LSTM PRHM model that only used the EMG features classified properties better for the existing dataset than the newly collected dataset. This can clearly be observed Figure 14.

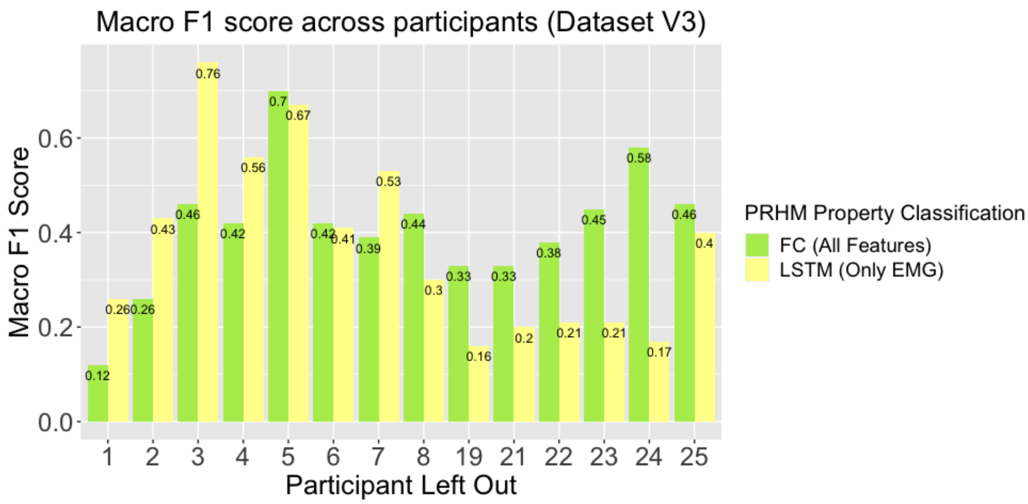


Figure 14: Macro F1 score across participants for the best performing models (Dataset V3)

5.1.2 PRHM Rating Classification Results

Tables 10 and 11 show the average weighted F1 score and percentage classification accuracy when the 18 PRHM Rating Classification models were used. On average, the models where only the EMG features (+ the one hot encoded properties) were used performed better than when all 180 features were used.

The LSTM PRHM Rating Classification model classifies ratings the best for all three variations of the dataset. Further, the confusion matrices for both neural network PRHM Rating Classification models were similar for all three dataset variations. Therefore, the confusion matrices of the RF PRHM Rating Classification model and the LSTM PRHM Rating Classification model will be compared for each dataset variation.

Using Dataset V1 (3 window segmentation)

On average, the PRHM Rating Classification models performed better using only the EMG (+ one hot encoded properties) for this dataset variation. The RF PRHM Rating Clas-

	Dataset V1		Dataset V2		Dataset V3	
	All Features	Only EMG	All Features	Only EMG	All Features	Only EMG
RF PRHM Rating Classification	0.38	0.37	0.38	0.37	0.37	0.37
LSTM PRHM Rating Classification	0.72	0.73	0.67	0.62	0.59	0.62
FC PRHM Rating Classification	0.62	0.64	0.53	0.59	0.6	0.6

* The corresponding properties were one hot encoded and added as input features

Table 10: Average weighted F1 scores when predicting property ratings*

	Dataset V1		Dataset V2		Dataset V3	
	All Features	Only EMG	All Features	Only EMG	All Features	Only EMG
RF PRHM Rating Classification	50.6%	49.3%	48.7%	49.1%	47.0%	47.9%
LSTM PRHM Rating Classification	76.5%	78.6%	71.6%	65.3%	62.2%	63.3%
FC PRHM Rating Classification	67.1%	70.6%	59.5%	64.2%	65.0%	64.6%

* The corresponding properties were one hot encoded and added as input features

Table 11: Average percentage classification accuracy when predicting property ratings*

sification had an average weighted F1 score of 0.37, whereas the LSTM PRHM Rating Classification had an average weighted F1 score of 0.73.

From Table 12 we can see that although the RF Rating Classification model classifies ratings that fall into the medium category correctly 94% of the time, this model performs very poorly when classifying low and high ratings. On the contrary, the LSTM Rating Classification model classified medium and high ratings very well (94% and 60% respectively) and low ratings accurately 31% of the time (see Table 13).

		PREDICTED Classification			
		Low	**Medium**	**High**	**Total**
ACTUAL Classification	**Low**	**11 (4%)**	277 (92%)	12 (4%)	300 (100%)
	Medium	28 (4%)	**601 (94%)**	13 (2%)	642 (100%)
	High	8 (3%)	301 (95%)	**9 (3%)**	318 (100%)
	Total	47	1179	34	1260

Table 12: Confusion matrix - RF PRHM Rating Classification using only the EMG features + one hot encoded properties of Dataset V1

Using Dataset V2 (15 window segmentation with 3 timesteps)

The PRHM Rating Classification models performed better using all features (+ one hot encoded properties) for Dataset V2. Similar to Dataset V1, the RF Rating Classification

		PREDICTED Classification			
		Low	Medium	High	Total
ACTUAL Classification	Low	**184 (31%)**	356 (59%)	60 (10%)	600 (100%)
	Medium	30 (2%)	**1212 (94%)**	42 (3%)	1284 (100%)
	High	20 (3%)	234 (37%)	**382 (60%)**	636 (100%)
	Total	234	1802	484	2520

Table 13: Confusion matrix - LSTM PRHM Rating Classification using only EMG features + one hot encoded properties of Dataset V1

model classified medium ratings well but poorly classified low and high (see Table 14. The LSTM PRHM Rating Classification model classifies low, medium and high ratings very well (51%, 85% and 63% of the time).

		PREDICTED Classification			
		Low	Medium	High	Total
ACTUAL Classification	Low	**244 (5%)**	3772 (84%)	484 (11%)	4500 (100%)
	Medium	175 (2%)	**8512 (88%)**	942 (10%)	9629 (100%)
	High	31 (1%)	4301 (90%)	**438 (9%)**	4770 (100%)
	Total	450	16585	1864	18899

Table 14: Confusion matrix - RF PRHM Rating Classification using all features + one hot encoded properties of Dataset V2

		PREDICTED Classification			
		Low	Medium	High	Total
ACTUAL Classification	Low	**1488 (51%)**	1008 (35%)	408 (14%)	2904 (100%)
	Medium	656 (10%)	**5528 (85%)**	328 (5%)	6512 (100%)
	High	320 (10%)	848 (27%)	**1960 (63%)**	3128 (100%)
	Total	2464	7384	2696	12544

Table 15: Confusion matrix - LSTM PRHM Rating Classification using all features + one hot encoded properties of Dataset V2

Using Dataset V3 (15 window segmentation with 10 timesteps)

The PRHM Rating Classification models performed better using only the EMG (+ one hot encoded properties) of Dataset V3. The RF Rating Classification model only classified medium ratings well. The LSTM Rating Classification model classified medium and high ratings very well (78% and 57% of the time) and low ratings well (37% of the time).

		PREDICTED Classification			
		Low	Medium	High	Total
ACTUAL Classification	Low	**645 (4%)**	13624 (91%)	731 (5%)	15000 (100%)
	Medium	2220 (7%)	**28432 (89%)**	1437 (4%)	32089 (100%)
	High	312 (2%)	14472 (91%)	**1116 (7%)**	15900 (100%)
	Total	3177	56528	3284	62989

Table 16: Confusion matrix - RF PRHM Rating Classification using EMG features + one hot encoded properties of Dataset V3

		PREDICTED Classification			
		Low	Medium	High	Total
ACTUAL Classification	Low	**1064 (37%)**	1488 (51%)	352 (12%)	2904 (100%)
	Medium	856 (13%)	**5104 (78%)**	552 (8%)	6512 (100%)
	High	328 (10%)	1024 (33%)	**1776 (57%)**	3128 (100%)
	Total	2248	7616	2680	12544

Table 17: Confusion matrix - LSTM PRHM Rating Classification using only EMG features + one hot encoded properties of Dataset V3

5.2 Can the PRHM Property Classification Model Generalise to New Data?

The PRHM Property Classification model was built using the existing data and its performance was assessed on the newly collected data. This was done to observe if the model could be generalised to a new type of garment that it had never seen before. The two best PRHM property classification models from the previous section was used for this purpose.

5.2.1 FC PRHM Property Classification using all 180 features of the 15 window segmentation with 10 timesteps data

This model had a macro F1 score of 0.22, an average percentage classification accuracy of 25.2% and generalised well for properties such as smoothness, warmth and flexibility. It accurately predicted smoothness, warmth and flexibility 39%, 25% and 46% of the time (Table 18). However, the model struggled to predict the thickness and softness properties correctly.

		PREDICTED Classification					
		Smoothness	Thickness	Warmth	Flexibility	Softness	Total
ACTUAL Classification	Smoothness	4067 (39%)	702 (7%)	2796 (27%)	2588 (25%)	197 (2%)	10350 (100%)
	Thickness	2447 (24%)	1080 (10%)	2573 (25%)	3933 (38%)	317 (3%)	10350 (100%)
	Warmth	2702 (26%)	1249 (12%)	2629 (25%)	3641 (35%)	268 (3%)	10489 (100%)
	Flexibility	2433 (23%)	1040 (10%)	1727 (16%)	4823 (46%)	477 (5%)	10500 (100%)
	Softness	2144 (20%)	1172 (11%)	2450 (23%)	4200 (40%)	534 (5%)	10500 (100%)
	Total	13793	5243	12175	19185	1793	52189

Table 18: Confusion matrix - FC PRHM Property Classification using all 180 features of the 15 window segmentation with 10 timesteps data

5.2.2 LSTM PRHM Property Classification using all 180 features of the 15 window segmentation with 3 timesteps data

This model had a macro F1 score of 0.23, an average percentage classification accuracy of 25.8% and generalised well for some properties. It accurately predicted smoothness, flexibility and softness 36%, 53% and 27% of the time (Table 18). However, the model struggled to predict the thickness and warmth properties correctly.

		PREDICTED Classification					
		Smoothness	Thickness	Warmth	Flexibility	Softness	Total
ACTUAL Classification	Smoothness	195 (36%)	29 (5%)	53 (10%)	152 (28%)	111 (21%)	540 (100%)
	Thickness	85 (16%)	35 (6%)	39 (7%)	230 (43%)	151 (28%)	540 (100%)
	Warmth	74 (14%)	13 (2%)	36 (7%)	259 (48%)	158 (29%)	540 (100%)
	Flexibility	89 (16%)	11 (2%)	10 (2%)	285 (53%)	145 (27%)	540 (100%)
	Softness	138 (26%)	16 (3%)	29 (5%)	212 (39%)	145 (27%)	540 (100%)
	Total	581	104	167	1138	710	2700

Table 19: Confusion matrix - LSTM PRHM Property Classification using all 180 features of the 15 window segmentation with 3 timesteps data

Chapter 6

Conclusion

This project successfully implemented a novel machine learning method to infer tactile affect in the context of fabric handling. 18 (3 dataset variations $\times$ 3 model architectures $\times$ 2 sets of features) Property Rating Hierarchical Models (PRHM) were built to predict the property assessed and the corresponding rating based on EMG, accelerometer and quaternion data.

On average, the PRHM Property Classification models where all 180 features were used performed better than when only the EMG features were used. The Fully-Connected (FC) PRHM Property Classification using all 180 features of Dataset V3 (15 window segmentation with 10 timesteps) predicted properties the best. This model resulted in a macro F1 score of 0.41 and overall classification accuracy of 45%. This model also had a classification accuracy of 49%, 48%, 24%, 65% and 40% for the smoothness, thickness, warmth, flexibility and softness properties, respectively.

The Long Short-Term Memory (LSTM) PRHM Rating Classification models performed well for all three dataset variations. The LSTM PRHM using only the EMG features (+ one hot encoded properties) of Dataset V1 (3 window segmentation) had the best overall classification performance. This model yielded a weighted F1 score of 0.73, overall classification accuracy of 78.6% and predicted low, medium and high ratings with a classification accuracy of 31%, 94% and 60%, respectively. The Random Forest (RF) PRHM Rating Classification models had a similar classification performance for all three dataset variations. These models classified medium ratings well (90% of the time) but failed to classify low and high ratings.

Therefore, the Property Rating Hierarchical Models (PRHM) built could accurately predict the property assessed and the corresponding rating based on pre-processed EMG, accelerometer and quaternion data. PRHM Property Classification model could also generalise to a new type of garment that it had never seen before. The PRHM Property Classification models accurately classified the smoothness, warmth and flexibility properties but failed to classify the thickness and softness properties. As the PRHM Property Classification model could generalise to new clothes, it can give consumers a better understanding of a garment's textual and physical properties when shopping for clothes. With further research, this model has the potential to reduce wastage caused in the fashion industry.

The limitations of this project include a small sample size as the dataset used was small. Further, some models had to be simplified as the PRHM required sizeable computational power. Future research could focus on further existing the available dataset. This study could also be replicated in the future by factoring in what fabric the assessed clothes are made of, as there will be a significant difference in the tactile experience between different fabrics. Future research could also be conducted to develop the mentioned chatbot and create a two-part PRHM that uses the predicted properties as input for classifying the ratings.

Bibliography

[1] Rashmila Maiti. *Fast Fashion and Its Environmental Impact*. https://earth.org/fast-fashions-detrimental-effect-on-the-environment/. (accessed: 07/08/2022). June 2022.

[2] Rachael Dottle and Jackie Gu. *The Global Gut of clothing Is An Environmental Crisis*. https://www.bloomberg.com/graphics/2022-fashion-industry-environmental-impact/#xj4y7vzkg. (accessed: 09/08/2022). Feb. 2022.

[3] Waste2Fresh. *Dye Pollution in the Textile Industry*. https://waste2fresh.eu/dye-pollution-in-the-textile-industry/. (accessed: 09/08/2022). June 2021.

[4] Maja Pantic et al. "Affective Multimodal Human-Computer Interaction". In: *Proceedings of the 13th Annual ACM International Conference on Multimedia*. New York, NY, USA: Association for Computing Machinery, 2005, pp. 669–676. ISBN: 1595930442. DOI: 10.1145/1101149.1101299. URL: https://doi.org/10.1145/1101149.1101299.

[5] Olivia Lai. *What is Fast Fashion?* https://earth.org/what-is-fast-fashion/. (accessed: 07/08/2022). Nov. 2021.

[6] Solene Rauturier. *What Is Fast Fashion and Why Is It So Bad?* https://goodonyou.eco/what-is-fast-fashion/. (accessed: 07/08/2022). Apr. 2022.

[7] United Nations. *Fashion Industry, UN Pursue Climate Action for Sustainable Development*. https://unfccc.int/news/fashion-industry-un-pursue-climate-action-for-sustainable-development. (accessed: 09/08/2022). Jan. 2018.

[8] Morgan McFall-Johnsen. *The fashion industry emits more carbon than international flights and maritime shipping combined. Here are the biggest ways it impacts the planet.* https://www.businessinsider.com/fast-fashion-environmental-impact-pollution-emissions-waste-water-2019-10?r=US&IR=T. (accessed: 09/08/2022). Oct. 2019.

[9] Gary Cook and Maya Rommwatt. *FASHION FORWARD: A Roadmap to Fossil Free Fashion*. https://www.stand.earth/sites/stand/files/standearth-fashionforward-roadmaptofossilfreefashion.pdf. (accessed: 09/08/2022).

[10] Julien Boucher and Damien Friot. *Primary Microplastics in the Oceans: A Global Evaluation of Sources*. Jan. 2017. ISBN: 978-2-8317-1827-9. DOI: 10.2305/IUCN.CH.2017.01.en.

[11] The United Nations Economic Commission for Europe. *Fashion and the SDGs: what role for the UN?* https://unece.org/fileadmin/DAM/RCM_Website/RFSD_2018_Side_event_sustainable_fashion.pdf. (accessed: 09/08/2022). Mar. 2018.

[12] Elizabeth Reichart and Deborah Drew. *By the Numbers: The Economic, Social and Environmental Impacts of "Fast Fashion"*. https://www.wri.org/insights/numbers-economic-social-and-environmental-impacts-fast-fashion. (accessed: 09/08/2022). Jan. 2019.

[13] United Nations Environment Programme. *Putting the brakes on fast fashion*. https://www.unep.org/news-and-stories/story/putting-brakes-fast-fashion. (accessed: 09/08/2022). Nov. 2018.

[14] India Morrison. "ALE meta-analysis reveals dissociable networks for affective and discriminative aspects of touch". In: *Human Brain Mapping* 37 (Feb. 2016). DOI: 10.1002/hbm.23103.

[15] Joann Peck and Terry Childers. "To Have and To Hold: The Influence of Haptic Information on Product Judgments". In: *Journal of Marketing - J MARKETING* 67 (Apr. 2003), pp. 35–48. DOI: 10.1509/jmkg.67.2.35.18612.

[16] Sylvia C. Mooy and Henry S. J. Robben. "Managing consumers' product evaluations through direct product experience". In: *Journal of Product & Brand Management* 11 (2002), pp. 432–446.

[17] Roberta L. Klatzky and Susan J. Lederman. "There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision". In: *Journal of Experimental Psychology* 116.4 (1987), pp. 356–369. DOI: 10.1509/jmkg.67.2.35.18612.

[18] Susan J. Lederman, Georgie Thorne, and Bill Jones. “Perception of texture by vision and touch: multidimensionality and intersensory integration.” In: *Journal of experimental psychology. Human perception and performance* 12.2 (1986), pp. 169–80.

[19] Morris B. Holbrook. “Aims, concepts, and methods for the representation of individual differences in esthetic responses to design features”. In: *Journal of Consumer Research* 13 4 (1986), pp. 337–347. DOI: `https://doi.org/10.1086/209073`.

[20] Deborah Brown Mccabe and Stephen M. Nowlis. “The Effect of Examining Actual Products or Product Descriptions on Consumer Preference”. In: *Journal of Consumer Psychology* 13 (2003), pp. 431–439.

[21] Bianca Grohmann, Eric R. Spangenberg, and David E. Sprott. “The Influence of tactile input on the evaluation of retail product offerings”. In: *Journal of Retailing - J RETAIL* 83 (Apr. 2007), pp. 237–245. DOI: `10.1016/j.jretai.2006.09.001`.

[22] Alka Varma Citrin et al. “Consumer need for tactile input: An internet retailing challenge”. In: *Journal of Business Research* 56.11 (2003). Strategy in e-marketing, pp. 915–922. ISSN: 0148-2963. DOI: `https://doi.org/10.1016/S0148-2963(01)00278-8`. URL: `%5Curl%7Bhttps://www.sciencedirect.com/science/article/pii/S0148296301002788%7D`.

[23] Deborah H. Lester, Andrew M. Forman, and Dolly Loyd. “Internet Shopping and Buying Behavior of College Students”. In: *Services Marketing Quarterly* 27 (2005), pp. 123–138.

[24] Aliya Ram. *UK retailers count the cost of returns*. `https://www.ft.com/content/52d26de8-c0e6-11e5-846f-79b0e3d20eaf`. (accessed: 15/08/2022). Jan. 2016.

[25] Lili Lin. “Inferring Fabric-Hand Perception from Wearable Sensors [Unpublished MSc dissertation]”. In: (2021).

[26] Lucy Hughes et al. “Crowdsourcing an Emotional Wardrobe”. In: *CHI ’12 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’12. Austin, Texas, USA: Association for Computing Machinery, 2012, pp. 231–240. ISBN: 9781450310161. DOI: `10.1145/2212776.2212801`. URL: `https://doi.org/10.1145/2212776.2212801`.

[27] Chayapa Katawetawaraks and Cheng Lu Wang. “Online Shopper Behavior: Influences of Online Shopping Decision”. In: *eBusiness & eCommerce eJournal* (2013).

[28] Michele A. Burton et al. "Crowdsourcing subjective fashion advice using vizwiz: Challenges and opportunities". English (US). In: *ASSETS'12 - Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility.* ASSETS'12 - Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility. 14th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2012 ; Conference date: 22-10-2012 Through 24-10-2012. 2012, pp. 135–142. ISBN: 9781450313216. DOI: 10.1145/2384916.2384941.

[29] Matthew J. Hertenstein et al. "The communication of emotion via touch." In: *Emotion* 9.4 (2009), pp. 566–73.

[30] Matthew J. Hertenstein et al. "Touch communicates distinct emotions." In: *Emotion* 6.3 (2006), pp. 528–33.

[31] Matthew J. Hertenstein. "Touch". In: vol. 3. Sage Publications, 2005.

[32] Matthew Hertenstein and Joseph Campos. "Emotion Regulation Via Maternal Touch". In: *Infancy* 2 (Sept. 2001), pp. 549–566. DOI: 10.1207/S15327078IN0204_09.

[33] Stanley E. Jones and A. Elaine Yarbrough. "A naturalistic study of the meanings of touch". In: (1985).

[34] Mark L. Knapp and Judith A. Hall. *Nonverbal Communication in Human Interaction.* 1997.

[35] Yuan Gao, Nadia Bianchi-Berthouze, and Hongying Meng. "What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay?" In: *ACM Trans. Comput.-Hum. Interact.* 19.4 (Dec. 2012), 31:1–31:30. ISSN: 1073-0516. DOI: 10.1145/2395131.2395138. URL: %5Curl%7Bhttp://doi.acm.org/10.1145/2395131.2395138%7D.

[36] Hillary Elfenbein and Nalini Ambady. "On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis". In: *Psychological bulletin* 128 (Apr. 2002), pp. 203–35. DOI: 10.1037/0033-2909.128.2.203.

[37] Klaus R. Scherer, Tom Johnstone, and Gundrun Klasmeyer. "Vocal expression of emotion". In: Oxford University Press, 2003.

[38] Sachin Shah, J. Teja, and Samit Bhattacharya. "Towards affective touch interaction: predicting mobile user emotion from finger strokes". In: *Journal of Interaction Science* 3 (Dec. 2015). DOI: 10.1186/s40166-015-0013-z.

[39] Lie Lu, Dan Liu, and Hong-Jiang Zhang. "Automatic mood detection and tracking of music audio signals". In: *Audio, Speech, and Language Processing, IEEE Transactions on* 14 (Feb. 2006), pp. 5–18. DOI: 10.1109/TSA.2005.860344.

[40] KuanTing Liu and Roger Reimer. "Social playlist: enabling touch points and enriching ongoing relationships through collaborative mobile music listening". In: Jan. 2008, pp. 403–406. DOI: 10.1145/1409240.1409299.

[41] Nadia Bianchi-Berthouze and Christine Lisetti. "Modeling Multimodal Expression of User's Affective Subjective Experience". In: *Bianchi-Berthouze, N. and Lisetti, C.L. (2002) Modeling Multimodal Expression of User's Affective Subjective Experience. User Modeling and User-Adapted Interaction, 12 (1). pp. 49-84. ISSN 09241868* 12 (Feb. 2002). DOI: 10.1023/A:1013365332180.

[42] Michelle Thrasher et al. "Mood Recognition Based on Upper Body Posture and Movement Features". In: Jan. 2011, pp. 377–386. ISBN: 978-3-642-24599-2. DOI: 10.1007/978-3-642-24600-5_41.

[43] Petra Sundström, Anna Ståhl, and Kristina Höök. "In situ informants exploring an emotional mobile messaging system in their everyday practice". In: *International Journal of Human-Computer Studies* 65 (Apr. 2007), pp. 388–403. DOI: 10.1016/j.ijhcs.2006.11.013.

[44] Duncan Graham-Rowearchive page. *A Smart Phone that Knows You're Angry.* https://www.technologyreview.com/2012/01/09/188503/a-smart-phone-that-knows-youre-angry/. (accessed: 07/08/2022). Jan. 2012.

[45] Sidney D'Mello and Art Graesser. "The half-life of cognitive-affective states during complex learning". In: *Cognition & emotion* 25 (Sept. 2011), pp. 1299–308. DOI: 10.1080/02699931.2011.613668.

[46] Panagiotis Tsinganos et al. "Deep Learning in EMG-based Gesture Recognition". In: Sept. 2018, pp. 107–114. DOI: 10.5220/0006960201070114.

[47] Kaixuan Chen et al. "Deep Learning for Sensor-based Human Activity Recognition". In: *ACM Computing Surveys (CSUR)* 54 (2021), pp. 1–40.

[48] Praneeth Vepakomma et al. "A-Wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities". In: *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (2015), pp. 1–6.

[49] Eoin Brophy et al. "An Interpretable Machine Vision Approach to Human Activity Recognition using Photoplethysmograph Sensor Data". In: *ArXiv* abs/1812.00668 (2018).

[50] Tâm Huynh and Bernt Schiele. "Analyzing features for activity recognition". In: *sOc-EUSAI '05*. 2005.

[51] Jessica Lin et al. "A symbolic representation of time series, with implications for streaming algorithms". In: *DMKD '03*. 2003.

[52] Yoshua Bengio and Yann LeCun. "Scaling learning algorithms towards AI". In: 2007.

[53] Philipp V. Rouast, Marc Thomas Philipp Adam, and Raymond Chiong. "Deep Learning for Human Affect Recognition: Insights and New Developments". In: *IEEE Transactions on Affective Computing* 12 (2021), pp. 524–543.

[54] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[55] Yoshua Bengio, Aaron C. Courville, and Pascal Vincent. "Representation Learning: A Review and New Perspectives". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013), pp. 1798–1828.

[56] Nils Y. Hammerla, Shane Halloran, and Thomas Plötz. "Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables". In: *ArXiv* abs/1604.08880 (2016).

[57] Klaus Greff et al. "LSTM: A search space odyssey". In: *IEEE transactions on neural networks and learning systems* 28 (Mar. 2015). DOI: `10.1109/TNNLS.2016.2582924`.

[58] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-term Memory". In: *Neural computation* 9 (Dec. 1997), pp. 1735–80. DOI: `10.1162/neco.1997.9.8.1735`.

[59] Kyunghyun Cho et al. "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation". In: *EMNLP*. 2014.

[60] Haodong Guo et al. "Wearable sensor based multimodal human activity recognition exploiting the diversity of classifier ensemble". In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2016).

[61] John Hopkins Medicine. *Electromyography (EMG)*. `https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electromyography-emg`. (accessed: 20/08/2022).

[62] Kejia Wang. "Using EMG and Accelerometer data to Capture the Tactile Experience with Fabrics [Unpublished MSc dissertation]". In: (2017).

[63] T. Scott Saponas et al. "Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces". In: *CHI*. 2008.

[64] Lesley Cary. "Exploring a language of gestures and emotional responses to textiles [Unpublished MSc dissertation]". In: (2013).

[65] Shah Rukh Humayoun, Munir Ahmad, and Achim Ebert. "3D Accelerometer-based Gestures for Interacting with Mobile Devices". In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (2016).

[66] Yuhan Gao. "Automatic Recognition of Properties Assessed in the Tactile Exploration of Textiles [Unpublished MSc dissertation]". In: (2019).

[67] OYMotion. *gForcePro/gForcePro+ Armband and gForceOct Module User Guide*. `https://oymotion.github.io/gForcePro/gForcePro/`. (accessed: 21/08/2022). Aug. 2020.

[68] VectorNav. *What is an Inertial Measurement Unit?* `https://www.vectornav.com/resources/inertial-navigation-articles/what-is-an-inertial-measurement-unit-imu`. (accessed: 21/08/2022).

[69] Jianyu Yang. "Chapter 4 - Bistatic SAR parameter estimation". In: *Bistatic Synthetic Aperture Radar*. Elsevier, 2022, pp. 185–215. ISBN: 978-0-12-822459-5. DOI: `https://doi.org/10.1016/B978-0-12-822459-5.00004-9`. URL: `https://www.sciencedirect.com/science/article/pii/B9780128224595000049`.

[70] OYMotion. *gForce EMG Armband User Guide*. `https://oymotion.github.io/assets/downloads/gForce-EMG-ARMBAND-User-Guide-202108.pdf`. (accessed: 21/08/2022).

[71] Moti Ben-Ari. *A Tutorial on Euler Angles and Quaternions*. `https://www.weizmann.ac.il/sci-tea/benari/sites/sci-tea.benari/files/uploads/softwareAndLearningMateria` `quaternion-tutorial-2-0-1.pdf`. (accessed: 22/08/2022).

[72] Jonathan Strickland. *What is a gimbal – and what does it have to do with NASA?* `https://science.howstuffworks.com/gimbal.htm`. (accessed: 21/08/2022).

[73] Adrian Popa. *What is meant by the term gimbal lock?* `http://www.madsci.org/posts/archives/aug98/896993617.Eg.r.html`. (accessed: 22/08/2022).

[74] Motorolla. *Motorolla - motog^9 POWER*. https://www.motorola.co.uk/smartphones-moto-g-power-gen-9/p. (accessed: 21/08/2022).

[75] *SQLite Home Page*. https://www.sqlite.org/index.html. (accessed: 22/08/2022).

[76] Julia Kho. *Why Random Forest is My Favorite Machine Learning Model*. https://towardsdatascience.com/why-random-forest-is-my-favorite-machine-learning-model-b97651fa3706. (accessed: 03/09/2022). Oct. 2018.

[77] Jasom Brownlee. *How to Develop a Random Forest Ensemble in Python*. https://machinelearningmastery.com/random-forest-ensemble-in-python/. (accessed: 03/09/2022). Apr. 2020.

[78] Bradley C. Boehmke. *Feedforward Deep Learning Models*. http://uc-r.github.io/feedforward_DNN#ff. (accessed: 05/09/2022).

[79] Turing. *Understanding Feed Forward Neural Networks With Maths and Statistics*. https://www.turing.com/kb/mathematical-formulation-of-feed-forward-neural-network. (accessed: 05/09/2022).

[80] Pooja Mahajan. *Fully Connected vs Convolutional Neural Networks*. https://medium.com/swlh/fully-connected-vs-convolutional-neural-networks-813ca7bc6ee5. (accessed: 05/09/2022).

[81] Benoit Liquet, Sarat Moka, and Yoni Nazarathy. *The Mathematical Engineering of Deep Learning*. https://deeplearningmath.org/general-fully-connected-neural-networks.html5. (accessed: 05/09/2022).

[82] Jason Brownlee. *A Gentle Introduction to Long Short-Term Memory Networks by the Experts*. https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/. (accessed: 05/09/2022).

[83] AI Sangam. *Difference between Feed Forward Neural Network and RNN*. https://www.aisangam.com/blog/difference-between-feed-forward-neural-network-and-rnn/. (accessed: 05/09/2022).

[84] M Raez, Md Hussain, and Faisal Mohd-Yasin. "Techniques of EMG signal analysis: Detection, processing, classification and applications". In: *Biological procedures online* 8 (Feb. 2006), pp. 11–35. DOI: 10.1251/bpo115.

[85] David M. Thompson. *Electromyography (EMG)*. https://ouhsc.edu/bserdac/dthompso/web/pk/emg/emg.htm. (accessed: 15/09/2022).

[86] Leandro R. Altimari et al. "Influence of Different Strategies of Treatment Muscle Contraction and Relaxation Phases on EMG Signal Processing and Analysis During Cyclic Exercise". In: 2012.

[87] William Rose. *Electromyogram analysis.* https://www1.udel.edu/biology/rosewc/kaap686/notes/EMG%20analysis.pdf. (accessed: 15/09/2022).

[88] Alexandre Chalard et al. "Impact of the EMG normalization method on muscle activation and the antagonist-agonist co-contraction index during active elbow extension: Practical implications for post-stroke subjects". In: *Journal of Electromyography and Kinesiology* 51 (2020), p. 102403. ISSN: 1050-6411. DOI: https://doi.org/10.1016/j.jelekin.2020.102403. URL: https://www.sciencedirect.com/science/article/pii/S1050641120300183.

[89] Gregory J. Lehman and Stuart M. McGill. "The importance of normalization in the interpretation of surface electromyography: A proof of principle". In: *Journal of Manipulative and Physiological Therapeutics* 22.7 (1999), pp. 444–446. ISSN: 0161-4754. DOI: https://doi.org/10.1016/S0161-4754(99)70032-1. URL: https://www.sciencedirect.com/science/article/pii/S0161475499700321.

[90] Martin Ling. *Quaternions in numpy.* https://quaternion.readthedocs.io/en/latest/. (accessed: 10/09/2022).

[91] Brinnae Bent. *Step-by-Step Guide to leave-one-person-out Cross Validation with Random Forests in Python.* https://medium.com/analytics-vidhya/step-by-step-guide-to-leave-one-person-out-cross-validation-with-random-forests-in-python-34b2eaefb628. (accessed: 20/09/2022). June 2020.

[92] Jason Brownlee. *LOOCV for Evaluating Machine Learning Algorithms.* https://machinelearningmastery.com/loocv-for-evaluating-machine-learning-algorithms/. (accessed: 15/08/2022). July 2020.

[93] Jason Brownlee. *How to Choose Loss Functions When Training Deep Learning Neural Networks.* https://machinelearningmastery.com/how-to-choose-loss-functions-when-training-deep-learning-neural-networks/. (accessed: 15/08/2022). Aug. 2020.

[94] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[95] Jason Brownlee. *What is a Confusion Matrix in Machine Learning*. https://machinelearningmastery.com/confusion-matrix-machine-learning/. (accessed: 21/09/2022). Nov. 2016.

[96] Margherita Grandini, Enrico Bagli, and Giorgio Visani. "Metrics for Multi-Class Classification: an Overview". In: *ArXiv* abs/2008.05756 (2020).

[97] Jason Brownlee. *How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification*. https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/. (accessed: 15/08/2022). Jan. 2020.

[98] Purva Huilgol. *Precision vs. Recall – An Intuitive Guide for Every Machine Learning Person*. https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning/. (accessed: 21/09/2022). Nov. 2016.

[99] Ina S. Markham and Terry R. Rakes. "The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression". In: *Comput. Oper. Res.* 25 (1998), pp. 251–263.

[100] Jayme Garcia Arnal Barbedo. "Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification". In: *Comput. Electron. Agric.* 153 (2018), pp. 46–53.

[101] Eric W. Weisstein. *Diagonal*. https://mathworld.wolfram.com/Diagonal.html. (accessed: 22/09/2022).

Appendix A: Documents related to the newly collected dataset

A.1 Information Sheet

UCL Interaction Centre
2nd floor, 66 – 72 Gower St
London WC1E 6EA

INFORMATION SHEET FOR PARTICIPANT

Project Title: TCC: Affective sensing technology: **The affective tactile language in clothing attachment**

Department: University College London Interaction Centre

Researcher: Yuanze Gan yuanze.gan.20@ucl.ac.uk, Alice Sansoni, a.sansoni@ucl.ac.uk, Nihara Warawita nihara.warawita.21@ucl.ac.uk,

Principal Researcher: Prof Nadia Berthouze, nadia.berthouze@ucl.ac.uk

Data Protection Email: data-protection@ucl.ac.uk

This study has been approved by the UCLIC Research Ethics Committee: UCLIC_2021_018_Berthouze_PE

1. Invitation Paragraph

You are being invited to take part in a research project. Before you decide whether to take part, it is important for you to understand why the research is being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

2. What is the project's purpose?

This project is part of the Textile Circularity Centre (TCC) funded by UK Research & Innovation, to understand important factors in sustainable fashion consumption and facilitate the development of a textile circular economy.

In this project, we explore the role of touch when buying clothes, or its role in building and expressing attachment. Specifically, we want to understand how gestures and movements with items correspond to different material properties and how much the cloth represents that property by using sensing technology. The extracted data could be visualised help material understanding. Our findings will inform the design of technology which encourages consumers to connect with and reflect upon their clothing items, to facilitate sustainable and circular textile consumption.

Information about the larger Textile Circularity Centre project can be found at: https://www.rca.ac.uk/research-innovation/research-centres/materials-science-research-centre/textiles-circularity-centre/.

3. Why have I been chosen?

We are asking people caring about the sense of touch when interacting with clothes, living in London and are at least 20 years-old to take part in our study.

4. Do I have to take part?

It is up to you to decide whether or not to take part in this study. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form. You can withdraw from the study at any time and without giving a reason. You will not in any way be penalised for withdrawing from the study.

You can additionally withdraw data collected during the study up till 2 days after your study session, without giving a reason. To withdraw from the study or withdraw your data, you should contact the researcher, using the contact details at the top of this information sheet. We will no longer use data that you withdraw if it is withdrawn no more than 2 days after the last data we collected from you. If you choose to withdraw from the study, you will be asked to revoke your original consent by signing a consent change form.

5. What will happen to me if I take part?

You will be asked to wear two Armbands [GForcePro+] (http://www.oymotion.com/en/product32/149). This armband measures the muscle activity in your arm and your arm movement. They do not harm your skin or hurt you in anyway. We use these devices to measure the muscle activity and movement of your arm while touching the six sample clothes you will be asked to engage with during the experiment. You will be asked to freely touch 6 items assessing different properties (e.g., its softness) of the fabrics. A mobile app will guide you in the use of the sensors and allow you to complete a questionnaire during the experience. The questionnaire will ask you to rate the properties of the clothes material after touching it. We will take a video of you while touching the fabrics, trying to avoid including your face, keeping only your hands and arms. The goal of this study is to use EMG data to build a AI-system that can infer your perception of the fabric qualities. The study will last 1h 15min.

The study will follow COVID-19 safety measures (see at the end of this information sheet for covid-safety procedure)

DEVICES:

- All of our study devices (smartphone, EMG band) are cleaned and disinfected after each participant.
- The researchers will use gloves to give you the devices

6. Will I be recorded and how will the recorded media be used?

NO ONE outside of our research team will be allowed access to any written or verbal information which can be used to personally identify you that you have provided in your communication with us or when giving your consent unless you have given us permission to do so (e.g., showing your video).

Your Responses - What we will record:

a. The gathered movement and physiological (e.g., muscle activity) data from sensors, video/image, audio, and questionnaire and diary recordings of your activities, as described above, will be used for our analysis. We would further use examples from the sensor recordings of your activities and your responses to our questions, neither of which can be used to personally identify you, in scientific publications (including academic theses) and presentations.

b. Questionnaires data - Your responses to our questionnaires will be recorded in digital format either as online forms or other electronic documents, e.g. pdf forms.

c. Video/Audio data – We will video record the study to capture yours hand and arms movement. We will try to avoid to capture your face.

How we will use what we record:

- The data we record (letter, sensor data, questionnaire ratings and transcribed interview) will be used for analysis. Anonymized data will be shared with our collaborators UKRI Textile Circularity Centre (https://www.rca.ac.uk/research-innovation/research-centres/materials-science-research-centre/textiles-circularity-centre/) and related project (e.g. CX) (https://www.rca.ac.uk/research-innovation/research-centres/materials-science-research-centre/textiles-circularity-centre/) that the principal researcher Nadia Berthouze is part of.

- ADDITIONAL USE OF DATA THAT CAN BE USED TO IDENTIFY YOU: We will never disclose your name or contact details with the data gathered. We will not share the video or the audio outside our UCL research group unless you have given us written permission to do so.

- ADDITIONAL USE OF DATA THAT CANNOT BE USED TO IDENTIFY YOU If you give us permission, we will share anonymised data (sensor data, questionnaires responses), which CANNOT be used to identify you, with the wider research community, without your names or contact details included. This will further support advance in better understanding of movement and the design of related AI-technology.

7. What are the possible benefits of taking part?

As a thank you for your time, you will receive £15 as payment. There are no immediate benefits for those people participating in the project. It is hoped that this work will shape future research on movement and touch sensing technology and technology to automatically detect affective experiences from movement and touch. We also hope our research will contribute to design technology that foster wellbeing and create a more sustainable and circular economy.

8. What if something goes wrong?

Extreme care will be taken in this research. However, if you wish to complain or have any concerns that are not addressed by the researcher, you should contact Prof Nadia Berthouze (nadia.berthouze@ucl.ac.uk) who is the Principal Researcher on the project. If you further feel that your complaint has not been handled to your satisfaction, you can contact the Chair of the UCL Research Ethics Committee (ethics@ucl.ac.uk).

9. Will my taking part in this project be kept confidential?

Your name and contact details will be kept strictly confidential. We would further only share DE-IDENTIFIED questionnaire, transcripts and sensor data with other researchers, and only if you give us permission to.

10. Limits to confidentiality

Confidentiality will be respected unless there are compelling and legitimate reasons for this to be breached. If this happens, we will inform you of any decisions that might limit your confidentiality.

11. What will happen to the results of the research project?

The findings of our analysis of the data collected from the participants of the research project will be published in reports and articles and presented at public engagement and research talk venues. You will be able to access academic publications of these findings on the project website: https://uclic.ucl.ac.uk/people/nadia-berthouze. You will not be identifiable in these publications and presentations. Your name and contact details will never be included in publications and presentations.

The questionnaire and sensor data we collect from you, which CANNOT be used to identify you, will be made open for use by other researchers, for the benefit of scientific and technology development; but only if you give us permission to do so in the consent form.

12. Local Data Protection Privacy Notice

Notice: The controller for this project will be University College London (UCL). The UCL Data Protection Officer provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk

This 'local' privacy notice sets out the information that applies to this particular study. Further information on how UCL uses participant information can be found in our 'general privacy notice: https://www.ucl.ac.uk/legal-services/privacy/ucl-general-research-participant-privacy-notice

The categories of personal data used will be as follows:

- Name, Age, Gender, contact details
- Muscle and movement activity
- Video

The lawful basis that would be used to process your personal data will be performance of a task in the public interest. The lawful basis used to process special category personal data will be for scientific and historical research or statistical purposes.

Your personal data will be processed so long as it is required for the research project. If we are able to anonymise or pseudonymise the personal data you provide we will undertake this and will endeavour to minimise the processing of personal data wherever possible. If you give us permission to make open the data to the wider research community or use it in other research projects we are involved in, processing of the personal data will continue beyond the end of the research project.

If you are concerned about how your personal data is being processed, or if you would like to contact us about your rights, please contact UCL in the first instance at data-protection@ucl.ac.uk.

Thank you for reading this information sheet and for considering taking part in this research study.

A.2 Consent Form

Information and Consent

AAA

INFORMATION SHEET

Thank you for your interest in our research project.
Please carefully read the Information Sheet attached below,
then proceed to complete the Consent Form below.

Attachment: Information Sheet.pdf (0.14 MB)

CONSENT FORM

Project Title: TCC: Affective sensing technology: The affective tactile language in clothing attachment

Department: University College London Interaction Centre

Researcher: Yuanze Gan yuanze.gan.20@ucl.ac.uk, Alice Sansoni, a.sansoni@ucl.ac.uk, Nihara Warawita nihara.warawita.21@ucl.ac.uk,

Principal Researcher: Prof Nadia Berthouze, nadia.berthouze@ucl.ac.uk

Data Protection Email: data-protection@ucl.ac.uk

This study has been approved by the UCL Interaction Centre Research Ethics Committee: UCLIC_2021_018_Berthouze_PE

Thank you for considering taking part in this research. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time. Before the study starts, you will be able to ask further information during a video conference call and you will still have the option to opt out. Please note that if you answer No to some of the questions you may not be elegible for this study.

1) **I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me. I have also had the opportunity to ask questions which have been answered to my satisfaction.**

* must provide value

- Yes
- No

2) **I confirm that I understand the recruiting criteria and I also confirm to be at least 20 years old.**

* must provide value

- Yes
- No

3) **I consent to participate in the study. I understand that my personal information, if applicable, (name, age, address) will be used to contact me, send me any materials, or used for the purpose explained to me in the Information Sheet.**

* must provide value

- Yes
- No

4) **I understand that I will be audio and video recorded to enable accurate analysis of the data. I understand that my personal information, survey data (text question responses, think aloud), sensors data (physiological data and movement data), interview transcript, audio recordings will be**

used for the purposes explained to me. I understand that according to data protection legislation, 'public task' will be the lawful basis for processing for my personal data and the lawful basis used to process any special category personal data will be for scientific and historical research or statistical purposes.

* must provide value

- Yes
- No

5) **I understand that my participation is voluntary and that I am free to withdraw from the study at any time. I understand that if I decide to withdraw from the study, I can revoke my consent by signing a consent change form. I understand that I can also withdraw data I have provided, up till the 2nd day after the experiment session I took part in is completed, without giving a reason. If I choose to also withdraw my data, the data will no longer be used.**

* must provide value

- Yes
- No

6) **I am aware of who I should contact if I wish to lodge a complaint.**

* must provide value

- Yes
- No

7) **I understand that identifiable information collected will remain confidential and that all efforts will be made to ensure that I cannot be identified except where explicit consent is given below.**

* must provide value

- Yes
- No

8) **FURTHER USE OF MY DATA (Tick the boxes to indicate you agree with the related use of your data)**

- [ ] I agree for my sensors data [body/hand movement or related physiological data], text question responses, interview transcriptions, which CANNOT be used to identify me, to be made open to other researchers and for future research or secondary analysis, to facilitate research and innovation.
- [ ] I agree for PHOTOS recorded from me (e.g., during the study) to be included in written publications used to disseminate the project findings. I understand that I could be recognised from such images and the audience may make copies of the image(s) and that the researcher will not have control over such copies.
- [ ] I agree for PHOTOS/VIDEOS/AUDIO recorded from me to be used in presentations used to disseminate the project findings. I understand that I could be recognised from such photos/videos/audio and the audience may record the presentation and that the researcher will not have control over such recordings.
- [ ] I agree for my non personal information (sensors data, text question responses, interview transcript, diary) to be used for secondary analysis in other studies that the researchers of this study are involved in, to facilitate research and innovation.
- [ ] I agree to be contacted to participate in follow up studies to this project, or in future studies of a similar nature. I understand that my name and contact details will not be shared with anyone who is not a member of this research team.
- [ ] I agree for PHOTOS/VIDEOS/AUDIO recorded from me to be shared with Alice Sansoni (University of Milano, Italy) and used in her thesis and related presentations.

9) **Preferred Contact Details (email):**

* must provide value

10) Full Name of Participant

* must provide value

11) Date

* must provide value

12) Signature

* must provide value

Submit

A.3 Order of properties and socks

Participant 1: PARTICIPANT ID 19

Cloth 3:

1^{st} prop: Softness

2^{nd} prop: Flexibility

3^{rd} prop: Smoothness

4^{th} prop: Thickness

5^{th} prop: Warmth

Enjoyment

Cloth 2:

1^{st} prop: Warmth

2^{nd} prop: Thickness

3^{rd} prop: Softness

4^{th} prop: Smoothness

5^{th} prop: Flexibility

Enjoyment

Cloth 4:

1^{st} prop: Smoothness

2^{nd} prop: Thickness

3^{rd} prop: Flexibility

4^{th} prop: Softness

5^{th} prop: Warmth

Enjoyment

Cloth 6:

1^{st} prop: Softness

2^{nd} prop: Thickness

3^{rd} prop: Flexibility

4^{th} prop: Smoothness

5^{th} prop: Warmth

Enjoyment

Cloth 5:

1^{st} prop: Warmth

2^{nd} prop: Softness

3^{rd} prop: Smoothness

4^{th} prop: Flexibility

5^{th} prop: Thickness

Enjoyment

Cloth 1:

1^{st} prop: Thickness

2^{nd} prop: Flexibility

3^{rd} prop: Warmth

4^{th} prop: Smoothness

5^{th} prop: Softness

Enjoyment

Participant 2: PARTICIPANT ID 22

Cloth 1:

1st prop: Softness

2nd prop: Warmth

3rd prop: Flexibility

4th prop: Thickness

5th prop: Smoothness

Enjoyment

Cloth 5:

1st prop: Softness

2nd prop: Smothness

3rd prop: Warmth

4th prop: Flexibility

5th prop: Thickness

Enjoyment

Cloth 4:

1st prop: Softness

2nd prop: Flexibility

3rd prop: Thickness

4th prop: Smoothness

5th prop: Warmth

Enjoyment

Cloth 6:

1st prop: Softness

2nd prop: Warmth

3rd prop: Smoothness

4th prop: Flexibility

5th prop: Thickness

Enjoyment

Cloth 2:

1st prop: Smoothness

2nd prop: Thickness

3rd prop: Smoothness

4th prop: Warmth

5th prop: Flexibility

Enjoyment

Cloth 3:

1st prop: Softness

2nd prop: Thickness

3rd prop: Flexibility

4th prop: Warmth

5th prop: Smoothness

Enjoyment

Participant 3: PARTICIPANT ID 23

Cloth 1:

1st prop: Thickness

2nd prop: Flexibility

3rd prop: Smoothness

4th prop: Warmth

5th prop: Softness

Enjoyment

Cloth 6:

1st prop: Softness

2nd prop: Flexibility

3rd prop: Smoothness

4th prop: Thickness

5th prop: Warmth

Enjoyment

Cloth 4:

1st prop: Thickness

2nd prop: Softness

3rd prop: Flexibility

4th prop: Warmth

5th prop: Smoothness

Enjoyment

Cloth 2:

1st prop: Warmth

2nd prop: Softness

3rd prop: Thickness

4th prop: Flexibility

5th prop: Smoothness

Enjoyment

Cloth 3:

1st prop: Warmth

2nd prop: Thickness

3rd prop: Softness

4th prop: Flexibility

5th prop: Smoothness

Enjoyment

Cloth 5:

1st prop: Flexibility

2nd prop: Smoothness

3rd prop: Softness

4th prop: Warmth

5th prop: Thickness

Enjoyment

Participant 4: PARTICIPANT ID 21

Cloth 4:

1st prop: Warmth

2nd prop: Softness

3rd prop: Thickness

4th prop: Smoothness

5th prop: Flexibility

Enjoyment

Cloth 6:

1st prop: Flexibility

2nd prop: Warmth

3rd prop: Softness

4th prop: Thickness

5th prop: Smoothness

Enjoyment

Cloth 1:

1st prop: Warmth

2nd prop: Smoothness

3rd prop: Flexibility

4th prop: Thickness

5th prop: Softness

Enjoyment

Cloth 5:

1st prop: Softness

2nd prop: Warmth

3rd prop: Thickness

4th prop: Flexibility

5th prop: Smoothness

Enjoyment

Cloth 3:

1st prop: Thickness

2nd prop: Warmth

3rd prop: Softness

4th prop: Smoothness

5th prop: Flexibility

Enjoyment

Cloth 2:

1st prop: Thickness

2nd prop: Flexibility

3rd prop: Smoothness

4th prop: Softness

5th prop: Warmth

Enjoyment

Participant 5: PARTICIPANT ID 25

Cloth 4:

1st prop: Thickness

2nd prop: Warmth

3rd prop: Flexibility

4th prop: Smoothness

5th prop: Softness

Enjoyment

Cloth 3:

1st prop: Smoothness

2nd prop: Warmth

3rd prop: Thickness

4th prop: Flexibility

5th prop: Softness

Enjoyment

Cloth 5:

1st prop: Softness

2nd prop: Flexibility

3rd prop: Thickness

4th prop: Smoothness

5th prop: Warmth

Enjoyment

Cloth 2:

1st prop: Flexibility

2nd prop: Softness

3rd prop: Smoothness

4th prop: Thickness

5th prop: Warmth

Enjoyment

Cloth 1:

1st prop: Flexibility

2nd prop: Warmth

3rd prop: Thickness

4th prop: Smoothness

5th prop: Softness

Enjoyment

Cloth 6:

1st prop: Smoothness

2nd prop: Warmth

3rd prop: Softness

4th prop: Flexibility

5th prop: Thickness

Enjoyment

Participant 6: PARTICIPANT ID 24

Cloth 5:

1st prop: Smoothness

2nd prop: Thickness

3rd prop: Flexibility

4th prop: Warmth

5th prop: Softness

Enjoyment

Cloth 2:

1st prop: Smoothness

2nd prop: Thickness

3rd prop: Softness

4th prop: Warmth

5th prop: Flexibility

Enjoyment

Cloth 6:

1st prop: Flexibility

2nd prop: Smoothness

3rd prop: Warmth

4th prop: Thickness

5th prop: Softness

Enjoyment

Cloth 4:

1st prop: Softness

2nd prop: Smoothness

3rd prop: Thickness

4th prop: Flexibility

5th prop: Warmth

Enjoyment

Cloth 1:

1st prop: Flexibility

2nd prop: Softness

3rd prop: Warmth

4th prop: Thickness

5th prop: Smoothness

Enjoyment

Cloth 3:

1st prop: Smoothness

2nd prop: Softness

3rd prop: Warmth

4th prop: Thickness

5th prop: Flexibility

Enjoyment

Appendix B: Distribution of the rating variable across properties

B.1 Grouping the rating variable for Dataset V1

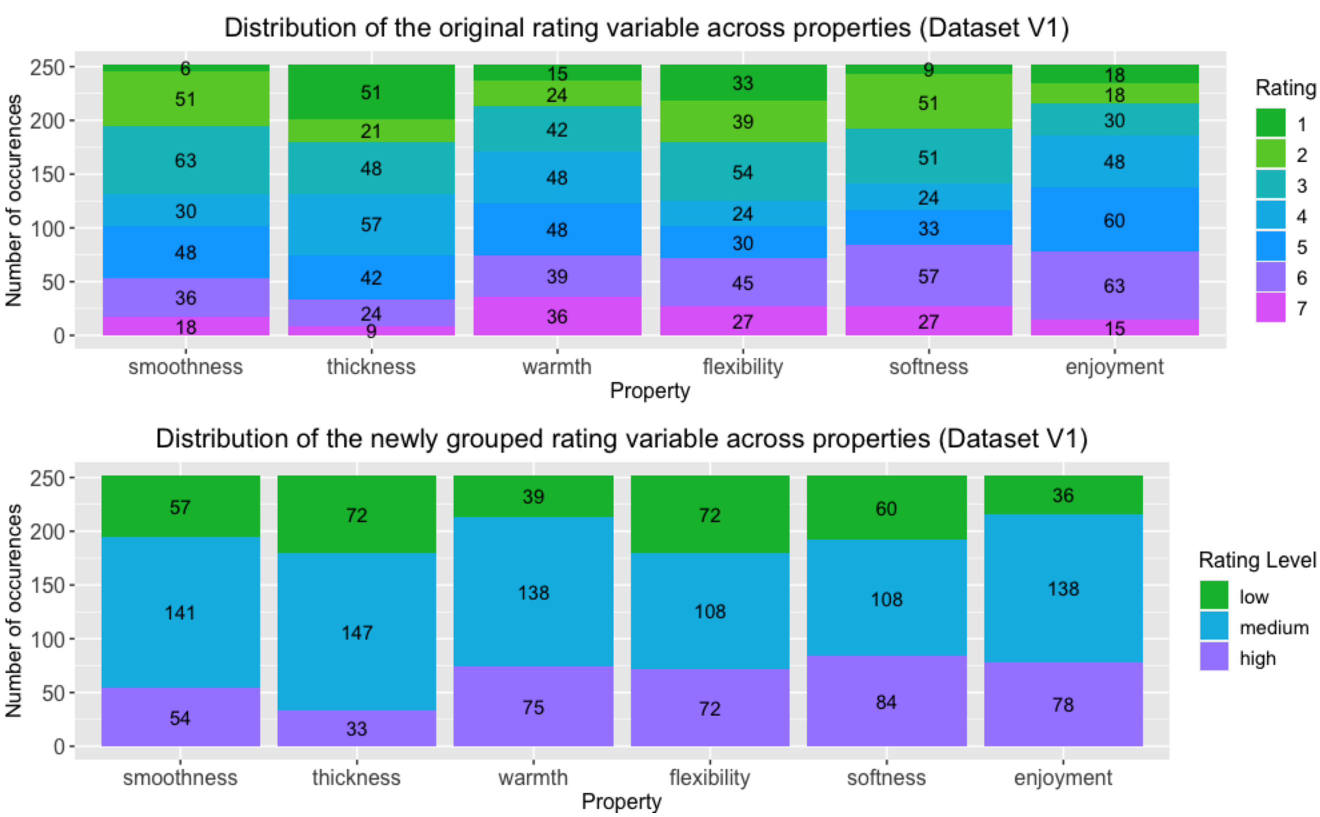


Figure 15: Distribution of the rating variable across properties for Dataset V1

Note: Dataset V1 was created by combining The first dataset and the second dataset. The first dataset included the EMG, accelerometer and quaternion data collected by Lin [25] and the second dataset consisted of the newly collected EMG, accelerometer and quaternion data.

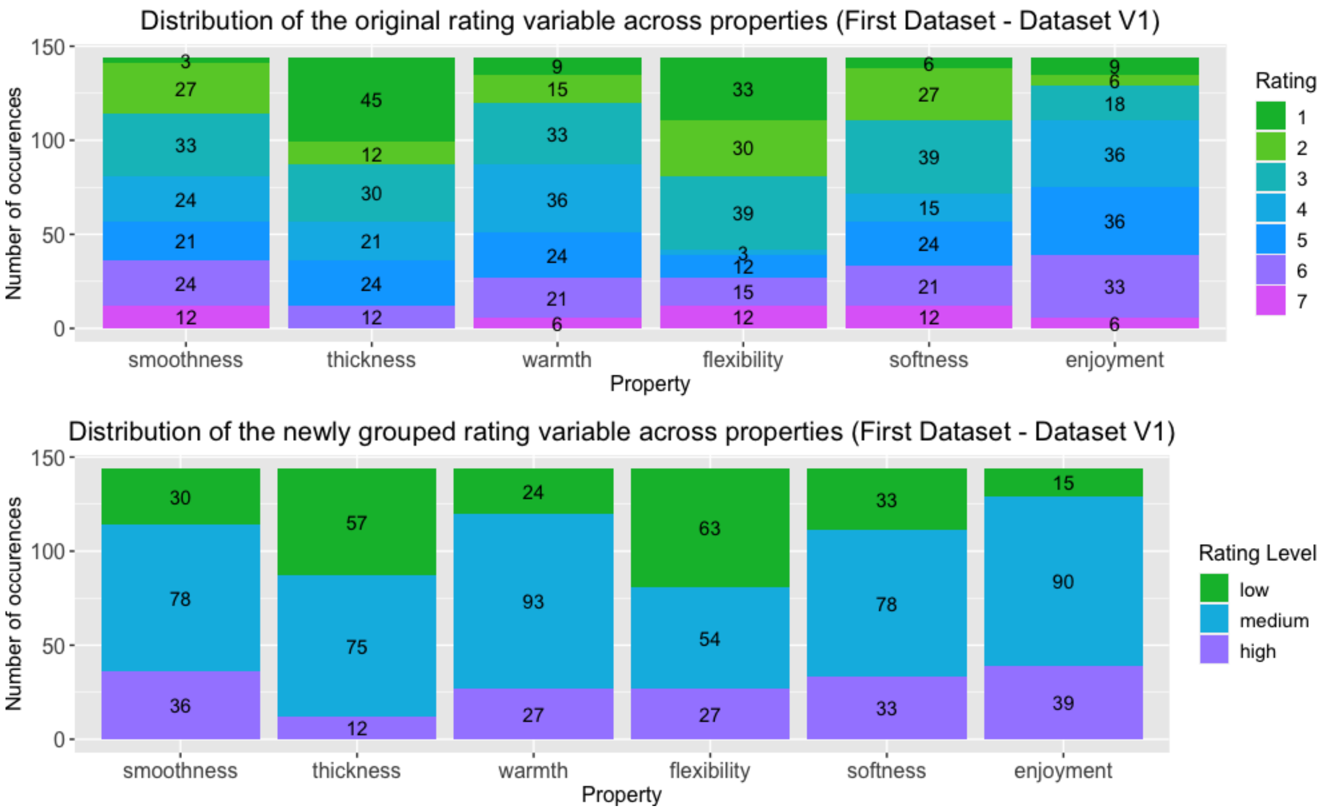


Figure 16: Distribution of the rating variable across properties for the first dataset (Dataset V1)

B.2 Grouping the rating variable for Dataset V2

Note: Dataset V2 was created by combining The first dataset and the second dataset. The first dataset included the EMG, accelerometer and quaternion data collected by Lin [25] and the second dataset consisted of the newly collected EMG, accelerometer and quaternion data.

B.3 Grouping the rating variable for Dataset V3

Note: Dataset V3 was created by combining The first dataset and the second dataset. The first dataset included the EMG, accelerometer and quaternion data collected by Lin [25]

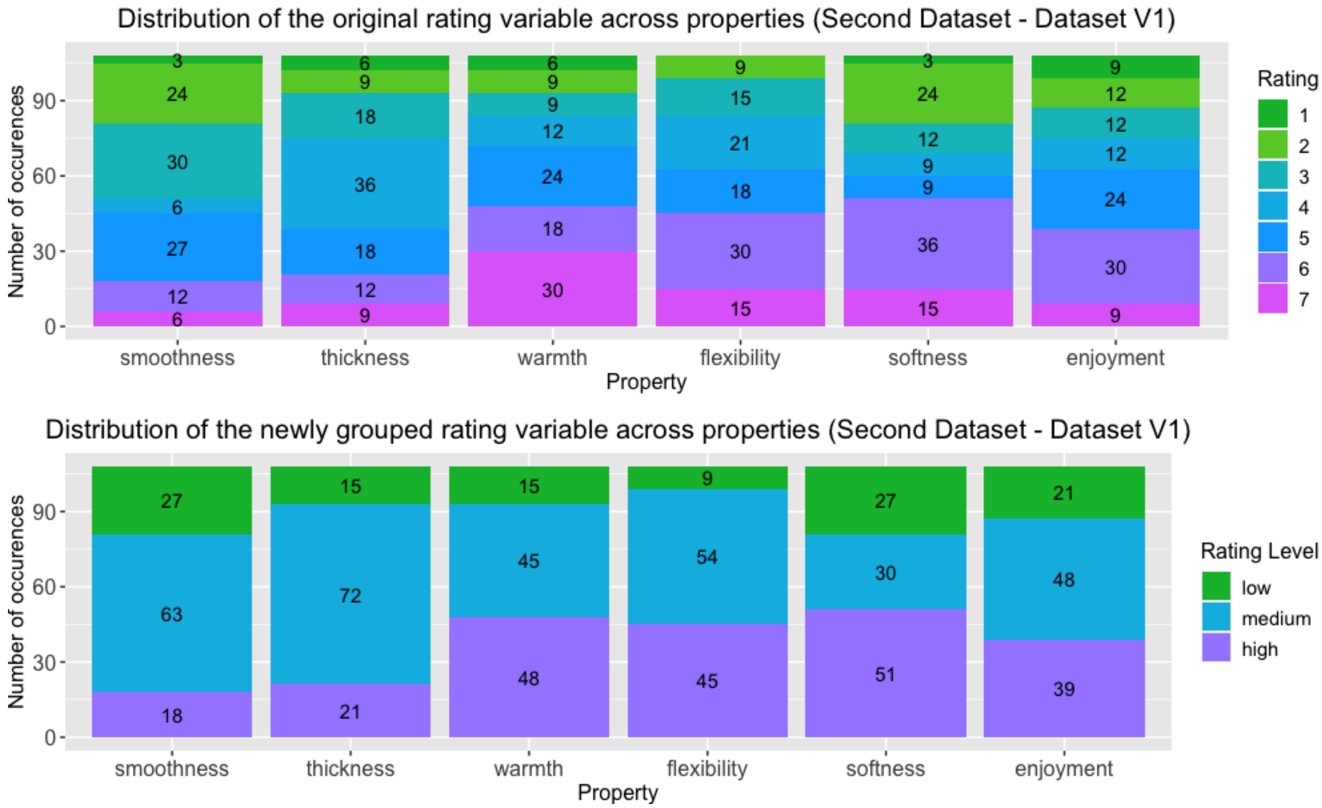


Figure 17: Distribution of the rating variable across properties for the second dataset (Dataset V1)

and the second dataset consisted of the newly collected EMG, accelerometer and quaternion data.

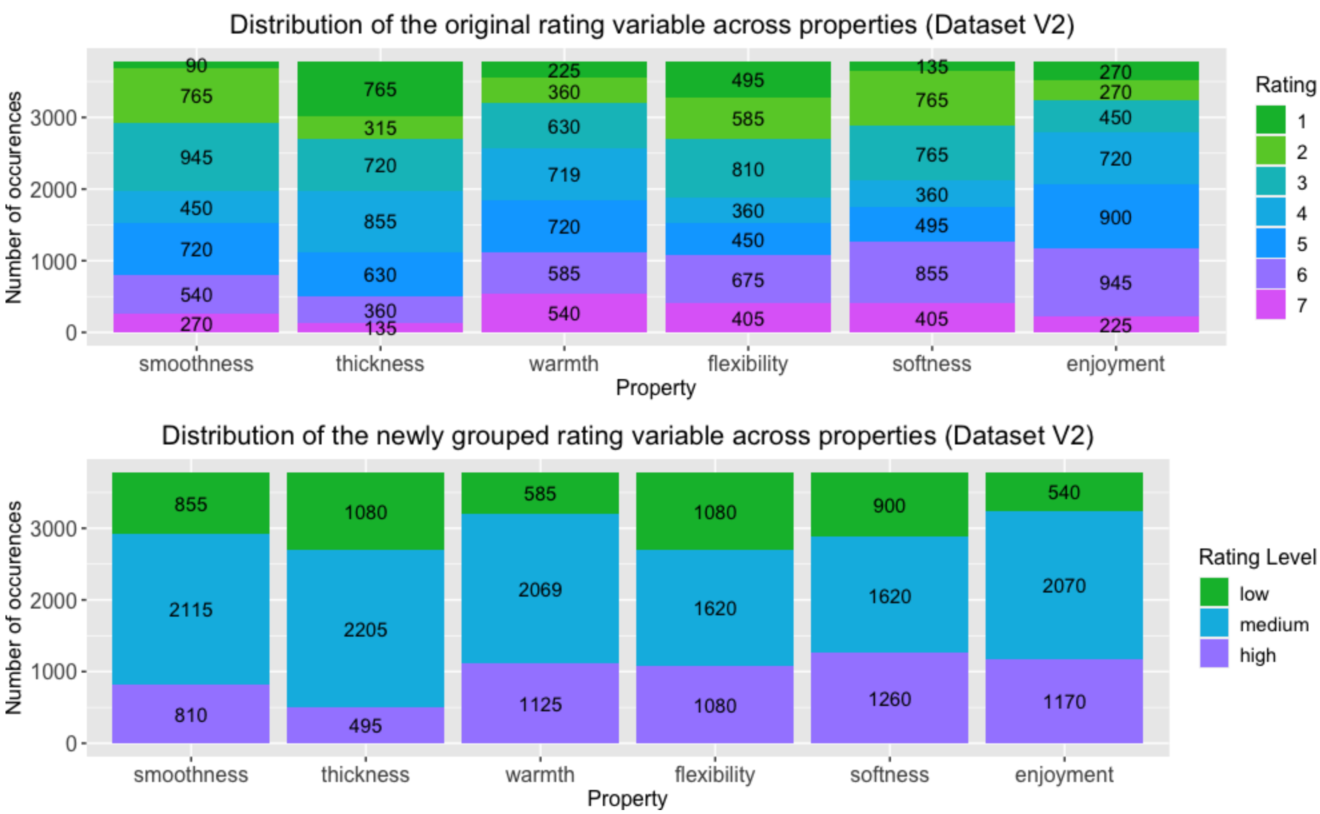
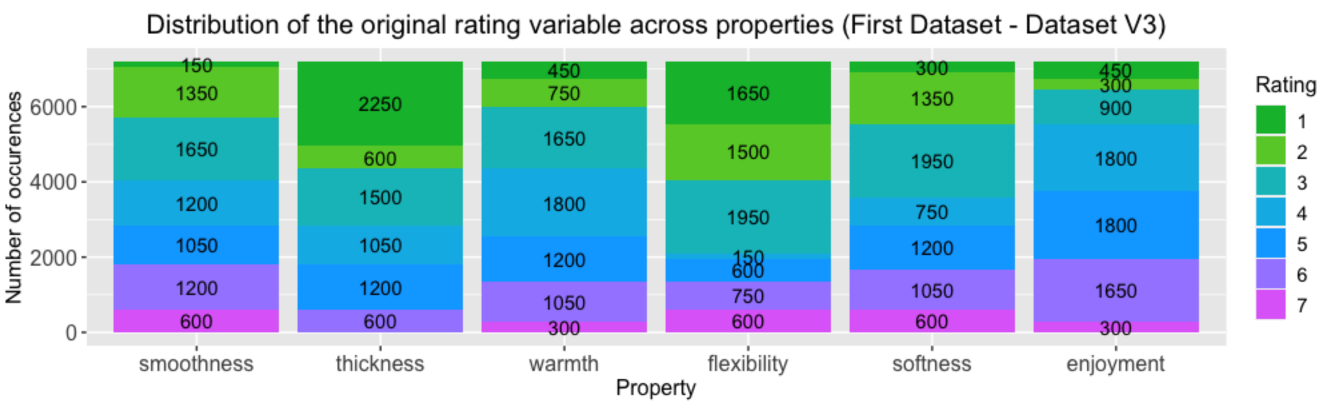


Figure 18: Distribution of the rating variable across properties for Dataset V2

Figure 19: Distribution of the rating variable across properties for the first dataset (Dataset V2)

Figure 20: Distribution of the rating variable across properties for the second dataset (Dataset V2)



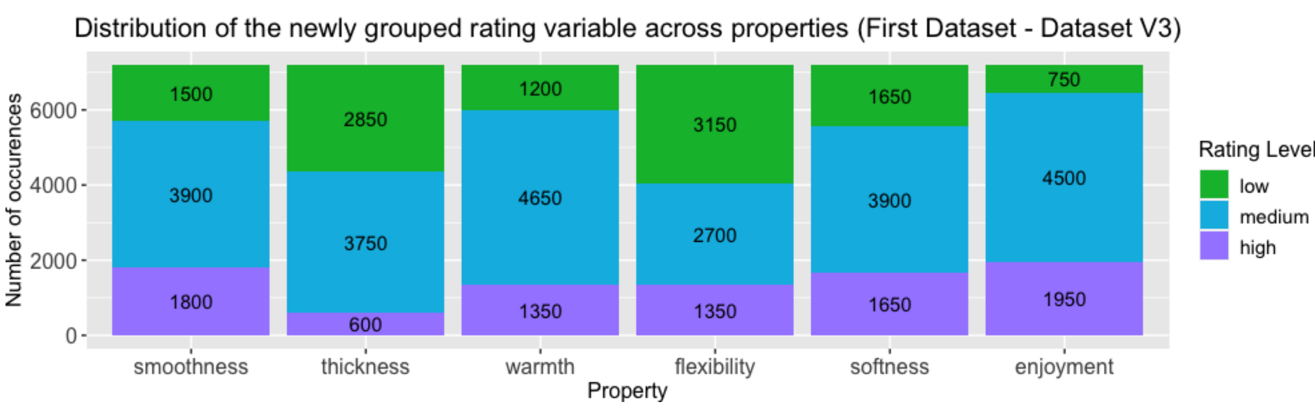
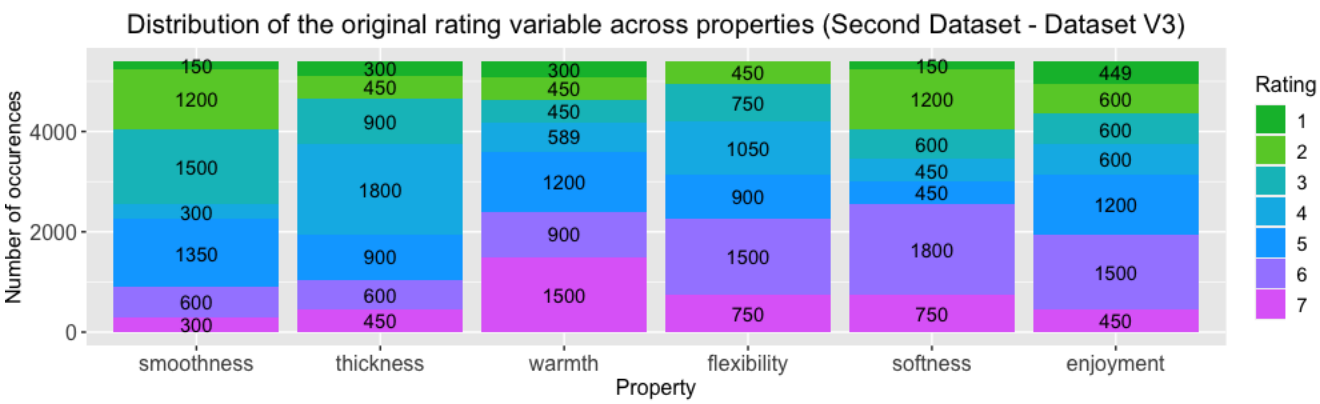


Figure 21: Distribution of the rating variable across properties for the first dataset (Dataset V3)



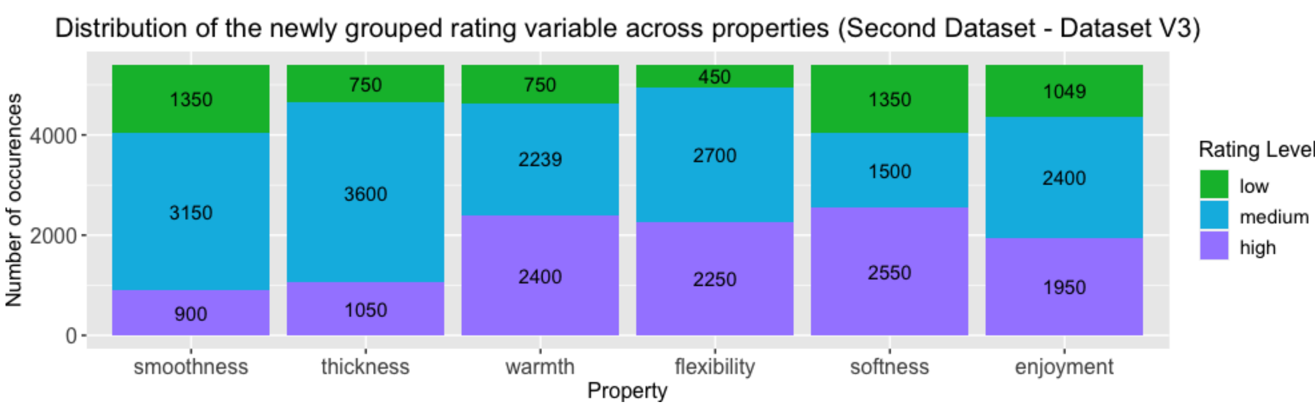


Figure 22: Distribution of the rating variable across properties for the second dataset (Dataset V3)

Appendix C: Additional Results

C.1 PRHM Property Classification Results

	Dataset V1		Dataset V2		Dataset V3	
	All 180 Features	**Only EMG**	**All 180 Features**	**Only EMG**	**All 180 Features**	**Only EMG**
Random Forest PRHM	0.37	0.33	0.34	0.3	0.33	0.28
LSTM PRHM	0.23	0.23	0.42	0.39	0.33	0.41
Fully-connected PRHM	0.33	0.25	0.39	0.4	0.45	0.35

Table 20: Average micro F1 scores when predicting properties

C.2 PRHM Rating Classification Results

	Dataset V1		Dataset V2		Dataset V3	
	All Features	**Only EMG**	**All Features**	**Only EMG**	**All Features**	**Only EMG**
Random Forest PRHM	0.51	0.49	0.49	0.49	0.47	0.48
LSTM PRHM	0.77	0.79	0.72	0.65	0.62	0.63
Fully-connected PRHM	0.67	0.71	0.6	0.64	0.65	0.65

Table 21: Average micro F1 scores when predicting property ratings