FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Human Activities Recognition: A Transfer Learning Approach

Paulo Barbosa

U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Mestrado em Engenharia de Software

Supervisor: João Moreira

Second Supervisor: André de Carvalho

September 25, 2018

Human Activities Recognition: A Transfer Learning Approach

Paulo Barbosa

Mestrado em Engenharia de Software

Parte das atividades de investigação descritas nesta dissertação foram realizadas no âmbito do projeto CONTEXTWA financiado pela FCT - Fundação para a Ciência e a Tecnologia, projeto POCI-01-0145-FEDER-016883.

September 25, 2018

Abstract

To classify human activities a Human Activity Recognition (HAR) system recurs to typical machine learning (ML) techniques that assume the input data characteristics, more specifically its distribution, will never change contrary to reality. For instance, users' tend to perform activities differently over time, and even sensors are prone to misplacement. Furthermore, building classifiers requires users' to label considerable quantities of data, making it impractical and nonsense in this era of Big Data with effortless access to public repositories. As such, we propose surpassing these two limitations by exploring the application of Transfer Learning within HAR systems.

We here commit to exploring HAR based on wearables only and to a specific scenario of TL, named Unsupervised Domain Adaptation, that is frequent in HAR, where the source dataset has similar feature and label spaces as the target one, but both differ only in marginal distributions. Our experiments address distribution mismatch caused by three circumstances: different sensor positions; different user's physiology due to age; and different environments. As for the TL techniques, we recur to two feature representation approaches, Transfer Component Analysis (TCA) and Subspace Alignment (SA), and two more instance-weighting approaches, Kernel Mean Matching (KMM) and Nearest Neighbour Weighting (NNeW). From these four techniques, we also apply a majority voting ensemble composed of three of those (KMM, NNeW, SA). As baseline methods, we employ both supervised, and semi-supervised techniques.

In total, from two HAR datasets, we devise twenty-seven unique experiments grouped into five different scenarios. The yielded results in most tests reveal a useful transfer of knowledge and more importantly the convenience of TL within HAR. Apart from the delineated experiments, our work also contributes to the field of TL in general through an exhaustive survey of TL works within HAR based on wearables, and a complete taxonomy.

Keywords: Human Activity Recognition, Transfer Learning, Unsupervised Domain Adaptation.

Resumo

Para classificar atividades humanas um sistema de reconhecimento de atividades humanas (HAR) recorre a técnicas de *Machine Learning* (ML) que assumem que as características dos dados, mais especificamente a sua distribuição, nunca se altera ao contrário da realidade. Por exemplo, os utilizadores tendem a realizar atividades de maneira diferente ao longo do tempo, e até mesmos os sensores são propensos a derrapagens. Além disso, o treino de classificadores implica que os utilizadores etiquetem quantidades consideráveis de sinais ao ponto de se tornar impraticável, e de não fazer sentido no atual contexto de *Big Data* com acesso, sem dificuldades, a vastos repositórios públicos. Como tal, nós propomos ultrapassar estas limitações explorando a utilização de TL em sistemas de HAR.

Comprometemo-nos aqui a explorar HAR, com base apenas em wearables, e um cenário específico de TL, denominado de *Unsupervised Domain Adaptation*, que é frequente em HAR, em que o conjunto de dados de origem possui semelhantes espaços de etiqueta e de *feature* em relação aos do alvo, mas ambos diferem nas distribuições marginais. As experiências realizadas abordam a diferença de distribuição causada por três circunstâncias: diferentes posições de sensores; diferente fisiologia dos utilizadores devido à idade; e diferentes ambientes de treino. Em relação às técnicas de TL, recorremos a duas abordagens de *feature-representation*, *Transfer Componenent Analysis* (TCA) e *Subspace Alignment* (SA), e mais duas abordagens de *importance-weighting*, *Kernel Mean Matching* (KMM) e *Nearest Neighbour Weighting* (NNeW). A partir destas quatro técnicas, também desenvolvemos um *ensemble* por maioria de votos composto por três destas (KMM, NNeW, SA). Como métodos de controlo, aplicamos técnicas supervisionadas e semi-supervisionadas.

No total, a partir de dois repositórios HAR, concebemos vinte e sete experiências únicas agrupadas em cinco cenários diferentes. Os resultados obtidos, na sua maioria, revelam a transferência útil de conhecimento e, mais importante, a conveniência de TL no HAR. Além das experiências delineadas, o nosso trabalho também contribui para o campo de TL em geral através de uma exaustiva pesquisa de trabalhos de TL no domínio de HAR baseado em wearables apenas e de uma taxonomia completa.

Keywords: Reconhecimento de Atividades Humanas, Transfer Learning, Unsupervised Domain Adaptation

Acknowledgements

I want to thank all the professors of MESW for the beautiful given opportunity, especially the directors prof. Dr Ana Paiva and prof. Dr Nuno Flores. These last few years have been enriching, full of challenges and learning. What I learned in FEUP, not only intellectually but also as in a personal level, made me a more mature person and better prepared for the upcoming future. It should be noted that this dissertation would not have been possible without prof. Dr João Moreira and prof. Dr André de Carvalho who proposed this project. For them, and to Kemilly Dearo I am forever in gratitude, they were vital for the success of the project.

Paulo Barbosa

Contents

List of Figures

List of Tables

Abbreviations

ADL	Activity of Daily Living
AI	Artificial Intelligence
GPS	Global Positioning System
HAR	Human Activities Recognition
HMM	Hidden Markov Model
KMM	Kernel Mean Matching
KNN	K-Nearest Neighbours
ML	Machine Learning
NNeW	Nearest Neighbour Importance Weighting
PCA	Principal Component Analysis
SA	Subspace Alignment
SVM	Support Vector Machine
TCA	Transfer Component Analysis
TL	Transfer Learning

Chapter 1

Introduction

A Human Activity Recognition (HAR) system to classify human activities recurs to typical machine learning (ML) techniques that assume the data distribution will never change contrary to reality. For instance, users' tend to perform activities differently over time, and even sensors are prone to misplacement. Furthermore, each problem requires high quantities of training data and takes significant time to build the ML model, making it impractical and nonsense in this era of Big Data with effortless access to public repositories. As such, we propose exploring Transfer Learning (TL) to leverage the performance and the efficiency of HAR systems.

1.1 Context

HAR has the potential to help revolutionise human-machine interaction by providing the ability to comprehend the context in which a person is involved. Just like humans comprehend when others' need help, it is a goal of this field to replicate this comprehension mechanism to an artificial machine. Though researchers have been exploring the field since the 1990s, only with the mass utilisation of smartphones and wearables the field gained a new momentum. The considerable technological advancements in smartphones, shortly after 2005, made it as a valuable rich data source, equipped with a wide variety of sensors and considerable processing power.

Furthermore, the diverse applicability of HAR proves that the field has the potential to increase, in a significant manner, everyone's quality of life. For instance, in Healthcare HAR can alleviate health-care costs by monitoring patients. With sensing technologies, physiological characteristics, such as Heart Rate, offer a better insight into the patient's context and provide a better understanding of his health condition.

1.2 Motivation and Goals

However, HAR still has challenges to overcome, including the subject sensitivity and weak model adaptability. The performance of human activities is profoundly influenced by the physiological features of each person, for instance, a young person performs ambulating activities differently than older people or even disabled people [1]. Though an obvious solution would be developing a specialised model for each individual, this is unpractical, making it necessary to create models as most user generic as possible. Another significant limitation is that traditional ML techniques assume that the distribution of the data is the same between the training and testing samples. Considering that the higher the physiological user features are, the more different is the dataset distribution between users, a model trained with a specific group of users will not accurately classify the activities of other different users. Therefore, the problem facing here is a combination of two considerable limitations, the difficulty of collecting labelled data and the limitation of ML techniques of not being able to avail datasets of different users to leverage the performance of a model.

For surpassing these limitations, we here commit to exploring Transfer Learning (TL) within HAR. TL is a ML sub-area that focuses on the re-utilisation of knowledge between ML tasks, because it allows the dataset distributions, and even their tasks used in training and testing be different. The learning paradigm behind TL is motivated by the fact people while performing tasks, not only learn from it, but they learn how to generalise it and apply in related tasks. For instance, we naturally assume that having two persons learning the piano the one with violin experience will outperform the other with no musical background. Although TL is not widely utilised in Industry, its research dates back to the 1980s and, nowadays, it remains as a research topic [2].

The goals we expect to achieve by applying TL in HAR are the following:

1. **Reducing the effort of capturing labelled data:** by re-utilising knowledge from previous experience it is no longer required to capture high quantities of labelled data.

2. **Leveraging different, but related, datasets to boost performance:** from the re-utilisation of knowledge of previous experiences improve the performance of the target domain

1.3 Outline

This work is structured into five additional chapters:

Chapter 2: "Human Activities Recognition" chapter provides an overview of the field, describes its application in four domains, explains each of the three stages that compose a typical HAR system, and, finally, concludes by listing the challenges the field still has to overcome.

Chapter 3: "Transfer Learning" chapter starts by explaining the basic terminologies necessary to understand TL along with a comparison between TL and similar fields, sometimes miscomprehended in literature. Then, it delineates a taxonomy resorting to related works, and, finally, concludes with an extensive survey of HAR works, based on wearables, applying TL.

Chapter 4: "Methodology" chapter focuses on unsupervised domain adaptation, a specific category of TL. It firstly explains its four different approaches to then concluding with a description of each of the four selected techniques for the experiments.

Chapter 5: "Experiments" chapter describes how the experiments were delineated, with what datasets it resorted to, and the data pre-processing techniques applied. It concludes with an analysis of the yielded results per implemented experiment scenario, and with an overall results discussion.

Chapter 6: "Conclusions" chapter provides an insight of the work performed throughout the dissertation, its limitations, and the envisioned future works.

Chapter 2

Human Activities Recognition

An artificial system that recognises human activities and comprehends the context in which a person is involved, understanding when help is needed, is the next step in human-machine interaction [3]. The field of Human Activities Recognition (HAR) has a crucial role in allowing these technological advancements through the recognition of human activities. Hence, researchers have long been exploring this field and experimenting with real scenarios, including in Healthcare.

This chapter starts by describing HAR, proceeding to its applications in real and potential use-cases. Afterwards, it explains how typical HAR systems are designed, providing an overview of three sequential stages: Data Acquisition, Data Manipulation, and Machine Learning. Finally, this chapter ends by describing four challenges that remain to be overcome in HAR.

2.1 Overview

Researchers have actively studied the field of HAR since the 1990s [4]. The focus of HAR is to collect, and detect real-life activities being performed by one human or a group, and to understand the environmental context surrounding humans. Just like humans comprehend when others' need help, HAR tries to replicate this comprehension mechanism to an artificial machine.

Since HAR has the potential to help revolutionise how humans nowadays interact with computers, it is considered as one promising branch of the Human-Computer Interaction field [5][6]. In order to make computers' look like humans and interact alike, it is necessary that computers become aware of the context of humans, and act accordingly.

HAR relies on the deployment of pervasive but non-intrusive sensors in the environment surrounding humans, which is also the focus of another field, Ubiquitous Computing. Ubiquitous Computing is also widely researched, however, the fact that sensors need to be almost non-noticeable and not disruptive during activities performance constituted severe challenges during the 1990s when the field of Ubiquitous Computing was in its infancy [4]. Only two decades after, with the technological advancements in processing power, wireless communications, and energy storage [7] mobile devices rapidly became integrated into each person's daily life. People now spend more time on the phone than other device [8], and are equipped with a wide variety of electronic devices, such as smartphones, smartwatch, and laptop, that offer an opportunity to HAR researchers of efficiently extracting context-aware data from sensors in a non-intrusive manner. Having captured the data, HAR also benefits from the integration of activity patterns discovery techniques. Pattern discovery is useful for finding unknown patterns and can help to define activities that can be later tracked during activity recognition.

2.2 Human Activities Recognition Applications

The diverse applicability of HAR proves that the field has the potential to increase, in a significant manner, the quality of life of people. This chapter presents the applications of HAR that we consider the most relevant to a wide range of people and could impact health-related domains and other public-related issues, such as authentication, and smart-Homes. However, it should be noted that one current popular applicability of HAR is in fitness, where it significantly helps the workout effectiveness by monitoring fitness activities performance [9]. Though fitness is essential for every person's quality of life, it focuses on a smaller target audience compared to other public-related issues, such as health, and therefore it is not described in detail in this subchapter.

2.2.1 Biometric Authentication

Conventional authentication relies mostly on unique identifiers such as a passcode, something a person knows and might forget, or tokens, something a person has and might lose. Though easy to configure, these authentication mechanism lack enough accuracy to avoid identity theft (e.g. four-pin passcode has an accuracy of $1/10^3$). Therefore, biometric authentication, as an alternative method, is starting to gain momentum due to its inherent high accuracy recognition (e.g. Apple's [1] Facial Recognition has an accuracy of $1/10^6$ [10].

Biometrics as the science that studies the capability to identify a person based on human characteristics can make use of HAR recognition techniques to capture unique behavioural properties of a user, like motion capture signatures [11], to prevent unauthorised access of a device. Today's biometric authentication is based mainly on physiological features of individuals, however, physiological features pose severe concerns regarding privacy and HAR could be seen as a viable alternative, acting only as a behavioural biometric mechanism.

2.2.2 Health Assistance

The current world population has a better quality of life and a higher life expectancy that continues to rise decade after decade, resulting in an increased number of senior citizens (it is estimated that in 2030 the proportion of the population, in the US, over sixty-five years old will increase from 12.4%, in 2000, to 19.6% [12]). This constitutes a challenge to medical institutions that are responsible for providing care to patients because the senior citizens are the most vulnerable to

[1] https://www.apple.com/en/

diseases and disability conditions. Furthermore, considering that the health-care costs of senior citizens are expensive, it is estimated to be three to five times greater compared with citizens under sixty-five years old [13], health-care institutions must necessarily self-adapt by finding cost-effective solutions.

HAR has the potential to alleviate health-care costs, and already active research works are applying HAR in health-care institutions. The integration of HAR can be performed in two different sub-domains: health-care staff monitoring (e.g. medics, nurses, surgeons); and patient monitoring. The former domain focuses on the interaction between medical staff through the inference of their activities to improve the efficiency of hospital processes and staff allocation. For instance, the work done by Osmani et al. [14] studies the efficiency of doctor-nurse collaboration by monitoring and recognising the activities performed by both. The latter domain focuses on the automation of health monitoring of patients, providing health-staff rapid guidance during emergencies. With sensing technologies, physiological characteristics, such as heart rate, offer a better insight into the patient's context and provide a better understanding of health condition [15].

2.2.3 Smart Homes

The next generation of households will, inevitably, be smart and automated able to monitor and control different appliances. The idea behind smart homes is to decrease costs by improving the efficiency of consumption of energy, and water and at the same time provide a secure and comfortable environment to its dwellers [16]. Moreover, a smart home should be able to learn and adapt itself to its users' habits. For instance, by learning the time when its habitant arrives from work, the house could provide a more comfortable environment increasing the house's temperature (when its cold) and even prepare the kitchen for meals cooking.

Therefore, if the future house must be able to understand its habitants' context, it necessarily has in its system an embedded HAR feature that recognises real human activities. In this regard, MIT [2] has already been researching this subject under laboratory and semi-naturalistic environments studying human typical household activities, such as cooking, and sleeping [17].

2.2.4 Government Surveillance

Countries such as China and the United States, nowadays, use activity recognition in their video surveillance systems recurring to non-intrusive biometric technologies in crowded public-spaces [18]. Non-intrusive means that the user cannot detect the technology and no permission is required, and these non-intrusive biometric technologies are becoming more common within smart surveillance systems, widely adopted at public spaces, and private areas. Indeed, surveillance systems permit not only perform identity tracking, with facial recognition but also activity recognition, providing the ability to authorities react rapidly and accordingly in specific contexts (e.g. riots).

[2] Massachusetts Institute of Technology , http://web.mit.edu/

2.3 Human Activities Recognition Systems

Though each HAR system follows different approaches, all systems have in common three sequential stages, the data acquisition, data manipulation and ML stage. Therefore, the following sections describe these three stages.

1. **Data Acquisition** (see Chapter 2.4): this stage involves the extraction, in short (real-time) or long-time intervals, of sensor data. From the sensors, the data is forward, via wireless communications or wire, to a single device that has more powerful hardware capabilities.

2. **Data Manipulation** (see Chapter 2.5): this stage includes an analysis and manipulation of the data to guarantee there no invalid values that could compromise the learning stage. This stage usually begins by removing noise values, proceeded by a segmentation of the dataset into time windows to apply then feature selection and extraction that extrapolate more valuable data. The techniques applied during this stage have a profoundly influential role in the performance of the following stage.

3. **Machine Learning** (see Chapter 2.6): this stage comprehends a training mechanism (either offline or online) to help build a classifier model with labelled or unlabeled data. Eventually, several different classification models could be used, even ensembled, and for determining what's the most accurate solution performance evaluation mechanisms are used.

2.4 Data Acquisition

To acquire real data when an individual performs activities it is necessary to deploy sensors measuring both the individual and the scenario surrounding him. Therefore, sensors in HAR systems can be classified into two categories:

1. **External**: sensor deployed outside the monitoring target, usually deployed in a fixed position (e.g. ceiling, building entrance). This category is advantageous when monitoring multiple individuals, but is restricted to a limited area in which the target must be present.

2. **Wearables**: sensor attached to the monitoring target, body-worn or internally, from which directly measures signals. Sensors of this category behave as portable devices with energy batteries, consequently having a finite, sometimes inadequate, execution duration. It is necessary that these devices must be as less intrusive as possible to avoid any disturbance during activities performance.

2.4.1 Sensor Taxonomy

Each sensor has different specificities, and each one has to deal with different noise agents. As an example, accelerometers are affected by gravity. Therefore, when deploying various sensors, it is essential to understand each sensor to extract as much valuable data as possible. Next, the most utilised sensors for motion, location, and external data capture for HAR systems are described.

2.4.1.1 Accelerometer

An accelerometer sensor measures the physical acceleration in three axes of an object. The accelerometer data of each axis is represented in units of g-force usually with a corresponding timestamp of the internal clock of the sensor. Being a wearable device, it is convenient for recognising body motion activities and as research shows, it is the most widely used sensor for this type of activities [19]. However, the gravitational earth's field, thought constant, interferes with the streamed data that makes it necessary applying filtering techniques, during Data Manipulation stage, to remove it.

2.4.1.2 Gyroscope

A gyroscope sensor measures the orientation of an object in angular speed (*rad/s*) by detecting its roll, pitch, and yaw motions. Thus, it is commonly employed in aerial applications, such as navigation systems, as a wearable device. Unlike accelerometer sensors, the earth's gravity does not contaminate the gyroscope data, and researchers recommend its use combined with accelerometers to compensate the accelerometer's limitations (e.g. a gyroscope is useful when recognising ascending and descending activities since the accelerometer performs poorly in these situations)[19].

2.4.1.3 Barometer

A barometer sensor measures the atmospheric pressure (in *atmosphere* units) and has the ability to assist localisation systems since air pressure varies regarding different earth's locations. Thus, it is most useful in situations of altitude variation and weather prediction. Since it only collects data regarding the environment, this sensor can be utilised as a wearable device, or as an external one.

2.4.1.4 Light

A light sensor measures the illumination level of the environment surrounding the sensor. The data is usually represented in *lux* units and determines if the sensor is in a dark or bright environment. It is useful for the recognition of outside activities or even activities such as sleeping, usually performed in dark environments. Since it only collects data regarding the environment, this sensor can be utilised as a wearable device, or as an external one.

2.4.1.5 Compass

A compass or magnetic sensor measures its position relative to the north Earth's pole. The sensor data is in degrees units ranging from 0^{o} to 360^{o}, considering the absolute north points to 0^{o}. This sensor is vital for situations requiring geo-location of the user since it aids when orienting maps. It is utilised as a wearable device.

2.4.1.6 Video Camera

A camera captures visual data, and it is usually deployed in fixed positions with a direct line of vision to a monitoring target. The captured data is processed recurring to image processing techniques to perform vision-based activity recognition. Depending on the image's capture frequency and the image processing techniques, it can be very accurate in situations like intrusion detection. However, the main issue is concerning privacy since not everyone is willing to be continuously monitored by cameras [4]. It is typically utilised as an external device.

2.4.1.7 Global Positioning System

The global positioning system measures the geographical location of the sensor in latitude and longitude. Though it is very accurate in outdoor environments, it has low accuracy underground and inside buildings. The GPS is useful when determining the context of the user since human activities can be associated with locations (taking a walk or running are more probably in parks). Like video cameras, the constant localisation monitoring of a user has privacy issues related, which makes it mandatory for the encryption of the streamed data. It is utilised as a wearable device.

2.4.2 Smartphone as a Wearable

Since the first mobile phone in the 1970s [20], there has been a considerable technology evolution that made phones more than a messaging device also becoming a working and personal device. We consider that there are three main reasons why mobile phones are a reliable and rich data source:

- **Hardware Capabilities:** The first mobile phones offered only elementary telephony services such as short and multimedia messaging services and had a limited battery. However, with the increase of processing power chips, benefiting from Moore's Law [21], and battery duration, enabled by the introduction of Li-Ion batteries of high energy density, mobile phones began to offer a wide range of services, the same as the desktop with more embedded devices and sensors. Nowadays, mobile phones are capable of dealing with high complexity 3D games, high image quality capture and video recording, and fast accessibility to web pages.

- **Ubiquity:** Statistics show that almost everyone has a mobile device, and there is a clear trend of people spending more and more time using mobile devices [22]. Therefore, the ubiquity of mobile phones provides a perfect opportunity to HAR systems by providing valuable and rich data source of how, when, and where a user carries his daily activities.

- **Diverse collection of sensors:** Considering that most applications on mobile phones require the use of sensors to enhance user experience (games usually require an accelerometer and a gyroscope), phones had to adapt and necessarily integrated a considerable number of embedded sensors. The most common sensors on a mobile phone are the accelerometer, gyroscope, light sensor, and compass.

2.5 Data Manipulation

The raw data streamed from the sensors if not adequately analysed and validated can do more harm than good to the overall performance of a HAR system. Therefore, it is vital to comprehend how each sensor works and apply proper techniques that extract knowledge, remove the effect of orientation and sensor position, or even select the most valuable data features to reduce noisy data and take as much valuable data as possible. Additionally, there are other techniques than the previously referred that are specific to the HAR domain, such as Data Segmentation.

2.5.1 Feature Extraction and Feature Selection

In machine learning the concept of data feature is related to the input attributes that compose the data streamed from sensors. Feature Extraction and Selection techniques are designed to change the representation of the data obtained from the sensors in order to improve the performance of models. Thus these two techniques are most necessary when resolving machine learning problems. However, every machine learning problem has a specific domain and requires different data representations applying appropriate Feature Extraction and Selection techniques [23].

Feature Extraction:

Feature Extraction, also known as Feature Construction, is a key process that constructs new features from the sensor raw data. Its goal is extracting more instructive values from the raw data, whether recurring to human expertise (e.g. from a date attribute it is possible to extract more valuable information such as holiday, weekend, season) or even by recurring to automatic Feature Extraction techniques as the following:

- **Principal Component Analysis:** reduces multidimensional data to a lower-dimensional space while retaining the same information. This is useful to avoid increasing the complexity of the learning model and its performance [24].
- **Time-Domain Features:** consists in the applications of basic statistical techniques that process value variation within time (Mean, Maximum, Minimum, Variation, Standard Variation, Correlation) [25].
- **Frequency-Domain Features:** consists in the application of statistical techniques, based on Fourier Transformation technique that calculate the periodicity of signals (Energy, Entropy, Time Between Peak, Binned Distribution) [25].

Feature Selection:

Feature Selection techniques are designed to tackle problems like data redundancy and dimensionality complexity by only selecting relevant features. For instance, standard features like name, and identifiers can negatively affect ML models by increasing its complexity unnecessarily.

- **Wrappers:** the selection of features is conceived by combining all data features into different subsets and assigning to each one a score based on how the model performed with the corresponding subset. Recursive Feature Elimination [26] can be used as a wrapper technique.

- **Filters:** ranks each data feature with a score and suppresses the least ranked features independently of the selected ML model. The Chi-Squared Test [27] can be used as a filtering technique.

2.5.2 Data Segmentation

Data segmentation consists in dividing the continuous data streamed from the sensors into time windows segments. The selection of the time windows has a significant impact on the overall system accuracy, and should take into consideration various factors, including the typical duration of the activities to be recognised, the attributes measured, and the capabilities of the system. Selecting short time windows triggers more frequently the HAR system transmitting more overhead and could not provide sufficient data to describe the performed activity. On the other side, with a long-time window, there's a chance that a single window has more than one performed activity. Furthermore, research works show that the accuracy decreases as the window size increases [28]. Data Segmentation in HAR can be categorised into two different categories:

- **Segmentation with overlapping:** two or more segmentation windows can be activated at the same time sharing the same sensor events. This category handles more accurately activities transitions and reduces the error caused by transition state noise [29].

- **Segmentation without overlapping (disjoint):** each segmentation window corresponds to different non-overlapping time intervals. It is commonly used in HAR systems due to its reduced computational complexity compared to the overlapping category [29].

Furthermore, besides each category, in data segmentation there are two different segmentation scenarios regarding the size of the time windows:

- **Fixed Time Slots:** every time windows consist of an equal and fixed size. With equal time windows, there is less computational complexity however it is vital to define an optimal windows size to avoid prejudicing the performance of the HAR system.

- **Dynamic Time Slots:** the time windows length is defined during run-time depending on different factors, such as activity inference and sensors events. There are different approaches, some scenarios shrunk and expand the time windows at run-time, while others use variable-sized time windows in which the length of each window is sequentially multiple of the initial window [30].

2.5.3 Orientation-Independency

When recurring to sensors, such as accelerometer or gyroscope, the orientation of the device significantly affects the output of these sensors. For instance, signals collected when a user is lying can be similar in a standing position if the sensor is placed in a different orientation. Also, considering that it is very intrusive obligating users to place the sensor device in a specific orientation, it becomes necessary applying techniques that automatically suppress the orientation effect.

Research works have been exploring two different manners of obtaining orientation-independent values. The first is by selecting data obtained from orientation-independent features, for instance using only the accelerometer magnitude and ignoring the three-axis values or even by utilising the standard deviation value of the accelerometer magnitude [31]. The other solution is by transforming signal values to orientation-independent values. One standard technique is converting the coordinate system of the capturing device into a global Earth coordinate system [32].

2.5.4 Position-Independency

The data streamed from motion sensors, such as the accelerometer, significantly depends on the position of the sensor. If the sensor is placed on a wrist, the captured data is different than if placed inside a pocket, mainly because the pocket limits the sensor movements, consequently capturing less motion than on the wrist. Though the most effective solution is to lock the sensor on a specific position of the user's body, commonly the pocket or the wrist, it is intrusive and burdensome to guarantee that the sensor is correctly placed. Therefore, the ideal solution is one that works independently regarding the location of the sensor.

Research works have been exploring different techniques, such as applying a generalised classifier trained using data from all positions [31] [33] (thought very utilised it is not recommended because it requires significant effort when collecting training data), or even by recurring to magnetometer sensors. The data from the magnetometer can be used to help converting the accelerometer data to the Earth coordinating axes [34].

2.6 Machine Learning

After manipulating the data captured from the sensors, the final stage is applying ML techniques that learn from the data and shape a mathematical model capable of identifying correctly human activities. In this stage, HAR systems can adopt different approaches, such as Supervised Learning, Unsupervised or Semi-Supervised and depending on the context, the learning process can either be Offline Learning or Online Learning. For supervised classification, there are available many ML algorithms to select from, like Hidden Markov Model, Decision Trees, Support Vector Machine. While supervised learning is commonly adopted within this domain, unsupervised learning is significantly less utilised. And as such we here focus only on supervised and semi-supervised learning. Furthermore, after the model is created, it is essential to evaluate its performance, having available in ML literature many metrics to opt from.

2.6.1 Supervised Learning

Supervised Learning is a type of a learning process that recurs only to labelled training data. Labelled means that the training data is composed of two elements, the input vector (observation values) paired with the target element (outcome). Both elements can be represented in quantitative and qualitative values. If the labelled data is represented quantitatively, the learning problem is considered as a regression problem (e.g. temperature forecasting). On the other hand, if represented qualitatively, the learning problem consists in a classification problem (e.g. activity recognition) [35].

Undeniably, the core of a supervised learning system is the ML algorithm. The algorithm receives all the training labelled data to shape a mathematical function that classifies or predicts an outcome for each data instance. The learning process works by estimating a relationship between the input data and the target label, which during learning, is continuously modified by always evaluating its performance between the original and the algorithm generated output. At the end of the learning process, the shaped relationship should be accurate enough for its classification of prediction problem.

A typical scenario where supervised learning is applied is email spam identification. In this scenario, the training data consists in a collection of labelled emails, in which the input vector is represented by email information, such as text and the sender, paired with a categorical label, spam or no-spam, that represents the target element.

Though very useful, the process of data labelling is very costly since it involves the presence of an artificial or human agent that labels the data according to a pre-defined labelling procedure. Moreover, in some scenarios the acquisition of labelled data is merely impossible, motivating the utilisation of other techniques that are more adapted to these conditions such as unsupervised and semi-supervised learning.

2.6.2 Semi-Supervised Learning

The main idea behind semi-supervised learning is replicating how human learning works since it is experimentally proved that humans when learning work with also labelled and unlabelled data [36]. During the learning process, there is, usually, a higher amount of unlabelled data along with a much lower amount of labelled data, since in some scenarios the labelling task is not feasible. The process of labelling data is complicated in many different scenarios, including in HAR systems, because it requires the presence of a person responsible for annotating data whenever each activity is performed. Indeed, the labelling task requires a considerable level of effort and user's cooperation, and it is even more difficult whenever the ML system has a high heterogeneity of target labels.

2.6.3 Offline Vs. Online Learning

ML models have two different approaches regarding the learning process, in an offline manner or online.

Offline Learning induces a learning model in a non-incremental manner. The learning model receives a static batch of instances only once and after the capturing task is done. The model is created based on the collected instances and, usually, afterwards is not improved. Offline learning systems are designed for non-real-time operations recurring to computationally demanding algorithms that require significant processing time. Most HAR studies have an offline learning approach however it is not recommended because human activities tend to be performed differently over time [37].

On the other hand, online learning is based on the process of creating a learning model incrementally, by iteratively reshaping itself with data instances. Online learning systems are intended for scenarios where there is a short time constraint, requiring real-time outcome feedback from the model. Furthermore, considering that in HAR systems people perform activities in different manners over time, online learning is seen as a potential robust solution for these possible alterations of data distribution [38].

2.6.4 Performance Evaluation

As the final stage of ML systems, the model predictive performance evaluation is a necessary stage, no less relevant than the other referred previous stages. Through the experimental data, it provides feedback on how the model correctly predicts or classifies the target label. The ability to assess the performance is vital for guiding the Model Selection phase, which is the process of selecting the algorithm and the values of its parameters. It primarily consists in a "trial-and-error" strategy, training different algorithms with different parameters. Each created model must be evaluated and compared with others. In the end, the highest accurate model is chosen based on estimative values, such as the generalisation error [39] which provides a sense of how the model performs with unseen data.

When estimating the performance, it is important to avoid overfitting, which can occur using the same data instances used during the model training, or else the evaluation becomes "biased". Meaning that, an overfitted model is only fit to its training data but incapable of correctly predicting future observations. The recommended approach for model evaluation is to separate the data into three sets:

- **Training Set:** used only for the model creation (usually is the largest set of the three, consisting of more than 50% of all dataset);
- **Validation Set:** used only to estimate prediction error;
- **Test Set:** used only to estimate the generalization error.

However, in scenarios where the data available is scarce, it is preferable than dividing into three sets to utilise techniques like cross-validation. Cross-Validation divides the data into two segments only, the training data and validation data. It performs successive rounds of validation, in a way that both segments cross over validating against each other [40]. Performing cross-validation also has the advantage of providing an indication of how well the model will generalise to an unseen dataset reducing the problem of overfitting.

For the comparison and evaluation of ML models, many performance metrics are available however one must ponder which metric is best for his problem. As such, we here present some of the most basic and commonly utilised metrics:

- **True Positive (*TP*):** number of instances correctly accepted, or predicted/classified as positive.
- **True Negative (*TN*):** number of instances correctly rejected, or predicted/classified as negative.
- **False Positive (*FP*):** number of instances incorrectly accepted, or predicted/classified as positive.
- **False Negative (*FN*):** number of instances incorrectly rejected, or predicted/classified as negative.
- **Accuracy:** proportion of true results (both positive and negative)

$$\frac{TP+TN}{TP+TN+FP+FN} \tag{2.1}$$

- **Sensitivity/Recall:** true positive rate

$$\frac{TP}{TP+TN} \tag{2.2}$$

- **Specificity:** true negative rate

$$\frac{TN}{TN+FP} \tag{2.3}$$

- **ROC AUC:** known as Area Under the Receiver Operating Characteristic Curve, it is the probability that a classifier will rank a randomly chosen positive example higher than a randomly chosen negative example [41].
- **Balanced F1 Score:** harmonic mean of precision and recall

$$2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2.4}$$

2.6.5 Machine Learning Algorithms

Selecting which algorithm to classify the activities is of vital importance to the success of the work since some are prone to a different type of problems. However, we here confine to some typical algorithms in HAR [4]: K-Nearest Neighbour; Hidden Markov Model; Decision Tree; and Support Vector Machine.

2.6.5.1 K-Nearest Neighbours

It is a supervised deterministic learning approach, based on distance measurements between data points. It works by searching in the dataset for the k closest samples to decide the outcome of a given sample [42]. The selection of the value of k depends on the problem domain and the dataset, for instance, a model with a high value of k has very low noise but distinguishes fewer classes. On the other hand, when k is 1, the technique becomes a 1-nearest neighbour algorithm essentially searching only for the closest data point. KNN is capable of working in both classification and regression problems. In classification problems, the technique recurs to vote majority to assign the target class. In regression problems, the target class is calculated by doing an average of the k nearest neighbours' data points.

2.6.5.2 Hidden Markov Model

Hidden Markov Model consists in a generative probabilistic model, based on the Markov Chain model, in which human activities are represented as hidden events, and the sensor streamed data as the observed events. It is used in a wide range of working fields, including speech recognition, bioinformatics, and text processing [43]. In HAR systems it has the advantage of efficiently capturing the transition states between activities, though it has difficulties in scenarios where there is concurrency of activities.

2.6.5.3 Decision Tree Induction Algorithm

Decision Trees induction algorithms induce a model by a decision tree, which is represented by hierarchical rules, easily interpretable, composed of two elements: decision nodes and branches. Activities are represented by the nodes, decided by the input data that is represented by the branches. A considerable advantage of using decision trees is the simple and straightforward interpretation of its learned rules from its structure. For an online learning approach, Yang et al. [44] have proposed a very fast decision tree that incrementally reads the dataset, self-adjusting its decision tree per incoming data.

2.6.5.4 Support Vector Machine

Support Vector Machine is a deterministic supervised learning approach that is capable of solving both classification and regression problems. The technique works by plotting each data instance as a coordinate point to then design a frontier, called hyperplane, to distinguish all the task classes. Moreover, the technique's purpose is maximising the distance between the hyperplane to the nearest instances, to diminish the generalisation error of the classifier. Regarding the field of HAR, SVM is one of the most utilised ML algorithms.

2.7 Human Activities Recognition Challenges

Finally, based on the research done, we conclude that it is necessary to overcome four particular challenges within HAR to facilitate its mass utilisation: activity complexity; energy constraint; user's privacy; and subject sensitivity and adaptability.

2.7.1 Activity Complexity

One of the fundamental concerns when designing HAR systems is the categorisation of human activities. Currently, there is no standard or even research consensus guiding on how human activities should be categorised. One factor that adds even more complexity is the fact humans are multi-task agents being able to perform many activities at the same time. This creates an interpolation of human activities that difficult even more the recognition process [45].

When categorising human activities, there are many factors to consider, including activity duration, number of participants, cyclic actions, motion gestures, environment context. For instance, the work [37] categorises concerning the activity's duration, and complexity (Short Events, Basic Activities; and Complex Activities). Other works consider the transition states, like sitting-to-standing positions, as usual activities, like sitting or standing [46]. Thus, considering that research works tend to adopt different approaches when categorising human activities, it creates the obstacle of difficulting analysing these works and comparing them to determine which has the effective approach.

2.7.2 Energy Constraint

Mobile phones offer a vast and rich data-source to HAR systems, but some challenges limit the extraction potential. One major challenge is the limited battery life of phones. Since HAR systems consist of continuous sensing applications, if not properly design could rapidly drain the energy of the battery making the system unusable. Therefore, HAR systems must be energy efficient capable of operating as long as possible.

One significant factor when designing a HAR system is the sampling frequency. A low-frequency sampling system consumes less energy than a with high-frequency sampling. However, there is a less quantity of training data that can jeopardise the system becoming unusable. Therefore, there is a stand-off between system accuracy and energy efficiency that must be taken into account when designing a HAR systems [47].

Subsequently, research works have been proposing energy efficiency strategies such as Dynamic and Adaptive Sensor Selection. This strategy is designed to enhance energy consumption efficiency, turning off and on sensors regularly in real-time, to only utilise the most useful ones in specific contexts. For instance, since the GPS is not accurate in indoor scenarios, this strategy proposes that the GPS sensor is dynamically turned off when a user is performing indoor activities [48].

2.7.3 Users Privacy Constraint

HAR systems' performance dramatically improves when working with a vast and diverse deployment of sensors, especially with biometric sensors, that directly measure physiological data. However, biometric sensors as camera sensors and location sensors, inevitably, raise serious privacy concerns.

With physiological information, it is possible to infer user health information, which is sensitive and dangerous in the wrong hands of particular individuals. Security reports have been showing that health profiles are more valuable than credit-cards [49] and consequently this incentives professional hackers to obtain possession of it. Indeed, physiological features have the disadvantage of providing further information, non-relevant for HAR, about the health, religion, race of its users which has the dangerous potential of resulting in discrimination and racism situations.

Since it is essential to guarantee the privacy of its users, HAR systems must apply security techniques that guarantee the privacy of sensitive data in storage and communication. As such, techniques like encryption, obfuscation and anonymisation can avoid these malicious actions. It should be noted that even when applying techniques that try to suppress user identity, these are weak in situations where a malicious user has a priori knowledge of its victim. Already knowing the victim movements' patterns, it is trivial to deanonymize the data [50]. Thus in these scenarios encryption techniques become a necessary security technique.

Nonetheless, applying too many security techniques increases complexity and overhead in data prejudicing the system's performance. Thus, it is essential to have a balanced application of security techniques.

2.7.4 Subject Sensitivity and Model Adaptability

The performance of human activities depends on the physiological features of each person, furthermore, time, age and culture are primary factors that shape each person physiology. It is expected that a young person performs differently motion activities than older people or even disabled people. Therefore HAR systems must be flexible enough for each person.

Currently, typical HAR systems are unable to accurately recognise activities when the spatial data distribution alters, meaning that if the user starts to perform his activities differently, the system will not be able to adapt itself and correctly recognise his activities. This is a considerable limitation because people tend to perform activities differently over time, some more others less, due to ageing, or disabilities.

Though an immediate solution to this problem is training specific models to each person, it is a costly task for the user. As such, research works have been focusing on developing subject-independent models, some suggest collecting a highly diverse training data from different people with different ages and body types, while others propose cross-person activity recognition models [29].

Chapter 3

Transfer Learning

Transfer Learning (TL) is a sub-area of ML that focuses on the re-utilisation of knowledge between ML tasks. Research in this area has increased with the advent of Big Data, which allows the use of the wide availability of public repositories to boost the performance of classifiers, and diminishing the effort in each ML tasks of capturing data and the time necessary to build the ML model.

This chapter provides an overview of this sub-area, explains the basic terminologies necessary to understand TL and similar ML fields that are in some way similar to TL. Afterwards, considering the relevance of dataset shift within TL as the root cause for the mismatch of distribution between datasets, we categorise its possible causes and classify its different genres. Next, we pose a taxonomy of TL complemented with the recent researching works for each classification. Finally, we explain how TL has been applied in the domain of HAR with a survey analysis on wearables.

3.1 Transfer Learning Overview

Typical ML models assume that both training and test datasets have the same distribution and feature space, hence given any sort of mismatch the models stop working effectively and have to be retrained from scratch. For instance, a model trained with a specific category of books and tested with another category has worse performance, since each category is characterised by a particular set of words and sentences [51]. This assumption is indeed a significative limitation since its proven [52] that real problems suffer distribution mismatch over time and the test error is in proportion to the given difference.

Our biological learning process inspires the learning paradigm behind TL. The more experience we have better prepared are we when confronting novel conditions [53]. For instance, we naturally assume that having two persons learning the piano the one with violin experience will outperform the other with no musical background. As such, humans, while performing tasks, not only learn from it, but they learn how to generalise it and apply in related tasks. It is intended the same with algorithms that become capable of learning, improving their performance with their experience [54]. Hence, as a solution to this limitation of conventional ML algorithms, TL pretends to replicate this learning paradigm mathematically.

Besides, the utilisation of knowledge more than helping increase the model's performance, it decreases the effort of continually capturing high quantities of training data. Thus, it can prove very useful in scenarios where training data are insufficient, due to the associated costs of collecting it or where it is just impossible to obtain it.

With today's advent of Big Data, TL has been an especially attractive field since more and more datasets of various domains are becoming abundantly available. One can now leverage the knowledge of similar domains to boost the target domain performance, whether through the re-utilisation of data instances or even through the model's parameters. Past difficulties when building ML systems such as capturing labelled data can with TL be overcome by merely re-utilising the knowledge of different but related domains.

The following section addresses the scientifical meaning of TL, what it means, along with some ML terms that are mutually referred with TL within the literature, sometimes causing some misconceptions. Afterwards, an overview of TL is presented, by first presenting its history and the field founding researchers, concluding with the present challenges the field still faces.

3.1.1 Background and Notations

The field since its beginning has been identified with different names, such as knowledge consolidation, knowledge transfer, learning to learn. However, in this work its always referred as TL.

TL implies that both source and target datasets must be different because transferring knowledge from same datasets makes it a case of simple ML problem and not a TL scenario. Hence it is essential to know how to distinguish between datasets and acknowledge when they are equal and when are different. A dataset is a collection of data characterised by a specific domain and a specific task. In a TL scenario, the source and target dataset have to be different whether regarding the source and target domains or both tasks, and the respective cause of the difference is essential to know since it dictates what approach of TL can be applied.

The domain, as represented in 3.1, consists of two elements, feature space and a marginal probability distribution. The feature space is related to the variables that compose the dataset, except the target variable if present. These variables, also called features, when all combined define an n-dimensional feature space, n being the number of variables. Regarding the marginal probability distribution, the literature, in unison, defines it as the probability distribution of the variables contained in the datasets. In a way, while the feature space is about the features of the dataset, the marginal distribution concerns the value of each feature. For instance, in a document classification scenario, the feature space could be the language of the documents (e.g. English, Portuguese, Spanish) and the marginal probability distribution the topic approached by the document (e.g. politics, science, product reviews).

The task, as represented in 3.2, consists of two elements, a label space and an objective predictive function. The label space is composed of the target variables of the dataset, namely the set of labels. The objective predictive function consists of a mathematical function learned from the training data. For instance, in a binary classification scenario, the label space consists of all possible label values, “True” or “False”, and the predictive function learned from the training sample, predicts the label value for each test instance.

On a more scientific point of view, we relied the formal definition of Transfer Learning on the research survey of Pan and Yang [2] represented in 3.3.

Domain: *D*; ***Feature Space:*** *X*; ***Marginal Probability Distribution:*** *P(X)*;

$$then, D = \{X, P(X)\} \tag{3.1}$$

Task: *T*; ***Label Space:*** *Y*; ***Objective Predictive Function:*** *f(.)*;

$$then, T = \{Y, f(.)\} \tag{3.2}$$

Definition of Transfer Learning: *Given a source domain D_s and learning task T_s, a target domain D_t and learning task T_t, transfer learning aims to help improve the learning of the predictive function $f(.)$ in D_t using the knowledge in D_s and T_s, where $D_s \neq D_t$, or $T_s \neq T_t$* [2]. (3.3)

Having then at least one source domain, TL can leverage the experience to boost the performance of the target domain. By sharing similarities, both source and target datasets have familiar concepts that when isolated from the particular concepts can help the target domain’s task. For instance, regarding the computer vision domain, the recognition of edges is considered as a general concept. Hence, the rule is the more similarities both datasets share more useful knowledge is utilised and even higher is the performance of the target domain’s task.

TL, as described above, is a relatively vast field which consequently diverges into different sub-fields and it is often used interchangeably with other terms, sometimes incorrectly. Many ML terms are similar but not equal to TL while others are a branch of it. Therefore we call the attention to a couple of ML terms that can be easily confused with TL by explaining the fundamental similarities and differences.

Multitask Learning:

Multitask learning is a ML field that works by learning the knowledge acquired in different, but related, tasks to simultaneously leverage the performance of all these same tasks. This field is appropriate in scenarios where there are slight distribution differences that make it worth instead of building a generalised model, building a set of separate models which combined improve the overall performance.

For instance, it has been studied the application of multitask learning in web search ranking domain [55] where search queries traditionally have been treated independently, building specialised region-based models, due to different specificities in queries across regions. Since some of these queries are nevertheless region-insensitive sharing some commonalities, with multitask learning its models can be enhanced working cooperatively.

Though TL and multitask learning fields are similar in the sense that both are bias-driven methods, they differ because TL focuses on only leveraging the performance of a specific task, the target task, while multitask learning focuses on all tasks.

Self-Taught Learning:

This recent ML field was introduced in 2007 [56] to address the problems of acquiring labelled training data, like TL does, but also, and more importantly, tackle large-scale problems where traditional algorithms cannot parallelise the processing tasks taking significant time to resolve these problems. For instance, some works in this field address the processing problem by utilising parallel algorithms, namely sparse coding, that reduces the learning time significantly from weeks to days [57].

The ideal scenario where this field is most appropriate is when there is very few labelled data of the desired target class and having at the same time fewer difficulties obtaining different, but related, unlabelled data. Essentially, this field leverages the similarities of these various target classes to enhance the model performance given all these limitations.

In the domain of computer vision, where this field has been more applied, one appropriate scenario is when trying to recognise pine trees with few images of pine trees. One solution is to randomly select unlabelled pictures of another type of tree or in the worst case any other domain. Since all these pictures contain basic visual patterns, they can be used to enhance the model's performance.

This field is similar to semi-supervised learning because it works with labelled and unlabeled data. However, it does not assume that both datasets have identical distributions and the same task's class labels. In a way, self-taught learning can be considered as a genre of both semi-supervised learning and TL, where it is only available, though limitedly, the target domain labels, making its founding researchers naming it also as "unsupervised inductive transfer learning" [56].

Domain Adaptation:

Domain adaptation is a branch of TL where it is assumed a distribution mismatch between the source and target domain, both with the same task. In a way, this field is similar to semi-supervised learning in the sense of dealing with both labelled and unlabeled data.

Domain adaptation diverges into two different approaches: supervised domain adaptation, where it is available labelled data from both source and target domain; and unsupervised domain adaptation, where it is available labelled data only from the source domain.

When tackling the distribution mismatch, the literature about domain adaptation presents two distinct approaches, conservative domain adaptation and non-conservative. The former tries to build a model capable of performing well in both target and source domains, the latter approach focus only on performing well in the target domain.

Domain adaptation is widely used in various domains, one of which is sentiment analysis since classification performance is significantly influenced by the topic of discussion (e.g. movie reviews, product reviews, book critic). For instance, a person that reviews DVDs will utilise different words than when talking about kitchen compliances. In each domain, a specific set of words is more frequent than others (e.g. “dramatic”, “boring” in the former and “reliable” and “spacy” in the latter) that cause a distribution mismatch between domains. Hence, domain adaptation is a reliable solution since it discovers intermediate abstractions shared across all these domains to enhance the model’s performance [58].

Zero-Shot Learning:

Zero-Shot learning is a ML sub-area that focuses on classifying unseen classes through the utilisation of semantic embedding space, that is, recurring to high-level descriptions of these unobserved classes. Through this intermediate level, new classes can be recognised without any training example, the same way humans can recognise unseen objects by just reading a description of it [59].

This problem is prevalent in domains where new classes are always being introduced, like in e-commerce where there is a constant introduction of new products that can’t be captured in training time. In recent years, zero-shot learning has been extensively applied in computer vision, for instance, Ding et al. [60] explored within the domain of animal recognition, where new species are always being introduced. Their solution links visual features with a respective semantic representation to capture features across different observed classes, and for handling unseen classes, an ensemble strategy is applied. Experiments show that their model can reasonably capture unseen categories, and outperforms human manual categorisation.

Zero-Shot learning relates to the TL field since both utilise the knowledge acquired in the source domain to leverage the target domain function performance. However, Zero-Shot learning differs by not having available any target instance, whether labelled and unlabelled.

Continuous Learning:

Continuous Learning focuses on addressing the problems of distribution mismatch between different domains and of concept drift, that is the change of distribution of the target variable over time. The goal is using the knowledge of different domains to leverage the performance of the model and over time continuously adapt the model through incremental learning.

This research field is far from mature, and it is known by various names such as lifelong learning, incremental learning, and never-ending learning [61]. However, there are already available some research studies focusing on deep-learning strategies. For instance, Kirkpatrick et al. [62] developed a neural-network-based algorithm that is capable of learning new tasks sequentially without forgetting older ones in contrast to traditional deep-learning approaches. Seff et al. [63] and Zenke et al. [64] utilise a similar approach but for digit recognition, where the model learns digits, from zero to ten, in a sequential manner. Indeed, this field has broad applicability and potential to revolutionise AI but still faces some challenges that need to be overcome like guaranteeing efficient knowledge retention, an effective learning process, and flawless knowledge transfer [61].

Comparing with TL, the only considerable difference is that continuous learning is restricted to scenarios with a constant variation of distribution of the target variable over time. As such, CL can be considered as a specific type of TL, one that is more focused on adaptability over time.

3.1.2 Transfer Learning History

Research on TL dates back from the 1980s with the work of Mitchell [65] and Utgoff [66]. Utgoff worked on the induction of bias in ML, highlighting that it is a significant part of the learning task. The work is based on the human learning mechanism and states that bias must guide programs' learning process. Like humans learn concepts from examples, guided not only by the example that they observe but also by bias and having the ability to determine which concept is to be considered as best from the observations, bias must also guide programs. Later in 1997, Thrum and Lorien surveyed the field of TL, providing a collection of algorithms that "learn to learn". Both demonstrated just as humans often generalise correctly after a small number of training examples by transferring knowledge acquired in other tasks, the presented algorithms also often produce superior results to those that are not given extra information that comes from other tasks [54].

Nowadays, TL has been actively researched, for instance, since 2005 [67] it has been widely discussed in the Advances in Neural Information Processing Systems (NIPS) Workshop until now. Nevertheless, with almost 30 years of research, the field of TL still has no sufficient maturity to be massively utilised in the Industry. Furthermore, other ML areas have gained more research focus, such as Unsupervised and Reinforcement Learning, contributing to slower field advancements and breakthroughs. However, because of the current Industry success and massive utilisation of ML researchers predict that TL could benefit from it and become an essential technique in the future [68].

3.1.3 Transfer Learning Challenges

One significant challenge that currently has not been actively researched in TL is correctly deciding when and what knowledge should be transferred, to avoid negative learning transfer. This term is based on scenarios where TL instead of increasing the performance of the predictive model, it prejudices even more.

In fact, there have been researching studies [69] [70] empirically proving that though TL often helps in many cases, when having a significant dissimilarity between the source and target data TL hinders the performance. Therefore, to prevent the occurrence of negative transfer, there must be techniques that can give an insight on how similar the knowledge transferred is to the target and indicate whether the utilised TL approach has an adverse effect. The technique, in the form of a mathematical function, must compare and measure the similarities between the source and target data to pre-determine, in a quantifiable manner, how likely the transfer of knowledge will have positive or negative consequences on the performance. Ideally, the indicator should be domain-independent making it generic enough for any domain and should also be able to discriminate the knowledge into parts, to determine which parts are useful, instead of transferring the complete source knowledge [2].

3.2 Dataset Shift

To better comprehend TL it is necessary to understand what dataset-shift is about. Therefore, this section focuses specifically on the topic of dataset-shift in machine learning. First, it describes the causes of it by categorising into two: by design; by a dynamic environment. Afterwards, based on the literature this section describes three main variations to which dataset-shift can diverge to: covariate-shift, prior-probability-shift, and domain-shift.

3.2.1 Causes

Of the various causes of dataset shift, some can be categorised by design causes, which take place in the design process of training data sampling. Others can be categorised as environmental causes since the cause of shift is due to the inevitable changes over time of the environment's characteristics.

3.2.1.1 By Design

Dataset shift caused by ML system design is the most commonly analysed in the literature [71], where the captured training sample does not accurately represent the environment of study. This problem can be due to an unintentional flawed data capturing, making it a sample selection bias problem, or it can be due to the constraints of the environment imposed to the developers, making it an imbalanced data problem.

Sample Selection Bias

This type of shift occurs when the selected training sample does not accurately represent the distribution of the test sample.

This type of shift is very related to survey designs, where specific groups of subjects are left out, accidentally, of the experience, and hence, the survey becomes biased. For instance, a survey performed only in work-hours will probably address groups of people that are not working, like retired persons and therefore it is biased against typical working class people.

Another typical cause is during the pre-processing of data, for instance in character recognition, the cleaning of unintelligible characters makes the model accurate only for recognising intelligible characters. However, this approach makes the model useless, since typically characters are written in a non-intelligible way.

Imbalanced Data

Imbalanced Data is similar to sample selection bias since both refer to the problem of models becoming biased due to the distribution of the training sample. However, the difference is that in imbalanced data this bias is already known and is related to severe class distribution skew, which according to the literature, typically surpasses the order of 100:1 [72].

This scenario is very typical in many domains, where some classes, considered as rare events, appear a lot less than other domain remaining classes. In health-care, mainly in detecting cancer, this problem is recurrent since most datasets have a significative number of non-cancerous patients. To even complicate this situation, the consequence of incorrectly classifying cancerous people as non-cancerous is worse than the other way around [73] making it a critical problem.

3.2.1.2 By Dynamic Environments

A dynamic environment is one where its characteristics change over time, causing a mismatch in the distribution between the training sample and the future sample. It should be noted that these changes are particular to each domain. However, some of the most common are related to the drift of sensors' calibration, seasonality effects, and market trendings.

One typical example of applying ML in a dynamic environment is of recommendation systems. The problem here is that users' interest and behaviours continuously change over time, due to personal needs, group trends, amongst others, [74]. Therefore, to address the environmental characteristics, the recommendation system must continuously adapt itself to its users, since a model built for a user is useless given a period if it does not adapt to its user [75].

3.2.2 Taxonomy

The taxonomy here described is based on the research work of Moreno-Torres et al. [71] and the book of Lawrence et al. [76]. Both documents are relevant research studies that address the topic of dataset shift in ML and are widely cited, especially the latter. To understand the differences between each type of dataset shift, it is relevant first to present the basic terminologies from which the utilised definitions are based on:

- **X $\rightarrow$ Y Problem:** the input values determine the value of the class label. One example, of this kind of problem, is the detection of credit fraud, where the users' behaviour determine if there is fraud;

- **Y $\rightarrow$ X Problem:** the class label determines the input feature values. One example of this kind of problem is the medical diagnosis of patients, where the diseases, as the class label, determine the symptoms, as the input values;

- **Joint Distribution P(y,x):** represents the relationship between two variables, it is defined as P(y|x)P(x) in X $\rightarrow$ Y problems, on the other hand, in Y $\rightarrow$ X problems it is defined as P(x|y)P(y).

Based on these terminologies, we can now define the following three types of dataset shift: covariate shift; prior probability shift; and domain shift.

Covariate Shift: as one of the most studied types of shift in the literature [71] it is related with the change of distribution of the input between the training and the test samples while the conditional probability remains the same, as depicted in 3.4. Meaning that it happens when the domain characteristics change affecting the input values distribution.

To better demonstrate this type of shift, we provide the following example based on medicine drugs discovery. To predict chemical properties of interest researchers build learning systems to aid this process. However, covariate-shift is common during this process since researchers, on the beginning focus on particular chemical series, and as the project progresses, focus on newer and different chemical series. This mismatch of distribution between chemical series prejudices the prediction performance [77].

Covariate Shift: *Occurs only in* $X \rightarrow Y$ *problems where* $P_{source}(y|x)=P_{target}(y|x)$ *and* $P_{source}(x) \neq P_{target}(x)$. (3.4)

Prior Probability Shift: this type of shift occurs when sampling is dependent on the class label and independent of the feature space [78], on other words, whenever there are class distribution changes between training and test samples, as depicted in 3.5.

Prior Probability Shift: *Occurs only in* $Y \rightarrow X$ *problems where* $P_{source}(x|y)=P_{target}(x|y)$ *and* $P_{source}(y) \neq P_{target}(y)$. (3.5)

Domain Shift: also known in the literature as Concept-Drift [71], it is related to changes in the measurement process. This happens when there is a change of the conditional probability between the training and test samples and can take place when sampling is dependent on the class label and the contrary also, as depicted in 3.6. In other words, this type of shift happens when it is the process of decision itself that has changed, and not the input values.

Domain Shift: *Occurs in* $Y \rightarrow X$ *problems whenever* $P_{source}(x|y) \neq P_{target}(x|y)$ *and* $P_{source}(y) = P_{target}(y)$. *It can also occur in* $X \rightarrow Y$ *problems whenever* $P_{source}(y|x) \neq P_{target}(y|x)$ *and* $P_{source}(x)=P_{target}(x)$. (3.6)

3.3 Transfer Learning Taxonomy

This section starts by describing five categories of knowledge that can be transferred between the source and target domains. Then, it concludes with the different classifications of TL according to four criteria: similarity of feature space and label space, and characteristics of the source and target data samples.

3.3.1 What To Transfer

An essential question of TL is what to transfer, and based on the work of Weiss et al. [79] we categorise five types of transferable knowledge:

1. **Instance-Based Knowledge Transfer:** In this scenario, the source data instances are re-utilised in the target domain recurring to instance reweighting techniques, importance sampling, and other similar techniques that measure the source instances to decide which are more useful for the target domain.

2. **Feature-Representation Knowledge Transfer:** In this scenario, the source data features representation is transferred to the target domain in order to learn a common feature structure between both datasets. The feature representation can be accomplished by mapping the source features to the target features or in other cases, firstly introducing meta-features into the feature space proceeding to the mapping representation [80].

3. **Parameter Knowledge Transfer:** In this scenario, it is assumed that both source and target tasks share model parameters or prior distributions of the models' hyperparameters. Hence the knowledge transferred across tasks is encoded into the shared model parameters.

4. **Relational-Knowledge Transfer:** This scenario assumes there is some level of relationship between the source and target domains, and therefore the transferred knowledge is in the form of relational patterns. To better describe this type of transfer we present the example of the work by Mihalkova et al. [81] in the domain of sentiment analysis. Here the transferred knowledge is based on grammatical and sentence structure patterns in order to boost the target domain performance. However, the only problem ingrained in this type of transfer, especially in this work, is that the developed algorithms are tightly coupled to the respective domains of application.

5. **Hybrid-Base Knowledge Transfer:** This scenario is different from the above since the transfer performed is of various types of knowledge and not one only. Though this approach is least common amongst the literature, we can explain it better by describing the work by Xai et al. [82] that proposes the transfer of knowledge based on instances and feature representation. This research study focuses on domain adaptation in the domain of sentiment analysis in four different areas: book; movies; electronic products; and kitchen. The transfer of instances is done through a sample selection process in which there is a selection of a specific subset of the source domain whose distribution is most close to the target domain. Then, this subset is utilised as training data. On the other hand, the transfer of feature representation knowledge is done through a feature-ensemble architecture consisted of four individual classifiers. Each classifier is assigned to one of four different features that correspond to the users' speech (adverbs, verbs, nouns, and others). Experiments show the effectiveness of this approach.

3.3.2 Transfer Learning Classification

The literature classifies TL into various categories in different manners [2] [83], as such, we decided to restrict to only four different taxonomy criteria, the ones we considered the most relevant, which are feature space, label space, source data and target. The first two criteria we compare

Table 3.1: Transfer Learning Taxonomy

Feature Space	Label Space	Source Data	Target Data	Scenario
Same	Same	Labelled	Labelled	Supervised Domain Adaptation
Same	Same	Labelled	Labelled and Unlabelled	Semi Supervised Domain Adaptation
Same	Same	Labelled	Unlabelled	Unsupervised Domain Adaptation
Same	Same	Labelled	Unavailable	Domain Generalisation
Different	Same	Labelled	Labelled	Heterogeneous Supervised Domain Adaptation
Different	Same	Labelled	Labelled and Unlabelled	Heterogeneous Semi Supervised Domain Adaptation
Different	Same	Labelled	Unlabelled	Heterogeneous Unsupervised Domain Adaptation
Same	Different	Labelled	Labelled	Few-Shot Learning
Same	Different	Labelled	Unlabelled	Unsupervised Transfer Learning
Same	Different	Labelled	Unavailable	Zero-Shot Learning
Same	Different	Unlabelled	Labelled	Self-Taught Learning
Different	Different	Labelled	Labelled	Heterogeneous Transfer Learning

them between the source and target dataset to verify if they are the same or different. The latter criteria depict if the datasets are: labelled, unlabelled, labelled and unlabelled, or unavailable.

Therefore, based on the established criteria, we traced twelve different classifications, depicted in Table 3.2 that already had research studies exploring it.

3.3.2.1 Supervised Domain Adaptation

This scenario is characterised by identical feature spaces and label spaces between the source and target datasets, both differing only in the marginal distributions. Regarding the data availability, there are available labelled from the two datasets. However, the target sample only is not sufficient for an accurate classification task hence the need for the utilisation of the source datasets.

For this scenario, various techniques have been proposed such as the hybrid-knowledge transfer TrAdaBoost [84], and the parameter-based transfer adaptive structural SVM (A-SSVM) [85].

3.3.2.2 Semi-Supervised Domain Adaptation

This scenario is characterised by identical feature spaces and label spaces between the source and target datasets, both differing only in the marginal distributions. Regarding the data availability, there are available only labelled from the source dataset and both labelled and unlabelled data from the target dataset.

For this scenario, various techniques have been proposed such as the self-training framework named CODA (Co-training for domain adaptation) [86] and the work of Csurka et al. [87] for object recognition which proposes a generic adaptive learning technique.

3.3.2.3 Unsupervised Domain Adaptation

This scenario is characterised by identical feature spaces and label spaces between the source and target datasets, both differing only in the marginal distributions. Regarding the data availability, there are available labelled from the source dataset and only unlabelled data from the target dataset.

For this scenario, various techniques have been proposed such as the feature-representation knowledge transfer named Transfer Component Analysis, or TCA, by Pan et al. [88], the instance-based transfer technique by Gong et al. [89] that automatically locates latent domain from multiple sources. There is also the hybrid knowledge transfer work by Hsu et al. which re-weights instances and maps features to a shared space to reduce domain divergence [90].

3.3.2.4 Domain Generalisation

This scenario is characterised by identical feature spaces and label spaces between the source and target datasets, both differing only in the marginal distributions. Regarding the data availability, there are available labelled data from the source dataset and none from the target dataset. Hence, the goal here is to capture a wide variety of source datasets to be able to learn generalised invariant knowledge.

Given the characteristics of this scenario, the proposed works focus on only feature-representation transfer from multiple sources, such as the work of Li et al. [91] and the work of Ghifary et al. [92] that develops a Scatter Component Analysis method.

3.3.2.5 Heterogeneous Supervised Domain Adaptation

This scenario is characterised by identical label spaces and different feature spaces and marginal distributions between the source and target datasets. Regarding the data availability, there are available data labelled from both source and target datasets.

For this scenario, various techniques have been proposed, the most recent are the parameter-based transfer techniques by Zhou et al. [93] and the feature-representation transfer work by Sukhija et al. [94] that learns shared label distributions and relationships among feature spaces.

3.3.2.6 Heterogeneous Semi-Supervised Domain Adaptation

This scenario is characterised by identical label spaces and different feature spaces and marginal distributions between the source and target datasets. Regarding the data availability, there are available labelled data from the source dataset and both labelled and unlabelled data from the target dataset.

For this scenario, only the following techniques have been proposed: the hybrid knowledge transfer technique by Tsai et al. [95] recurring to data instances and feature knowledge transfer among different datasets; and the feature-representation based transfer by Xiao and Guo [96].

3.3.2.7 Heterogeneous Unsupervised Domain Adaptation

This scenario is characterised by identical label spaces and different feature spaces and marginal distributions between the source and target datasets. Regarding the data availability, there are available labelled data from the source dataset and both labelled and unlabelled data from the target dataset.

For this scenario, various techniques have been proposed. The most recent are the feature representation transfer techniques by Gupta et al. [97] and Yang et al. [98]. The former focuses on computer vision problems, mainly on unlabelled RGB images, the latter, of more general applicability, recurs to Markov Chain Monte Carlo method where each node corresponds to the domain and each edge weight the conditional dependence from one domain to another.

3.3.2.8 Few Shot Learning

This scenario is characterised by identical feature spaces and different label spaces and marginal distributions between the source and target datasets. Regarding the data availability, there are available labelled data from the source dataset and very few labelled data from the target dataset, typically one sample per category. This type of scenario is also called in the literature as one-shot learning [99].

For this scenario, various techniques have been proposed, the most recent are the works by Larochelle et al. [100] and Snell et al. [101]. Both works recur to parameter-based knowledge transfer. The former proposes a meta-learning approach and the latter through the use of a neural network learns non-linear mapping of the input into an embedding.

3.3.2.9 Unsupervised Transfer Learning

This scenario is characterised by identical feature spaces and different label spaces and marginal distributions between the source and target datasets. Regarding the data availability, there are available labelled data from the source dataset and only unlabelled data from the target dataset.

For this scenario, the proposed techniques restrict to feature representation-based transfer, the most recent are the work of Saligrama et al. [102] and Li et al. [103]. Both focus on object recognition problems, though the former develops a max-margin framework to learn the similarity between source and target. While the latter's framework tackles the target prediction function directly without introducing intermediate prediction, being also capable of working with semantic label information from the sources datasets.

3.3.2.10 Zero Shot Learning

This scenario is characterised by identical feature spaces and different label spaces and marginal distributions between the source and target datasets. Regarding the data availability, there are available labelled data from the source dataset and only auxiliary information, such as semantic information [104], from the target dataset to help classify unseen categories.

For this scenario, various techniques have been proposed, some of the most recent are the works of Reed et al. [105], Gan et al. [106] and Bucher et al. [107]. All the three focus on object recognition problems and leverage the semantic embedding of images so that the classifiers can make inference on this auxiliary information.

3.3.2.11 Self Taught Learning

This scenario is characterised by identical feature spaces and different label spaces and marginal distributions between the source and target datasets. Regarding the data availability, there are only available unlabelled data from the source dataset and labelled data from the target dataset.

For this scenario various techniques have been proposed, the most recent are the works of Kumagai et al. [108], and Kuen et al. [109]. Both focus on object recognition, though the former utilises parameter knowledge-based transfer, where only suitable parameters of feature mapping are learned to the objective task. While the latter work utilises the same type of knowledge recurring to a logistic regression classifier which distinguishes target object from the background, and is self-adaptive to any sort of concept-drift, such as the change of user appearance, through online learning.

3.3.2.12 Heterogeneous Transfer Learning

This scenario is characterised by different feature spaces, label spaces and marginal distributions between the source and target datasets. Regarding the data availability, there are available labelled data from both source and target dataset.

For this scenario, only two different approaches have been researched, feature representation-based transfer, and hybrid knowledge transfer. Regarding the former approach, the work of Jia et al. [110] employs a latent low-ranking tensor transfer learning to recognise human action from video frames. Regarding the latter approach, the work of Yang et al. [111] addresses this type of problem on sensor networks.

3.4 Transfer Learning in Human Activities Recognition

TL as a sub-field of ML has the same broad applicability as ML. Nonetheless, there is one main application where TL is being intensively applied. The healthcare domain, where it TL proves useful in helping improve the accuracy of medical diagnosis and, consequently, everyone's quality of life. HAR as inherently associated with human health-care proves to be an exciting field for applying TL.

A fundamental step towards understanding how TL has been utilised in HAR is surveying the literature. However, by studying the literature, one realises at the very first moment that research works diverge into three different sensor modalities: camera sensors; house network sensor; wearables. Given that our work focuses only on wearables, we restricted the survey on wearables only.

Hence, we gathered fourteen research studies, and categorised them based on the researched scenario and learning approach, if online or offline, as depicted in Table 3.2. From the analysis made, we first notice the need for research in TL approaches due to its low number of research studies. Furthermore, the surveyed works are considerably widespread through nine different scenarios.

This sub-chapter starts by differentiating the possible shift causes in a HAR system making use of the surveyed works. Then it proceeds by describing three possible types of transferable knowledge - instance, feature, and parameter-based - in the HAR domain. Since no work applying relational or even hybrid knowledge could be found, these categories were discarded.

Table 3.2: Existing work in HAR with TL approaches using wearables

Scenario	**Online/Offline Learning**	**Paper**
Heterogeneous Supervised Domain Adaptation	Offline	[112]
Heterogeneous Supervised Domain Adaptation	Online	[113]
Heterogeneous Unsupervised Domain Adaptation	Online	[114]
Few-Shot Learning	Offline	[115] [116]
Unsupervised Transfer Learning	Offline	[117]
Supervised Domain Adaptation	Online	[118] [119]
Unsupervised Domain Adaptation	Online	[120] [121] [122]
Unsupervised Domain Adaptation	Offline	[123] [124] [125]

3.4.1 Possible Shift-Causes

Before pondering what dataset shift causes there may be in a HAR system, one must first discriminate its agents that have a considerable role in this type of system. As such, we delineate two agents: the sensor; and the user.

A system to recognise activities it must first capture user data through one or a network of sensors. Then, considering the characteristics and the functioning of signal sensors, such as accelerometers, we realise that various possible sensor related shift causes are prevalent. One is the sensor placement. A sensor if positioned differently than the training position causes a different sample distribution making more difficult the task of classifying activities. This situation is likely to happen since sensor placement can sometimes be intrusive to the user causing him to change the position voluntarily. On the other hand, if even the sensor is placed in a comfortable position, it can be misplaced involuntarily during the activities, especially the ambulant ones. Furthermore, another shift cause can be related to the modification of the sensing infrastructure, for instance through the addition of new sensors or even the disconnection of sensors for an energy-saving purpose. Indeed various works have been addressing these type of problems, and based on the performed survey, depicted in Table 3.2, we can highlight the works by Calatroni et al. [112], Khan et al. [116], Rokni [114], Krishnan [119], Kurz et al. [113], Roggen et al. [120], and Zhao et al. [125].

Then, having captured the data, the classifier must recognise the activities performed by the user. The problem here is that the more composed are the activities more likely is that users perform it differently, difficulting the process of generalising the model to a wide range of users. For instance, an injured person may likely perform ambulating activities differently than a non-injured person, or even a young person compared to a much older person. Hence, the physiology of users must be thought of when designing a HAR system, to prevent training a model with users different than the target user. One solution might be developing a specialised model for each subject, but this requires significative effort that in many cases is just impractical. Other shift-causes are also prevalent such as the effect of ageing, more related to long-term models, and the proficiency of tasks by users, that is likely to change over time also, though it has a higher effect on highly composite activities, as studied in the work of Winter et al. [126]. For user-related shift problems, many works have been proposed, such as the works of Deng et al. [122], Zhao et al. [124], Chen et al. [117], Hachiya et al. [123], Diethe et al. [118], and Fallahzadeh [121].

Though we consider the last shift causes the most relevant in HAR, it should be noted that there are many others. For instance, one is the environment itself, typical in smart homes HAR, where models are trained with laboratory data that do not accurately represent a real-world scenario. As research works have shown, laboratory environments generally restrict and influence subject activity patterns [127].

3.4.2 Instance-Based Transfer

The work of Hachiya et al. [123] explores an instance-based transfer approach under a covariate shift problem to recognise motion activities using only accelerometers from a mobile device.

The proposed method, named Importance-Weighted Least-Squares Probabilistic Classifier (IWLSPC), consists in new importance weighted variant of the least-squares probabilistic classifier combined with a sample re-weighting approach. The importance weights are controlled with flattening parameters that allow weight smoothness contributing to the accuracy of the model.

Thought the proposed method has the limitation of relying on the assumption that the class-posterior probability does not change between training and test phases, experiments on real-scenarios illustrated the usefulness of the method.

3.4.3 Feature-Based Transfer

The work of Rashidi and Cook [128] explores a feature-based transfer approach to recognise human activities in smart environments. The work's approach consisted in collecting data from six smart houses during a three month period for three houses and a six month period for the other three. Each house has a different layout, for instance, some have two bedrooms, while others have only one, with different residents' daily routines.

Regarding the sensing equipment, every house has a wide variety of sensors, including motion sensors to monitor doors' status, scattered throughout all the house. So in total, the work experiments and compares six different TL scenarios, one for each house, considering that for each case the knowledge is transferred from the rest five houses (source domain) to the target house (target domain). Due to the houses different layouts, sensors inevitably have different location positions and different properties than the target sensors. Therefore the work maps between the target and source domain the activities and sensors based on their similarity.

As a solution, two matrices are created, one for the activities similarity and other for the sensor similarity. In both matrices, the similarity is measured in probabilistic value, from 0 to 1. The final work's phase is labelling target activities based on activity mapping probabilities, in which only the label with the highest weighted vote is selected.

The work successfully reduces the differences between the source and target feature spaces recurring to mapping techniques. Though the proposed method is not capable of handling with multi-resident scenarios, the results prove the method can successfully recognise activities in target domains where there is no labelled data.

3.4.4 Parameter-Based Transfer

The work of Kasteren et al. [129] explores a parameter-based transfer approach to recognise the daily activities of elderly people, such as bathing and cooking, to automate the monitoring of health care tasks. The capturing of data is done only by external sensors scattered throughout all the house that measure the interaction of the person with the houses, that detect if doors are being opened, measure the flushing of toilets and temperature of the house and much more.

Since the labelling of the data is very expensive, the developed work proposes a method using labelled data from source domains with unlabelled data from the target domain, to learn the parameters of the activity model of the target domain without labelling the captured data. The researchers of this work utilise the Hidden Markov Model (HMM) because it allows incorporating the labelled data with unlabelled data and the model, being a probabilistic temporal model, successfully handles noisy data.

The work first task is mapping the source and target domain data into a matrix because the source houses have a different layout than the target house and could prejudice the model's performance if no measures are taken. Then, both unlabelled and labelled data are combined into a single dataset, applying semi-supervised learning to find the maximum-likelihood parameters for the target model. The proposed method achieved an accuracy higher than 90% when recognising daily activities such as drinking, sleeping, showering and toileting.

Chapter 4

Methodology

Having surveyed the literature to understand how researchers have been addressing TL in HAR, we decided to explore from the twelve different scenarios presented in Section 3.3.2 Unsupervised Domain Adaptation. The main reason that underlies this choice is related to the fact that it constitutes a frequent scenario when designing HAR systems. It is impractical requiring users to label their data and, at the same time, is nonsense when having easy access to repositories with a considerable amount of labelled data. These two conditions impulse the utilisation of TL techniques to leverage this past-knowledge, or in other words, experience, to boost the task of accurately classify our target's activities. Furthermore, this thesis is part of a more extensive research project with other co-workers, one that focuses on exploring novelty detection in HAR systems. As such, this forces the dissertation to explore a scenario where the target dataset is composed only of unlabeled data, hence unsupervised domain adaptation.

This chapter has the goal to explore the field of unsupervised domain adaptation in machine learning comprehensively. Therefore, the first section presents the four approaches that techniques of this field tend to diverge into (instance-weighting, self-labelling, feature representation, and cluster-based). For each of these approaches, we specify some implementations proposed by researchers. The final section focuses on the four selected techniques we selected to explore TL in HAR: Transfer Component Analysis, Subspace Alignment, Nearest Neighbour-Based Importance Weighting, Kernel Mean Matching.

4.1 Unsupervised Domain Adaptation

Unsupervised domain adaptation assumes the availability of labelled data only from the source domain, and unlabelled data from the target domain. The two domains have the same feature space and label space but differ regarding the marginal distributions (e.g. different users, sensor's positions, surrounding environments). Notice that semi-supervised and unsupervised domain adaptation are not very different since both deal with both labelled and unlabeled data. However, unsupervised domain adaptation assumes a distribution mismatch between the unlabelled data (target domain) and labelled (source domain), while semi-supervised assumes an equality.

This section objective is to perceive how researchers have been addressing unsupervised domain adaptation by briefly describing some of the techniques proposed by researchers. Hence, based on the survey work of Margolis [130], we firstly notice four different approaches to tackle this field: instance-weighting; self-labelling, feature representation; cluster-based. An instance weighting approach assigns weights to the source samples depending on their similarity with the target samples. A self-labelling approach focuses on adapting classifiers trained from the source samples to the new target domain. An approach of feature representation constructs an abstract representation of the data, while a cluster-based one assumes that from high-density regions we can perceive the similarity between the two domains. As a concluding note, from the available research works, we notice that researchers have been focusing more on the feature representation and instance weighting approaches due to a more significant number of available research works.

4.1.1 Self-Labelling Approach

A self-labelling approach, also known as classifier adaptation, addresses domain adaptation by adapting a pre-trained classifier from the source to the desired target domain. Typically consists in an iterative method that trains an initial model based on the labelled source data only, use that to estimate labels on target data, and then use the estimated labels for building another model [130].

There are various techniques available in the literature such as the work of Yang et al. [131] introduce a method called Adaptive Support Vector Machine (A-SVM) that adapts a SVM classifier to a target dataset. For instance, the work of Jiang et al. [132] also proposes a similar technique named cross-domain SVM that is based on support vector machines. It adapts previously learned support vectors from a source domain to aid the classification of the target domain. Rastrow et al. [133] proposes a method based on entropy stability combined with conditional entropy minimisation, helping in selecting parameters settings corresponding to stable decision boundaries. Duan et al. [134] introduce a method named FastDam which employes a sparsity regularizer to learn a sparse target classifier with the support vectors only from the target domain, making the prediction very fast.

4.1.2 Instance Weighting Approach

An instance Weighting approach reweights the source instances that are more similar to the target instances. Determined by statistical methods the weights are to be assigned to a loss function minimising the expected loss over the distribution of data [130].

By surveying the literature, we can notice many research works proposing various instance weighting techniques. For instance, Bickel et al. [135] introduce a discriminative technique deriving from a Newton gradient descent procedure. One characteristic of their technique is that it tackles the distribution mismatch by characterising the existing divergence between the source and target domain without the intermediate step of estimating training and test distribution. On the other hand, Schapire et al. [136] propose three techniques that incorporate sample selection bias in a common density estimation based on the principle of maximum entropy - it is related with the estimation of probability distribution from data, meaning that the distribution should be as uniform as possible, that is, have maximal entropy [137]. Sugiyama et al. [138] introduce a variant of the Cross Validation selection technique called importance weighted CV (IWCV) that specifically addresses the problem within model selection under covariate-shift. The same authors proposed another method called Kullback-Leibler Importance Estimation Procedure (KLIEP) based on the minimisation of the Kullback-Leibler divergence metric. Tsuboi et al. [139] introduce a new variation of KLIEP more computationally efficient which reformulates the optimisation problem as an unconstrained convex problem and also modifies the function for modelling the importance function to a log-linear model instead of a linearly one. Loog [140] introduces a nonparametric estimator based on a Voronoi tessellation of the space, named Nearest Neighbour Weighting (NNeW). As such, the weights are determined through a nearest-neighbours search for every target data point among the source data points. Huang et al. [141] propose the Kernel Mean Matching (KMM) method, which consists in a weighted estimator based on the Maximum Mean Discrepancy - the distance between the mean of two sets of samples. Without performing density estimation, the weights are assigned to the instances in a reproducing kernel Hilbert space.

4.1.3 Feature Representation Approach

A feature representation approach implies an alteration of the feature representation of the data because it assumes specific original features are domain-specific. As such, this approach assumes the existence of hidden features (latent features) from the original features that contain common knowledge shared between the source and target domains. A mapping is then created, through matrix factorisation, from the original feature space to a latent one.

The literature is overwhelmed with a considerable number of feature representation works. Therefore we here only highlight some of the most utilised by researchers. Such as the work of Gong et al. [142] introduces a novel kernel-based method, named geodesic flow kernel, which constructs a set of subspaces that characterise changes in geometric and statistical properties from the source to the target domain. Then, the method extracts only the subspaces that are domain invariant. Sun et al. [143] propose an elementary method, named Correlation Alignment (CORAL), developed to be easily integrated into applications without requiring any specialised knowledge about domain adaptation. The method aligns the marginal distributions of the source and target domains by minimising their second-order statistics difference. Baktashmotlagh et al. [144] introduces a technique that is able to work without and with labels of both domains during the training process. Regardless, the technique projects the data to a low-dimensional latent space where the dissimilarity between the source and target distribution is measured recurring to the Maximum Mean Discrepancy. Long et al. [145] proposes a sparse representation method, name Transfer Sparse Coding (TSC), that focuses on cross-distribution image classification problems. It uses the nonparametric Maximum Mean Discrepancy to minimise the difference of distributions between the domains. Furthermore, this metric is also incorporated into the objective function of sparse coding to make new representations of the source and target images. The same authors also [146] developed another method named Joint Distribution Adaptation, which focuses on jointly adapting the marginal and conditional distributions of both domains in a principled dimensionality reduction procedure. The technique extends the Maximum Mean Discrepancy to measure the distribution differences and integrates with the Principal Component Analysis to construct a new feature representation of the data. Jordan et al. [147] introduce a method, named Deep Adaptation Network (DAN), based on deep convolutional neural networks. The method's approach focuses on enhancing the feature transferability in the task-specific layers of the network by reducing the domains' discrepancy. Tzeng et al. [148] proposes a method, named Deep Domain Confusion (DDC), composed by a convolutional neural network architecture that can work not only on unsupervised domain adaptation but also on a supervised scenario with available labelled training data. One of the layers composing the method consists of an adaptation layer based on the Maximum Mean Discrepancy metric, that automatically learns a representation that is both semantically meaningful and domain invariant. Pan et al. [88] introduce the Transfer Component Analysis technique which tries to learn some components across domains in a reproducing kernel Hilbert space using maximum mean discrepancy. Fernando et al. [149] propose an algorithm where the source and target domains are represented by subspaces spanned by eigenvectors. This technique, named Subspace Alignment, focuses on learning a mapping function to align the source subspace to the target one.

4.1.4 Cluster-Based Approach

Cluster-based methods determine the labels of data points based on the density regions, that is, two data points are likely to have the same label in the condition that between them exists a high-density path. This assumption is similar to semi-supervised clustering which the data is partitioned into a group of subsets, named clusters, such that observations within a cluster are more similar to one another than observations in other clusters [150]. However, the cluster based approach here discussed differs from semi-supervised clustering since it assumes a distribution mismatch within the data.

Regarding any published implementation, we only found one proposal by Dai et al. [151]. It introduces a co-clustering method that focuses on learning text data across domains in an iterative manner. It starts by propagating label information to define the class structure from the source domain data. Then co-clustering is extended for the target domain data to constrain domain-invariant data shared among the two domains.

4.2 Employed Techniques

The thesis goal is to explore TL in HAR in a general manner, that is, to utilise as much TL techniques as possible and not a specific one. We were able to utilise two feature representation TL techniques, Transfer Component Analysis, and Subspace Alignment. And two instance-weighting techniques, Nearest Neighbour Weighting, and Kernel Mean Matching. Though we had access to other techniques, such as GFK and DIP, the selected were the only ones available whose implementation language was the same as the one used in the pre-processing of the datasets, Python[1].

This section is composed of four subsections, each one dedicated to one of the TL techniques used. The four selected techniques are thoroughly described with the support of a pseudo-algorithm illustrating its steps. To get a sense of the prevalence of each technique within researchers, we complement the descriptions with the research works where the techniques were utilised.

4.2.1 Transfer Component Analysis

Transfer Component Analysis (TCA) is an unsupervised domain adaptation technique with a feature-representation based approach. It assumes that there exists a transformation that can equal the marginal distributions of the source and target domains. Hence, the fundamental goal here is finding a transformation mapping that minimises the distance between the marginal distributions, and at the same time, preserving essential properties of the source and target domains. To better understand how TCA operates a pseudo-algorithm is depicted in Algorithm 1.

Algorithm 1 Transfer Component Analysis Algorithm

Input: source domain dataset, $D_s = \{(x_{s_i}, y_{src_i})\}_{i=1}^{n_S}$, and target domain dataset, $D_t = \{(x_{t_j})\}_{j=1}^{n_T}$.
Output: transformation matrix W .

1: **procedure** TCA(D_s, D_t)
2: embed both the source and target domain into a shared low-dimensional latent space, $K_{S,T}$.
3: learn a nonlinear transformation, W, that maps the source subspace to the target one as close as possible using the divergency-metric Maximum Mean Discrepancy.
4: **return** transformation matrix W .

To minimise the distributions distances, the technique, instead of explicitly finding a transformation mapping, it recurs to a dimensionality reduction method, Maximum Mean Discrepancy Embedding (MMDE)[152], that learns a low-dimensional latent space common to both domains. This way, it embeds both source and target domains into a shared low-dimensional latent space using a nonlinear mapping function to learn then a corresponding kernel matrix portrayed in equation 4.1. The kernel matrix is defined on all data by minimising the distance, measured with Maximum Mean Discrepancy metric, between the projected two domains while maximising the embedded data variance. Having found a transformation matrix, a classifier can finally be trained with it to predict the target domain accurately.

$$W = \begin{vmatrix} W_{S,S} & W_{S,T} \\ W_{T,S} & W_{T,T} \end{vmatrix} \in \mathbb{R}^{(n_S+n_T)\times(n_S\times n_T)} \tag{4.1}$$

By reviewing the literature to determine the usability of TCA amongst researchers, we can notice that there is a considerable number of works that utilise TCA as a domain adaptation technique. Besides, since its introduction in 2011, TCA has been widely used in many domains,

[1] Python 3.0 release, https://www.python.org/download/releases/3.0

including human behaviour prediction [153], diagnosis of gearbox fault [154], object detection [155], and sentiment classification [82].

4.2.2 Subspace Alignment

Subspace Alignment (SA) is an unsupervised domain adaptation technique with a feature-representation based approach. As represented in Algorithm 2 two main steps compose the technique's approach, the first is the generation of subspaces by representing the source and target domains into subspaces spanned by eigenvectors. The final step is the alignment of the two subspaces to decrease the discrepancy between both source and target domains.

Algorithm 2 Subspace Alignment Algorithm

Input: source domain dataset, D_s=$\{(x_{s_i}, y_{src_i})\}_{i=1}^{nS}$, and target domain dataset, D_t=$\{(x_{t_j})\}_{j=1}^{n_T}$.
Output: transformation matrix, W .

1: **procedure** $\mathrm{SA}(D_s, D_t)$
2: project each source, $\{ x_{s_i}\}_{i=1}^{nS}$, and target data, $\{ x_{t_j}\}_{j=1}^{n_T}$, to its respective subspaces, X_S and X_T.
3: learn a linear transformation, W, that maps the source subspace to the target one as close as possible using the divergency-metric Bregman.
4: **return** transformation matrix W .

Rather than manipulating the original source and target data, the technique represents both domains in a more robust manner via subspaces. The subspaces are first created by transforming every data to a D-dimensional z-normalised vector. Then, using PCA, the d eigenvectors corresponding to the d largest eigenvalues are selected to be used as vector bases of the source and target subspaces ($X_S, X_T \in \mathbb{R}^{D \times d}$). As for the alignment of the subspaces, the goal is to align the source data to a target aligned source subspace (X_A) and the target data to the target subspace. This is achieved by learning a linear transformation that through a transformation matrix M maps the vector bases of the source subspace to the target ones. The technique for measuring the distance between the subspaces uses the Bregman divergence metric, essential to the minimisation of subspaces divergency. Finally, after aligning the subspaces, a classifier can be trained with the source data projected into the target aligned source subspace and the target data into the target subspace.

By surveying the literature, we explicitly observe that the majority of research studies using SA focus on the domain of computer vision, such as the work of Fernando et al. [156] which annotates old photographs with the respective correct location label given a set of labelled recent photos. There are also the works of Mittal et. [157], Cao et al. [158], and Gupta et al. [159] that focus on particular visual domain shifts, such as illumination, image blurring, and specific type of optical devices (spectroscopic, photometric) respectively. However, there are other domains where the SA technique has also been used, including wifi localisation through the work of Fernando et al. [156].

4.2.3 Nearest Neighbour Weighting

Nearest Neighbour Weighting (NNeW) is an unsupervised domain adaptation technique with an importance weighting based approach. As represented in Algorithm 3, the technique applies a Voronoi tessellation of the space to help to determine the weights for each source sample. Hence, the weight on each sample is dependent on the number of target neighbour samples. As the final step, it is then applied Laplace smoothing for regularising the weights to avoid some getting exaggeratedly biased towards the test set.

Algorithm 3 Nearest Neighbour Weighting Algorithm

Input: source domain dataset, D_s=$\{(x_{s_i}, y_{src_i})\}_{i=1}^{ns}$, and target domain dataset, D_t=$\{(x_{t_j})\}_{j=1}^{n_T}$.
Output: source domain weights, $\{\widehat{w}_{S_i}\}_{i=1}^{ns}$.
1: **procedure** NNEW(D_s, D_t)
2: employ a Voronoi tessellation to estimate weights of every source sample, x_{S_i}, by counting the number of target samples, x_{T_j}, that are whithin its associated Voronoi cell.
3: **return** source domain weights $\{\widehat{w}_{S_i}\}_{i=1}^{ns}$.

For estimating a weight function, w, NN performs data interpolation through Voronoi tessellation to determine the influence of the target samples on the source samples. As such, the idea is to create a schema of Voronoi cells that represent regions of influence associated with the source samples. Each weight is then estimated by counting the number of target samples that are within its respective cell, V_i, as depicted in equation 4.2. The advantage of adopting a Voronoi tessellation is that it guarantees no target data is wasted as the cells cover the whole space and every target data is helpful in the decision of the weight of at least one source sample. Finally, in order to avoid the existence of too many empty Voronoi cells and an exaggerated bias of the weight function towards the sample set, the reached tessellation is regularised with Laplace Smoothing by adding one to each cell.

$$\widehat{w}(x_{S_i}) = |V_i \cap \{x_{T_j}\}_{j=1}^{N_T}| \tag{4.2}$$

Since importance weighting techniques are widely applicable in machine learning [140], there is a considerable number of different approaches proposed by researchers regarding domain adaptation. This makes the adoption of NN in research studies shyer comparing to the feature representation techniques. As such we can only highlight the work of Kouw et al. [160] in the domain of healthcare, more specifically in the diagnosis of heart diseases in patients. Here, the NN is essential to address the problem of covariate-shift caused by geographically biased sampling of patients.

4.2.4 Kernel Mean Matching

Kernel Mean Matching (KMM) is an unsupervised domain adaptation technique with an importance weighting based approach. It directly gives estimates of the source weights without performing density estimation as most algorithms in this type of scenario [141]. As represented in

Algorithm 4, the technique recurs to the technique Maximum Mean Discrepancy to help to define the weights in a way that the divergence between the source and target domain is minimised.

Algorithm 4 Kernel Mean Matching Algorithm

Input: source domain dataset, D_s=$\{(x_{s_i}, y_{src_i})\}_{i=1}^{ns}$, and target domain dataset, D_t=$\{(x_{t_j})\}_{j=1}^{n_T}$.
Output: source domain weights, $\{\widehat{w}_{S_i}\}_{i=1}^{ns}$.

1: **procedure** KMM(D_s, D_t)
2: estimate weights of every source sample such that the means of the the source and target points in a reproducing kernel Hilbert space are close.
3: **return** source domain weights $\{\widehat{w}_{S_i}\}_{i=1}^{ns}$.

To overcome the divergence of the marginal distribution between the source and target domain, KMM firstly transposes the data points in a reproducing kernel Hilbert space. Then, the weights for the source samples are defined by minimising the distances to the target samples measured with the Maximum Mean Discrepancy metric. Finally, the reweighted samples can then be incorporated into a classifier algorithm.

By surveying the literature, we observe the broad applicability of the KMM technique in various domains. For instance, in the work of Hassan et al. [161] the technique is used for addressing the problem of covariate-shift within the domain of speech emotion recognition. In the work of Feng et al. [162] it is used to address JPEG mismatched steganalysis while in the work of Zhu et al. [163] it is used in the diagnosis of Alzheimer's Disease due to the existing distribution mismatch between the training data collected in laboratory and the samples collected in real clinic setting. As proven, KMM is indeed utilised in a wide range of domains.

Chapter 5

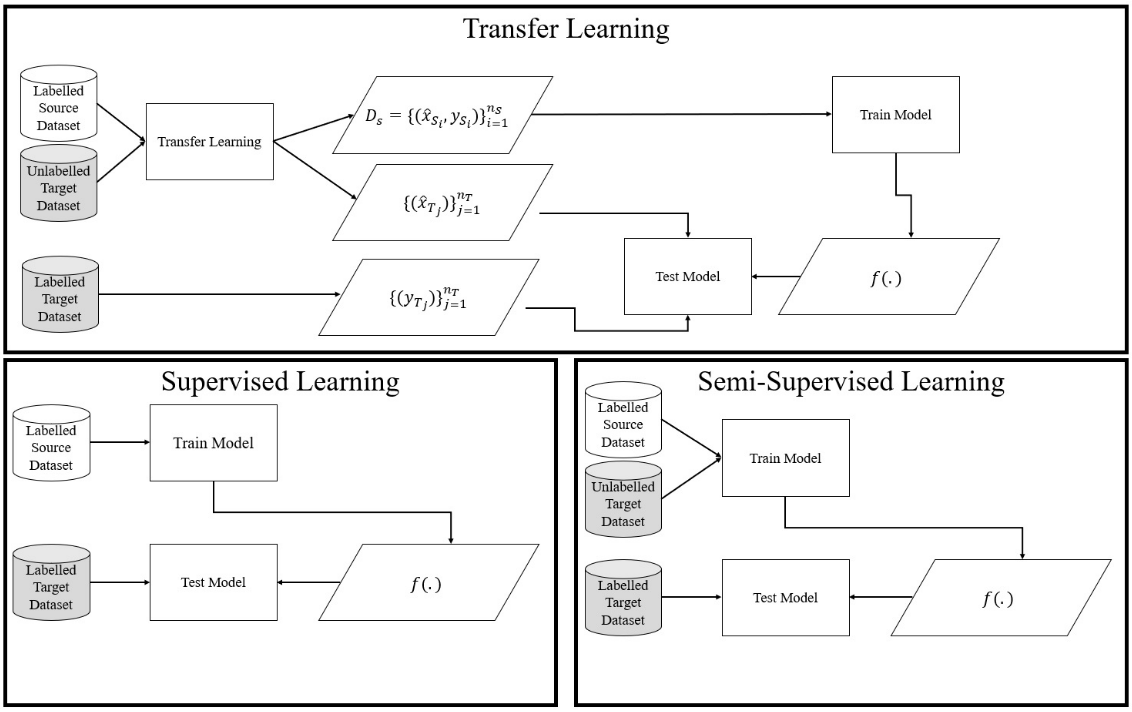
Experiments

Having determined four different TL techniques (TCA, SA, KMM, NNeW), we intended to explore scenarios where there is a mismatch of distribution between the source and target datasets caused by three motives: user's age (section 5.4); sensor misplacement (section 5.5); and environment (section 5.6). In order to perform these scenarios, we employed two HAR datasets, described in section 5.2, here named as PAMAP2 [164] and ANSAMO [165], collected by researchers and made publicly available. After preparing the data (section 5.3) we constructed twenty-two different approaches (section 5.1), sixteen TL-based, four supervised, and two semi-supervised. The supervised and semi-supervised approaches serve as baseline methods to help understand the presence of negative transfer within the TL approaches. The sixteen different TL approaches were implemented because for each TL approach we employ four different classifiers to minimise the classifier's influence. This also allowed us to understand with which classifier the TL techniques worked better, for instance, KMM has, in overall, superior accuracy with decision trees while SA and NNeW work better with logistic regression. Since each TL technique has its unique characteristics, we also developed a majority voting ensemble composed by the TL techniques with the highest overall performance along the experiments (KMM with a decision tree, SA with logistic regression, and NNeW with logistic regression). Finally, we conclude the experiments by analysing the obtained results through analytical tests (section 5.7).

5.1 Overall Design

In total, we delineated twenty-seven unique experiments grouped into five different scenarios (AA-AGE, AA-POS, AP-ENV, PA-ENV, PP-POS). The name given to a scenario is based on the initial letter of the employed datasets, for instance, a scenario named AA-AGE utilises the dataset ANSAMO as both source and target while AP-ENV utilises as source the ANSAMO dataset and PAMAP2 as target one. The AA-AGE scenario explores the influence of age on the performance of human activities. It makes use of labelled data of young individuals (with less than 30 years old), to help identify the activities of older people (more than 50 years old). The AA-POS and PP-POS scenarios are for exploring the misplacement of sensors by employing labelled data of a particular position (e.g. wrist) and unlabelled data of another location (e.g. chest). Finally, the AP-ENV and PA-ENV scenarios apply two datasets, PAMAP2 and ANSAMO, collected by similar users and sensors but collected in different environments. While labelled data of ANSAMO dataset compose AP-ENV source dataset, PA-ENV uses PAMAP2 as the source dataset and ANSAMO as the target one.

Figure 5.1: Overall design of the experiments.



In each experiment besides employing transfer learning techniques, we also utilise two more different learning approaches as baseline methods as depicted in Figure 5.1. One is semi-supervised learning where it trains a model with the labelled source and unlabelled target samples. The other is supervised learning where a model is taught only with the source and tested with the target sample. These two approaches are necessary to help identify negative transfer in the TL results. Regarding the TL techniques, it should be noted that these methods only manipulate the source and target data to conjugate their distributions, and do not make the classifications itself. As such, a classifier technique is also necessary to make the classifications with the data outcome of the TL techniques. To minimise the influence of the classifier on the accuracy we employed four different ones, Logistic Regression [166], Bernoulli Naive Bayes [167], Decision Tree [168], and Linear Support Vector Machine [169]. From these four classifiers, we then select the one with the highest accuracy to determine the TL's accuracy for each experiment, making it best-case scenario approach. Therefore, by combining with each TL technique four classifiers we implemented, in total, sixteen unique TL approaches. Furthermore, the same classification techniques are applied in the supervised approach, and regarding the semi-supervised learning, we adopt two methods, Label Propagation [170], and Label Spreading [171]. As such, for each experiment we developed twenty-two unique approaches, sixteen are TL-based, four supervised, and two semi-supervised.

After analysing the results yielded by the TL techniques, we decided to employ a majority voting ensemble composed by the three most accurate TL approaches, KMM with Decision Tree, NNeW with Logistic Regression, and SA with Logistic Regression. It should be noted that the TCA technique was not employed given that in the majority of the experiments it experienced negative transfer.

5.2 Datasets

In order to devise the experiments, we firstly performed a survey of HAR datasets made publicly available by researchers. We then selected two, PAMAP2 [165] and ANSAMO [164], that had a set of similar activities and made use of accelerometer, gyroscope, and magnetometer sensors, as depicted in Table 5.1. Since one scenario we committed to required creating distribution mismatch by age, we necessarily had to select a dataset that had a wide age range, from 14 to 55 years old, hence ANSAMO. While regarding the PAMAP2 dataset, it was selected a-posteriori since it was very similar to the ANSAMO with similar sensors, activities, and user's physiological characteristics.

Regarding the ANSAMO dataset, it was designed to determine the importance of wearable mobility sensors' position and network for discriminating human falls from activities of daily living (ADL). The dataset data for every user was captured in the same domestic environment, more specifically in an apartment block. The seventeen tested users, six females and eleven males, had ages from 14 to 55 years old, with weights between 50 and 93 kilograms and heights from 155 to 195 cm. Of all subjects, two had more than 50 years old while the remaining had an age below 30. In overall, the dataset includes nine different ADLs, walking, jogging, bending, hopping, ascending stairs, descending stairs, lying down, getting up, and sitting down. Every subject executed these activities at least three times, expect two (those older than 50 years) that did not perform jogging. Regarding the utilised sensors, every subject when performing the activities transported a network of five sensors. Four had a chip composed by a triaxial accelerometer, gyroscope, and magnetometer and were attached through elastic bands at the ankle, ankle, wrist, and chest of the subjects. Additionally, a smartphone located in the pocket was responsible for

Table 5.1: Description of the employed datasets.

Name	Year	Users Characteristics	Activities	Type of Sensors	Sensors Placement
PAMAP2	2012	08M 01F 24-32 years-old 169-194 cm 65-92 Kg	12	acc, gyro, magn, heart-rate	chest, wrist, ankle
ANSAMO	2016	11M 06F 14-55 years-old 155-195 cm 50-93 Kg	9	acc, gyro, magn,	chest, waist, wrist, ankle

activating the four wearables and receiving the wearables data via Bluetooth low energy. As for the sampling rate, in the smartphone was 200 Hz while in the wearable due to hardware limitations was 20 Hz.

Regarding the PAMAP2 dataset, it was designed to capture a wide range of activities, including sports and household, and help study complex activities. Young 9 individuals comprise the dataset, 8 males and 1 female, with similar physiological characteristics ageing from 24 to 32 years old, with heights ranging from 169 to 194 cm, and weight values between 65 and 92 kilograms. In total the dataset contains 18 different activities, some are more basic (walking, running, cycling, ascending stairs, and descending stairs,), some posture-related (lying, sitting, and standing), a few more of everyday sort (ironing, vacuum cleaning), others more physical-related (rope jumping, Nordic Walking, rope jumping). Most of these activities were performed following a protocol which the users had to execute with a duration of approximately 3 minutes for each activity. As for the hardware setup, each user had attached 3 wearables onto the chest, ankle, and wrist, composed by three-axial accelerometer, gyroscope, and magnetometer sensors. Additionally, on the chest was attached a heart-rate monitor, especially useful for intensity estimation when distinguishing activities with similar movements. The data was sampled at 100 Hz in the three wearables while the heart-rate sensor provided values with approximately 9 Hz.

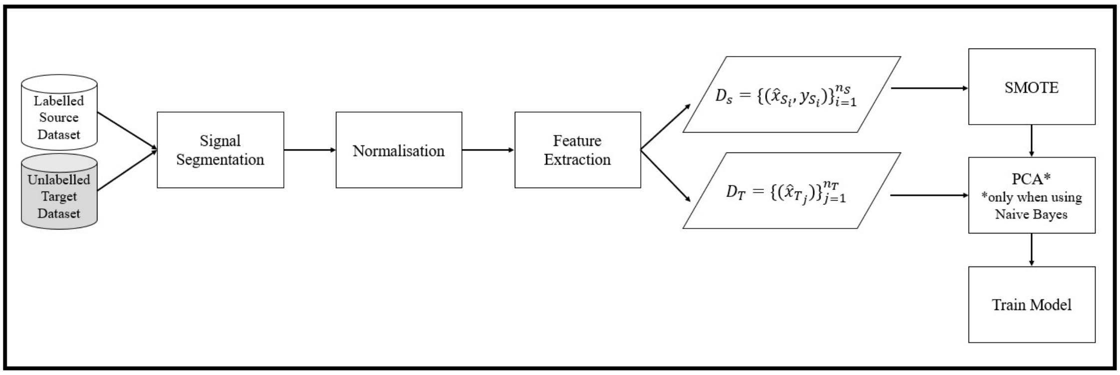
5.3 Data Preparation

From the employed datasets we have access only to raw signal data of three type of sensors, accelerometer, gyroscope, and magnetometer. Each data instance is therefore composed of three type of signals distributed within three axes and the time instance from when it was captured.

As any HAR system, the first necessary step here is to apply signal segmentation to allocate the data in time windows. The problem is that signal segmentation is of crucial importance for the performance of a HAR system, and there is no clear consensus among researchers of the best approach. By resorting to the literature studies addressing this problem we, based on the work of Banos et al. [172], decided to apply a non-overlapping fixed sliding window of three seconds. A sliding window approach proves to be the most useful for the recognition of periodic (e.g. walking) and static activities (e.g. sitting).

After segmenting the data, we proceed to its normalisation and then its feature extraction. For all signals, we apply twelve time-domain techniques (average, maximum, minimum, variance, centroid, standard deviation, root mean square, interquartile range, mean absolute deviation, auto-correlation, skewness, and kurtosis). Finally, to avoid class imbalance, we apply the oversampling technique SMOTE [173] which will create synthetic samples for the minority classes. Finally, only on the approaches using Naive Bayes (five from the twenty-two), we utilise the dimension reduction PCA to decorrelate features.

Figure 5.2: Data preparation design.



5.4 Age as shift cause

We intend in this scenario to explore the age as the leading cause for distribution mismatch between datasets. It is expected that a young person performs activities differently, maybe quicker than older people. As such, we resorted to the ANSAMO dataset since it has both young (<30 years old) and older subjects (>50 years old). Because the dataset has a limited number of subjects with more than 50 years old, only two, we could only implement a scenario where we would use these two subjects as target dataset only. This meant younger users would compose the scenario's source dataset to help classify the activities of the older people, as the target unlabelled dataset. In total we implemented 3 scenarios, each trying to classify within different positions (ankle, chest, and wrist) 6 activities (bending, hopping, sitting, ascending stairs, and walking).

We applied the 22 approaches and evaluated their performance with two metrics, accuracy and f1-score, the latter for a more realistic result in situations with an imbalanced test set. In overall, as illustrated in Table 5.2, the ensemble approach had the highest average score while the semi-supervised the worst. From the individual TL techniques, TCA was the only one suffering from negative transfer since the supervised approach had a higher score by 19 percentage points. The remaining TL techniques had a higher score than the supervised approach but with less than 10 percentage points of difference.

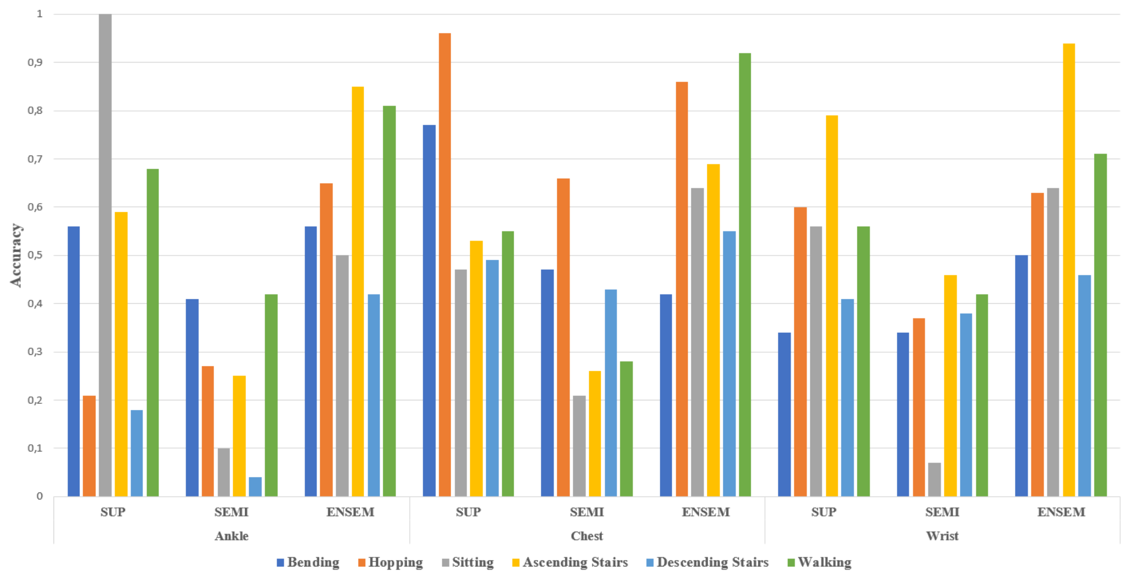
Table 5.2: Overall results (accuracy, and f1-score) of the different employed approach types in the AA1-AGE scenario.

	ACC	**F1**
SUP	0,52	0,52
SEMI	0,3	0,31
KMM	0,58	0,58
NNeW	0,57	0,57
SA	0,54	0,53
TCA	0,35	0,33
ENSEM	<u>0,6</u>	<u>0,6</u>

Figure 5.3 illustrates even in more detail the accuracy per activity of the ensemble and the

baseline approaches. We can notice right away that the activities TL leveraged the most in all three experiments were ascending stairs, descending stairs, and walking. In fact, of all six activities these are the more dynamic while the others are static (hopping, bending) or posture related (sitting).

Figure 5.3: Accuracy obtained in each three experiments composing the AA1-AGE scenario of the supervised, semi-supervised, and transfer learning approach (ensemble).



5.5 Position as shift cause

We intended in this scenario to explore the misplacement of sensors as the leading cause for distribution mismatch between datasets. Each experiment consists of using a source dataset with data of a certain position (e.g. wrist) to help identify the activities of a target dataset captured from a different position (e.g. ankle). We devised two scenarios, one using the ANSAMO, named AA4-POS, and the other the PAMAP2 dataset, named PP1-POS. Since the ANSAMO dataset had data captured from four different positions (ankle, chest, waist, wrist), and the PAMAP from three (ankle, chest, wrist), we devised 18 unique experiments. The former scenario comprehends 12 experiments while the latter 6. As for the activities, the AA4-POS scenario had 7 (bending, hopping, running, sitting, ascending stairs, descending stairs, and walking) and PP1-POS had 8 (lying, rope jumping, running, sitting, ascending stairs, descending stairs, standing, and walking).

Table 5.3 illustrates both accuracy and f1 scores per approach. Though in scenario AA1-POS the ensemble technique had the highest score in the other scenario suffered from negative transfer with less six percentage points than the supervised approach. Regarding the TL techniques individually, TCA obtained in PP1-POS scenario the worst performance with a significant difference, and in scenario PP1-POS none TL technique achieved an effective knowledge transfer (only SA had a similar performance with the supervised approach).

Table 5.3: Overall results of the different employed approach types in the AA4-POS and PP1-POS scenarios.

	AA4-POS		PP1-POS	
	ACC	**F1**	**ACC**	**F1**
SUP	0,25	0,19	0,57	0,55
SEMI	0,24	0,22	0,33	0,32
KMM	0,35	0,29	0,5	0,46
NNeW	0,34	0,3	0,55	0,52
SA	0,24	0,19	0,57	0,55
TCA	0,27	0,21	0,19	0,13
ENSEM	0,36	0,34	0,54	0,49

Tables 5.4 and 5.5 discriminate the performance of the ensemble and baseline methods for each experiment (also including the ones with the same source and training positions). In all AA4-POS experiments (excluding the ones with the same source and training positions) the ensemble technique is able to leverage, though not significantly, the performance by having higher accuracy than the baseline methods. As for the PP1-POS scenario, the ensemble technique in the majority of the experiments suffered from negative transfer, mainly when the ankle samples were utilised as source dataset.

Table 5.4: AA4-POS scenario accuracy results of the supervised, semi-supervised, and transfer learning approaches.

		Target											
		Ankle			Chest			Waist			Wrist		
		SUP	SEMI	ENSEM	SUP	SEMI	ENSEM	SUP	SEMI	ENSEM	SUP	SEMI	ENSEM
Source	**Ankle**	0,89	0,39	0,83	0,23	0,19	0,31	0,08	0,11	0,21	0,22	0,38	0,47
	Chest	0,19	0,17	0,24	0,94	0,37	0,8	0,39	0,31	0,52	0,46	0,27	0,5
	Waist	0,13	0,16	0,24	0,25	0,32	0,43	0,88	0,37	0,8	0,24	0,36	0,36
	Wrist	0,2	0,32	0,36	0,29	0,13	0,31	0,33	0,2	0,4	0,87	0,37	0,69

Table 5.5: PP1-POS scenario accuracy results of the supervised, semi-supervised, and transfer learning approaches.

		Target								
		Ankle			Chest			Wrist		
		SUP	SEMI	ENSEM	SUP	SEMI	ENSEM	SUP	SEMI	ENSEM
Source	**Ankle**	0,84	0,6	0,78	0,42	0,3	0,37	0,65	0,34	0,56
	Chest	0,67	0,21	0,64	0,76	0,43	0,73	0,61	0,31	0,59
	Wrist	0,57	0,54	0,47	0,49	0,25	0,58	0,68	0,47	0,7

5.6 Environment as shift cause

Here, the goal was to explore the environment as the primary cause for distribution mismatch between datasets. We delineated two scenarios, AP1-ENV and PA1-ENV, in which the former uses the ANSAMO dataset and PAMAP2 as source and target datasets, respectively. The latter employs the PAMAP2 dataset to leverage the activity recognition in the ANSAMO dataset. Every implemented experiment classifies 5 activities (running, sitting, ascending stairs, descending stairs, and walking) in 3 different positions (ankle, chest, wrist). Therefore, in total there are 6 experiments, 3 in AP1-ENV and 3 in PA1-ENV.

Table 5.6 demonstrates the average obtained accuracy and f1 score per approach in the two scenarios. The ensemble approach obtained the highest score in the PA1-ENV with a gap of 11 percentage points between the supervised. However, in the other scenario KMM approach was the TL technique that leveraged the most with a 7 percentage points difference than the supervised approach. TCA suffered from negative transfer in both scenarios.

Table 5.6: Overall results (accuracy, and f1-score) of the different employed approach types in the AP1-ENV and PA1-ENV scenarios.

	AP1-ENV		**PA1-ENV**	
	ACC	**F1**	**ACC**	**F1**
SUP	0,65	0,64	0,71	0,7
SEMI	0,49	0,44	0,46	0,43
KMM	<u>0,72</u>	<u>0,71</u>	0,77	0,75
NNeW	0,68	0,67	0,78	0,77
SA	0,68	0,65	0,76	0,76
TCA	0,27	0,17	0,25	0,1
ENSEM	0,65	0,6	<u>0,83</u>	<u>0,82</u>

Figures 5.4 and 5.5 discriminate the results obtained by the ensemble and baseline approaches per activity. In AP1-ENV the ensemble approach had the highest performance in the ankle experiment where outperformed all activities compared with the baseline methods. Furthermore, from the AP1-ENV three experiments, we can notice that the activities from which the TL technique leveraged the most were ascending stairs, descending stairs, and sitting. As for the PA1-ENV scenario, the ensemble obtained a similar performance in the three positions, with higher accuracy than the baseline methods in the activities descending and ascending stairs.

Figure 5.4: AP1-ENV scenario accuracy results of the supervised, semi-supervised, and transfer learning approach (only ensemble).

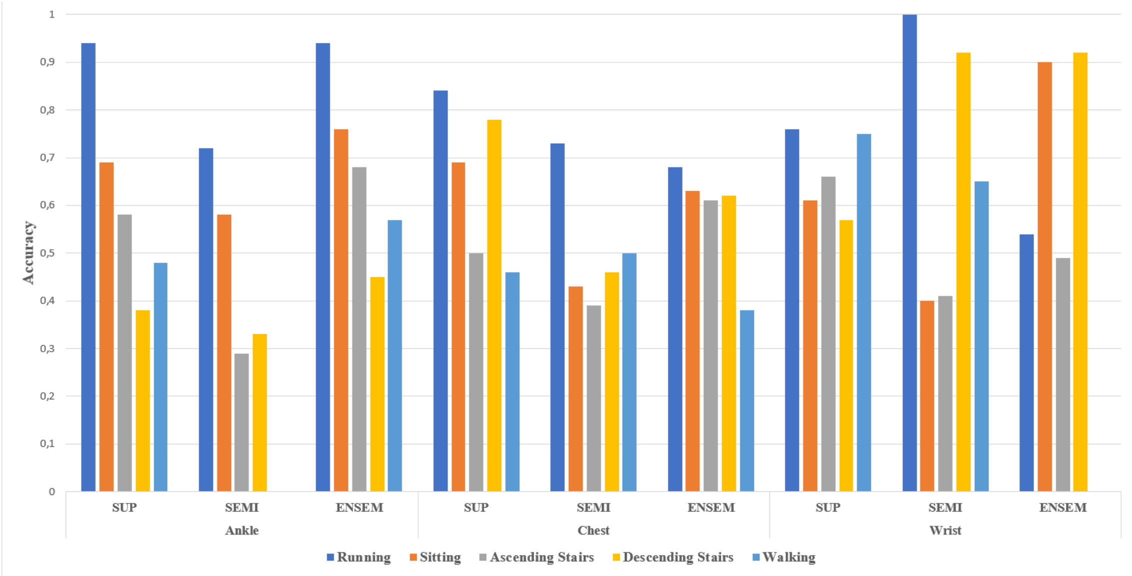
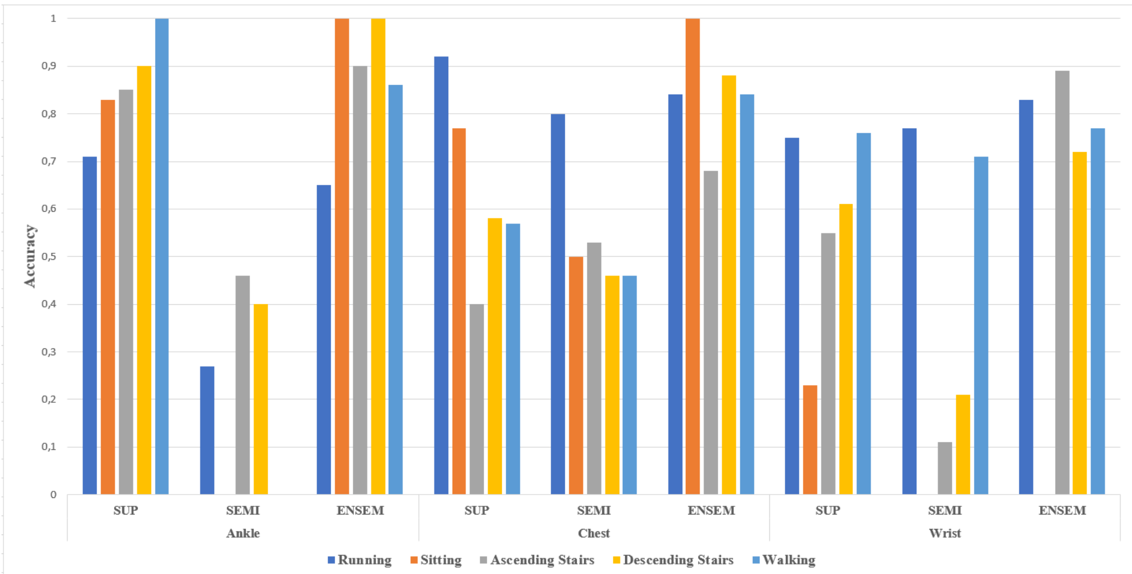


Figure 5.5: PA1-ENV scenario accuracy results of the supervised, semi-supervised, and transfer learning approach (only ensemble).



5.7 Discussion

Altogether, we tested twenty-two different approaches in twenty-seven unique experiments grouped into five scenarios. The goal now is to focus on analysing the performance of the ensemble method comprehensively, that was composed of three different TL techniques, comparing to the baseline approaches, the supervised and semi-supervised ones. In three of the five scenarios, the ensemble approach had a significant transfer of knowledge gaining 8 percentage points in AA1-AGE, 15 in AA4-POS, and 12 in PA1-POS comparing with the supervised approach. In fact, looking at the obtained performance per activity we can notice that TL leveraged the most especially in ambulating activities, such as ascending and descending stairs and running. However, in PP1-POS and AP1-ENV the ensemble suffered negative transfer, having an inferior performance by 7 and 4 points, in f1-score, than the supervised approach.

To statistically understand if there is a significant difference regarding the obtained results, depicted in Table 5.7, between the ensemble and baseline methods, we resorted to the Friedman hypothesis test [174]. Having performed the test, the results indicated a significant difference (Friedman chi-squared = 28.43, df = 2, p-value = 6.707e-07, p-value < 0.01). We then proceeded to the application of the Nemenyi post-hoc test [175] to identify the existing differences among the performances of the three approaches. As such, according to test, there is a highly significant difference between the ensemble and semi-supervised approach (p = 3.3e-07, $p < 0.01$) and a significant difference (p = 0.022, $p < 0.05$) between the ensemble and supervised approach.

Furthermore, with Figure 5.6, which compares the performance variation between ensemble and supervised, and between ensemble and semi-supervised, we notice a positive median variance in the two comparisons. While the ensemble has a median positive variance relative to the semi-supervised of about 23 percentage points, it has a minor variation relative to the supervised approach with 9 percentage points. In overall, the results show the usefulness of applying TL in scenarios where there is a considerable distribution mismatch.

Figure 5.6: Accuracy's variation between the ensemble technique and the supervised approach, and between the ensemble technique and the semi-supervised approach.

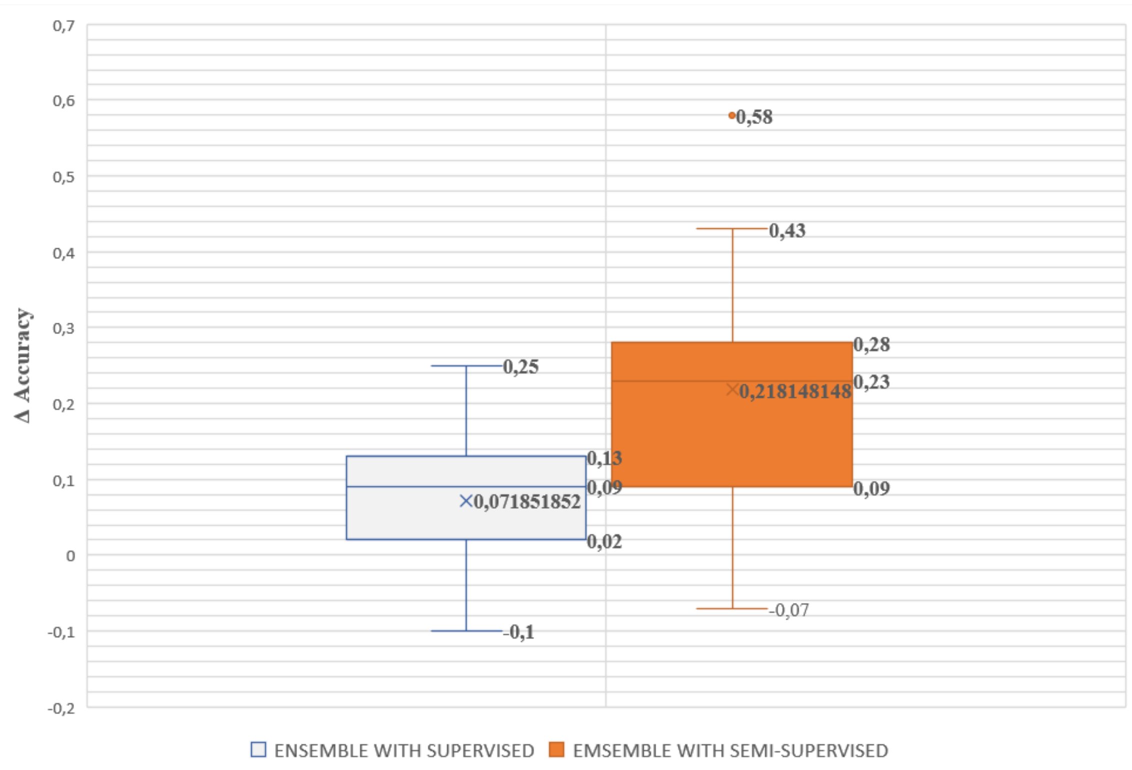


Table 5.7: Accuracy obtained from the ensemble, and baselines methods per experiment.

		SUP	SEMI	ENSEM
AA1-AGE	**Ankle**	0,48	0,21	0,58
	Chest	0,6	0,4	0,64
	Wrist	0,5	0,31	0,59
AA4-POS	**Ankle-Chest**	0,23	0,17	0,31
	Ankle-Waist	0,08	0,11	0,21
	Ankle-Wrist	0,22	0,38	0,47
	Chest-Ankle	0,19	0,17	0,24
	Chest-Waist	0,39	0,31	0,52
	Chest-Wrist	0,46	0,27	0,5
	Waist-Ankle	0,13	0,16	0,24
	Waist-Chest	0,25	0,32	0,43
	Waist-Wrist	0,24	0,36	0,36
	Wrist-Ankle	0,2	0,32	0,36
	Wrist-Chest	0,29	0,13	0,36
	Wrist-Waist	0,33	0,2	0,4
AP1-ENV	**Ankle**	0,71	0,31	0,89
	Chest	0,92	0,58	0,86
	Wrist	0,67	0,5	0,76
PA1-ENV	**Ankle**	0,87	0,31	0,89
	Chest	0,66	0,58	0,86
	Wrist	0,67	0,5	0,76
PP1-POS	**Ankle-Chest**	0,42	0,3	0,37
	Ankle-Wrist	0,65	0,34	0,56
	Chest-Ankle	0,67	0,21	0,64
	Chest-Wrist	0,61	0,31	0,59
	Wrist-Ankle	0,57	0,54	0,47
	Wrist-Chest	0,49	0,25	0,58

Chapter 6

Conclusions

In overall, we accomplished the dissertation delineated objectives. Our work not only contributes to unsupervised domain adaptation field but to TL as well through the extensive performed survey. As for future work, we envision the study of this same domain but with an online learning approach.

This chapter starts with section 6.1 by describing, in overall, the work performed throughout the dissertation along with the accomplished objectives. It concludes with section 6.2 by stating the limitations and the envisioned future work.

6.1 Accomplished objectives

We undertook this dissertation to explore TL in the HAR domain with two objectives in mind: reducing the effort of capturing labelled data; and utilise different, but related, datasets to boost performance. We started by studying the crucial topics within HAR, from the different sensor architecture approaches to the typical ML techniques responsible for the classification of activities. Then, we proceed to research the field of TL, firstly studying its basic terminologies, fundamental to understand the existence of dataset shift, and its extensive taxonomy. We conclude state of the art with an extensive survey of the HAR research works applying TL. As for the methodology, we determined to delve into unsupervised domain adaptation since it constitutes as a common HAR scenario, one where we have unlabelled data from target user/s and effortless access to considerable labelled data sample/s of different distributions. Through another survey, we selected four techniques, TCA, KMM, NNeW, and SA to experiment with both individually and through a majority-voting ensemble (composed with KMM, NNeW, and SA) in series of 5 different experiments scenarios. One scenario, AA1-AGE explores the user's age as the central cause of distribution mismatch, AA4-POS and PP1-POS explore the sensor's misplacement while AP1-ENV and PA1-ENV explore the environment. Each scenario is composed of a series of experiments, making in total 27 unique experiments we implemented. For each experiment, we compare the performance of the TL techniques with supervised and semi-supervised approaches, both serving as baseline methods. The yielded results indicate a highly significant difference between the

TL ensemble approach with the semi-supervised approach and a less but significant difference with the supervised approach. While the performance variation between the ensemble and semi-supervised is around positive 23 percentage points, with the supervised is of 9. Nevertheless, the obtained results show clearly the usefulness of TL within HAR.

Since the literature is scarce in research works (14) applying TL in HAR using wearables, the dissertation's content is of great relevance to the field. In general, we highlight the most significant work's contributions:

- **A comprehensive study on unsupervised domain adaptation:** while a typical research work focuses only on a single distribution mismatch, our work focuses on three different ones.

- **An extensive survey on HAR works applying TL:** we performed an intensive survey, collecting in total 14 research studies which we analysed and categorised them according to five different criteria (feature and label space similarity, source and target data characteristics, and learning approach - if online or offline). This information is significantly useful to understand the literature tendencies within this specific domain and for newcome researchers since it takes considerable time to execute it.

6.2 Limitations and Future Work

Having obtained in experiments negative transfer by some TL techniques, especially with TCA, it would be better to use multiple datasets as source instead of using one dataset. This multi-source dataset approach would likely have higher success since it increases the chance of discovering useful transferable knowledge between the source and target datasets. Furthermore, since the applied data pre-processing techniques, such as the sliding window and feature extraction techniques, have a significant influence regarding the model's outcome it would be interesting to adopt deep learning approaches. This is because with a deep learning approach there is no need to hand-craft features, removing any possible negative influence from the pre-processing techniques.

Since we explored unsupervised domain adaptation with an offline learning approach, we consider as the next work's phase shifting it to an online learning approach. This would be interesting to investigate because the system would instantly classify the data and update itself over time. It is expected, initially, a low accuracy but after a sufficient time period, the system could obtain a high performance just like an offline approach.

Appendix A

Experiments' Results

Figure A.1: AA4-POS scenario accuracy results of the supervised, semi-supervised, and transfer learning approaches.

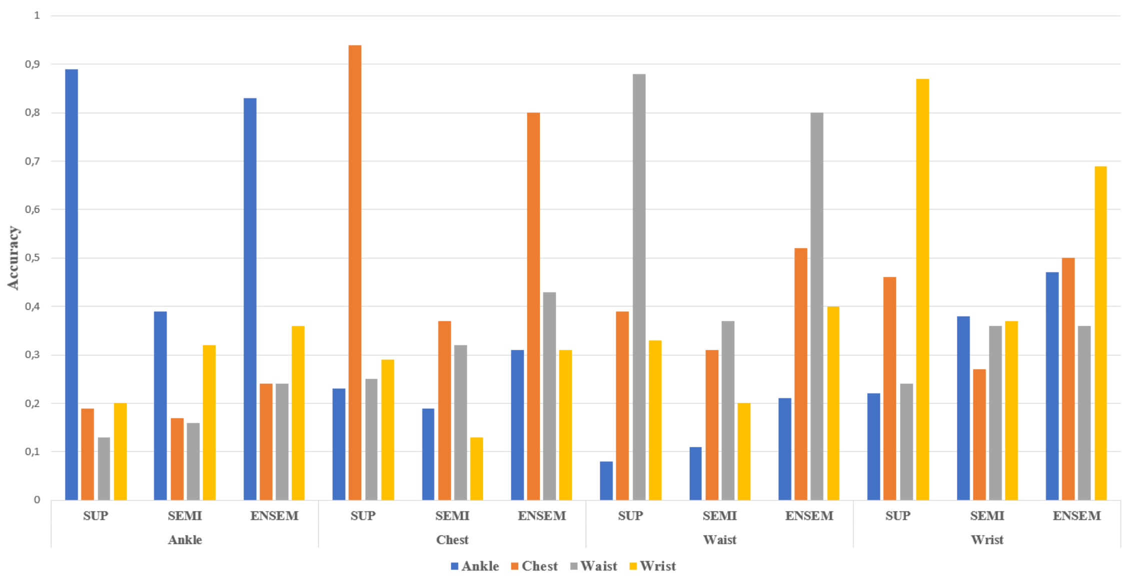


Table A.1: AA1-AGE scenario accuracy results of the supervised, and transfer learning techniques per utilised classifier.

	Ankle				Chest				Wrist			
	DT	LR	NB	SV	DT	LR	NB	SV	DT	LR	NB	SV
SUP	0,4	0,48	0,32	0,21	0,6	0,6	0,5	0,3	0,39	0,5	0,26	0,13
SA	0,27	0,46	0,21	0,44	0,28	0,63	0,3	0,35	0,26	0,52	0,25	0,3
TCA	0,26	0,31	0,31	0,25	0,63	0,41	0,32	0,39	0,23	0,33	0,28	0,32
KMM	0,62	0,38	0,32	0,35	0,61	0,56	0,5	0,35	0,51	0,38	0,25	0,35
NN	0,4	0,57	0,3	0,43	0,8	0,64	0,29	0,38	0,46	0,51	0,25	0,36

Figure A.2: PP1-POS scenario accuracy results of the supervised, semi-supervised, and transfer learning approaches.

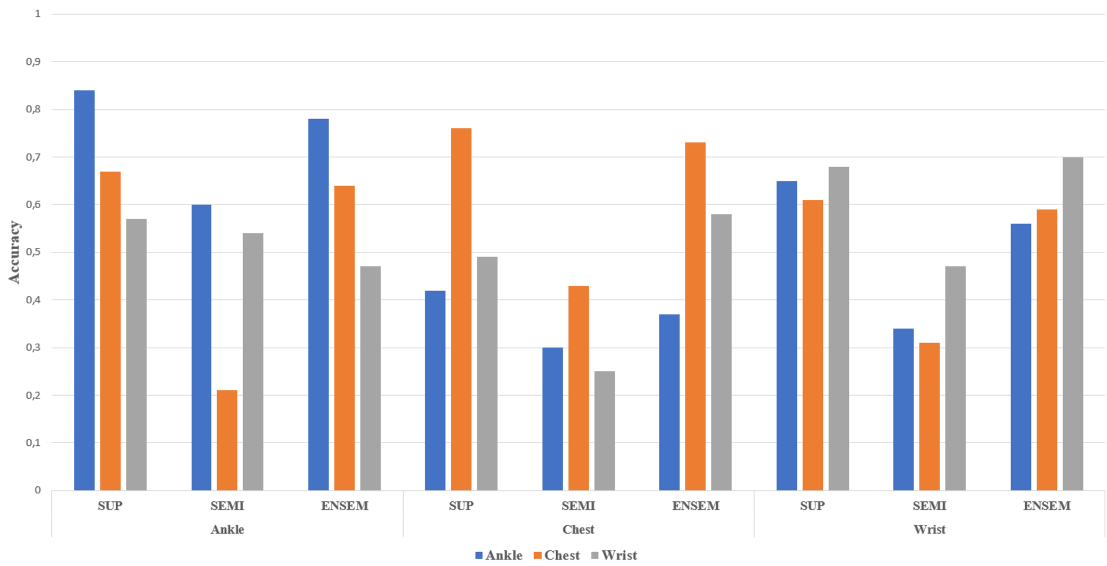


Table A.2: AA4-POS scenario accuracy results of the transfer learning techniques.

		Source															
		Ankle				**Chest**				**Waist**				**Wrist**			
		TCA	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**
Target	**Ankle**	0,56	0,93	0,81	0,73	0,27	0,19	0,28	0,17	0,9	0,13	0,21	0,37	0,14	0,21	0,33	0,28
	Chest	0,16	0,23	0,33	0,29	0,39	0,94	0,95	0,86	0,16	0,25	0,38	0,58	0,15	0,28	0,4	0,38
	Waist	0,25	0,08	0,29	0,22	0,22	0,31	0,45	0,61	0,43	0,88	0,89	0,91	0,22	0,32	0,4	0,38
	Wrist	0,24	0,22	0,14	0,26	0,18	0,49	0,4	0,24	0,2	0,23	0,42	0,41	0,41	0,87	0,88	0,72

Table A.3: AA1-AGE scenario accuracy results of each transfer learning technique per activity.

	Ankle				Chest				Wrist			
	TCA	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**
Bending	0,47	0,55	0,6	0,59	0,44	0,71	0,53	0,58	0,36	0,39	0,31	0,4
Hopping	0,31	0,17	0,5	0,81	0,74	0,86	0,93	0,69	0,33	0,62	0,57	0,66
Sitting	0,17	0,86	1	0,75	0,19	0,5	0,58	0,64	0,12	0,56	0,67	1
Ascending Stairs	0,5	0,62	0,79	0,46	0,33	0,55	0,68	0,58	0,67	0,81	0,93	0,58
Descending Stairs	0,12	0,17	0,41	0,81	0,43	0,56	0,55	0,58	0,25	0,4	0,4	0,54
Walking	0,39	0,61	0,58	0,73	0,29	0,69	0,72	0,68	0,38	0,62	0,69	0,4
Total	0,31	0,46	0,57	0,62	0,41	0,63	0,64	0,61	0,33	0,52	0,51	0,51

Table A.4: AA4-POS scenario average accuracy results of the supervised, and transfer learning approaches per utilised classifier.

	DT	**LR**	**NB**	**SV**
SUP	0,36	0,41	0,41	0,31
SA	0,28	0,41	0,25	0,3
TCA	0,25	0,29	0,19	0,26
KMM	0,46	0,36	0,39	0,24
NN	0,47	0,35	0,18	0,24

Table A.5: AP1-ENV scenario accuracy results of each transfer learning technique per activity.

	Ankle				Chest				Wrist			
	TCA	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**
Running	0,31	0,87	0,71	0,91	0	0,79	0,84	0,91	0	0,99	0,68	0,89
Sitting	0,46	0,76	0,63	0,73	0,32	0,53	0,79	0,77	0,02	0,81	0,81	0,79
Ascending Stairs	0	0,69	0,6	0,77	0	0,57	0,55	0,57	0,26	0,41	0,69	0,46
Descending Stairs	0	0,72	0,89	0,68	0	0,53	0,67	0,83	0	0,91	0,65	0,94
Walking	0	0,86	0,43	0,83	0,17	0,36	0,69	0,62	0,39	0,75	0,88	0,13
Total	0,37	0,77	0,61	0,76	0,24	0,59	0,69	0,72	0,21	0,67	0,73	0,68

Table A.6: AP1-ENV scenario average accuracy results of the supervised, and transfer learning approaches per utilised classifier.

	Ankle				Chest				Wrist			
	DT	**LR**	**NB**	**SV**	**DT**	**LR**	**NB**	**SV**	**DT**	**LR**	**NB**	**SV**
SUP	0,67	0,62	0,51	0,68	0,64	0,66	0,51	0,66	0,61	0,66	0,57	0,53
SA	0,5	0,6	0,3	0,77	0,6	0,49	0,26	0,59	0,67	0,68	0,26	0,67
TCA	0,27	0,21	0,2	0,37	0,3	0,23	0,2	0,24	0,2	0,15	0,11	0,21
KMM	0,76	0,58	0,47	0,52	0,72	0,6	0,39	0,67	0,68	0,62	0,42	0,58
NN	0,61	0,54	0,2	0,51	0,69	0,59	0,2	0,59	0,73	0,6	0,2	0,57

Table A.7: AP1-ENV scenario results of the supervised, semi-supervised, and transfer learning approaches.

	Ankle			Chest			Wrist		
	SUP	**SEMI**	**ENSEM**	**SUP**	**SEMI**	**ENSEM**	**SUP**	**SEMI**	**ENSEM**
Running	0,94	0,72	0,94	0,84	0,73	0,68	0,76	1	0,54
Sitting	0,69	0,58	0,76	0,69	0,43	0,63	0,61	0,4	0,9
Ascending Stairs	0,58	0,29	0,68	0,5	0,39	0,61	0,66	0,41	0,49
Descending Stairs	0,38	0,33	0,45	0,78	0,46	0,62	0,57	0,92	0,92
Walking	0,48	0	0,57	0,46	0,5	0,38	0,75	0,65	0
Total	0,62	0,48	0,68	0,66	0,48	0,63	0,66	0,52	0,65

Table A.8: PA1-ENV scenario results of the supervised, semi-supervised, and transfer learning approaches.

	Ankle				Chest				Wrist			
	TCA	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**
Running	0,11	0,74	1	1	0,37	0,94	0,83	0,83	0,23	0,9	0,86	0,8
Sitting	0,12	0,71	0	0	0,03	1	0,5	0	0,06	0,09	0	0
Ascending Stairs	0,38	0,96	0,86	0,87	0,14	0,94	0,83	0,83	0,19	0,94	0,72	0,77
Descending Stairs	0,28	0,74	0,78	0,88	0,19	0,75	0,76	0,62	0,11	0,79	0,79	0,68
Walking	0,21	1	0,73	0,71	0,32	0,84	0,76	0,73	0,26	0,61	0,69	0,73
Total	0,15	0,81	0,83	0,86	0,3	0,87	0,77	0,76	0,3	0,61	0,73	0,7

Table A.9: PA1-ENV scenario average accuracy results of the supervised, and transfer learning approaches per utilised classifier.

	Ankle				Chest				Wrist			
	DT	**LR**	**NB**	**SV**	**DT**	**LR**	**NB**	**SV**	**DT**	**LR**	**NB**	**SV**
SUP	0,68	0,84	0,46	0,62	0,52	0,68	0,46	0,44	0,56	0,63	0,23	0,43
SA	0,53	0,53	0,32	0,81	0,79	0,71	0,16	0,87	0,62	0,66	0,3	0,61
TCA	0,23	0,23	0,15	0,24	0,21	0,12	0,3	0,16	0,15	0,05	0,3	0,1
KMM	0,65	0,86	0,53	0,67	0,74	0,76	0,37	0,79	0,7	0,7	0,37	0,43
NN	0,64	0,83	0,06	0,73	0,72	0,77	0,27	0,71	0,71	0,73	0,1	0,38

Table A.10: PA1-ENV scenario results of the supervised, semi-supervised, and transfer learning approach (only ensemble).

	Ankle			**Chest**			**Wrist**		
	SUP	**SEMI**	**ENSEM**	**SUP**	**SEMI**	**ENSEM**	**SUP**	**SEMI**	**ENSEM**
Running	0,71	0,27	0,65	0,92	0,8	0,84	0,75	0,77	0,83
Sitting	0,83	0	1	0,77	0,5	1	0,23	0	0
Ascending Stairs	0,85	0,46	0,9	0,4	0,53	0,68	0,55	0,11	0,89
Descending Stairs	0,9	0,4	1	0,58	0,46	0,88	0,61	0,21	0,72
Walking	1	0	0,86	0,57	0,46	0,84	0,76	0,71	0,77
Total	0,84	0,37	0,86	0,68	0,49	0,85	0,63	0,51	0,8

Table A.11: PP1-POS scenario accuracy results of the transfer learning techniques.

		Source											
		Ankle				**Chest**				**Wrist**			
		TCA	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**	**TCA**	**SA**	**NNeW**	**KMM**
Target	**Ankle**	0,27	0,81	0,81	0,83	0,21	0,7	0,67	0,48	0,24	0,57	0,58	0,65
	Chest	0,2	0,44	0,37	0,25	0,15	0,74	0,71	0,73	0,17	0,45	0,53	0,59
	Wrist	0,19	0,71	0,52	0,49	0,16	0,58	0,61	0,52	0,22	0,6	0,61	0,69

Table A.12: PP1-POS scenario average accuracy results of the supervised, and transfer learning approaches per utilised classifier.

	DT	**LR**	**NB**	**SV**
SUP	0,52	0,63	0,49	0,59
SA	0,31	0,62	0,28	0,58
TCA	0,15	0,2	0,2	0,15
KMM	0,51	0,58	0,43	0,46
NN	0,52	0,6	0,23	0,5

Table A.13: Pairwise comparisons between ensemble, supervised, and semi-supervised approaches using Nemenyi multiple comparison test.

	SUP	**SEMI**
SEMI	0.022	-
ENSEM	0.022	3.3e-07

References

[1] K Malbut-Shennan and A Young. The physiology of physical performance and training in old age. *Coronary artery disease*, 10(1):37—42, 1999.

[2] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

[3] Pramila Rani, Changchun Liu, Nilanjan Sarkar, and Eric Vanman. An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications*, 9(1):58–69, 2006.

[4] Oscar D. Lara and Miguel A. Labrador. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys & Tutorials*, 15(3):1192–1209, 2013.

[5] Addwiteey Chrungoo, S. S. Manimaran, and Balaraman Ravindran. Activity recognition for natural human robot interaction. In Michael Beetz, Benjamin Johnston, and Mary-Anne Williams, editors, *Social Robotics*, pages 84–94, Cham, 2014. Springer International Publishing.

[6] D. Gehrig, P. Krauthausen, L. Rybok, H. Kuehne, U. D. Hanebeck, T. Schultz, and R. Stiefelhagen. Combined intention, activity, and motion recognition for a humanoid household robot. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4819–4825, Sept 2011.

[7] F. Tartanoglu F. Sailhan R. Chibout N. Levy A. Talamona V. Issarny, D. Sacchetti. Developing ambient intelligence systems: a solution based on web services. *Automated Software Engineering*, September 2009.

[8] Statista. Weekly time spent by u.s. users watching video content on smartphones. Available at `https://www.statista.com/statistics/323922/us-weekly-minutes-smartphone-video-age`, July 2017. Accessed: 2017-12.

[9] Ian Anderson, Julie Maitland, Scott Sherwood, Louise Barkhuus, Matthew Chalmers, Malcolm Hall, Barry Brown, and Henk Muller. Shakra: Tracking and sharing daily activity levels with unaugmented mobile phones. *Mobile Networks and Applications*, 12(2):185–199, Jun 2007.

[10] Apple. About face id advanced technology. Available at `https://support.apple.com/pt-pt/HT208108`, December 2017. Accessed: 2017-12.

[11] Muhammad Ehatisham-ul Haq, Muhammad Awais Azam, Jonathan Loo, Kai Shuang, Syed Islam, Usman Naeem, and Yasar Amin. Authentication of Smartphone Users Based on Activity Recognition and Mobile Sensing. *Sensors*, 17(9):2043, 2017.

[12] US Census Bureau. Population by age and sex. Available at `http://www.census.gov/ipc/www/idbagg.html`, 2007. Accessed: 2017-12.

[13] Koplan JP and Fleming DW. Current and future public health challenges. *JAMA*, 284(13):1696–1698, 2000.

[14] Venet Osmani, Sasitharan Balasubramaniam, and Dmitri Botvich. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of Network and Computer Applications*, 31(4):628–655, 2008.

[15] Xi Liu, Lei Liu, Steven J. Simske, and Jerry Liu. Human Daily Activity Recognition for Healthcare Using Wearable and Visual Sensing Data. *Proceedings - 2016 IEEE International Conference on Healthcare Informatics, ICHI 2016*, pages 24–31, 2016.

[16] Rezaul Begg and Rafiul Hassan. *Artificial Neural Networks in Smart Homes*, pages 146–164. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[17] Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson. *Activity Recognition in the Home Using Simple and Ubiquitous Sensors*, pages 158–175. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[18] Reuter. Commentary: China's 'big brother' tech gives surveillance new reach. Available at `https://www.reuters.com/article/wu-apps-china-commentary/commentary-chinas-big-brother-tech-gives-surveillance-new-reach-idUSKBN1CL29A`, October 2017. Accessed: 2017-12.

[19] Muhammad Shoaib, Stephan Bosch, Ozlem Incel, Hans Scholten, and Paul Havinga. A Survey of Online Activity Recognition Using Mobile Phones. *Sensors*, 15(1):2059–2085, 2015.

[20] M. Cooper, R.W. Dronsuth, A.J. Leitich, J.C.N. Lynk, J.J. Mikulski, J.F. Mitchell, R.A. Richardson, and J.H. Sangster. Radio telephone system, September 16 1975. US Patent 3,906,166.

[21] Robert R Schaller. Moore's law: past, present and future. *IEEE spectrum*, 34(6):52–59, 1997.

[22] Inc ComScore. Us mobile app report of 2014. Available at `http://www.ella.net/pdfs/comScore-US-Mobile-App-Report-2014.pdf`, 2014. Accessed: 2017-12.

[23] Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lotfi A. Zadeh. *Feature Extraction: Foundations and Applications (Studies in Fuzziness and Soft Computing)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[24] Sasan Karamizadeh, Shahidan M. Abdullah, Azizah A. Manaf, Mazdak Zamani, and Alireza Hooman. An Overview of Principal Component Analysis. *Journal of Signal and Information Processing*, 04(03):173–175, 2013.

[25] Xing Su, Hanghang Tong, and Ping Ji. Activity recognition with smartphone sensors. *Tsinghua Science and Technology*, 19(3):235–249, 2014.

[26] Max Kuhn. Recursive feature elimination. Available at `https://topepo.github.io/caret/recursive-feature-elimination.html`, September 2017. Accessed: 2017-12.

[27] The Pennsylvania State University. Chi-square tests. Available at `https://onlinecourses.science.psu.edu/statprogram/node/158`, September 2017. Accessed: 2017-12.

[28] Liang Wang, Tao Gu, Xianping Tao, and Jian Lu. A hierarchical approach to real-time activity recognition in body sensor networks. *Pervasive and Mobile Computing*, 8(1):115 – 130, 2012.

[29] X. Su, H. Tong, and P. Ji. Activity recognition with smartphone sensors. *Tsinghua Science and Technology*, 19(3):235–249, June 2014.

[30] George Okeyo, Liming Chen, Hui Wang, and Roy Sterritt. Dynamic sensor data segmentation for real-time knowledge-driven activity recognition. *Pervasive and Mobile Computing*, 10(PART B):155–172, 2014.

[31] P. Siirtola and J. Röning. Ready-to-use activity recognition for smartphones. In *2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 59–64, April 2013.

[32] Hong Lu, Jun Yang, Zhigang Liu, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, SenSys '10, pages 71–84, New York, NY, USA, 2010. ACM.

[33] S. Thiemjarus, A. Henpraserttae, and S. Marukatat. A study on instance-based learning with reduced training prototypes for device-context-independent activity recognition on a mobile phone. In *2013 IEEE International Conference on Body Sensor Networks*, pages 1–6, May 2013.

[34] Yunus Emre Ustev, Ozlem Durmaz Incel, and Cem Ersoy. User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pages 1427–1436, New York, NY, USA, 2013. ACM.

[35] Sotiris B Kotsiantis, I Zaharakis, and P Pintelas. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160:3–24, 2007.

[36] Xiaojin Zhu. Semi-Supervised Learning Literature Survey Contents. *SciencesNew York*, 10(1530):10, 2008.

[37] J.L.R. Ortiz. *Smartphone-Based Human Activity Recognition*. Springer Theses. Springer International Publishing, 2015.

[38] Hugo Cardoso and João Mendes-Moreira. Improving Human Activity Classification through Online Semi-Supervised Learning. pages 1–12.

[39] University of Washington. Generalization error: what we really want. Available at `https://pt.coursera.org/learn/ml-regression/lecture/CDx5h/generalization-error-what-we-really-want`. Accessed: 2017-12.

[40] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[41] D. Ghosh, D. Giri, R.N. Mohapatra, E. Savas, K. Sakurai, and L.P. Singh. *Mathematics and Computing: 4th International Conference, ICMC 2018, Varanasi, India, January 9-11, 2018, Revised Selected Papers*. Communications in Computer and Information Science. Springer Singapore, 2018.

[42] S Kaghyan and Hakob Sarukhanyan. Activity recognition using k-nearest neighbor algorithm on smartphone with triaxial accelerometer. 1:146–156, 01 2012.

[43] Christopher M Bishop. *Pattern Recognition and Machine Learning*, volume 4. 2006.

[44] Hang Yang, Simon Fong, Guangmin Sun, and Raymond Wong. A very fast decision tree algorithm for real-time data mining of imperfect data streams in a distributed wireless sensor network. *International Journal of Distributed Sensor Networks*, 2012, 2012.

[45] Ozlem Durmaz Incel, Mustafa Kose, and Cem Ersoy. A Review and Taxonomy of Activity Recognition on Mobile Phones. *BioNanoScience*, 3(2):145–171, 2013.

[46] Mridul Khan, Sheikh Ahamed, Miftahur Rahman, and Roger Smith. A feature extraction method for realtime human activity recognition on cell phones. 12 2017.

[47] Z. Yan, V. Subbaraju, D. Chakraborty, A. Misra, and K. Aberer. Energy-efficient continuous activity recognition on mobile phones: An activity-adaptive approach. In *2012 16th International Symposium on Wearable Computers*, pages 17–24, June 2012.

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

[49] Reuters. Your medical record is worth more to hackers than your credit card. Available at `https://uk.reuters.com/article/us-cybersecurity-hospitals/your-medical-record-is-worth-more-to-hackers-than-your-credit-card-idUSKCN0HJ21` 2014. Accessed: 2017-12.

[50] Kuan Lun Huang, Salil S. Kanhere, and Wen Hu. Preserving privacy in participatory sensing systems. *Computer Communications*, 33(11):1266 – 1280, 2010.

[51] Oscar Day and Taghi M. Khoshgoftaar. A survey on heterogeneous transfer learning. *Journal of Big Data*, 4(1), 2017.

[52] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of Frustratingly Easy Domain Adaptation. (Figure 1), 2015.

[53] James Stazicker. The structure of perceptual experience. *The Structure of Perceptual Experience*, 170:1–154, 2015.

[54] Sebastian Thrun and Lorien Pratt. *Learning to Learn*, volume 31. Springer US, Boston, MA, 1998.

[55] Olivier Chapelle, Pannagadatta Shivaswamy, Srinivas Vadrevu, Kilian Weinberger, Ya Zhang, and Belle Tseng. Multi-task learning for boosting with application to web search ranking. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '10*, page 1189, 2010.

[56] Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y Ng. Self-taught Learning : Transfer Learning from Unlabeled Data. *Proceedings of the 24th International Conference on Machine Learning (2007)*, pages 759–766, 2007.

[57] Rajat Raina. *Self-taught Learning*. PhD thesis, Stanford, CA, USA, 2009. AAI3382948.

[58] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach. *Proceedings of the 28th International Conference on Machine Learning*, (1):513–520, 2011.

[59] Bernardino Romera-Paredes and Philip H. S. Torr. An Embarrassingly Simple Approach to Zero-Shot Learning. 37:11–30, 2017.

[60] Zhengming Ding, Ming Shao, and Yun Fu. Low-rank embedded ensemble semantic dictionary for zero-shot learning. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-January:6005–6013, 2017.

[61] Daniel L Silver, Qiang Yang, and Lianghao Li. Lifelong Machine Learning Systems : Beyond Learning Algorithms. *AAAI Spring Symposium Series*, (Solomonoff 1989):49–55, 2013.

[62] Ronald Kemker, Marc McClure, Angelina Abitino, Tyler Hayes, and Christopher Kanan. Measuring Catastrophic Forgetting in Neural Networks. 114(13):3521–3526, 2017.

[63] Ari Seff, Alex Beatson, Daniel Suo, and Han Liu. Continual Learning in Generative Adversarial Nets. pages 1–9, 2017.

[64] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3987–3995, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

[65] Tom M. Mitchell. The need for biases in learning generalizations. Technical report, 1980.

[66] D Silver and Isabelle Guyon. ICML2011 Unsupervised and Transfer Learning Workshop. *Journal of Machine Learning Research*, pages 1–16, 2012.

[67] Chuong B Do and Andrew Y Ng. Transfer learning for text classification. *Advances in Neural Information Processing Systems*, 18(3):299–306, 2006.

[68] Sebastian Ruder. Transfer learning motivations. Available at `http://ruder.io/transfer-learning/index.html#whytransferlearningnow`, March 2017. Accessed: 2017-12.

[69] Michael T Rosenstein, Zvika Marx, Leslie Pack Kaelbling, Computer Science, and Thomas G Dietterich. To Transfer or Not To Transfer. *NIPS Workshop on Inductive Transfer*, pages 1–4, 2005.

[70] M M Hassan Mahmud and Sylvian R Ray. Transfer Learning using Kolmogorov Complexity: Basic Theory and Empirical Evaluations. *Advances in Neural Information Processing Systems 20*, pages 985–992, 2008.

[71] Jose G. Moreno-Torres, Troy Raeder, Rocío Alaiz-Rodríguez, Nitesh V. Chawla, and Francisco Herrera. A unifying view on dataset shift in classification. *Pattern Recognition*, 45(1):521–530, 2012.

[72] Haibo He and Edwardo A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.

[73] R. Bharat Rao, Sriram Krishnan, and Radu Stefan Niculescu. Data mining for improved cardiac care, 2006.

[74] Dwi H. Widyantoro, Thomas R. Ioerger, and John Yen. Tracking changes in user interests with a few relevance judgments. *Proceedings of the twelfth international conference on Information and knowledge management - CIKM '03*, page 548, 2003.

[75] Gregory Ditzler, Manuel Roveri, Cesare Alippi, and Robi Polikar. Learning in Nonstationary Environments: A Survey. *IEEE Computational Intelligence Magazine*, 10(4):12–25, 2015.

[76] Niall Adams. *Dataset Shift in Machine Learning*, volume 173. 2010.

[77] Georgia Mcgaughey, W Patrick Walters, and Brian Goldman. Understanding covariate shift in model performance [version 1 ; referees : 2 approved with reservations] Referee Status :. 5:1–8, 2017.

[78] David A. Cieslak and Nitesh V. Chawla. A framework for monitoring classifiers' performance: When and why failure occurs? *Knowledge and Information Systems*, 18(1):83–108, 2009.

[79] Karl Weiss, Taghi M. Khoshgoftaar, and Ding Ding Wang. *A survey of transfer learning*, volume 3. Springer International Publishing, 2016.

[80] Diane Cook, Kyle D. Feuz, and Narayanan C. Krishnan. Transfer learning for activity recognition: A survey. *Knowledge and Information Systems*, 36(3):537–556, 2013.

[81] Lilyana Mihalkova and Raymond J Mooney. Transfer learning by mapping with minimal target data. *Proceedings of the AAAI workshop on transfer learning for complex tasks*, (July):31–36, 2008.

[82] Rui Xia, Chengqing Zong, Xuelei Hu, and Erik Cambria. Feature ensemble plus sample selection: Domain adaptation for sentiment classification. *IJCAI International Joint Conference on Artificial Intelligence*, 2015-January:4229–4233, 2015.

[83] Jing Zhang, Wanqing Li, and Philip Ogunbona. Transfer Learning for Cross-Dataset Recognition: A Survey. V(July), 2017.

[84] Wenyuan Dai, Qiang Yang, Gui-Rong Xue, and Yong Yu. Boosting for transfer learning. *Proceedings of the 24th international conference on Machine learning - ICML '07*, pages 193–200, 2007.

[85] Jiaolong Xu. *Incremental Domain Adaptation of Deformable Part-based Models*, volume 1. 2014.

[86] Minmin Chen, Kilian Q Weinberger, and John C Blitzer. Co-training for domain adaptation. *International Conference on Neural Information Processing Systems*, pages 2456–2464, 2011.

[87] Michael M Bronstein and David Hutchison. *Computer Vision - ECCV 2014 Workshops*, volume 8927. 2015.

[88] Sinno Jialin Pan, Ivor W. Tsang, James T. Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2011.

[89] Boqing Gong and Kristen Grauman. Reshaping Visual Datasets for Domain Adaptation. *Neural Information Processing Systems*, (December):1–9, 2013.

[90] Tzu Ming Harry Hsu, Wei Yu Chen, Cheng An Hou, Yao Hung Hubert Tsai, Yi Ren Yeh, and Yu Chiang Frank Wang. Unsupervised domain adaptation with imbalanced cross-domain data. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 International Conference on Computer Vision, ICCV 2015:4121–4129, 2015.

[91] Wen Li, Zheng Xu, Dong Xu, Dengxin Dai, and Luc Van Gool. Domain generalization and adaptation using low rank exemplar svms. *IEEE transactions on pattern analysis and machine intelligence*, 40(5):1114–1127, 2017.

[92] Muhammad Ghifary, David Balduzzi, W. Bastiaan Kleijn, and Mengjie Zhang. Scatter component analysis: A unified framework for domain adaptation and domain generalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(7):1414–1430, 2017.

[93] Joey Tianyi Zhou, Ivor W Tsang, Sinno Jialin Pan, and Mingkui Tan. Heterogeneous Domain Adaptation for Multiple Classes. *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, 33:1095–1103, 2014.

[94] Sanatan Sukhija, Narayanan C. Krishnan, and Gurkanwal Singh. Supervised heterogeneous domain adaptation via random forests. *IJCAI International Joint Conference on Artificial Intelligence*, 2016-January:2039–2045, 2016.

[95] Yao-Hung Hubert Tsai, Yi-Ren Yeh, and Yu-Chiang Frank Wang. Learning Cross-Domain Landmarks for Heterogeneous Domain Adaptation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5081–5090, 2016.

[96] Min Xiao and Yuhong Guo. Feature space independent semi-supervised domain adaptation via kernel matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):54–66, 2015.

[97] Saurabh Gupta, Judy Hoffman, and Jitendra Malik. Cross Modal Distillation for Supervision Transfer. 2015.

[98] Liu Yang, Liping Jing, Jian Yu, and Michael K. Ng. Learning Transferred Weights from Co-Occurrence Data for Heterogeneous Transfer Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 27(11):2187–2200, 2016.

[99] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006.

[100] Sachin Ravi and Hugo Larochelle. Optimization As a Model for Few-Shot Learning. *Iclr*, pages 1–11, 2017.

[101] Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical Networks for Few-shot Learning. 2017.

[102] Ziming Zhang and Venkatesh Saligrama. Zero-shot learning via semantic similarity embedding. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 International Conference on Computer Vision, ICCV 2015:4166–4174, 2015.

[103] Xin Li, Yuhong Guo, and Dale Schuurmans. Semi-supervised zero-shot classification with label representation learning. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 International Conference on Computer Vision, ICCV 2015:4211–4219, 2015.

[104] Mark Palatucci, Geoffrey E Hinton, Dean Pomerleau, and Tom M Mitchell. Zero-Shot Learning with Semantic Output Codes. *Neural Information Processing Systems*, pages 1–9, 2009.

[105] Scott Reed, Zeynep Akata, Bernt Schiele, and Honglak Lee. Learning Deep Representations of Fine-grained Visual Descriptions. 2016.

[106] Chuang Gan, Tianbao Yang, and Boqing Gong. Learning Attributes Equals Multi-Source Domain Generalization. 2016.

[107] Maxime Bucher, Stéphane Herbin, and Frédéric Jurie. Improving semantic embedding consistency by metric learning for zero-shot classiffication. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9909 LNCS:730–746, 2016.

[108] Wataru Kumagai. Learning Bound for Parameter Transfer Learning. *Nips*, 2016.

[109] Jason Kuen, Kian Ming Lim, and Chin Poo Lee. Self-taught learning of a deep invariant representation for visual tracking via temporal slowness principle. *Pattern Recognition*, 48(10):2964–2982, 2015.

[110] Chengcheng Jia, Yu Kong, Zhengming Ding, and Yun Raymond Fu. Latent Tensor Transfer Learning for RGB-D Action Recognition. *Proceedings of the ACM International Conference on Multimedia - MM '14*, pages 87–96, 2014.

[111] Qiang Zhu, Yin and Chen, Yuqiang and Lu, Zhongqi and Pan, Sinno Jialin and Xue, Gui-Rong and Yu, Yong and Yang. Heterogeneous Transfer Learning for Image Classification. *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pages 1304–1309, 2011.

[112] Alberto Calatroni, Daniel Roggen, and Gerhard Tröster. Automatic transfer of activity recognition capabilities between body-worn motion sensors: Training newcomers to recognize locomotion. *Eighth International Conference on Networked Sensing Systems (INSS'11)*, 2011.

[113] Marc Kurz, Gerold Hölzl, Alois Ferscha, Alberto Calatroni, Daniel Roggen, and Gerhard Tröster. Real-Time Transfer and Evaluation of Activity Recognition Capabilities in an Opportunistic System. *Third International Conference on Adaptive and Self-Adaptive Systems and Applications (ADAPTIVE2011)*, (c):73–78, 2011.

[114] Seyed Ali Rokni and Hassan Ghasemzadeh. Autonomous Training of Activity Recognition Algorithms in Mobile Sensors: A Transfer Learning Approach in Context-Invariant Views. *IEEE Transactions on Mobile Computing*, pages 1–1, 2018.

[115] Ulf Blanke and Bernt Schiele. Remember and transfer what you have learned-recognizing composite activities based on activity spotting. *Wearable Computers (ISWC), 2010 International Symposium on*, pages 1–8, 2010.

[116] Md Abdullah Al Hafiz Khan and Nirmalya Roy. TransAct: Transfer learning enabled activity recognition. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2017*, pages 545–550, 2017.

[117] Hailin Chen, Shengping Cui, and Sebastian Li. Application of Transfer Learning Approaches in Multimodal Wearable Human Activity Recognition. 2017.

[118] Tom Diethe, Niall Twomey, and Peter Flach. Active transfer learning for activity recognition. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, (April):27–29, 2016.

[119] Narayanan C. Krishnan. A Computational Framework for Wearable Accelerometer-Based Activity and Gesture Recognition. (December):1–194, 2010.

[120] Daniel Roggen, Kilian Förster, Alberto Calatroni, and Gerhard Tröster. The adARC pattern analysis architecture for adaptive human activity recognition systems. *Journal of Ambient Intelligence and Humanized Computing*, 4(2):169–186, 2013.

[121] Ramin Fallahzadeh. Personalization without User Interruption : Boosting Activity Recognition in New Subjects Using Unlabeled Data Personalization without User Interruption : Boosting Activity Recognition in New Subjects Using Unlabeled Data. (April):293–302, 2017.

[122] Wan-Yu Deng, Qing-Hua Zheng, and Zhong-Min Wang. Cross-person activity recognition using reduced kernel extreme learning machine. *Neural Networks*, 53:1–7, 2014.

[123] Hirotaka Hachiya, Masashi Sugiyama, and Naonori Ueda. Importance-weighted least-squares probabilistic classifier for covariate shift adaptation with application to human activity recognition. *Neurocomputing*, 80:93–101, 2012.

[124] Zhongtang Zhao, Yiqiang Chen, Junfa Liu, Zhiqi Shen, and Mingjie Liu. Cross-people mobile-phone based activity recognition. *IJCAI International Joint Conference on Artificial Intelligence*, pages 2545–2550, 2011.

[125] Zhongtang ZHAO , Yiqiang CHEN , Junfa LIU , and Mingjie LIU . Cross-mobile ELM based Activity Recognition. *International Journal of Engineering and Industries*, 1(1):30–40, 2010.

[126] J; Winter, D; Patla, A; Frank. Biomechanical walking patterns in the fit and healthy elderly. *Physical Therapy*, 70(6):340–347, 1990.

[127] Kari Laasonen, Mika Raento, Hannu Toivonen, Alois Ferscha, and Friedemann Mattern. Pervasive Computing. *Pervasive Computing*, 3001:287–304, 2004.

[128] Parisa Rashidi and Diane J. Cook. Activity knowledge transfer in smart environments. *Pervasive and Mobile Computing*, 7(3):331 – 343, 2011. Knowledge-Driven Activity Recognition in Intelligent Environments.

[129] T. L M Van Kasteren, G Englebienne, and B. J A Kröse. Recognizing activities in multiple contexts using transfer learning. *AAAI Fall Symposium - Technical Report*, FS-08-02(Schmidt 2002):142–149, 2008.

[130] Anna Margolis. A literature review of domain adaptation with unlabeled data. *Tec. Report*, (March 2008):1–42, 2011.

[131] Jun Yang, Rong Yan, and Alexander G. Hauptmann. Cross-domain video concept detection using adaptive svms. *Proceedings of the 15th international conference on Multimedia - MULTIMEDIA '07*, page 188, 2007.

[132] Wei Jiang, E. Zavesky, Shih-Fu Chang, and A. Loui. Cross-domain learning methods for high-level visual concept classification. In *2008 15th IEEE International Conference on Image Processing*, pages 161–164, Oct 2008.

[133] Ariya Rastrow, Frederick Jelinek, Abhinav Sethy, and Bhuvana Ramabhadran. Unsupervised Model Adaptation using Information-Theoretic Criterion. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, (June):190–197, 2010.

[134] Lixin Duan, Dong Xu, and Ivor Wai Hung Tsang. Domain adaptation from multiple sources: A domain-dependent regularization approach. *IEEE Transactions on Neural Networks and Learning Systems*, 23(3):504–518, 2012.

[135] Steffen Bickel, Michael Brückner, and Tobias Scheffer. Discriminative learning for differing training and test distributions. *Proceedings of the 24th international conference on Machine learning - ICML '07*, pages 81–88, 2007.

[136] M Dudík, Steven J Phillips, Robert E Schapire, Miroslav Dudik, Robert E Schapire, Steven J Phillips, Miroslav Dudık, Robert E Schapire, and Steven J Phillips. Correcting sample selection bias in maximum entropy density estimation. *Advances in neural information processing systems 18*, 17:323–330, 2005.

[137] Kamal Nigam, John Lafferty, and Andrew Mccallum. Using Maximum Entropy for Text Classification. *IJCAI-99 Workshop on Machine Learning for Information Filtering*, pages 61–67, 1999.

[138] Masashi Sugiyama SUGI, Matthias Krauledat MATTHIASKRAULEDAT, and Firstfhgde Klaus-Robert uller KLAUS. Covariate Shift Adaptation by Importance Weighted Cross Validation. *Journal of Machine Learning Research*, 8:985–1005, 2007.

[139] Yuta Tsuboi, Hisashi Kashima, Shohei Hido, Steffen Bickel, and Masashi Sugiyama. Direct Density Ratio Estimation for Large-scale Covariate Shift Adaptation. *Journal of Information Processing*, 17(Cv):138–155, 2009.

[140] Marco Loog. Nearest neighbor-based importance weighting. *IEEE International Workshop on Machine Learning for Signal Processing, MLSP*, 1(4), 2012.

[141] Jianyuan Jiayuan Huang, Alexander J Smola, Arthur Gretton, Karsten J. M Borgwardt, Bernard Scholkopf, Bernhard Schölkopf, and Alexander J Smola. Correcting Sample Selection Bias by Unlabeled Data. *Advances in Neural Information Processing Systems 19*, 19:601–608, 2007.

[142] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2066–2073, June 2012.

[143] Baochen Sun, Jiashi Feng, and Kate Saenko. Correlation alignment for unsupervised domain adaptation. *Advances in Computer Vision and Pattern Recognition*, (9783319583464):153–171, 2017.

[144] Mahsa Baktashmotlagh, Mehrtash T. Harandi, Brian C. Lovell, and Mathieu Salzmann. Unsupervised domain adaptation by domain invariant projection. *Proceedings of the IEEE International Conference on Computer Vision*, pages 769–776, 2013.

[145] Mingsheng Long, Guiguang Ding, Jianmin Wang, Jiaguang Sun, Yuchen Guo, and Philip S. Yu. Transfer Sparse Coding for Robust Image Representation. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 407–414, 2013.

[146] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, and Philip S. Yu. Transfer feature learning with joint distribution adaptation. *Proceedings of the IEEE International Conference on Computer Vision*, pages 2200–2207, 2013.

[147] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I. Jordan. Learning Transferable Features with Deep Adaptation Networks. 37, 2015.

[148] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep Domain Confusion: Maximizing for Domain Invariance. 2014.

[149] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars. Subspace Alignment For Domain Adaptation. 2014.

[150] Eric Bair. Semi-supervised clustering methods. pages 1–28, 2013.

[151] Wenyuan Dai, Gui-Rong Xue, Qiang Yang, and Yong Yu. Co-clustering based classification for out-of-domain documents. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '07*, page 210, 2007.

[152] SJ Pan, JT Kwok, and Qiang Yang. Transfer Learning via Dimensionality Reduction. *AAAI Conference on Artificial Intelligence*, pages 677–682, 2008.

[153] Peng Dai, Shen Shyang Ho, and Frank Rudzicz. Sequential behavior prediction based on hybrid similarity and cross-user activity transfer. *Knowledge-Based Systems*, 77:29–39, 2015.

[154] Junyao Xie, Laibin Zhang, Lixiang Duan, and Jinjiang Wang. On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on Transfer Component Analysis. *2016 IEEE International Conference on Prognostics and Health Management, ICPHM 2016*, 2016.

[155] Fatemeh Mirrashed, Vlad I Morariu, Behjat Siddiquie, Rogerio S Feris, and Larry S Davis. Domain Adaptive Object Detection. pages 323–330, 2012.

[156] Basura Fernando, Tatiana Tommasi, and Tinne Tuytelaars. Joint cross-domain classification and subspace learning for unsupervised adaptation. *Pattern Recognition Letters*, 65:60–66, 2015.

[157] Ayush Mittal, Anant Raj, Vinay P. Namboodiri, and Tinne Tuytelaars. Unsupervised Domain Adaptation in the Wild: Dealing with Asymmetric Label Sets. pages 1–15, 2016.

[158] Xiaokang Xie, Zhiguo Cao, Yang Xiao, Mengyu Zhu, and Hao Lu. Blurred image recognition using domain adaptation. *2015 IEEE International Conference on Image Processing (ICIP)*, (January 2016):532–536, 2015.

[159] K.D. Gupta, R. Pampana, R. Vilalta, E.E.O. Ishida, and R.S. De Souza. Automated supernova Ia classification using adaptive learning techniques. *2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016*, 2017.

[160] Wouter M. Kouw and Marco Loog. On regularization parameter estimation under covariate shift. *Proceedings - International Conference on Pattern Recognition*, pages 426–431, 2017.

[161] Ali Hassan, Robert Damper, and Mahesan Niranjan. On acoustic emotion recognition: Compensating for covariate shift. *IEEE Transactions on Audio, Speech and Language Processing*, 21(7):1458–1468, 2013.

[162] C. Feng, X. Kong, M. Li, Y. Yang, and Y. Guo. Contribution-based feature transfer for jpeg mismatched steganalysis. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 500–504, Sept 2017.

[163] Yingying Zhu, Minjeong Kim, Xiaofeng Zhu, Jin Yan, Daniel Kaufer, and Guorong Wu. Personalized diagnosis for alzheimer's disease. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 205–213. Springer, 2017.

[164] Attila Reiss and Didier Stricker. Creating and benchmarking a new dataset for physical activity monitoring. *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '12*, page 1, 2012.

[165] José Antonio Santoyo-Ramón, Eduardo Casilari, and José Manue Cano-García. Analysis of a smartphone-based architecture with multiple mobility sensors for fall detection with supervised learning. *Sensors (Switzerland)*, 18(4):1–17, 2018.

[166] Aaron Defazio, Francis R. Bach, and Simon Lacoste-Julien. SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives. *CoRR*, abs/1407.0202, 2014.

[167] Vangelis Metsis and et al. Spam filtering with naive bayes – which naive bayes? In *THIRD CONFERENCE ON EMAIL AND ANTI-SPAM (CEAS*, 2006.

[168] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Unsupervised learning. In *The elements of statistical learning*, pages 485–585. Springer, 2009.

[169] sklearn. Linear support vector classification. Available at `http://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html`. Accessed: 2018-06.

[170] Xiaojin Zhu and Zoubin Ghahramani. Learning from labeled and unlabeled data with label propagation. 2002.

[171] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*, pages 321–328. MIT Press, 2004.

[172] Oresti Banos, Juan-Manuel Galvez, Miguel Damas, Hector Pomares, and Ignacio Rojas. Window Size Impact in Human Activity Recognition. *Sensors*, 14(4):6474–6499, 2014.

[173] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[174] Tony Caporale and Barbara McKiernan. High and variable inflation: further evidence on the friedman hypothesis. *Economics Letters*, 54(1):65–68, 1997.

[175] Alan C Elliott and Linda S Hynan. A sas® macro implementation of a multiple comparison post hoc test for a kruskal–wallis analysis. *Computer methods and programs in biomedicine*, 102(1):75–80, 2011.