INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Extraction of Biomedical Indicators from Gait Videos

João Pedro Sobreiro Machado

Master in Telecommunications and Computer Engineering

Supervisor:
Prof. Doctor Luís Eduardo de Pinho Ducla Soares, Associate Professor
ISCTE - Instituto Universitário de Lisboa

Co-Supervisor:
Prof. Doctor Paulo Luís Serras Lobato Correia, Associate Professor
Instituto Superior Técnico, Universidade de Lisboa

October, 2020

Acknowledgements

I want to start by thanking Prof. Luís Ducla Soares and Prof. Paulo Lobato Correia for their consistent support and time dedicated to my guidance. Pedro Albuquerque, with whom I shared part of this project, for all the shared ideas and support. I would like to thank FCCN for helping us record the videos for the dataset and for all the material made available for us. I would also like to thank Instituto de Telecomunicações, with whom I developed this work. Next, I want to thank my parents and my brother for everything they did for me. My girlfriend who has supported me from the beginning. Finally, I would like to thank to all my friends who supported me in this journey and everyone that guided me through the development of this thesis.

Resumo

A marcha tem sido um tema muito investigado nos últimos anos. Através da análise da marcha é possível detetar patologias, o que torna esta análise muito importante para avaliar anómalias e consequentemente, ajudar no diagnóstico e na reabilitação dos pacientes. Existem alguns sistemas para analisar a marcha, mas habitualmente, ou estão sujeitos a uma interpretação subjetiva, ou são sistemas usados em laboratórios especializados com equipamento complexo, o que os torna muito dispendiosos e inacessíveis. No entanto, tem havido um esforço significativo com o objectivo de disponibilizar sistemas mais simples e mais precisos para análise e classificação da marcha. Esta dissertação revê os sistemas de análise e classificação da marcha desenvolvidos recentemente, apresenta uma nova base de dados com videos de 21 sujeitos, a simular 4 patologias diferentes bem como marcha normal, e apresenta também uma aplicação *web* que permite ao utilizador aceder remotamente a um sistema automático de classificação e assim, obter a classificação prevista e mapas de características respectivos de acordo com a entrada dada. O sistema de classificação baseia-se no uso de imagens de representação da marcha como a *Gait Energy Image* (GEI) e *Skeleton Gait Energy Image* (SEI), que são usadas como entrada numa rede neuronal convolucional VGG-19 que é usada para realizar a classificação. Este sistema de classificação corresponde a um sistema baseado na visão. Em suma, a aplicação web desenvolvida tem como finalidade mostrar a utilidade do sistema de classificação, tornando possível o acesso a qualquer pessoa.

Palavras-chave: Análise da marcha, Classificação da marcha, Aprendizagem profunda, Aplicação Web, Servidor Web

Abstract

Gait has been an extensively investigated topic in recent years. Through the analysis of gait it is possible to detect pathologies, which makes this analysis very important to assess anomalies and, consequently, help in the diagnosis and rehabilitation of patients. There are some systems for analyzing gait, but they are usually either systems with subjective evaluations or systems used in specialized laboratories with complex equipment, which makes them very expensive and inaccessible. However, there has been a significant effort of making available simpler and more accurate systems for gait analysis and classification. This dissertation reviews recent gait analysis and classification systems, presents a new database with videos of 21 subjects, simulating 4 different pathologies as well as normal gait, and also presents a web application that allows the user to remotely access an automatic classification system and thus obtain the expected classification and heatmaps for the given input. The classification system is based on the use of gait representation images such as the Gait Energy Image (GEI) and the Skeleton Gait Energy Image (SEI), which are used as input to a VGG-19 Convolutional Neural Network (CNN) that is used to perform classification. This classification system is a vision-based system. To sum up, the developed web application aims to show the usefulness of the classification system, making it possible for anyone to access it.

Keywords: Gait analysis, Gait classification, Deep learning, Web Application, Web Server

Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
CSS	Cascading Style Sheets
CNN	Convolutional Neural Network
DL	Deep Learning
EMG	Electromyography
FCCN	Fundação para a Computação Científica Nacional
GEI	Gait Energy Image
GAN	Generative Adversarial Network
GRF	Ground Reaction Force
HS	Heel Strike
HTML	Hyper Text Markup Language
HTTP	Hypertext Transfer Protocol
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IMU	Inertial Measurement Unit
LDA	Linear Discriminant Analysis
LSTM	Long Short Term Memory
ML	Machine Learning
NN	Neural Network
PAF	Part Affinity Field
PCA	Principal Component Analysis

RNN	Recurrent Neural Network
SEI	Skeleton Energy Image
SE	Surface Electromyography
TO	Toe Off
VGRF	Vertical Ground Reaction Force
WSGI	Web Server Gateway Interface
WBB	Wii Balance Board

Chapter 1

Introduction

This chapter is divided into four different sections. In the first section, the context and motivation are presented, where it is explained why gait analysis is important, how over the years it has evolved and also the difficulty of analysing gait through older methods and how new methods, with the help of technology, can actually make the analysis more precise. In the second section, the objectives of this dissertation are presented and, in the third section, the contributions that were made are described. Finally, the fourth section is the outline section where an overview of the upcoming chapters is made.

1.1 Context and Motivation

Since this dissertation is about gait analysis, it makes sense to explain the meaning of the word "gait". The dictionary describes this word as "a particular way of walking". The way of walking has been studied since it was noticed that several pathologies are reflected in gait. Several researchers and medical doctors have recognized a relationship between human gait and diseases such as strokes, multiple sclerosis, Parkinson's disease, dementia and aging related impairments. Analysing gait by its indicators and monitoring it over time can help specialists with precise and early diagnosis of such disorders.

Neurodegenerative disorders [12, 13], aging [14] and injuries can affect gait and can lead to walking impairments or pathological gait. There are several indicators which can be extracted from people's gait, such as step length and joint angles, which can be analyzed in order to validate if gait is normal or impaired and, in case of impairment, they can be used to identify which type of pathology is affecting the subject's gait. These indicators can be very helpful to identify gait impairments.

As illustrated in Figure 1.1, the most common way of gait assessment consists of a medical expert observing a patient walking in controlled conditions, e.g. in a clinical environment, and sometimes can be followed by a survey, in which the patient is asked to give a subjective evaluation regarding her/his gait. Different indicators can help to identify different pathologies and these indicators can be extracted by a variety of systems in many ways. For example, indicators such as gait speed can help to identify poor balance and significantly slower paces which are symptoms of some neurological diseases, while step length can help to identify multiple sclerosis which often shows several gait alterations such as shorter steps and less gait speed.

FIGURE 1.1: Medical expert observing a patient walking.

Alternatively, there are automated systems, developed by specialized companies, that can vary from sensors in laboratories and sophisticated equipment, which can be very expensive and thus having it in a clinical environment may not be viable, to vision-based systems, which are simpler and easier to use, at least to have them in a clinical environment.

There has been a rise of new systems and technologies trying to solve the problem of obtaining objective measures from gait indicators and classifying gait, especially vision-based systems because of the lower cost and simplicity of calibration. These systems can be:

- **Sensor-based Systems** - which can be divided into systems with wearable sensors and systems with non-wearable sensors. The main goal of systems with wearable sensors is to measure gait kinematics and kinetics through the attached body sensors. The main advantage of wearable sensors is that they can operate in uncontrolled environments due to the fact that these sensors are attached to the body. The main disadvantage also comes from the fact that they are attached to the body, which can be intrusive for the patient and even affect their gait, which obviously can lead to inaccurate measurements. The systems with non-wearable sensors include floor sensors and force sensors. These sensors are limited in their length, which restricts the number of steps that can be measured and analyzed, but in return are not invasive for the person who is using them.

- **Vision-based Systems** - which also can be divided into appearance-based systems and model-based systems. These systems can vary in usage from fully equipped specialized laboratories to the use of a single video camera, which means less cost. These systems, based on a single camera have gained popularity, especially due to the success of automatic analysis systems based on deep learning vision systems. The main disadvantage of non wearable sensors, including vision-based systems, is the fact that they can only be operated in controlled environments. All automated gait classification systems, including those based on vision, require a training step that needs to

have access to a representative dataset of walking cycles from normal and impaired gait, in order to learn the classifier parameters. Such datasets can be really challenging to obtain, due to the fact that having samples of gait from real patients, that suffer from the pathologies to be detected is not easy, for instance due to the privacy and ethical issues involved. This is why all the publicly available datasets were captured from volunteers simulating different types of gait pathologies, and even the acquisition of such datasets can be very time consuming and the obtained gait samples will never fully replicate the impairments that affect a real patient.

The analysis of human gait has become a popular theme of research, as it can help people in many areas like sports, security/identification and medicine. For instance, in medicine, gait analysis can reveal key information about people's neurological diseases such as Parkinson's disease [15–17], multiple sclerosis [18–20], cardiopathies [21] and so on. In this field, the most common way to analyze a person's gait is the evaluation made by medical expert watching the patient walking. This method can lead to some uncertainty because it is based on human observation and it can be subjective. This dissertation work develops a web application that allows anyone with access to this application to evaluate the video of a person's gait. Such a web application could have a fundamental role, diagnosing less urgent patients and leaving only more urgent patients to medical doctors. A practical and current example of the usefulness of such web application is in the case of a pandemic outbreak (e.g., COVID-19), in which overcrowding hospitals and medical centers are to be avoided. However, the usefulness of this web application is not limited to pandemic times. Well after this crisis has passed, this type of web application could still help remote populations and people with less resources to have access to such classification systems, since only an internet connection and a simple camera are needed.

1.2 Objectives

The first step of this work is to perform a review of the state of the art regarding gait analysis and gait pathology classification systems. Then there are three main objectives to achieve with this dissertation work: (i) to acquire a new gait dataset replicating a selection of gait pathologies, notably parkinsonian, hemiplegic, diplegic and neurophatic gait; (ii) to develop a simple and non-intrusive solution for the automatic classification of gait videos as normal or as showing evidence of one of the gait pathologies to be included in the dataset; and (iii) to develop a web application allowing to remotely upload a gait video or a gait image representation of a walking cycle, and automatically compute a classification of the uploaded gait data into one of the classes included in the dataset. The last two objectives of this work are contained in one contribution, which is the gait classification web application.

1.3 Contributions

Aside from the state of the art review in which Sensor-based Systems and Vision-based Systems are discussed, there are two main contributions of this work:

- **Acquisition of a new gait dataset - "Gait-IT"** — The first contribution is the acquisition of a new gait dataset, containing sequences of 21 subjects simulating 4 types of pathologies as well as performing normal gait. Each subject performed 2 severity levels per pathological gait type and 4 sequences per severity level besides normal gait. This acquisition was made in a professional studio.

- **Gait classification web application** — The second contribution of this dissertation is a web application, allowing to remotely execute the gait pathology classification system, which is made available as a web service. This contribution also includes the gait pathology classification system, which is also an objective of this dissertation. The interface of this web application

was developed targeting two different types of users, making available two types of interfaces:

- **Basic** - who can upload a video of a patient walking and have access not only to the output of the gait pathology classification system, but also to a saliency map [22] and a heatmap of class activations [23] displayed for a Gait Energy Image (GEI), which can be helpful in the patient's diagnosis, as the important features extracted by the Convolutional Neural Network (CNN) are represented in these maps. This user can also have the analysis results sent to him/her by email, for storage and possibly for later analysis by a medical expert;
- **Advanced** - which displays more possibilities for the user, being possible to upload videos, GEIs and Skeleton Energy Image (SEI)s. In that way the user can test their own GEI and SEI in the classification system and they can also have access to the feature map corresponding to the chosen layer and channel of the deep learning model used. It is also possible to extract a saliency map and a heatmap of the uploaded gait representation, being possible to understand what features the CNN extracted from the uploaded gait representation.

1.4 Outline

This work is organized as follows. Chapter 1, which corresponds to the current chapter, briefly introduces the context and motivation for the current work, along with a set of objectives and contributions for the dissertation. In Chapter 2, the various systems currently used for gait analysis and classification will be reviewed and analyzed. Chapter 3 presents a new gait dataset called GAIT-IT dataset, including a complete description of how it was captured, what is included and how it is organized. Chapter 4 is divided into two sections: (i) the classification system which can be used to classify the input images between normal and different types of pathological gait; (ii) the proposed web application in which it is possible to upload a gait representation as input and to remotely execute the gait

pathology classification system. In Chapter 5, the experiments and results of the classification system are presented. Finally, Chapter 6 concludes the dissertation with a summary of the achievements of this work and also some suggestions for future work.

Chapter 2

State of the Art on Gait Analysis

This chapter presents an overview of the existing systems that extract biomedical indicators from gait videos, emphasizing the recent solutions using deep learning vision-based systems. Recent progress with the incorporation of new technologies has given rise to devices and techniques which allow an objective evaluation of several gait features, resulting in an efficient measurement and providing medical doctors with a large amount of reliable information about patients' gait. This reduces the error margin caused by subjective techniques. Automatic gait analysis systems can be classified according to two different types: sensor-based systems and vision-based systems, as illustrated in Figure 2.1. The gold standard for clinical evaluation is to use a so-called optoelectronic motion capture system [24], because of the accuracy of the features obtained. A disadvantage of this type of system is that it can only be operated in special laboratories due to the complex setup and the need for calibrations before use.

2.1 Sensor-based Systems

Sensors are devices which measure physical properties such as pressure, magnetism, or a particular motion. These systems use such devices to acquire signals representing human motion. Sensors can be attached to the body of the individual, such as the feet, knees or hips to measure different characteristics of the human

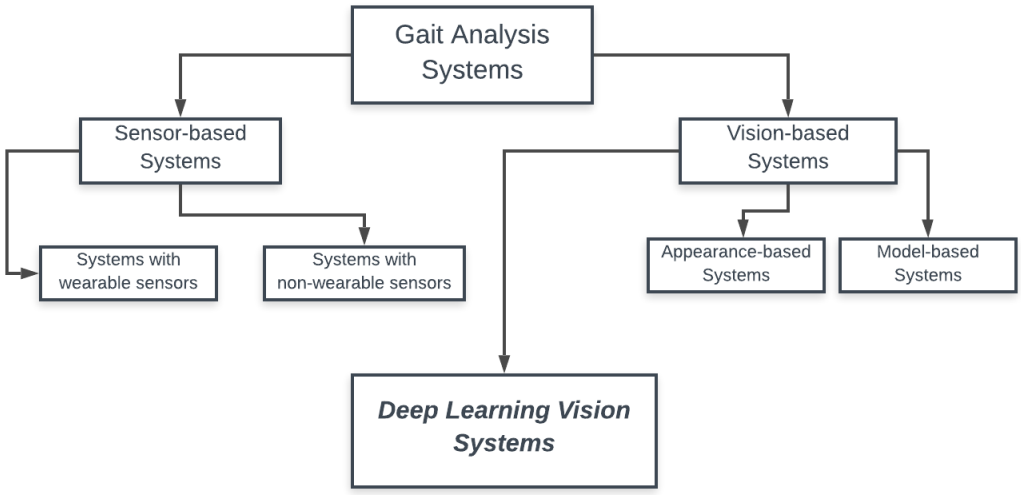


FIGURE 2.1: High-level overview of the different kind of systems in the field.

gait, or can be setup up on the floor. This section offers a brief overview of the different types of sensor-based systems used to acquire signals for gait analysis.

2.1.1 Systems with wearable sensors

These type of sensors can capture signals that can be interpreted by the systems to estimate certain indicators/features relating to: i) kinematics, focused on the movements of the lower limbs and joints; ii) kinetics, focused on forces involved in producing movement; iii) Electromyography (EMG) which makes it possible to obtain the resulting electric activity from muscle contractions during locomotion.

There are various advantages in using wearable sensor-based systems. They can acquire gait over long periods of time and can also operate in uncontrolled environments, so they are not restricted to special laboratories. One of the disadvantage of these systems is that setting up an individual with such systems requires clinical professionals, due to the fact that all sensors placed in the body (such as the feet, knees or hips) must be precise. This is described in several reviews [25, 26]. Additionally, they are also susceptible to noise and interference from external factors, caused by the possible uncontrolled environment.

Force Sensors

These sensors measure the forces involved in the production of movement (gait kinetics). Force sensitive resistors [27] measure the Ground Reaction Force (GRF) under the foot and can return a voltage proportional to the measured pressure. These sensors are typically attached to the insole of shoes [28, 29]. The weight placed on the sensors is inversely proportional to its resistance, and that provides the change in potential. A novel calibration method of insole sensors to estimate Vertical Ground Reaction Force (VGRF) during walking using theWii Balance Board (WBB) was presented in [1], as illustrated in Figure 2.2.

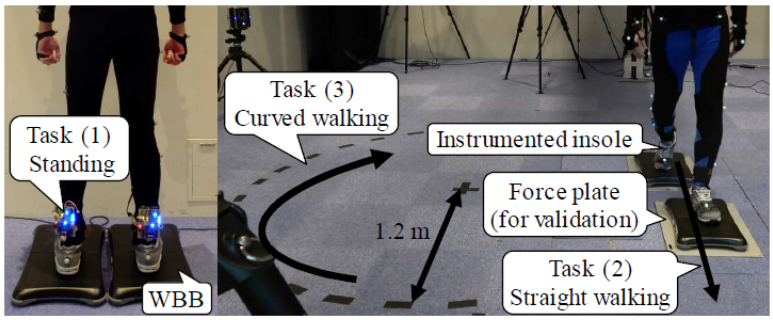


FIGURE 2.2: Task executions and walkway with measuring devices [1].

Inertial Sensors

Using a combination of gyroscopes and accelerometers and sometimes magnetometers, these sensors are electronic devices that can measure an object's velocity, acceleration, orientation, and gravitational forces. These systems are called Inertial Measurement Unit (IMU) systems and are one of the most widely used systems in gait analysis. In [30], a system with inertial sensors to quantify gait symmetry and gait normality was developed. A gyroscope is a device used for measuring orientation and angular velocity. Accelerometers calculate the acceleration by measuring the net force acting on them. In [31], an algorithm was presented to estimate gait features, from on-body mounted inertial sensors, showing a difference in step length below 5% when considering median values of the camera-based gold standard system. The miniaturization of these sensors makes it possible to integrate them on instrumented insoles for gait analysis, such as the Veristride insoles developed by Bamberg et al., which additionally include especially designed pressure sensors for distributed plantar force sensing, Bluetooth communication modules and an inductive charging system, as illustrated in Figure 2.3.

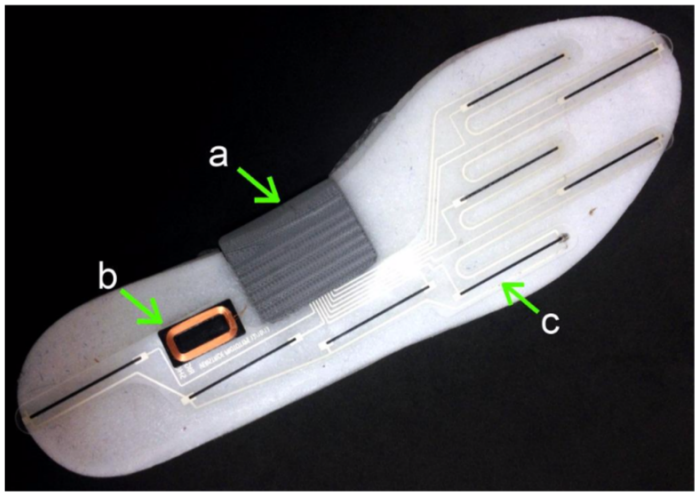


FIGURE 2.3: Instrumented insole: (a) inertial sensor, Bluetooth, microcontroller and battery module; (b) coil for inductive recharging; and (c) pressure sensors. Reproduced with permission from Stacy Morris Bamberg (Veristride, Salt Lake City, UT, USA). [2]

Electromyography (EMG)

The EMG is an electrical manifestation of contracting muscles. EMG is the process of recording electrical signals. The EMG signal can be obtained from the subject by either measuring invasively with wire or needle electrodes, or non-invasively with surface electrodes, as illustrated in Figure 2.4. In [32], the application of Surface Electromyography (SE) is useful for non-invasive assessment of relevant pathophysiological mechanisms potentially hindering the gait function. In a study performed by Wentink et al. [33], when the prosthetic leg is leading, it was determined that EMG signals measured at the prosthetic leg can be used for prediction of gait initiation. Compared to inertial sensors, EMG can predict initial movement up to 138ms in advance.

FIGURE 2.4: Brainquiry Wireless EMG/EEG/ECG system [2]

2.1.2 Systems with non-wearable sensors

These systems can extract features from people's gait without attaching any sensor to an individual's body. One of the advantages of these systems is that they are not invasive. This subsection describes a category of non-wearable sensors, called floor sensors, and the other category, called vision-based systems, are described in section 2.2.

Floor sensors

These sensors can be pressure mats [34], which are able to quantify the pressure patterns under the feet, or force platforms [35], that can quantify, along with the centre of pressure, the horizontal and shear components of the forces applied.

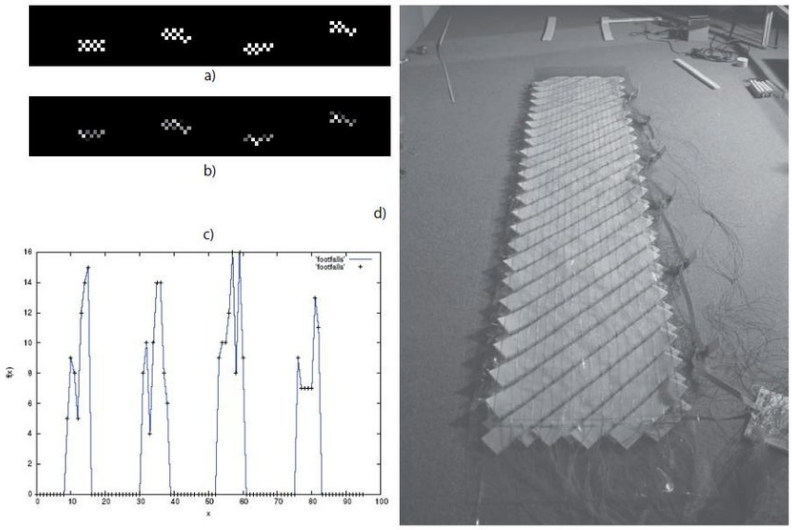


FIGURE 2.5: Floor sensor for gait analysis. (a) Steps recognized; (b) time elapsed in each position; (c) profiles for heel and toe impact; and finally (d) image of the prototype sensor mat on the floor. [2]

As illustrated in Figure 2.5, these systems are setup on the floor to capture information about the gait of individuals walking over it. Force plates are the gold standard for measuring GRF due to the high accuracy measurements they make possible, but they are usually expensive for clinical sites and they are also limited by the available space since they are placed on the ground.

2.2 Vision-based Systems

These systems can be divided into: i) model-based systems, which try to model the individuals body (by segmenting the different body parts into independent shapes, or in the form of a human skeleton model) to perform pathological classification; and ii) appearance-based systems, that rely on the spatiotemporal information obtained from motion patterns of the individuals.

2.2.1 Model-based Systems

Model-based systems use depth-sensing cameras or multiple calibrated 2D cameras. The duration of gait cycle, stance phase, swing phase [36], step length, step width and other indicators [37] are the features that these systems try to accurately estimate, using the captured 3D skeletal model. In [38], it was performed classification of gait as being either normal or impaired, capturing a 3D position of the skeletal joint during a gait cycle (2 steps) and a classification accuracy of 98% was achieved. Features such as the stride length, stride time and stride velocity [39] are used in multiple 2D calibrated cameras to perform classification of gait pathological types.

2.2.2 Appearance-based Systems

Appearance-based systems produce gait representations which do not contain any prior knowledge of human body. Typically, the original images are not suitable for analysis, and background subtraction is usually applied on the acquired image sequences, in order to isolate the observed subject and produce a sequence of binary silhouette shapes. These systems use a single 2D camera to perform classification of

gait across different gait related pathologies based on spatio-temporal information extracted. The features acquired may vary from biomechanical features, such as step length, leg angles and gait cycle time [3], to biometric gait representations, such as GEI [40] and SEI presented in [6].

In [3], after pre-processing the images to obtain a binary silhouette, appearance-based techniques are applied to detect Heel Strike (HS) and Toe Off (TO) events, as shown in Figure 2.6.

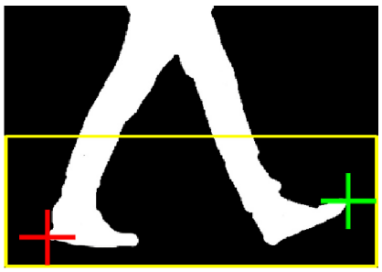


FIGURE 2.6: External heel and toe estimation [3].

These events can be detected by first defining a bounding box around the lower part of the silhouette, in which the foot length is the with of the bounding box when its value is maximum during the gait sequence. After this, external heel and toe of each frame are calculated as the minimum and maximum silhouette pixels in the horizontal axis (x). Next, using these points and the foot length, a bounding box for each foot is obtained and the internal heel and toe are estimated as the minimum and maximum values of x in each foot bounding box.

2.3 Deep Learning Vision Systems

The use of deep learning vision systems has been frequent in the field of gait recognition and gait pathological analysis. These methods, after processing raw input data (videos or images) and turning them into a pre-processed version of the input data (described further in this section), can execute image feature extraction and classification. This section starts by defining briefly what is deep learning and neural networks in a more general point of view. It is almost impossible to define deep learning without first defining Artificial Intelligence (AI) and Machine

Learning (ML). Usually AI is defined as the effort to automate intellectual tasks normally performed by humans [4]. ML is a sub-field of AI, as illustrated in Figure 2.7.

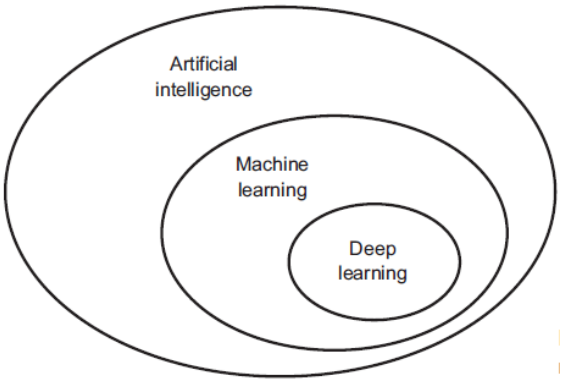


FIGURE 2.7: High-level structure of artificial intelligence, machine learning, and deep learning in [4]

.

ML can be defined, in fact, as an application of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning algorithms are often categorized as:

- **Supervised Machine Learning algorithms** - which can apply what has been learned in the past to new data using labeled examples to predict future events.
- **Unsupervised Machine Learning algorithms** - are used when the information used to train them is neither labeled nor classified.
- **Reinforcement Machine Learning algorithms** - is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning.

As illustrated in Figure 2.7, Deep Learning (DL) is a subfield of ML and uses algorithms inspired by the structure and function of the brain called Artificial Neural Network (ANN). ANN are composed of layers of nodes, which are designed to behave similarly to a neuron in the brain. The first layer of a Neural

Network (NN) is called the input layer, followed by hidden layers and finally the output layer, as shown in Figure 2.8.

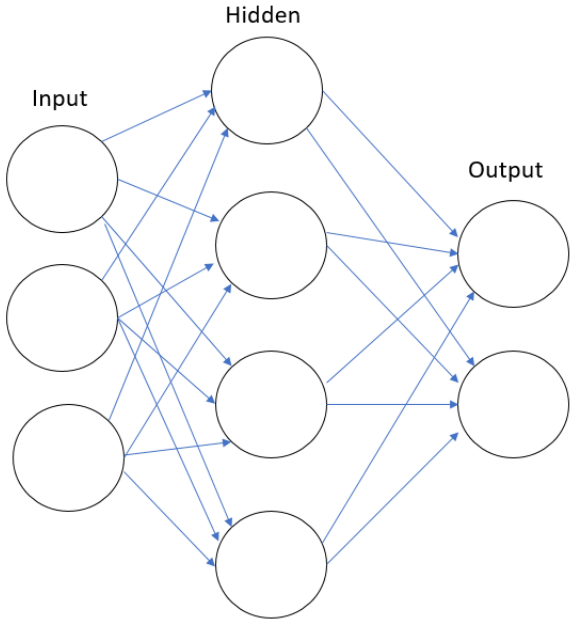


FIGURE 2.8: Artificial Neural Network architecture example showing the different layers.

2.3.1 Image Pre-processing

There are a lot of different forms in which an image or a video can be transformed into input data. Since raw data is rarely used directly as input data in this context, this is a fundamental stage. In this subsection, an overview about pre-processing techniques will be made, such as background subtraction, and different types of image representations, such as Gait Energy Images (GEIs), Skeleton Gait Energy Images (SEI), pose heatmaps and optical flow images.

Background Subtraction

Usually, background subtraction is the first step of Deep Learning Vision Systems when classifying gait. The main goal of such technique is to distinguish the static background from the detected foreground objects, storing a mask of the pixels belonging to the foreground as a binary image, representing the background

pixel values in black and the others in white (foreground). This technique is widely used for surveillance purposes. Mixture of Gaussians model is the most common approach and it was developed by Stauffer and Grimson [41]. Bouwmans et al. [42] made a survey of the numerous improvements of the original MOG.

Deep Learning Vision Systems, when used to performed gait classification using silhouettes, depend on the quality of the used silhouettes, which makes the background subtraction step of the utmost importance, due to the fact that it can have a significant influence in the quality of a silhouette.

Optical Flow Computation

The definition of optical flow is the motion of objects between consecutive frames of a sequence, caused by the relative movement between the object and camera. Optical Flow images can be an input for CNN and in [43], it was shown that it was advantageous, rather than using single frames only for the videos recognition taks.

Gait Energy Image (GEI) Computation

A very popular biometric gait representation when classifying different gait pathology types is GEI [44]. Such representation contains the dynamic gait information about a full gait cycle, compressed into one image. After obtaining the binary images corresponding to a full gait cycle, a GEI is computed as in 2.1.

$$GEI(x, y) = \frac{1}{N} \sum_{i=1}^{N} B_i(x, y) \tag{2.1}$$

as proposed by Han and Bhanu [44], where N represents the number of frames in one (or multiple) gait cycle(s) and $B_i(x, y)$ is a binary silhouette image, with x and y pixel coordinates. The main advantage of this representation is the fact that it represents relevant information such as body posture, amplitude of movement and gait symmetry from a full gait cycle, in one image.

FIGURE 2.9: Gait sequence through a series of key silhouettes, and the resulting Gait Energy Image (GEI) [5]

This gait representation is commonly used in deep learning systems, especially in the ones using CNNs.

Skeleton Gait Energy Image (SEI) Computation

This representation was proposed in [6], having achieved better results than GEI representation with the VGG-19 architecture and the GAIT-IST dataset to train. The first step to compute this representation is to obtain the skeleton joints from the videos of people's gait. To obtain this information, the OpenPose system [11] is used, as exlplained in more detail in Chapter 3. Such algorithm can take the captured videos as input and return, from each frame of the video, 2D coordinates from different parts of the body. After having all these coordinates for all the frames, the skeleton is drawn using the OpenCv library [45].

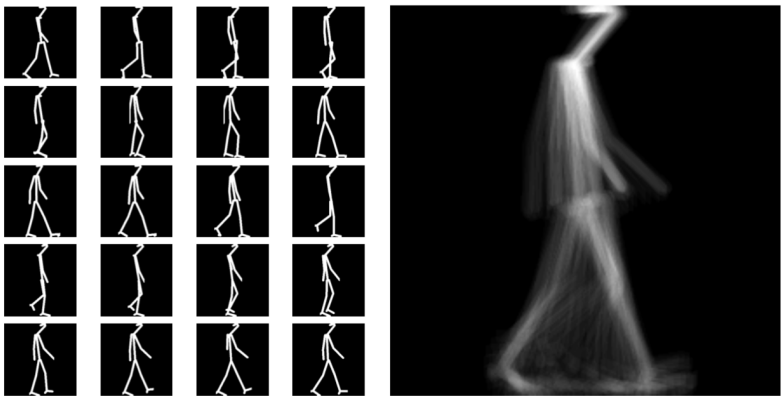


FIGURE 2.10: Down-sampled set of skeleton frames corresponding to one complete gait cycle (left) and respective SEI (right)[6].

The last step of computing this representation is similar to GEI. After having all gait cycles, a SEI is the mean image of all binary skeleton images corresponding to a full gait cycle, as illustrated in the right side of Figure 2.10.

2.3.2 Convolutional Neural Networks

CNNs are one of the most used deep learning architectures. They have proved to be very accurate in areas such as image recognition and classification. CNNs have gained popularity as a good image recognition system in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [46]. ImageNet is a large visual dataset and its purpose is to help develop visual object recognition software research. A CNN is a class of deep neural networks that can successfully capture the spatial dependencies in an image through the application of relevant filters. These filters capture more and more specific features, as the layers progress, so that more complex structures in the image are detected. The performance of silhouette-based gait recognition systems has increased since the use of deep CNNs, such as VGG-16 in [47]. In the medicine field, CNNs have also shown improvement in detecting Alzheimer's disease as seen in [48], with transfer learning. Due to the fact that there are not many datasets for gait pathology classification publicly available and the ones available are small datasets, training data can be insufficient for training. CNNs need a significant number of images for the model to be trained effectively. This recurring problem is expected to result in problems such as overfitting, but it can be resolved by:

- **Data Augmentation** - is one of the techniques that can be used to avoid insufficient data when training the model. Images are the input data for CNNs, so image augmentation is performed by flipping the image horizontally or vertically, rotating the image by a specified degree, shifting one part of the image like a parallelogram, zooming in or zooming out and changing brightness or contrast. Such a technique can be applied as a pre-processing step before training the model or can be applied in real time. In [7], image augmentation was performed over 160 GEIs on the training dataset by using

small shifts, shear, zoom and horizontal flipping, generating a total of 480 GEIs.

- **Transfer Learning** - is a machine learning technique where a model developed to address one issue is reused as a starting point to address another related issue. This is a popular approach in deep learning where pre-trained models are used as the starting point on computer vision tasks. In [7] proposed system, it is used a VGG-19 model pre-trained in ImageNet [46] and to fine-tune the model parameters, part of the network is retrained with gait GEIs from the available dataset.

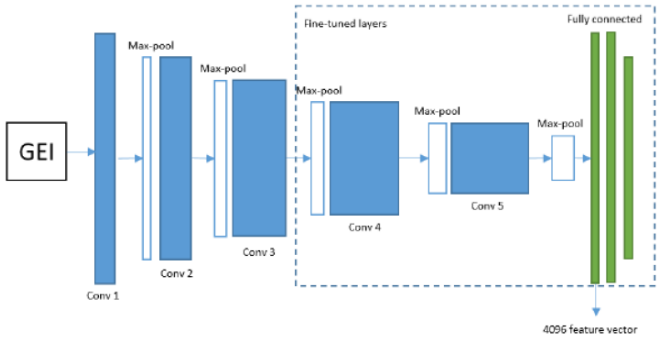


FIGURE 2.11: Feature extraction (VGG-19) architecture [7]

Many authors, like Verlekar et al. [7], used the CNN for feature extraction, fine-tuning a pre-trained VGG-19 network to extract features from GEI images, as illustrated in Figure 2.11. The first fully connected layer is taken as a feature vector and Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are used for classifcation between normal gait and four types of gait abnormality, achieving an accuracy of 95% using the DAI Gait Dataset 2 [40]. In [6], using the same feature extraction technique and LDA to classify the different pathologies instead of using the last layer of the CNN to classify showed several improvements. The overall validation accuracy improved from 43.3% to 76.7% performing a cross-dataset scenario in which GAIT-IST [6] was used to train the model and DAI Gait Dataset 2 [40] to test it.

2.3.3 3D Convolutional Neural Networks

3D Convolutional Neural Networks extend the convolutional filters to a third dimension, which corresponds to time, in that way they can take in consideration spatio-temporal information. In [8], one of the earliest implementations of such architectures was made, for action recognition purposes. As illustrated in Figure 2.12, the input is composed of seven frames, which represent the size of time dimension of the convolutional filters.

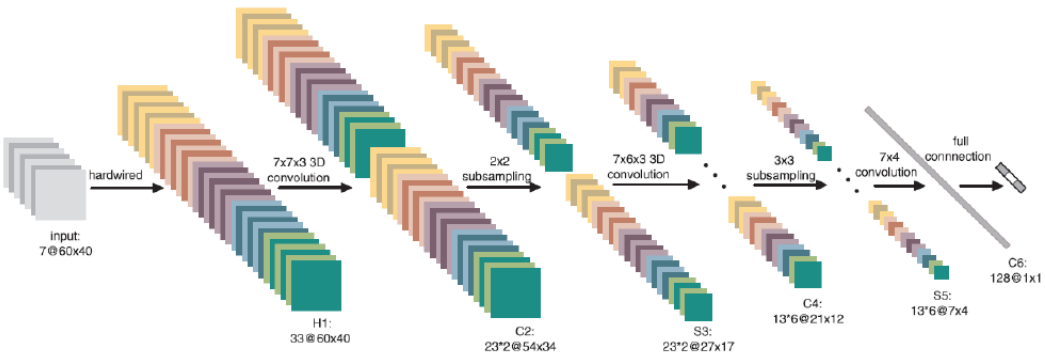


FIGURE 2.12: 3D CNN architecture for human action recognition [8].

In [49], the first large scale application of 3D convolutional networks was done with the objective of classifying sports. In the work of Simonyan and Zisserman

[50], two CNNs were used, one operating on individual RGB images and the other one operating on optical flow. By combining their softmax activations using an SVM good results were obtained. In [51], it was also used a 3D CNN, but with a deeper structure fully exploiting spatio-temporal features for video classification.

2.3.4 Recurrent Neural Networks (RNN)

The main difference between a Recurrent Neural Network (RNN) and the basic feed forward neural networks is that RNNs can learn from prior inputs while generating outputs. These networks are influenced not just by weights applied on inputs, but also by a "hidden" state vector representing the context based on prior input. Long Short Term Memory (LSTM) recurrent networks have gained popularity due to the fact that LSTM operations allow the LSTM to keep or forget information, so they have the capability of modeling both short and long term dependencies. This special RNN structure has proven to be powerful in several previous studies [52–55].

Usually these networks are used for forecasting problems, such as weather forecasting. In [56], a ConvLSTM network approach was presented in which the idea of FC-LSTM was extended. Most recently, this approach has been used for human gait recogition [9], where a variation of Gait Energy Images, i.e. frame-by-frame GEI (ff-GEI) was presented, which aims at expanding available and trainable gait data. This ConvLSTM network consists of three convolutional layers, three pooling layers, one fully connected (FC) layer, three LSTM layers, and one soft-max layer as illustrated in Figure 2.13. The ff-GEIs are fed into three two-tuple sets (convolutional layer and pooling layer). A CNN is used for feature extraction in each image frame and then LSTM is used to classify each frame, considering temporal information.

2.3.5 Generative Adversarial Networks

Generative Adversarial Network (GAN) are convolution based networks that rely on the generation of new data. Usually, in gait analysis, the generation of new data consists in changing the input view angle or eliminating the impact of

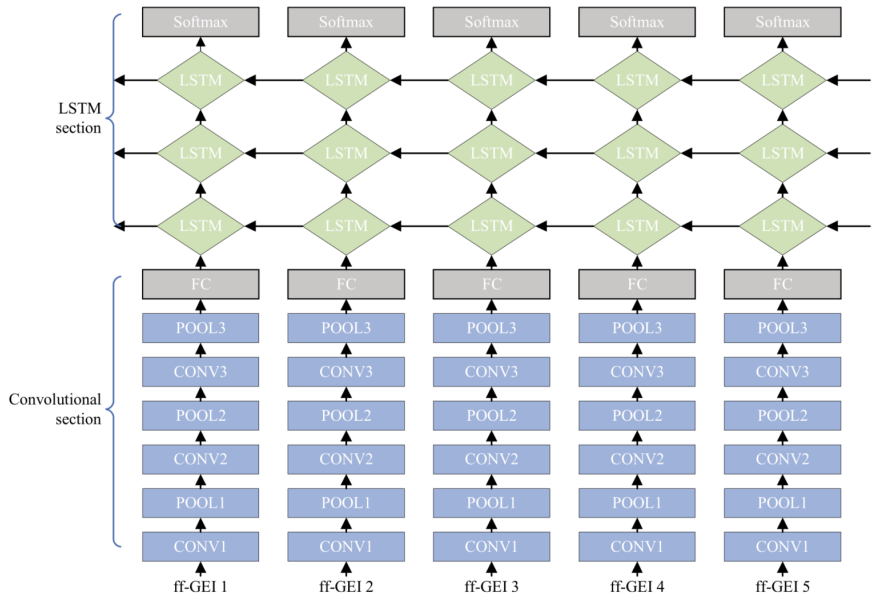


FIGURE 2.13: The architecture of the proposed gait recognition network based on Conv-LSTM [9] .

clothes and objects being carried by the subject, due to the fact that the shape of the silhouette can change.

A GAN has a discriminator network, which takes as input either generated data or real data and classifies it as being either real or generated, and also has a generative network which is made of an encoder and decoder that encodes the input data and then generates new data based on the distribution of training data. In [57], a method named GaitGAN was proposed. This approach differs from the traditional GAN, which has only one discriminator, in that GaitGAN contains two discriminators. One is a fake/real discriminator which can make the generated gait images to be realistic. Another one is an identification discriminator which ensures that the the generated gait images contain human identification information. In [58], a Two-Stream Generative Adversarial Network (TS-GAN) method is proposed for cross-view gait recognition purposes. For any view of gait representations, GAN can restore it to the corresponding standard view, to learn view invariant gait features, having the global-stream that can learn global contexts and the part-stream that can learn local details. The two streams were combined to learn final identities and the proposed method was evaluated on two

widely used gait datasets: CASIA-B and OU-ISIR.

Chapter 3

Gait-IT Dataset Acquisition

Due to the difficulty of obtaining gait sequences from real patients and also due to the privacy and ethical issues involved, there are currently not many publicly available gait datasets dedicated to the study of pathological gait. For this reason, it was decided to create a new dataset in the context of this dissertation, called the GAIT-IT dataset. The acquisition of this dataset was a collaboration with Pedro Albuquerque, a master's degree student of Electrical and Computer Engineering (MEEC) at *Instituto Superior Técnico - Universidade de Lisboa*, and *Fundação para a Computação Científica Nacional (FCCN)* which provided the recording studio and the necessary resources to make the dataset acquisition. This collaboration made it possible to have more volunteers for the experiment than would otherwise be available and the shared knowledge also allowed to better define the acquisition conditions (and, accordingly, instruct the subjects how to proceed).

This chapter first overviews the most important pathological gait datasets publicly available. This is followed by a description of each pathological gait that will be considered for the acquisition of this dataset, as well as a full description of how the dataset acquisition was made. Finally, the chapter ends with a description of the image pre-processing operations, for the reader to understand how the various components available in the dataset were obtained.

3.1 Existing Gait Datasets

The existing gait datasets were created with two different objectives, notably: i) for gait recognition purposes; or ii) for pathological gait analysis. For gait recognition purposes, subjects are typically required to walk normally, possibly at different speeds, with different types of shoes, with different clothing or carrying different items. Currently, there is a significant number of such datasets publicly available. For this dissertation, however, the focus is the study of pathological gait and, therefore, the following overview focuses on publicly available datasets for this purpose, which are summarized in Table 3.1 and are much fewer than those available for recognition purposes. The four pathological gait datasets included in Table 3.1 were all captured from a canonical viewpoint and recorded in controlled environments.

The first dataset, the DAI gait dataset [1] [59], contains binary silhouettes of 5 walking individuals, corresponding to a total of 30 gait sequences. It has 15 sequences considered normal gait, and another 15 sequences of random abnormal gait simulations. The individuals are captured walking over a distance of 3 m using the RGB camera of a Kinect sensor and also using a smartphone.

The second dataset, the DAI2 gait dataset [60], was also created considering 5 walking individuals, but contains a total of 75 gait sequences. Every person simulates four pathologies (Parkinson's, diplegia, hemiplegia and neuropathy), as well as a normal walking gait sequence. Each condition was recorded 3 times, while walking along a distance of 8 m.

[1] http://hdl.handle.net/10045/70567

Datasets	*Year*	*Individuals*	*Sequences*	*GaitType*	*Total*
DAI [59]	2016	5	3	2	30
DAI2 [60]	2017	5	3	5	75
INIT [5]	2018	10	2	8 (*limitations*)	80
GAIT−IST [6]	2019	10	4	5	360

TABLE 3.1: The most notable pathological gait datasets publicly available

The third dataset, the INIT gait dataset [2] [5], contains binary silhouettes of ten individuals (nine males, one female), consisting of a total of 160 sequences. Every subject is recorded 2 different times in a studio, at 30 fps, capturing multiple gait cycles and simulating seven different gait impairments (in addition to a normal gait sequence): i) right arm motionless; ii) half motion of the right arm; iii) left arm motionless; iv) half motion of the left arm; v) full body impairments; vi) half motion of the right leg; and vii) half motion of the left leg.

The fourth and largest dataset, the GAIT-IST gait dataset [3][6], simulated by 10 walking individuals, contains a total of 360 gait sequences. The dataset includes, for each gait type, sequences with 2 severity levels, 2 directions of walking and 2 repetitions per participant, except for the normal gait, due to the fact that it does not have different severity representations. The four pathological gait types are the same considered in DAI2. The video capture was done using a cellphone camera with a resolution of 1280×720 pixels, supported on a tripod at a height of about 1.5 m and at a distance of about 4 m from the target.

3.2 Proposed Gait Dataset

Due to the very limited number of publicly available gait datasets dedicated to the study of pathological gait, it was decided to acquire a new dataset for this purpose called GAIT-IT dataset. For this dataset, the abnormal types of gait chosen were the following: diplegic, hemiplegic, neuropathic and parkinsonian gaits. This choice was made based on the fact that there were already two datasets (i.e., DAI2 and Gait-IST) simulating these abnormal types of gait, making it possible to perform cross-dataset tests. These abnormal types of gait are also defined by a set of symptoms and movements that can be easily simulated. All volunteers were instructed to simulate two severity levels for each of the following pathological gait types, according to the given instructions:

[2] https://www.vision.uji.es/gaitDB/
[3] http://www.img.lx.it.pt/GAIT-IST/

- **Hemiplegic gait** - The subject stands with unilateral weakness on the affected side (which in this case is the right side), arm flexed, adducted and internally rotated. With ambulation, the right leg moves in a circular movement. The left side of the body remains normal. In the first severity level, the right arm is slightly flexed and leaning against the waist, and the affected leg makes smaller circular motions. In the second severity level, the affected arm is leaning against the chest, the hand is closed and the affected leg emphasizes the circular motion.

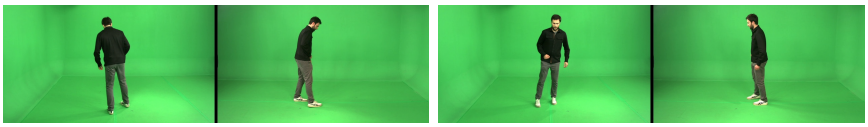


FIGURE 3.1: Frontal/side views example of the first severity level of the hemiplegic gait simulation.

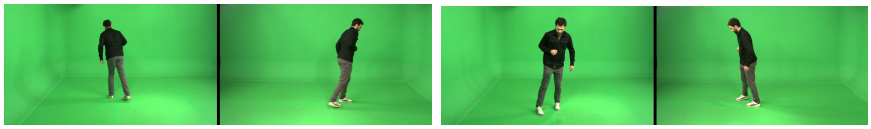


FIGURE 3.2: Frontal/side views example of the second severity level of the hemiplegic gait simulation.

- **Diplegic gait** - Also known as scissors gait, both sides of the body are affected. The subject walks dragging both legs in circular movements and scraping the toes against the floor. In the first severity level, the arms are flexed, the body is leaning slightly forward and the legs make a smooth circular motion. In the second severity level, the arms are leaning against the chest, knees are closer together and the legs make a bigger circular motion.

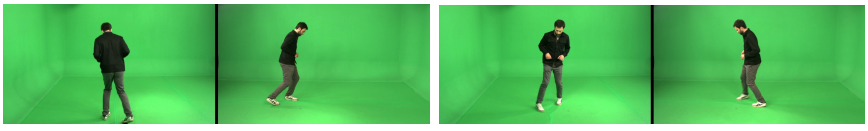


FIGURE 3.3: Frontal/side views example of the first severity level of the diplegic gait simulation.

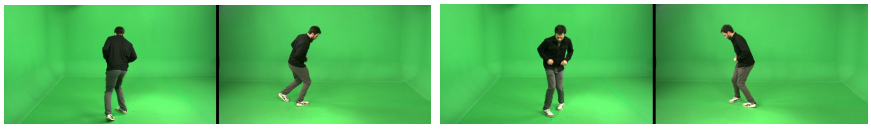


FIGURE 3.4: Frontal/side views example of the second severity level of the diplegic gait simulation.

- **Neuropathic gait** - Also known as steppage gait or equine gait, it is characterized by attempting to lift the legs high enough during walking so that the foot does not drag on the floor. The difference between severity levels is that in the second level of severity it almost seems that the subject whips his/her leg, because of the foot drop.

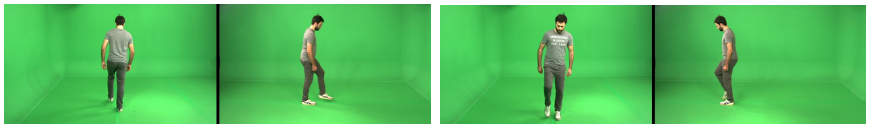


FIGURE 3.5: Frontal/side views example of the first severity level of the neuropathic gait simulation.

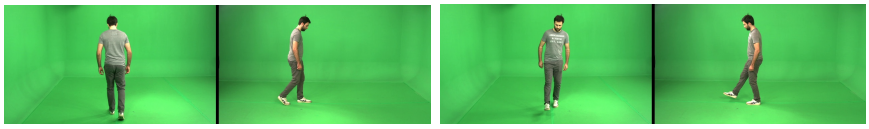


FIGURE 3.6: Frontal/side views example of the second severity level of the neuropathic gait simulation.

- **Parkinsonian gait** - The subject is stooped with the head and neck forward, with flexion at the knees. The subject walks with slow little steps, while slightly shaking his/her body. The second level of severity consists only in a slight exaggeration of these symptoms, with smaller steps and more inclination of the body.

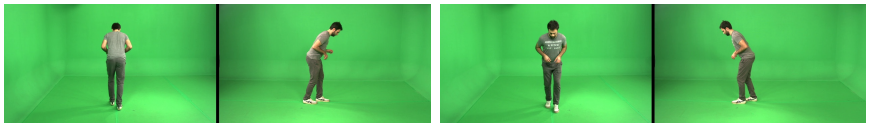


FIGURE 3.7: Frontal/side views example of the first severity level of the parkinsonian gait simulation.

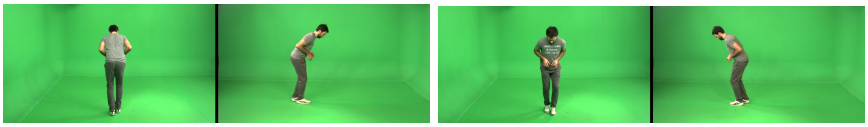


FIGURE 3.8: Frontal/side views example of the second severity level of the parkinsonian gait simulation.

3.3 Dataset Acquisition

A total of 21 subjects (19 males and 2 females) in the age range of 19-56 years old participated in the dataset acquistion, as illustrated in Figure 3.9. Each subject performed 2 severity levels per pathological gait type, 4 sequences per severity level besides normal gait. Considering that there were 2 subjects repeating the experience (on another day and with different clothes), this makes a total of 828 sequences.

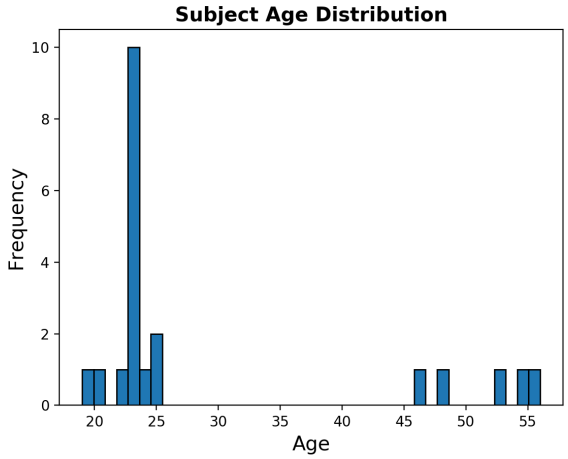


FIGURE 3.9: Histogram representing the age range of the 21 subjects who have participated in the GAIT-IT dataset aquisition

Sequences were acquired using three different cameras:

- The first camera (4K professional camera) was placed approximately 3 m from the target at 1.75 m from the ground, pointing at an empty chroma-keying background to capture the side view.
- The second camera (4K professional camera), at the same height as the first camera, was placed facing the target to capture a frontal view of the subject. This camera was synchronized with the previous one.

- Finally, the third camera was a cellphone camera with a resolution of 720p, supported on a tripod at a height of about 1.5 m, to also obtain a side view, but with a different quality. The scene was illuminated with artificial lighting all the time. All the studio information is available at FCCN's website [4].

In each take, the subject walks parallel to the first camera plane, two times from left to right and another two times from right to left, making a total of 4 sequences. In the first and third cameras, it was possible to capture at least 3 gait cycles and, in the second camera, it was only possible to capture at least 2 gait cycles, due to the fact that the frontal view does not capture the full body during all the recording time. The number of gait cycles extracted from the subjects' gait varied between gait types and subjects. The Gait-IT dataset is organized by gait type, including the 21+2 subjects, and contains 4 different representations: silhouettes, skeletons, GEIs and SEIs. These representations are saved by view (side view and frontal view) in 2 different folders, and contain 4 sequences per severity level in abnormal gait types and 4 sequences for the normal gait type.

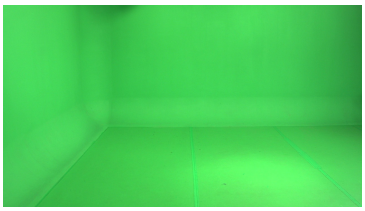
3.4 Image Pre-Processing

The obtained images have to go through several processing stages in order to obtain the desired representations, notably silhouettes, skeletons, GEI, and SEI. There are many possible representations, but the ones used in this work (in Chapter 5) are: i) the GEI since they were also used in [59], [60] and [6]; and ii) the SEI used in [6].

3.4.1 Binary Silhouette Extraction

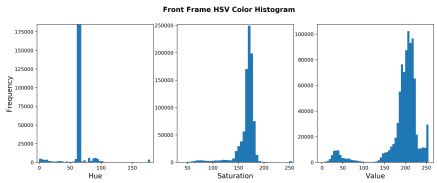
To obtain the binary silhouette representation, after extracting the raw frames from the captured videos, the extracted frames were converted into binary silhouette images. All the videos were captured in a recording studio as mentioned

[4]https://www.fccn.pt/en/collaboration/studio/

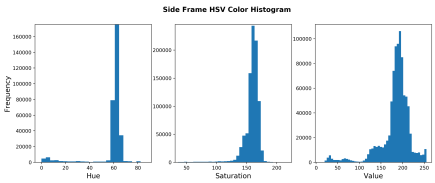


(a) Frontal view background

(b) Side view background



(c) Frontal view background HSV representation



(d) Side view background HSV representation

FIGURE 3.10: Side and frontal views of the chroma-key background and their respective HSV histograms.

before, so the background subtraction task was simplified since the uniform illumination and colour of the background allowed using segmentation techniques based on chroma-keying techniques.

For this task, an image of the chroma-keying background was used to extract the HSV (hue, saturation, value) representation of the background color, as shown in Figure 3.10. HSV is an alternative colour space representation of the original RGB (red, green, blue) color representation and it was possible to have the measured values of: i) *hue*, which is the most important component since it is the color appearance; ii) *saturation*, which determines how intense the color is; and iii) *value*, which determines the lightness of an image.

FIGURE 3.11: Background subtraction of a side view silhouette (left) and frontal view silhouette (right).

After having all the measured values from both background views (side and frontal views), it was possible to apply a filter in all silhouette frames and identify the pixel values within the range of the measured values in HSV representation. In

that way, as illustrated in Figure 3.11, it was possible to convert them into binary images, representing those pixel values in black (background) and the others in white (foreground).

3.4.2 Silhouette Size Normalization

After background subtraction, image resizing is the next stage and it is very important in the pre-processing stage, due to the fact that input sizes and shapes can vary, depending on the system that is going to be used after. The size normalization step is really important, there are different input sizes for different systems and approaches. For instance, typical CNNs usually require a fixed image input size. CNNs such as VGG-19, VGG-16, and ResNet50 require an image input size of 224×224, but Xception requires an input size of 299×299.

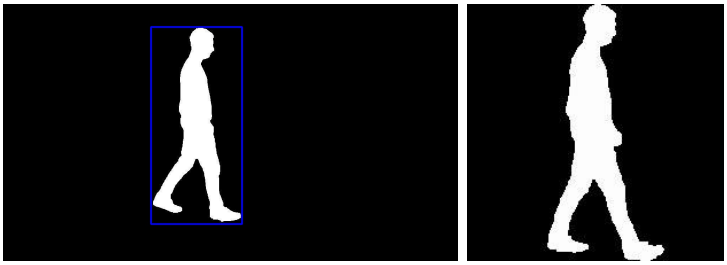


FIGURE 3.12: Computation of the bounding box of the silhouette (left) and its full resized silhouette (right).

As can be seen in Figure 3.12, by using a bounding box around the foreground object, a cropped image of the silhouette is obtained. After that, it is possible to measure the height of the image and then the image is padded with zeros from both sides until the with of the image is equal to the height, forming a perfect square. That way the image is prepared to be resized to any size because it will always be resized by its mass center. Finally, the image was resized to the desired width and height, 224×224 pixels, since this is the input size typically expected by existing CNNs (such as VGG-19).

3.4.3 Gait Energy Image (GEI)

After the resizing stage, it is now possible to obtain the gait energy image representation. To obtain such a representation, we need to obtain a full gait cycle from which a GEI will be extracted. Since a full gait cycle contains two walking steps, we need to count the steps of each subject. This is accomplished by measuring the step size and drawing the corresponding evolution along time, as illustrated in Figure 3.13.

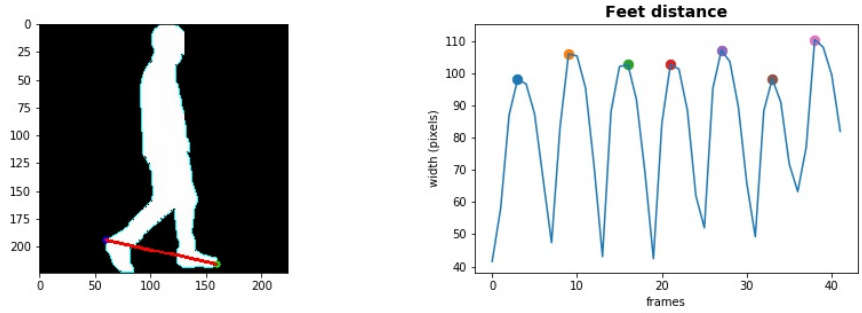


FIGURE 3.13: Final resized silhouette image with feet euclidean distance representation (left) and a plot of its width along a full gait sequence (right)

The step size is measured by applying the findContours() function from openCV library to the lower part of the silhouette, finding the contours of both legs and in that way it is possible to have the coordinates of the leftmost point of the left leg and the rightmost point of the right leg, illustrated in the left side of the Figure 3.13. These points correspond to the heel of the back leg and the toe of the front leg, respectively. The euclidean distance in a two dimensional space between points p and q, if $p = (p1, p2)$ and $q = (q1, q2)$, is given by:

$$d(p, q) = \sqrt{(q1 - p1)^2 + (q2 - p2)^2} \tag{3.1}$$

The euclidean distance between these two points was computed, as we can see in the right side of Figure 3.13. By measuring this this distance along time, makes it possible to count the steps of each gait sequence, since each step corresponds to the maximum distance between these two points. For each full gait cycle (i.e., two steps), a GEI is computed by the equation presented in Chapter 2, Section 2.3.1. The result is a grey-level image with the mean values of each pixel across all frames, comprising motion features of a full gait cycle in a single image.
As mentioned before, the frontal and side cameras were synchronized, so all the frontal and side frames were extracted with the same frame rate. In that way, it was also possible to obtain frontal view GEIs, as illustrated in Figure 3.14. There are less frontal GEIs than side GEIs, since the proximity between the subject and the frontal camera in some moments was such that the camera could not capture the full body of the subject.

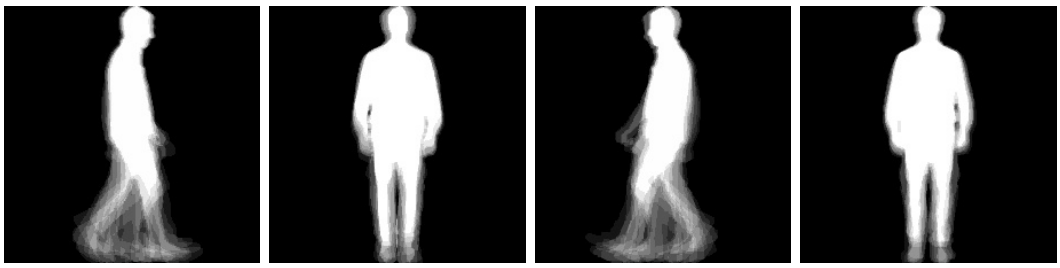


FIGURE 3.14: Side view GEI (left) and frontal view GEI (right) of the same subject at the same exact moment of the sequence (normal gait).

3.4.4 Skeleton Extraction

The OpenPose system [11] is used to compute skeleton joints information from gait videos. The algorithm can extract a set of 2D coordinates, for each frame, from videos of walking people, corresponding to different parts of the human body, as shown in Figure 3.17. The system architecture consists of a multi-stage CNN, as illustrated in Figure 3.15.

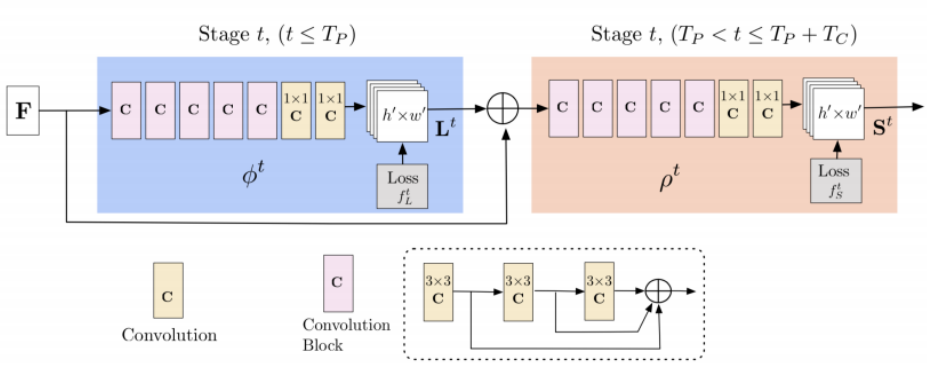


FIGURE 3.15: CNN architecture in [10].

The network predicts a set of 2D confidence maps, corresponding to the location of different body parts, as well as a set of Part Affinity Field (PAF)s able to encode the level of association between parts. The PAFs are a feature representation introduced by this work that preserves location and orientation information about the limbs. The confidence maps and PAFs are ultimately parsed to output 2D coordinates of the anatomical keypoints.

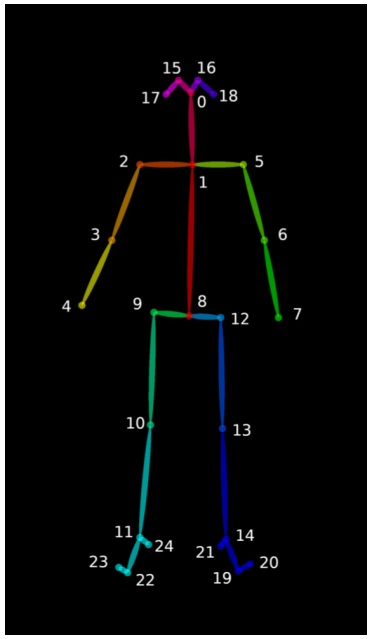
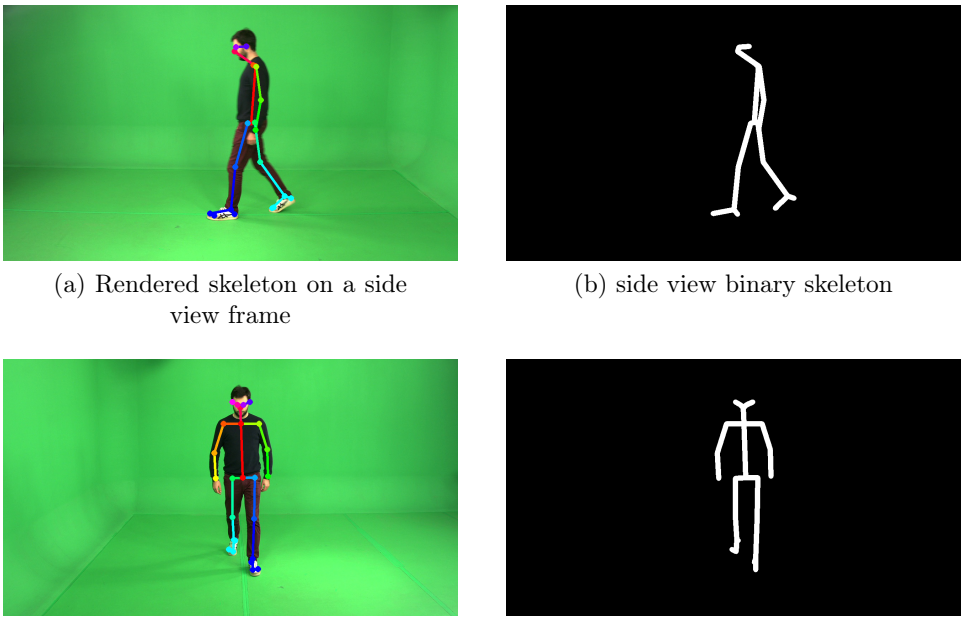


FIGURE 3.16: Pose output format of detected body parts using OpenPose [11]. Extracted coordinates: 0-Nose, 1-Neck, 2-RightShoulder, 3-RightElbow, 4-RightWrist, 5-Left Shoulder, 6-Left Elbow, 7-Left Wrist, 8-Hip (Middle), 9-Hip (Right), 10-Right Knee, 11-Right Ankle, 12-Left Hip, 13-Left Knee, 14-Left Ankle, 15-Right Eye, 16-Left Eye, 17-Right Ear, 18-Left Ear, 19-Left Big Toe, 20-Left Small Toe, 21-Left Heel, 22-Right Big Toe, 23-Right Small Toe, 24-Right Heel

After having a set of body part coordinates for each frame of the videos, as illustrated in Figure 3.16, the skeletons are then obtained by drawing a line between pairs of extracted keypoint coordinates according to these labels.



(a) Rendered skeleton on a side view frame

(b) side view binary skeleton

(c) Rendered skeleton on a frontal view frame

(d) frontal view binary skeleton

FIGURE 3.17: Side and frontal views of the openPose extraction coordinates and respective skeleton representation

3.4.5 Skeleton Gait Energy Image (SEI)

The SEI representation was first introduced in [61], showing better results in the classification task of the proposed system than the GEI representation. The SEI is obtained in the same way GEI is obtained. However, instead of silhouette frames, SEIs are obtained with skeleton frames. Steps are calculated through the euclidean distance between the leftmost point and the rightmost point of the skeleton legs, as done above for the GEI representation. Each step corresponds to the maximum measured euclidean distance between feet and, for each full gait cycle, one SEI is computed.

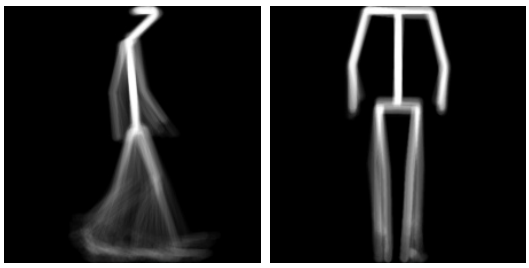


FIGURE 3.18: Side view SEI (left) and frontal view SEI (right) of the same subject at the same exact moment of the sequence (normal gait)

The frontal view SEI, as illustrated in the right side of Figure 3.18, was computed without the head of the subject. Due to the fact that there were many frames in which the head did not appear because of the proximity of the camera and its perspective. It was decided to remove the head of the subject for this representation, since it is not an important feature of this representation for gait classification.

3.5 Final Remarks

The GAIT-IT dataset was achieved due to the fact that there are currently not many publicly available gait datasets dedicated to the study of pathological gait. This dataset is organized by pathology gait type folders, where inside of each folder there is a list of 21+2 subjects. For each subject folder there are 5 more folders: i) GEI, which is further divided into a side view folder, containing all the side view GEIs by severity level, and a frontal view folder, containing all the frontal view GEIs by severity level; ii) SEI, which is also further divided into a side view folder, containing all the side view SEIs by severity level, and a frontal view folder, containing all the frontal view SEIs by severity level; iii) Silhouettes, where all the silhouette frames extracted from the recorded videos are stored. These silhouettes are organized by side view and frontal view; iv) Skeletons, which were extracted from the open pose system. The Skeletons are also organized by side view and frontal view; v) Pose, which is also organized by side view and frontal view, and contains all the body part coordinates for each skeleton frame.

Chapter 4

Proposed Gait Classification Web Application

This chapter presents the proposed gait classification application, which contains two main components:

- **Classification System** - that is able to accept a visual representation of a person's gait, in the form of a GEI or a SEI, and classify the gait as being either normal or impaired, and in the case of being impaired as either parkinsonian, hemiplegic, diplegic or neurophatic gait;
- **Web Interface** - able to accept a gait input and to remotely execute the gait pathology classification system, which is made available as a web service.

The ultimate goal of this work is to allow anyone with access to this application to evaluate the video of a person's gait, by uploading a video (which is converted into GEI), a GEI or a SEI gait representation.

4.1 Classification System Description

This classification system is divided into: image pre-processing, feature extraction and classification, as illustrated in Figure 4.1.

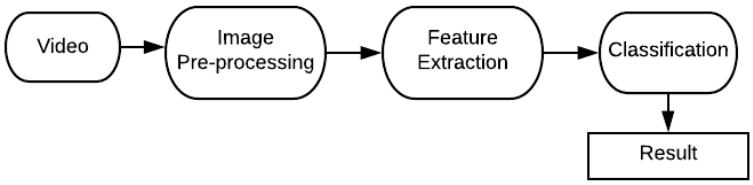


FIGURE 4.1: Diagram of the classification vision system steps.

As in [6], when using GEIs as input, fine-tuning involved re-training the last 3 convolutional blocks of the VGG-19 convolutional base. For the SEI representation, the best results were achieved by retraining all blocks except the first one. Fine-tuning is done using backpropagation, considering categorical cross entropy as the loss function, Stochastic Gradient Descent (SGD) with the Nesterov momentum variation as the optimizer and a learning rate of 0.0002.

The training and validating are made by the classification system. After the pre-processing step, described in Chapter 3, Section 3.4, the CNN performs feature extraction in the desired input images (GEI or SEI), as illustrated in Figure 4.2, in order to obtain a feature vector for every image input. After performing feature extraction, the last layers of the CNN perform classification between 4 pathologies and normal gait. There are two different input representations (GEIs and SEIs) for the CNN and in that way there are two CNNs, one trained and validated with GEIs and another one trained and validated with SEIs.

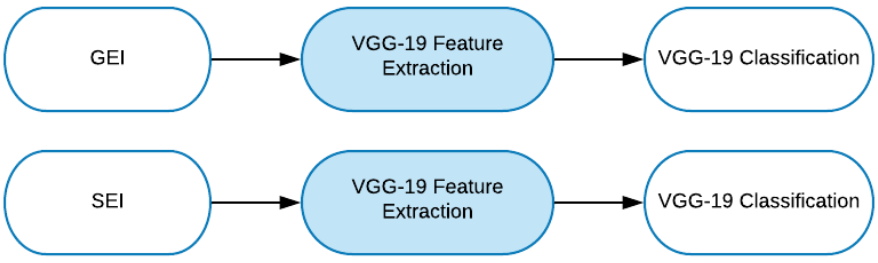


FIGURE 4.2: Diagram representing the high-level architecture of the classification system.

The classification step, represented in Figure 4.3, is made by the direct classification using the VGG-19, after the feature extraction step and after obtaining the feature vectors.

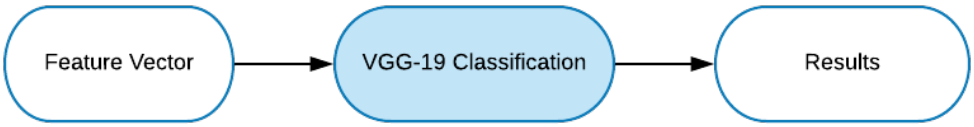


FIGURE 4.3: Diagram of the classification step implemented.

The architecture of the system proposed in [6] was used. Although the GAIT-IT dataset is the largest dataset publicly available in this field, problems such as overfitting are still likely to occur, due to the fact that the data is still not enough to be able to to completely retrain the network. Transfer leaning, explored in [62], [6] and [7], was the chosen technique to overcome this issue, using a VGG-19 deep neural network [63], pre-trained on a subset of the ImageNet dataset [64], and fine-tuned using the selected pathological gait dataset. The VGG-19 takes as input images of size 224×224 with 3 channels, which go through 5 convolutional blocks for feature extraction, each consisting of consecutive convolutional layers followed by a max-pooling layer, as illustrated in Figure 4.4. Classification is then made with a set of 3 fully connected (FC) layers at the end of the architecture. The first two FC layers have 4096 units, corresponding to the size of the flattened feature vector output by the fifth convolutional block. The third and last FC layer has a softmax activation function and was modified to have 5 units corresponding to the 5 gait classes considered here for classification: diplegic, hemiplegic, neuropathic, parkinsonian and normal gaits.

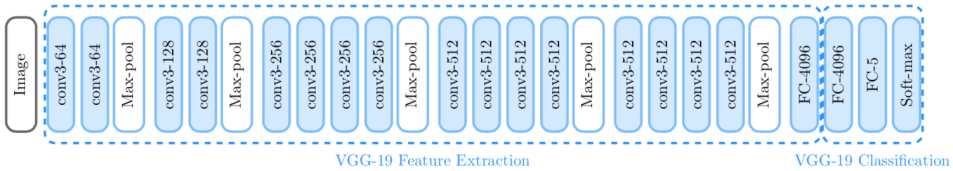


FIGURE 4.4: Architecture of the VGG-19 network, with the third fully connected layer modified. The first FC-4096 corresponds to the layer whose output is used as a feature vector [6].

4.2 Web Interface Description

The major advantage of this web interface is to have remote access to the classification system described in Section 4.1, thus making it possible for everyone with an internet connection to access it and test it. The gait classification web application has two different interface modes:

- **Basic** - The first interface mode could be used by medical doctors, nurses and technicians, who could use the web application in a clinical environment or a hospital. It could be used as a first stage of the diagnosis process or simply to help the medical doctor to detect and extract important features from the patient's gait. These users will be able to upload a video into the web application, generate a GEI and get the predicted pathology by the classification system. They will also have access to saliency maps and heatmaps to better understand what features the classification system gave more importance to. If the user so desires, all this information can be sent by email to the patient through the web application, automatically.

- **Advanced** - The second interface mode could be used by researchers in this field or students with interest in this field, with the necessary deep learning skills, but also healthcare professionals with expertise and experience in this web application. These users will have access to more features than the users in the previous group. Like the users in the previous group, they have the possibility to upload a video, but in addition to this they also have the possibility to directly upload a GEI or a SEI. This additional possibility is due to the fact that these users may want to use the web application to test

their own GEI or SEI in the classification system. They not only have access to saliency maps and heatmaps, but also to the feature map extracted from a certain layer and channel chosen by the user. This allows these users to better understand the filters applied by the CNN in each layer and channel.

The use of a web service allows an outside web application to have access to the classification system, even if it is not written in the same programming language. The web application was developed using Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS) and Javascript for the front-end and Python for the back-end. The main goal of such a web application is to allow the user to access the trained model and its pathology classification from an outside application and that is only possible using a web service. The web service used for this web application was Flask [1], which allowed to deploy the previous trained models. The fact that Flask is written in Python and allows using frameworks such as Tensorflow and Keras is fundamental since these frameworks were used before to train and validate the models. Flask is a lightweight Web Server Gateway Interface (WSGI), which is a specification that describes how a web server communicates with web applications and how web applications can be chained together to process one request. Deploying a model into a web service allows the user to access it from the web application. By uploading the models input into the web application, it is possible to make a Hypertext Transfer Protocol (HTTP) request to the web service, and the web service sends back a HTTP response to the web application, as shown in Figure 4.5. HTTP is a protocol designed to enable communications between clients and servers.

[1] https://flask.palletsprojects.com/en/1.1.x/

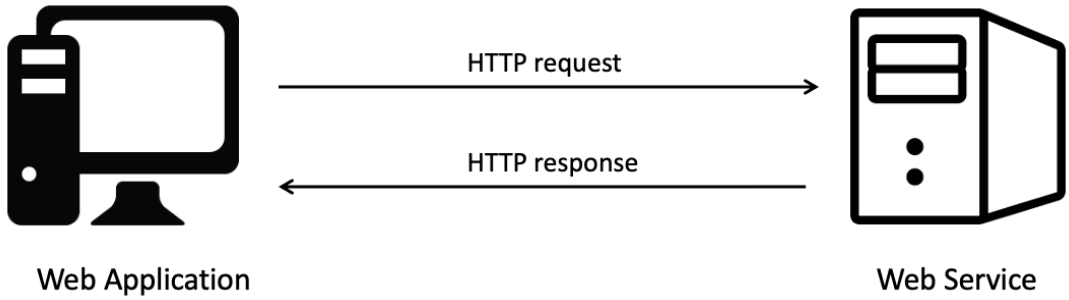


FIGURE 4.5: Web application request and response.

The communication between server and client is what allows the user of the web application to have access to the classification system and all the different maps representing the extracted features, since both can only be obtained by requesting it to the server.

4.3 Gait Classification Web Application Scenarios

In this section, two different scenarios will be presented, corresponding to the two possible modes, to help the reader better understand the gait classification web application and its functionality.

The first scenario is a simulation of the basic interface mode, in which the user uploads a video (of the patient's gait) into the web application. After uploading it, the user can ask the application to generate a GEI and get the automatic system classification result. The user will also be able to see a saliency map and a heatmap, highlighting which areas of the GEI contributed most to the classification decision. Finally, it is possible to automatically send all the information about the patient's diagnosis to the application user, or even to the patient, by email.

The second scenario simulates the advanced interface mode, in which the user can directly upload a GEI and/or a SEI, using the web interface in the same web page. Afterwards, the user will be presented the classification result obtained by the automatic gait classification system. For each uploaded gait representation, GEI and/or SEI, the corresponding pre-trained CNN model is used by the web service. After getting the classification result for the uploaded GEI and/or SEI, the

user can choose to visualize detailed information about the feature maps created by the CNN for this input - there is an option to select the CNN layer and a specific channel of the layer to be visualized. The user will also be presented with the saliency map and the heatmap, illustrating the areas of the GEI and/or SEI that contributed most to the obtained classification result. It is therefore possible to compare the maps and the extracted features for the different gait representations.

4.3.1 First Scenario: Basic Mode

In this scenario, the basic interface mode is considered. This mode is essentially for users such as medical doctors, nurses and technicians, to use it in a clinical environment or in a hospital for a first diagnosis of the patient or even to help the medical doctor to extract features from the patients gait. By clicking on the diagnosis button, presented in the home page, as shown in Figure 4.6, the user can start the patient's diagnosis on the Gait Classification Web Application.

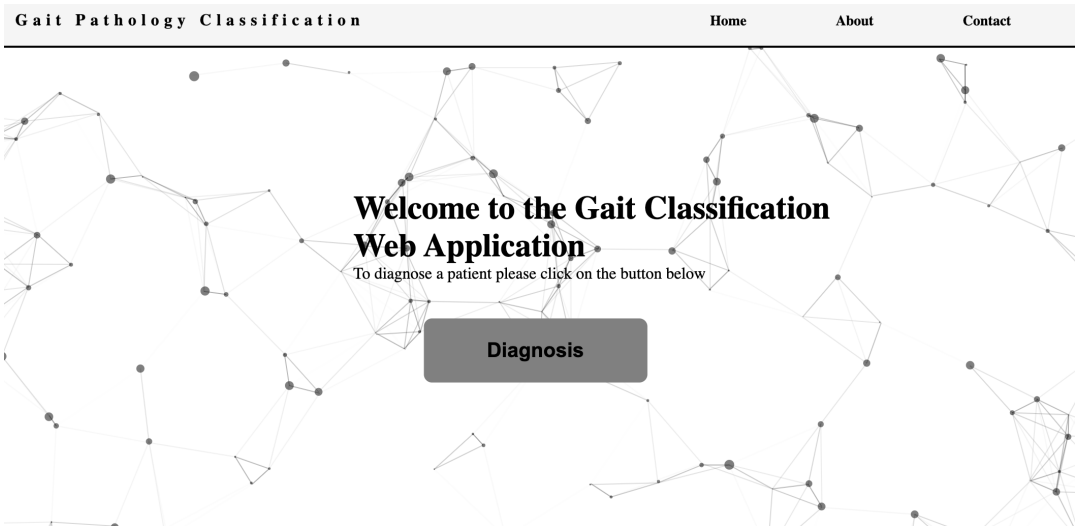


FIGURE 4.6: Gait Classification Web Application home page

To choose how the diagnosis will be done, the user has to choose which type of interface mode he/she wants to use. As mentioned before, the interface mode considered in this scenario is the basic mode, as illustrated in Figure 4.7 where that option is highlighted.

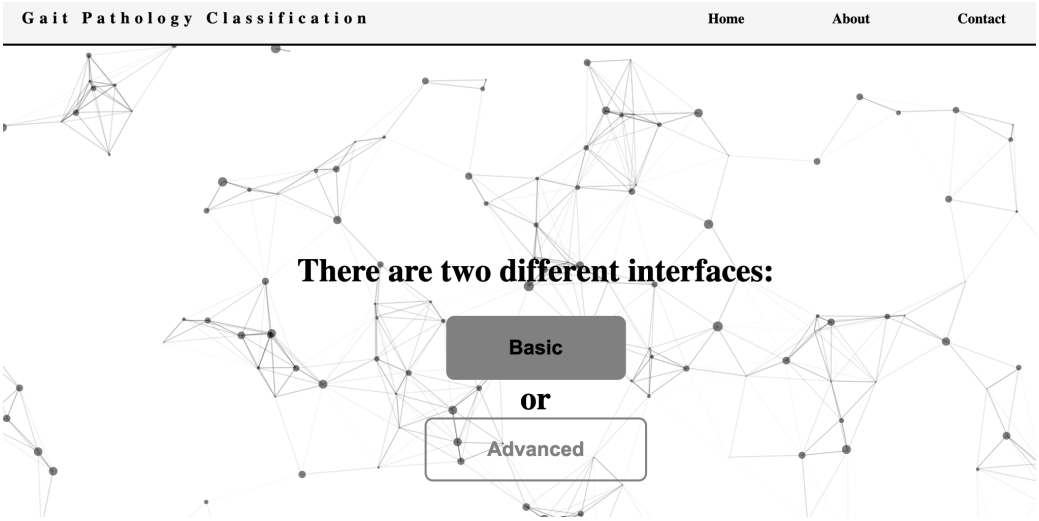


FIGURE 4.7: Selection of interface mode in the Gait Web Application: Basic.

After choosing the basic mode, an interface with all the possible interactions with the user is presented, as shown in Figure 4.8.

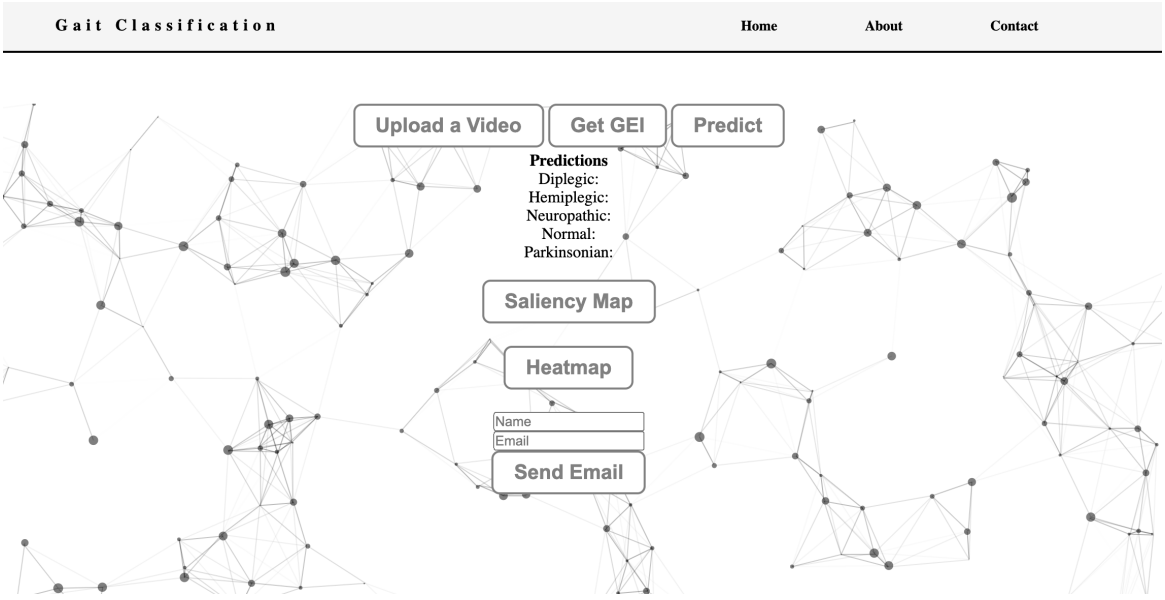


FIGURE 4.8: Web page with all the available interactions for the basic interface mode.

Here, the user can upload the patient's video into the web application and generate a GEI for it. After generating the GEI, the user is informed of the predicted classification for the uploaded gait video sequence. This sequence is presented in Figure 4.9.

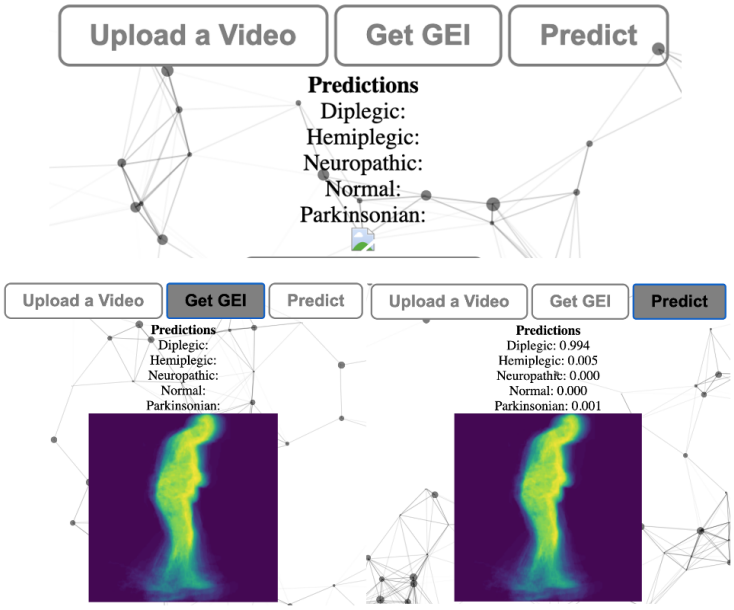
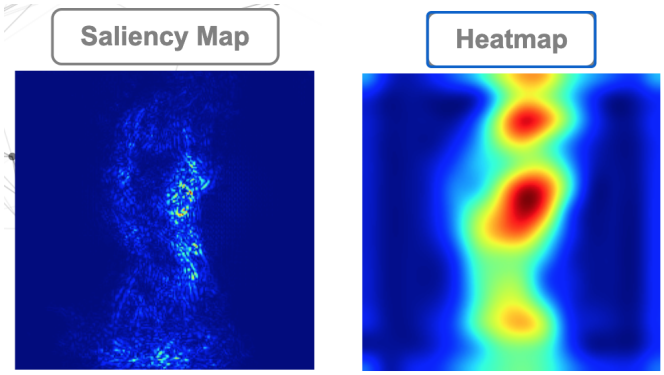


FIGURE 4.9: On top, it is possible to observe the uploaded video; On the bottom left, the generated GEI; On bottom right, the predicted classification.

In addition to this, the user is also shown the saliency map and the heatmap (described in Section 4.4) of the generated GEI, as illustrated in Figure 4.10. In this way, it is possible to observe which features were extracted by the CNN.



(a) (b)

FIGURE 4.10: (a) Saliency map of the generated GEI. (b) Heatmap of the generated GEI.

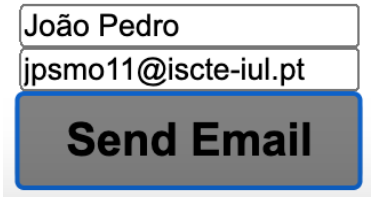


FIGURE 4.11: Form of the Gait Web Application to insert the email address of the patient to which the results will be automatically sent to.

Finally, the user can send an email to the patient with all the information of the diagnosis by just filling in the form with the patient's name and email, as illustrated in Figure 4.11. The email, of which an example is given in Figure 4.12, will be automatically sent by the server.

Dear João Pedro,

Your results from the Gait Web App are now available. Please check them below:

Diplegic:0.994;
Hemiplegic:0.005;
Normal:0.000
Neuropathic:0.000;
Parkinsonian:0.001.

Please find attached the image results from the WA diagnosis.

Best Regards,

Gait Web App

FIGURE 4.12: Email sent by the Gait Web Application automatically.

4.3.2 Second Scenario: Advanced Mode

In this scenario, the advanced interface mode is considered. This mode is for users such as researchers, students in the field or even experienced users of this web application which can understand the input gait representations (GEI and SEI). It could be used to compare not only the results from the same person and from different gait representations, but also to compare the extracted features from the saliency map and heatmap. This choice is made by selecting the advanced mode option available on the page, as shown in Figure 4.13. This interface mode will have more options available since the user can now upload a GEI and/or a SEI, instead of the original video, as illustrated in Figure 4.14.

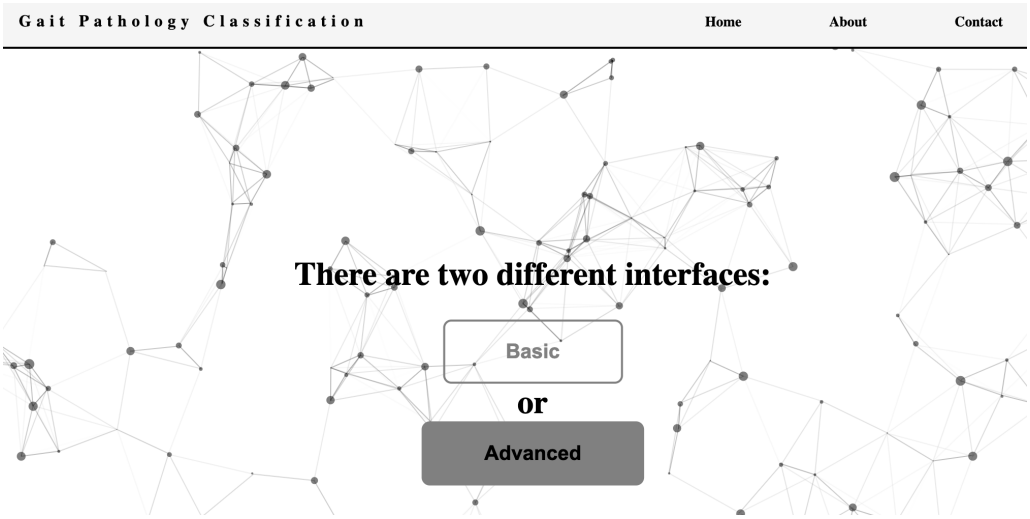


FIGURE 4.13: Selection of interface mode in the Gait Web Application: Advanced.

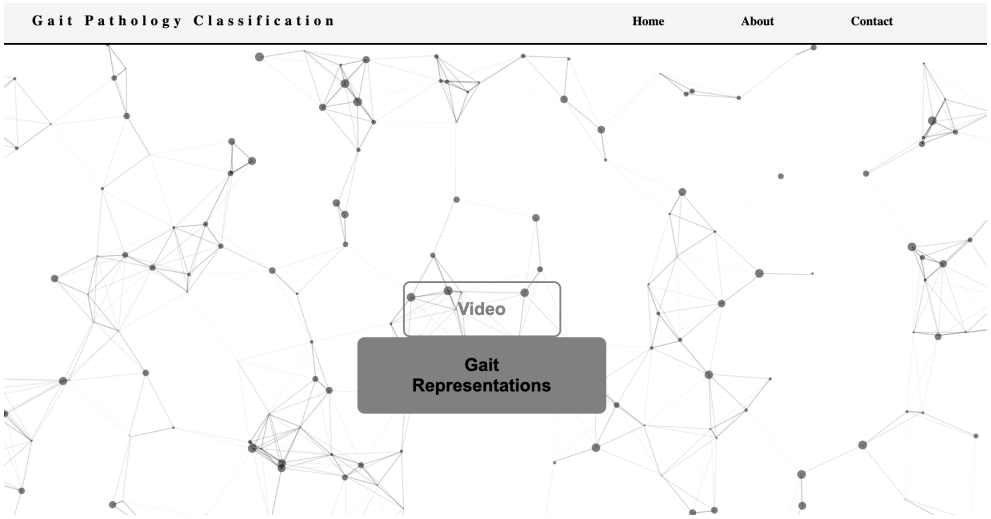


FIGURE 4.14: Web page showing the two possible selections for the advanced interface mode: video or gait representations.

The possibility of uploading GEIs and SEIs only makes sense for this type of users because these gait representations are only typically known by researchers/students in the field or someone used to this representations, not by general healthcare professionals. In the gait classification web application, to upload these representations, the user chooses "Gait Representations", as illustrated in Figure 4.14, having the possibility to upload a GEI and/or a SEI, as shown in Figure 4.15.

The user uploads a GEI and/or a SEI of the same person, as shown in Figure

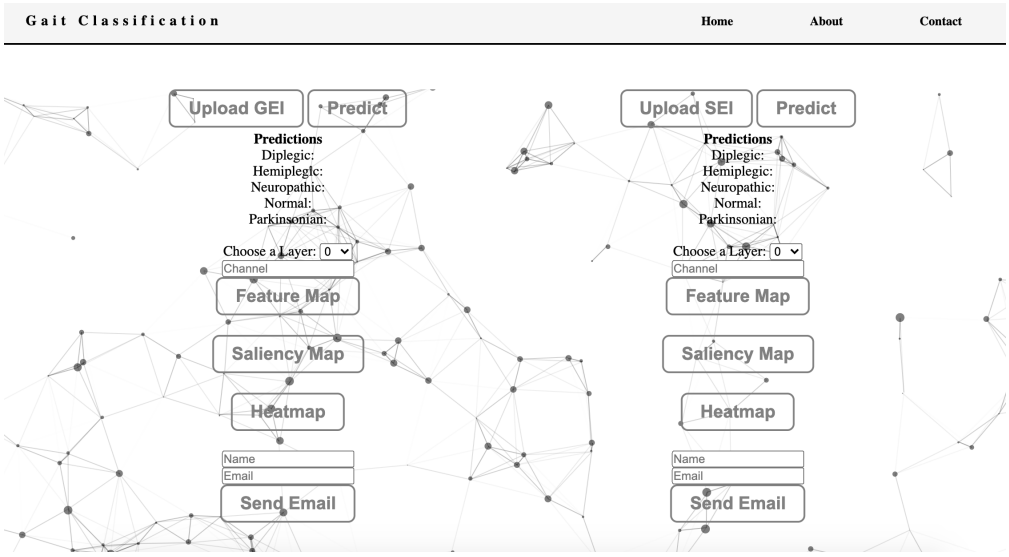


FIGURE 4.15: Web page showing the possible gait representations that can be uploaded into the Gait Web Application.

4.16 to check the differences in the predicted results and extracted features from the different maps.

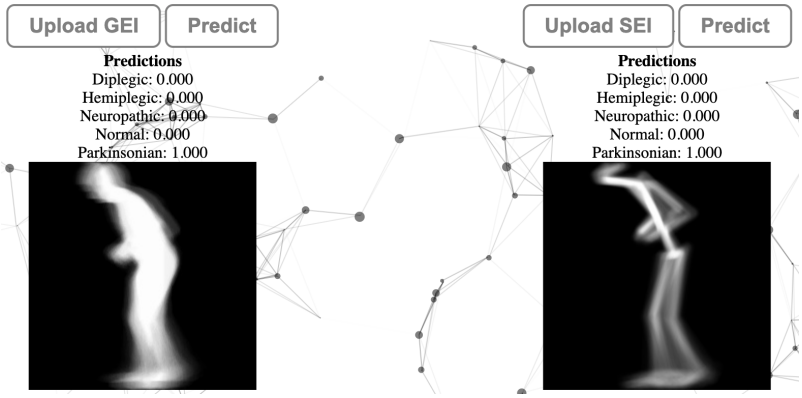


FIGURE 4.16: Upoaded GEI and SEI with the respectives predicted classifications.

After getting the predicted classifications, the user extracts the feature maps from both classifications, by choosing the desired layer and channel (described in Section 4.4), as illustrated in Figure 4.17.

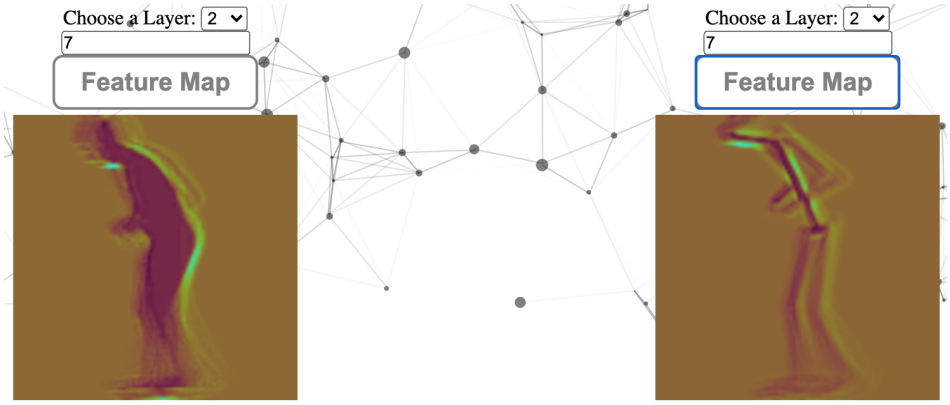


FIGURE 4.17: Feature maps of the second layer and seventh channel from both gait representations.

Finally, to better understand the different features extracted from the CNN, the user gets a saliency map and a heatmap for each gait representation. In Figure 4.18, it is possible to observe that in GEI the main focus was the torso orientation and hands, while in SEI the main focus was the elbows and the torso orientation. In both maps of the gait representations, the feet were also an extracted feature, although with less importance than the others.

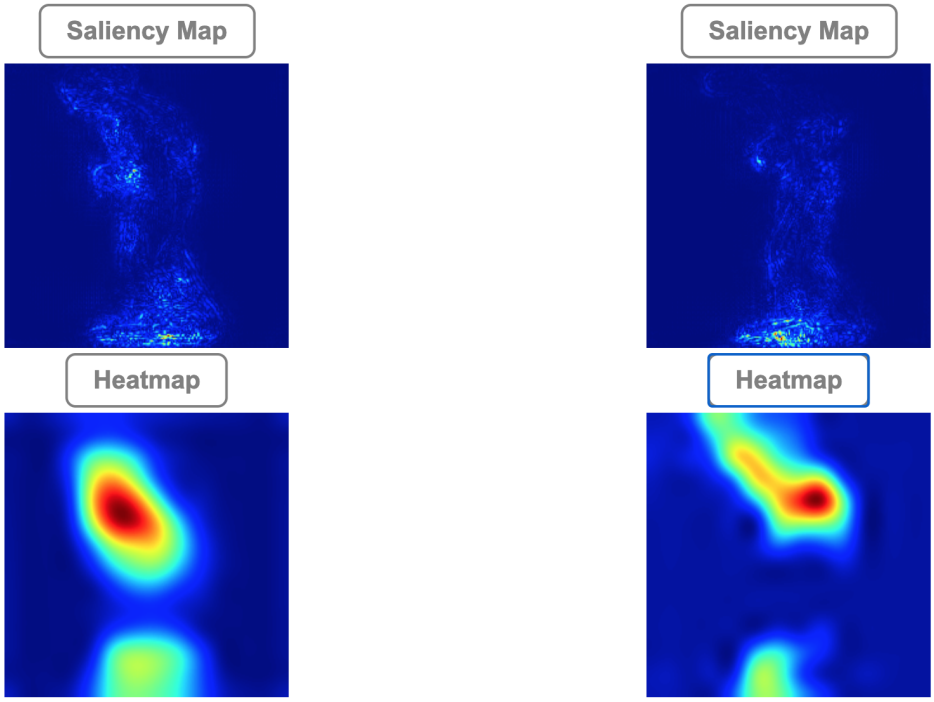


FIGURE 4.18: Saliency maps and heatmaps of the uploaded gait representations.

4.4 CNN Deep Feature Visualisation

NN are usually seen as "black boxes" because they are not always transparent and it can be confusing to understand how NN learn features and make decisions. For CNNs, however, this is not the case, due to the fact that the representations learned by CNNs are highly amenable to visualization since they correspond to representations of visual concepts. This web application gives users the possibility to access feature maps from different layers and channels, which are a visualization of intermediate activations. It also gives users the possibility to obtain saliency maps and heatmaps of class activation in an image. These representations, which are obtained through the Keras Visualisation Toolkit [65], will be discussed in the following sub-sections.

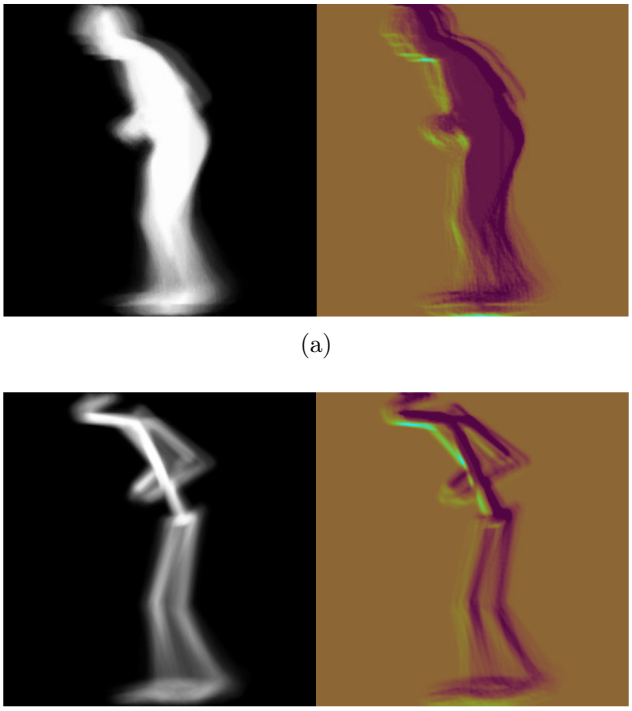
4.4.1 Visualizing Intermediate Activations

To visualize intermediate activations, the feature maps that are output by various convolution and pooling layers in a network are displayed. The output of a layer is often called its activation. This gives a view into how an input is decomposed by the different filters learned by the network. The visualized feature maps have three dimensions (or channels): width, height, and depth. Each of these channels encodes relatively independent features [4]. All the layers, output shapes (which are 4D arrays representing the batch size of the image, height, width and depth) and parameters of the trained CNN are presented in Figure 4.19.

Layer (type)	Output Shape	Param #
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 4096)	102764544
dropout (Dropout)	(None, 4096)	0
dense_1 (Dense)	(None, 4096)	16781312
dropout_1 (Dropout)	(None, 4096)	0
dense_2 (Dense)	(None, 5)	20485

FIGURE 4.19: Summary of all the layers and activation shapes of the classification system model.

It is possible to observe that the size of the output image and the number of channels of each layer vary significantly. The size of the output image varies from 224×224 pixels in the first convolutional layer to 7×7 pixels in the last pooling layer, while the number of channels varies from 64 in the first convolutional layer to 512 channels in the last pooling layer.



(a)

(b)

FIGURE 4.20: (a) Feature map of the first convolutional layer and seventh channel of the GEI. (b) Feature map of the first convolutional layer and seventh channel of the SEI.

As shown in Figure 4.20, from these two examples of feature maps, it is possible to observe that the seventh channel of the first convolutional layer appears to encode an edge detector.

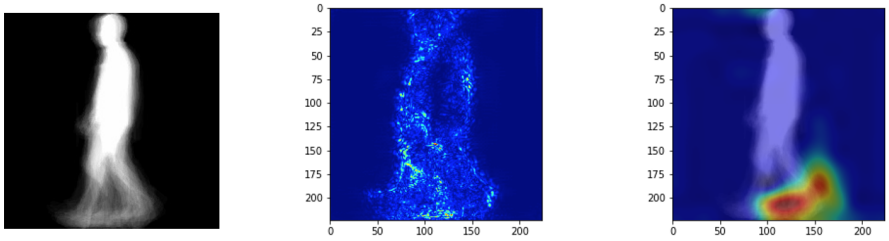
4.4.2 Visualizing Saliency Maps and Class Activation Maps

Both maps are a way of visualising what part of the image the network is paying more attention to. This way, it is possible to know if the model is learning the correct features for each class.

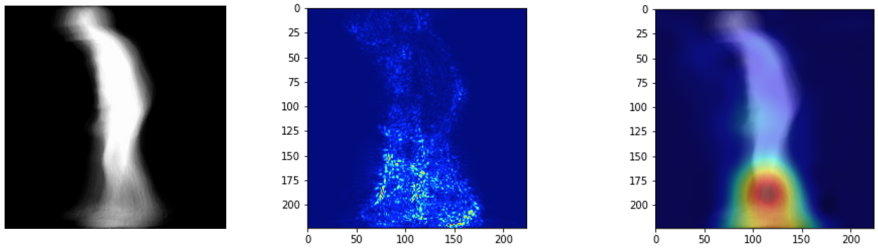
The saliency map is the computation of the gradient of an output class and the final result is an image where the pixels that would produce an increase in the output value are given more importance.

The other way of visualising what CNNs learn is through Class Activation Maps (grad-CAMs). This map instead of using the gradients with respect to the output, obtains spatial information on the last convolution layer.

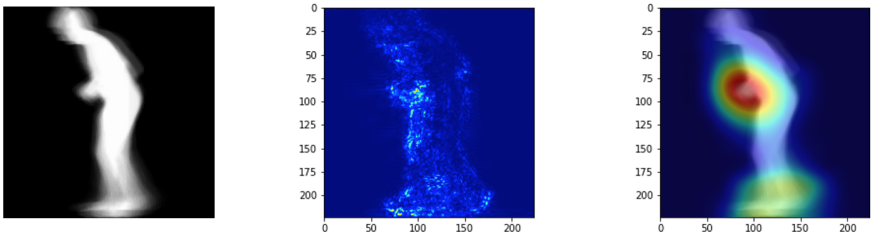
In figure 4.21, for each pathological and normal gait, a GEI, a Saliency Map and a grad-CAM is shown. The features extracted from these GEIs are almost all from the lower part of the body (normal, hemiplegic and neuropathic). Only in diplegic gait and parkinsonian gait, upper body features are extracted. In diplegic gait, hands are next to the hips. In parkinsonian gait, the torso orientation and the fact that the hands are almost together are the features that were extracted. In Figure 4.22, the same maps were extracted from SEIs. It is possible to observe that the features are slightly different. For normal gait, parkinsonsian gait and diplegic gait, the features are basically the same as in GEIs. For neuropathic gait, besides the legs, shoulders and hips were the locations from which features were extracted in SEI (but not in GEI). This is probably due to the fact that these features are more explicit in this gait representation. For a similar reason, in hemiplegic gait in SEI, the CNN looked at the affected arm. This happens because in SEI arms are more explicit than in GEI.



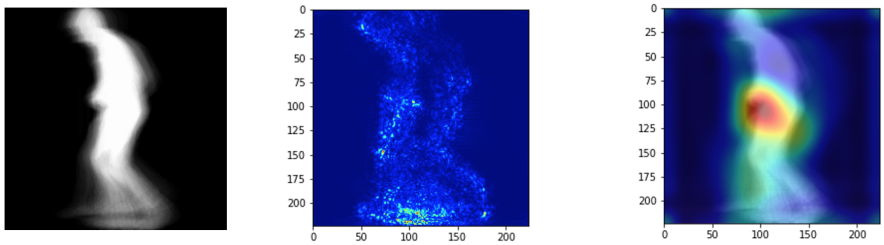
(a)



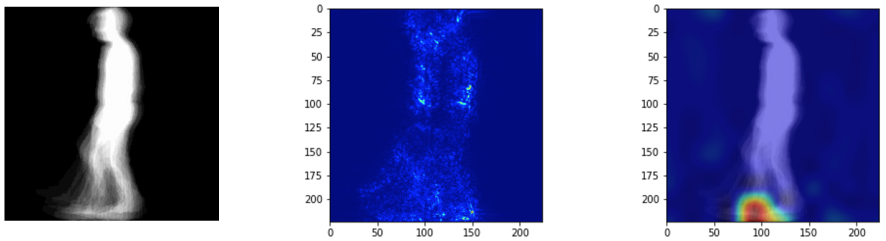
(b)



(c)

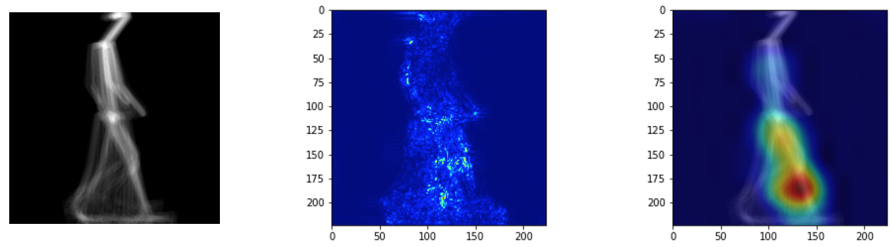


(d)

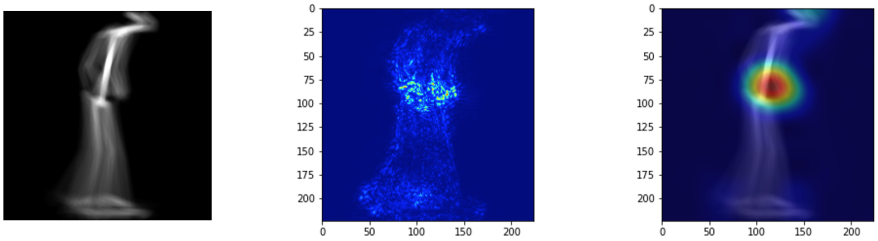


(e)

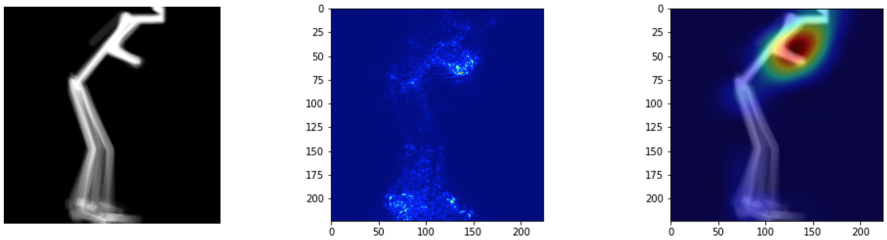
FIGURE 4.21: (a) Normal gait.. (b) Hemiplegic gait. (c) Parkinsonian gait. (d) Diplegic gait (e) Neuropathic gait.



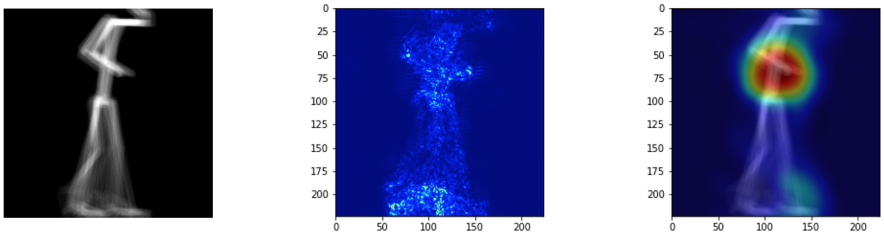
(a)



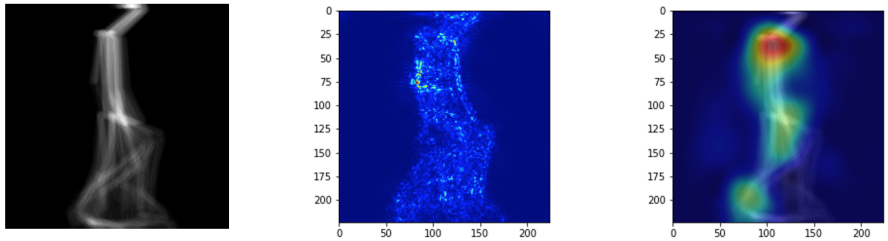
(b)



(c)



(d)



(e)

FIGURE 4.22: (a) Normal gait. (b) Hemiplegic gait. (C) Parkinsonian gait. (d) Diplegic gait. (e) Neuropathic gait.

4.5 Final Remarks

This chapter presents the gait classification web application, which is divided into two components, the classification system and the web interface. This web application makes such gait classification system more useful, due to the fact that having a web interface makes it accessible for all the people with an internet connection and not just for researchers. This web application has a basic interface mode and an advanced interface mode, allowing different users with different objectives to use it. For instance, the basic interface mode can help healthcare professionals to extract features from the patients gait which can help in the patients diagnosis, and the advanced interface mode can help more experienced users, researchers or specialists in the field to compare features extracted from different gait representations or even to compare the predicted results from the different gait representations of the same person.

Chapter 5

Classification Performance Evaluation

This chapter provides pathological gait classification results computed using the publicly available datasets (DAI2 and GAIT-IST) and the proposed GAIT-IT dataset. The initial set of results provided are obtained by training and testing with the same pathological gait dataset, using a cross-validation methodology. However, this Section shows that a larger and better quality dataset can help obtain improved classification results, especially when the trained algorithms have to operate in conditions different from those used for training. This is simulated with cross-dataset tests, used to compare and study the influence of the size and image acquisition quality of each dataset. All the reported results adopt the VGG-19 deep neural network [63], pre-trained on ImageNet dataset [64], and fine-tuned using the selected pathological gait datasets.

5.1 Cross-validation Experiments

This set of results corresponds to training and testing using the same pathological gait dataset. Classification performance results are reported for each gait representation (GEI and SEI), using 10-fold cross-validation, considering 3 datasets:

- **GAIT-IST** - for each fold 9 subjects were used for training and the remaining subject for testing, as done in [6], which means all the subjects ended up being used for validation.

- **GAIT-IT** - the test set for each fold is defined as $V_k = \{S_i, S_{i+1}, S_{i+2}\}$, where $i = 2 \times k - 1$, k is the fold iteration and S_i represents one of the 21 available subjects, following the numbered labels used for each subject in the dataset. This arrangement had the purpose of using all subjects in the test set at least once and providing a significant number of folds for the cross-validation.

- **GAIT-IST and GAIT-IT** - in each fold a different subject from GAIT-IST was used in the test set together with 3 subjects from GAIT-IT, making a total of 27 subjects for training and 4 subjects for testing. The test sets are defined as $V_k = \{V_{IST_k}, V_{IT_k}\}$ where $V_{IST_k} = \{S_{IST_k}\}$, $V_{IT_k} = \{S_{IT_i}, S_{IT_{i+1}}, S_{IT_{i+2}}\}$, $i = 2 \times k - 1$ and k is the fold iteration. The combination of these two datasets is made to test the impact of more quantity and two different dataset scenarios on training.

The classification accuracy on the test set was computed for each fold at the optimal training epoch and the overall accuracy was obtained by averaging over all folds.

		Tested on		
	Trained on	GAIT-IST	GAIT-IT	GAIT-IST+IT
GEI	GAIT-IST	94.2%	-	-
	GAIT-IT	-	94.8%	-
	GAIT-IST+IT	-	-	94.4%
SEI	GAIT-IST	98.4%	-	-
	GAIT-IT	-	93.6%	-
	GAIT-IST+IT	-	-	93.2%

TABLE 5.1: Summary of all classification accuracies across the same datasets using GEIs and SEIs as inputs and re-trained on VGG-19.

The reported results when training and testing on the same dataset are fairly good for all the considered datasets, always above 93% average classification accuracy, as illustrated in Table 5.1. The best classification accuracy was achieved by the GAIT-IST dataset using SEIs as input, reaching 98.4% classification accuracy.

5.2 Cross-Dataset Tests

Cross-dataset tests were conducted to evaluate the impact that training on a larger dataset can have in the trained model's generalization capability. Three scenarios were considered: i) training on the GAIT-IST dataset and testing on the GAIT-IT or the DAI2 datasets; ii) training on GAIT-IT and testing on GAIT-IST or DAI2; iii) training on the combined GAIT-IST and GAIT-IT datasets and testing on DAI2. Notice that, since some of the silhouettes provided in DAI2 include significant segmentation errors, this dataset was not considered for training in the cross-dataset tests, as it might lead the model to learn inaccurate features.

5.2.1 Training with GAIT-IST

When using the GEIs from the complete GAIT-IST dataset to train the VGG-19 network, the model achieved its best cross-dataset results on the GAIT-IT dataset, with 32 training epochs, and an overall accuracy of 72.4%. As expected, the classification accuracy was lower when testing with the DAI2 dataset (55.6%), due to the fact that DAI2 has poorly segmented silhouettes.

		Predicted Label				
		Diplegic	Hemiplegic	Neuropathic	Normal	Parkinson
True Label	Diplegic	0.49	0.20	0.05	0.00	0.26
	Hemiplegic	0.29	0.62	0.07	0.00	0.02
	Neuropathic	0.01	0.10	0.85	0.02	0.03
	Normal	0.01	0.17	0.03	0.80	0.00
	Parkinson	0.13	0.01	0.00	0.00	0.86

TABLE 5.2: Confusion matrix of the classification accuracy using GEI as input and re-trained on VGG-19 using GAIT-IST and tested on GAIT-IT.

As illustrated in Table 5.2, diplegic and hemiplegic gaits were the ones with worst classification accuracy when GEI is used as input and VGG-19 is re-trained using GAIT-IST and tested on GAIT-IT. The fact that these tests are made on different datasets can justify these two pathologies having a lower classification accuracy, due to the fact that they can have been simulated slightly different on each dataset.

		Predicted Label				
		Diplegic	Hemiplegic	Neuropathic	Normal	Parkinson
True Label	Diplegic	0.13	0.21	0.36	0.00	0.30
	Hemiplegic	0.04	0.79	0.06	0.06	0.05
	Neuropathic	0.00	0.70	0.24	0.06	0.00
	Normal	0.00	0.24	0.02	0.75	0.00
	Parkinson	0.02	0.03	0.09	0.00	0.87

TABLE 5.3: Confusion matrix of the classification accuracy using GEI as input and re-trained on VGG-19 using GAIT-IST and tested on DAI2.

Testing on DAI2, the pathologies with less classification accuracy were diplegic and neuropathic gaits, as shown in Table 5.3. The adoption of different walking styles when simulating these pathologies on each dataset can lead to a lower classification accuracy. It is also possible to observe that hemiplegic gait is predicted by

the network when neurophatic gait is the true label 70% of the times, which means that a similarity was detected in the way that these pathologies were simulated on GAIT-IST and DAI2. The similarity between hemiplegic gait and neurophatic gait can be explained by the fact that in side view it is possible to observe that both types of gait have similarities in the leg movement, which can make GEIs looking really close.

When using the SEI representation, the best performance was again achieved when testing on the GAIT-IT dataset, with a classification accuracy of 68.8%, for 23 training epochs and it is possible to observe that the classification accuracy doesn't vary substantially, as shown in Table 5.4, from the results presented on Table 5.2, but it is possible to observe an improvement for diplegic and neuropathic gaits. It was not possible to test it on DAI2, because the original videos are not available and in that way it is not possible to obtain the skeletons, thus it is not possible to obtain the SEI representation.

		Predicted Label				
		Diplegic	Hemiplegic	Neuropathic	Normal	Parkinson
True Label	Diplegic	0.37	0.17	0.01	0.00	0.45
	Hemiplegic	0.18	0.72	0.02	0.01	0.07
	Neuropathic	0.00	0.20	0.76	0.03	0.02
	Normal	0.00	0.17	0.13	0.71	0.00
	Parkinson	0.12	0.00	0.00	0.00	0.88

TABLE 5.4: Confusion matrix of the classification accuracy using SEI as input and re-trained on VGG-19 using GAIT-IST and tested on GAIT-IT.

5.2.2 Training with GAIT-IT

When using the GEIs from the GAIT-IT dataset to train the VGG-19 network, the best overall classification accuracy (86.4% with 24 training epochs) was obtained when testing on the GAIT-IST dataset. It is possible to observe that hemiplegic gait was the one with lower classification accuracy, with a value of 59%. For this pathology there was a 25% misclassification rate with normal gait, as illustrated in Table 5.5. All the other pathologies have classification accuracies above 90% which is really positive when performing a cross-dataset test.

		Predicted Label				
		Diplegic	Hemiplegic	Neuropathic	Normal	Parkinson
True Label	Diplegic	0.92	0.00	0.00	0.00	0.08
	Hemiplegic	0.06	0.59	0.10	0.25	0.00
	Neuropathic	0.00	0.02	0.95	0.03	0.00
	Normal	0.00	0.00	0.06	0.93	0.00
	Parkinson	0.07	0.00	0.00	0.00	0.93

TABLE 5.5: Confusion matrix of the classification accuracy using GEI as input and re-trained on VGG-19 using GAIT-IST and tested on GAIT-IST.

When testing on the DAI2 dataset and using the GEIs from the GAIT-IT dataset to train the VGG-19 network, it is possible to observe a significant improvement in the classification accuracy, specially in diplegic gait and neuropathic gait, from the one presented when using the GEIs from the GAIT-IT dataset to train. This significant improvement from 55.6% to 78.0% shows the truly impact of a bigger dataset when training, as illustrated in 5.6.

		Predicted Label				
		Diplegic	Hemiplegic	Neuropathic	Normal	Parkinson
True Label	Diplegic	**0.68**	0.11	0.11	0.01	0.08
	Hemiplegic	0.12	**0.68**	0.18	0.00	0.02
	Neuropathic	0.08	0.06	**0.85**	0.00	0.02
	Normal	0.00	0.02	0.27	**0.71**	0.00
	Parkinson	0.02	0.00	0.00	0.00	**0.98**

TABLE 5.6: Confusion matrix of the classification accuracy using GEI as input and re-trained on VGG-19 using GAIT-IT and tested on DAI2.

When using the SEI representation, the only results were presented testing with the GAIT-IST, with an accuracy of 92%, with 17 training epochs, which can be calculated by the mean accuracy of the classifications presented in Table 5.7.

		Predicted Label				
		Diplegic	Hemiplegic	Neuropathic	Normal	Parkinson
True Label	Diplegic	**0.88**	0.03	0.00	0.00	0.08
	Hemiplegic	0.00	**0.92**	0.05	0.03	0.00
	Neuropathic	0.00	0.08	**0.87**	0.04	0.00
	Normal	0.00	0.02	0.00	**0.98**	0.00
	Parkinson	0.05	0.00	0.00	0.00	**0.95**

TABLE 5.7: Confusion matrix of the classification accuracy using SEI as input and re-trained on VGG-19 using GAIT-IT and tested on GAIT-IST.

The impact of using a larger dataset for training is obvious when testing on the DAI2 dataset using the GEI gait representation (the SEI could not be computed from the DAI2 contents). The overall accuracy increased to 78.0%, with for 35 training epochs. This represents an increment exceeding 22% in the model's performance on this cross-dataset classification task, when compared to the results of the model trained with GAIT-IST. This confirms that the larger dataset used for training limited the model overfitting, a clear advantage of the proposed GAIT-IT dataset.

To further evaluate the impact of the high-quality silhouettes provided with the proposed GAIT-IT dataset, the VGG-19 was also trained using just the first 10 subjects of GAIT-IT (the same number available in the GAIT-IST dataset), and used this truncated version of the dataset, named GAIT-IT(10), for training the model in the same conditions considered when training with GAIT-IST. An improvement of 9.8% in the cross-dataset results was observed when testing with the DAI2 dataset, in comparison to the accuracy value obtained when training with the GAIT-IST dataset (65.4% vs. 55.6%). This improvement seems to confirm the advantage of the proposed GAIT-IT dataset, even when training with the same number of samples.

5.2.3 Training with GAIT-IST and GAIT-IT

Finally, both datasets were combined to train the VGG-19 network, corresponding to a total of 33 sets of input video sequences. This includes the 10 subjects from GAIT-IST and the 21 subjects (plus the 2 subject repetitions) from GAIT-IT. Besides further increasing the amount of available training data, this experiment provides the model with some variety in terms of the conditions in which the gait sequences were acquired.

As shown in Table 5.8, the best results were achieved by training the model for 19 epochs with an overall accuracy of 81.4%, an increment of more than 25%

and 3% in the classification performance when compared to training the model with just the GAIT-IST and GAIT-IT, respectively.

		Predicted Label				
		Diplegic	Hemiplegic	Neuropathic	Normal	Parkinson
True Label	Diplegic	0.59	0.09	0.26	0.00	0.06
	Hemiplegic	0.06	0.70	0.20	0.00	0.04
	Neuropathic	0.00	0.03	0.94	0.03	0.00
	Normal	0.00	0.06	0.06	0.88	0.00
	Parkinson	0.03	0.01	0.01	0.00	0.96

TABLE 5.8: Confusion matrix of the classification accuracy using GEI as input and re-trained on VGG-19 using GAIT-IST and GAIT-IT combined, tested on GAIT-IT.

5.3 Final Remarks

The VGG-19 deep neural network trained on the proposed GAIT-IT dataset was able to perform well on cross-dataset tests, including tests on the DAI2 dataset, which is known to present major difficulties to state-of-the-art methods, because of its poorly segmented silhouettes. Even though the use of the GEI gait representation allows to reduce this problem, the resulting gait features are still significantly affected and not suited to be used in a training stage of current deep learning solutions. A classification performance improvement of more than 22.0%, when testing on DAI2, was achieved on the cross-dataset test when comparing training on the proposed dataset and on the previous largest pathological gait dataset, the GAIT-IST. The best cross-dataset performance was observed when the GAIT-IT and GAIT-IST datasets were combined to train the VGG-19 network, thus providing a significant amount and variety of data for the application of a deep learning solution in the classification of abnormal gait patterns. The GEI and SEI gait representations compact the dynamic information of a gait sequence into a single image, allowing to reduce the problem of poorly segmented silhouettes, as observed for instance in the DAI2 dataset.

Chapter 6

Conclusions and Future Work

This chapter is divided into two sections: i) Achievements, where the steps taken to fulfil the goals of the dissertation are discussed; ii) Future Work, where proposals for further developments of the work reported in the dissertation presented.

6.1 Achievements

This dissertation work has three main objectives: i) to acquire a new dataset simulating a selection of gait pathologies; ii) to develop a simple and non-intrusive classification system of gait videos; iii) to develop a web application allowing to upload a gait input into a web interface and to remotely execute the gait pathology classification system, which is made available as a web service.

The above objectives were successfully accomplished. The main outputs of the work include:

- **GAIT-IT dataset** - A new gait dataset is acquired with sequences from 21 subjects (19 males and 2 females) simulating 4 pathologies, 2 severity levels per pathological gait type, 4 sequences per severity level besides normal gait, with the purpose of being used for gait analysis and classification systems.

- **Gait Classification Web Application** - which is able to accept a gait input in its web interface and to remotely execute the gait pathology classification system, which is made available as a web service. The main goal is to show that it is possible for people who are not specialists in the field to access this classification system and prove their usefulness, in this case for gait pathology classification and gait analysis purposes. This achievement has two components: the classification system and the web interface which makes two objectives accomplished.

6.2 Future Work

Having a large and representative dataset is crucial when the end goal is classification. It is extremely difficult to obtain data for abnormal gait, due to the difficulty of obtaining gait sequences from real patients and also due to the privacy and ethical issues involved. One of the most common issues when data is not enough is overfitting. Thus, the first proposal for future work would be the acquisition of a larger dataset having real patients, instead of simulations.

Although the GAIT-IT dataset is organized by pathology and severity level, the classification is made to differentiate only the pathologies. The second proposal is to develop a system that classifies gait not only as reflecting a certain pathology or corresponding to normal gait, but that would also be able to assign a severity level when each gait pathology is detected.

The third proposal is to run the web service in a cloud virtual machine such as: Google Cloud Platform [1], Amazon Web Services [2] or Microsoft Azure [3], among others. In this way, it is possible to have the memory and GPU needed in the environment and in terms of scalability is an advantage.

The fourth and last proposal is to allow the user of the web application to upload his/her own saved model into the web interface and be able to test all the functionalities of the classification system on it.

[1] https://cloud.google.com/
[2] https://aws.amazon.com/
[3] https://azure.microsoft.com/

Bibliography

[1] Ryo Eguchi and Masaki Takahashi. Accessible calibration of insole force sensors using the wii balance board for kinetic gait analysis. In *2018 IEEE SENSORS*, pages 1–4. IEEE, 2018.

[2] Alvaro Muro-De-La-Herran, Begonya Garcia-Zapirain, and Amaia Mendez-Zorrilla. Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors*, 14(2):3362–3394, 2014.

[3] Mario Nieto-Hidalgo, Francisco Javier Ferrández-Pastor, Rafael J Valdivieso-Sarabia, Jerónimo Mora-Pascual, and Juan Manuel García-Chamizo. Vision based gait analysis for frontal view gait sequences using rgb camera. In *International Conference on Ubiquitous Computing and Ambient Intelligence*, pages 26–37. Springer, 2016.

[4] François Chollet. *Deep Learning with Python*. Manning, November 2017.

[5] Javier Ortells, María Trinidad Herrero-Ezquerro, and Ramón A Mollineda. Vision-based gait impairment analysis for aided diagnosis. *Medical & biological engineering & computing*, 56(9):1553–1564, 2018.

[6] João Firmino. Using Deep Learning for Gait Abnormality Classification. Master's thesis, Instituto Superior Técnico, 2019.

[7] Tanmay Tulsidas Verlekar, Paulo Lobat, and Luís Ducla Soares. Using transfer learning for classification of gait pathologies. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2376–2381. IEEE, 2018.

[8] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye. Tensor completion for estimating missing values in visual data. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):208–220, 2012.

[9] Xiuhui Wang and Wei Qi Yan. Human gait recognition based on frame-by-frame gait energy images and convolutional long short-term memory. *International journal of neural systems*, pages 1950027–1950027, 2019.

[10] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[11] Zhe Cao, Gines Hidalgo Martinez, Tomas Simon, Shih-En Wei, and Yaser A. Sheikh. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[12] Mingjing Yang, Huiru Zheng, Haiying Wang, and Sally McClean. Feature selection and construction for the discrimination of neurodegenerative diseases based on gait analysis. In *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–7. IEEE, 2009.

[13] Manuel Montero-Odasso, Frederico Pieruccini-Faria, Robert Bartha, Sandra E Black, Elizabeth Finger, Morris Freedman, Barry Greenberg, David A Grimes, Robert A Hegele, Christopher Hudson, et al. Motor phenotype in neurodegenerative disorders: gait and balance platform study design protocol for the ontario neurodegenerative research initiative (ondri). *Journal of Alzheimer's Disease*, 59(2):707–721, 2017.

[14] Ugo H Buzzi, Nicholas Stergiou, Max J Kurz, Patricia A Hageman, and Jack Heidel. Nonlinear dynamics indicates aging affects variability during gait. *Clinical biomechanics*, 18(5):435–443, 2003.

[15] Olumide Sofuwa, Alice Nieuwboer, Kaat Desloovere, Anne-Marie Willems, Fabienne Chavret, and Ilse Jonkers. Quantitative gait analysis in parkinson's disease: comparison with a healthy control group. *Archives of physical medicine and rehabilitation*, 86(5):1007–1013, 2005.

[16] Jochen Klucken, Jens Barth, Patrick Kugler, Johannes Schlachetzki, Thore Henze, Franz Marxreiter, Zacharias Kohl, Ralph Steidl, Joachim Hornegger, Bjoern Eskofier, et al. Unbiased and mobile gait analysis detects motor impairment in parkinson's disease. *PloS one*, 8(2):e56956, 2013.

[17] Jens Barth, Jochen Klucken, Patrick Kugler, Thomas Kammerer, Ralph Steidl, Jürgen Winkler, Joachim Hornegger, and Björn Eskofier. Biometric and mobile gait analysis for early diagnosis and therapy monitoring in parkinson's disease. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 868–871. IEEE, 2011.

[18] Uri Givon, Gabriel Zeilig, and Anat Achiron. Gait analysis in multiple sclerosis: characterization of temporal–spatial parameters using gaitrite functional ambulation system. *Gait & posture*, 29(1):138–142, 2009.

[19] MG Benedetti, R Piperno, L Simoncini, P Bonato, A Tonini, and S Giannini. Gait abnormalities in minimally impaired multiple sclerosis patients. *Multiple Sclerosis Journal*, 5(5):363–368, 1999.

[20] Michelle H Cameron and Joanne M Wagner. Gait abnormalities in multiple sclerosis: pathogenesis, evaluation, and advances in treatment. *Current neurology and neuroscience reports*, 11(5):507, 2011.

[21] Chih-Hung Jen, Yu-Ming Lu, Yi-Min Huang, and Bo-Hua Lian. Construct a discriminative system for cardiopathy using electrocardiogram data.

[22] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[23] Rishab Gargeya and Theodore Leng. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, 124(7):962–969, 2017.

[24] SAAN Bolink, H Naisas, R Senden, H Essers, IC Heyligers, K Meijer, and B Grimm. Validity of an inertial measurement unit to assess pelvic orientation angles during gait, sit–stand transfers and step-up transfers: Comparison with an optoelectronic motion capture system. *Medical engineering & physics*, 38(3):225–231, 2016.

[25] Weijun Tao, Tao Liu, Rencheng Zheng, and Hutian Feng. Gait analysis using wearable sensors. *Sensors*, 12(2):2255–2283, 2012.

[26] Abdul Razak, Abdul Hadi, Aladin Zayegh, Rezaul K Begg, and Yufridin Wahab. Foot plantar pressure measurement system: A review. *Sensors*, 12(7):9884–9912, 2012.

[27] Adam M Howell, Toshiki Kobayashi, Heather A Hayes, K Bo Foreman, and Stacy J Morris Bamberg. Kinetic gait analysis using a low-cost insole. *IEEE Transactions on Biomedical Engineering*, 60(12):3284–3290, 2013.

[28] Julian Andres Ramirez-Bautista, Jorge Adalberto Huerta-Ruelas, Silvia Liliana Chaparro-Cárdenas, and Antonio Hernández-Zavala. A review in detection and monitoring gait disorders using in-shoe plantar measurement systems. *IEEE reviews in biomedical engineering*, 10:299–309, 2017.

[29] Ryo Eguchi, Ayanori Yorozu, Takahiko Fukumoto, and Masaki Takahashi. Estimation of vertical ground reaction force using low-cost insole with force plate-free learning from single leg stance and walking. *IEEE journal of biomedical and health informatics*, 24(5):1276–1283, 2019.

[30] Anita Sant'Anna, Nicholas Wickström, Helene Eklund, Roland Zügner, and Roy Tranberg. Assessment of gait symmetry and gait normality using inertial sensors: in-lab and in-situ evaluation. In *International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 239–254. Springer, 2012.

[31] A Ferrari, L Rocchi, J van den Noort, and J Harlaar. Toward the use of wearable inertial sensors to train gait in subjects with movement disorders. In *Converging clinical and engineering research on neurorehabilitation*, pages 937–940. Springer, 2013.

[32] Carlo Frigo and Paolo Crenna. Multichannel semg in clinical gait analysis: a review and state-of-the-art. *Clinical Biomechanics*, 24(3):236–245, 2009.

[33] EC Wentink, VGH Schut, EC Prinsen, Johan S Rietman, and Peter H Veltink. Detection of the onset of gait initiation using kinematic sensors and emg in transfemoral amputees. *Gait & posture*, 39(1):391–396, 2014.

[34] Kate E Webster, Joanne E Wittwer, and Julian A Feller. Validity of the gaitrite® walkway system for the measurement of averaged and individual step parameters of gait. *Gait & posture*, 22(4):317–321, 2005.

[35] Lee Middleton, Alex A Buss, Alex Bazin, and Mark S Nixon. A floor sensor system for gait recognition. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*, pages 171–176. IEEE, 2005.

[36] Moshe Gabel, Ran Gilad-Bachrach, Erin Renshaw, and Assaf Schuster. Full body gait analysis with kinect. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1964–1967. IEEE, 2012.

[37] Richard Zhi-Ling Hu, Adam Hartfiel, James Tung, Adel Fakih, Jesse Hoey, and Pascal Poupart. 3d pose tracking of walker users' lower limb with a structured-light camera on a moving platform. In *CVPR 2011 WORKSHOPS*, pages 29–36. IEEE, 2011.

[38] Alexandros Andre Chaaraoui, José Ramón Padilla-López, and Francisco Flórez-Revuelta. Abnormal gait detection with rgb-d devices using joint motion history features. In *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, volume 7, pages 1–6. IEEE, 2015.

[39] Erik E Stone and Marjorie Skubic. Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing. In *2011 Annual international conference of the IEEE engineering in medicine and biology society*, pages 6491–6494. IEEE, 2011.

[40] Mario Nieto-Hidalgo and Juan Manuel García-Chamizo. Classification of pathologies using a vision based feature extraction. In *International Conference on Ubiquitous Computing and Ambient Intelligence*, pages 265–274. Springer, 2017.

[41] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 2, pages 246–252. IEEE, 1999.

[42] Thierry Bouwmans, Fida El Baf, and Bertrand Vachon. Background modeling using mixture of gaussians for foreground detection-a survey. *Recent patents on computer science*, 1(3):219–237, 2008.

[43] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576, 2014.

[44] Ju Han and Bir Bhanu. Individual recognition using gait energy image. *IEEE transactions on pattern analysis and machine intelligence*, 28(2):316–322, 2005.

[45] Gary Bradski and Adrian Kaehler. Opencv. *Dr. Dobb's journal of software tools*, 3, 2000.

[46] J Deng, A Berg, S Satheesh, H Su, A Khosla, and L Fei-Fei. Imagenet large scale visual recognition competition. *ilsvrc2012*, 2012.

[47] Chao Li, Xin Min, Shouqian Sun, Wenqian Lin, and Zhichuan Tang. Deepgait: A learning deep convolutional representation for view-invariant gait recognition using joint bayesian. *Applied Sciences*, 7(3):210, 2017.

[48] Marcia Hon and Naimul Mefraz Khan. Towards alzheimer's disease classification through transfer learning. In *2017 IEEE International conference on bioinformatics and biomedicine (BIBM)*, pages 1166–1169. IEEE, 2017.

[49] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.

[50] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 568–576. Curran Associates, Inc., 2014.

[51] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497, 2015.

[52] Alex Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

[53] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[54] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318, 2013.

[55] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[56] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.

[57] Shiqi Yu, Haifeng Chen, Edel B. Garcia Reyes, and Norman Poh. Gaitgan: Invariant gait feature extraction using generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[58] Yanyun Wang, Chunfeng Song, Yan Huang, Zhenyu Wang, and Liang Wang. Learning view invariant gait features with two-stream gan. *Neurocomputing*, 339:245–254, 2019.

[59] Mario Nieto-Hidalgo, Francisco Javier Ferrández-Pastor, Rafael J Valdivieso-Sarabia, Jerónimo Mora-Pascual, and Juan Manuel García-Chamizo. Vision based extraction of dynamic gait features focused on feet movement using rgb camera. In *Ambient Intelligence for Health*, pages 155–166. Springer, 2015.

[60] M. Nieto-Hidalgo and J. M. García-Chamizo. Classification of pathologies using a vision based feature extraction. *In International Conference on Ubiquitous Computing and Ambient Intelligenceg*, Philadelphia, USA, Nov. 2017.

[61] Joao Loureiro and Paulo Lobato Correia. Using a skeleton gait energy image for pathological gait classification. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG)*, pages 410–414.

[62] Lorien Y Pratt, Jack Mostow, Candace A Kamm, and Ace A Kamm. Direct transfer of learned information among neural networks. In *Aaai*, volume 91, pages 584–589, 1991.

[63] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–14, 2015.

[64] L Fei-Fei, J Deng, W Dong, R Socher, L.-J. Li, and K Li. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[65] Raghavendra Kotikalapudi and contributors. keras-vis. `https://github.com/raghakot/keras-vis`, 2017.