# Using Deep Learning for Gait Abnormality Classification

João Loureiro
joao.firmino@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2019

**Abstract**

Gait has been the target of many recent research works. Many pathologies affect the normal gait patterns and the analysis of gait indicators can be a useful way of assessing the different gait disorders and their evolution, while also providing valuable information for rehabilitation. Current gait assessment is mostly either based on subjective analysis, that lacks precision and quantitative measurements, or done in dedicated laboratories with complex and expensive equipment setups, that can be impractical. However, recent studies have been made with the objective of obtaining precise gait analysis and classification using simpler systems, composed of a variety of sensors. This work presents a method for classification of abnormal gait types from videos of gait sequences. The proposed system is based on the use of gait representation images, like the Gait Energy Image (GEI) and a novel representation we call Skeleton Gait Energy Image (SGEI), that uses information of the skeleton, obtained with a pose machine. These representations are used as input to a VGG-19 convolutional neural network, that is used to perform classification or to extract a feature vector that is then classified using machine learning methods such as Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA). The system is trained and tested in a newly acquired gait dataset, the GAIT-IST dataset, that simulates four different types of abnormal gait and is also tested with another dataset, to assess its generalization ability.

**Keywords:** Gait analysis, Gait indicators, Gait disorders, Gait recognition, Abnormal gait, Gait classification, Deep learning, Transfer learning

## 1. Introduction

### 1.1. Motivation

Gait can be defined as a set of movements of the limbs that, in a cyclic manner, allows for the locomotion of humans [1]. Certain neurodegenerative disorders can affect this set of movements leading to an impaired or abnormal gait and, although each individual has its own unique gait, some gait indicators, such as step or stride length and joint angles, can be analyzed in order to distinguish normal from abnormal gait and even discriminate which disorder is present. The most common practice of gait assessment consists of a specialist simply observing these indicators. This approach lacks the precision and information that can be obtained with instrumented methods that provide quantitative and objective measures and can have a negative impact on the diagnosis and posterior treatment of the conditions. On the other side, there are solutions from specialized companies that use a wide variety of, mainly vision based, sensors in laboratories and sophisticated equipment that can acquire motion data with high accuracy. However, this uses high cost technology in specialized environments, which is not always available in clinical conditions and often require equipment setup and calibration procedures that can be complicated and time-consuming. As such, there is a need for reliable and accurate ways of assessing gait and its indicators and monitoring them over time, using simple device setups, for a quick, yet precise, gait analysis.

### 1.2. Contributions

Besides presenting an implementation of a system that is able to perform classification between normal gait and four different types of abnormal gait using the widely used Gait Energy Image (GEI) representation, there are two main contributions of this work. The first one is the acquisition of a new gait database, where 10 subjects simulate the four types of gait with different types of severity, which is important given the lack of datasets regarding pathological gait. The second main contribution is the usage of pose data obtained from a pose machine algorithm to compute a new gait representation we call the Skeleton Gait Energy Image (SGEI), which is not affected by the body shape or clothes being used, in contrast with silhouette based representations, and which outperforms the usage of GEIs on the same system.

### 1.3. State of the Art

To tackle the issue of gait analysis, there has been a rise of new technologies and methods aimed at solving the problem of obtaining objective measures of a varied selection of gait indicators and classifying gait between normal and abnormal and even between the different possible pathologies. These methods use a variety of sensors, that can be wearable or not, to measure the different gait indicators.

Wearable sensors are attached to parts of the body in order to measure gait kinetics and kinematics. Examples include electrogoniometers, textile sensors, ultrasonic, electromagnetic tracking systems, inertial sensors, force and pressure sensors and electromyographys. These methods, while being precise and not restricting their usage to a controlled environment, can be considered intrusive, as they may cause discomfort to the patients or even affect their gait, producing inaccurate measurements.

The non wearable sensors include floor sensors, whose main limitation is their length, which restricts how many steps can be taken and measured, and vision based sensors. The last ones are used in the specialized laboratories mentioned before, but, lately, there has also been an increased number of techniques using only a couple, or even just one camera. For instance, systems using less costly devices, such as the Kinect and basic RGB cameras are capable of obtaining precise results on gait indicators and perform an accurate classification of abnormal gait. These vision-based methods often rely on the use of algorithms that take gait video sequences as input. Depending on whether the gait video features are computed with algorithms that make specified measurements on the image or are extracted automatically using more sophisticated deep learning techniques, we can divide video base methods into the ones that use handcrafted features and learned features, respectively.

Handcrafted feature methods can be further classified into shape-based or model-based. Shape-base techniques analyse the shape of the subject, often times in the form of silhouettes extracted from the videos after background subtraction, to compute gait indicators. An example is the work in [2], who developed an algorithm to estimate Heel Strike (HS) and Toe Off (TO) events, in order to calculate indicators and classify between normal and abnormal gait. To tackle the self-occlusion problem accentuated in patients whose impairments reduce the step size significantly, a new way to determine step related indicators is proposed in [3], based on the detection of foot-flat (FF) events, where the foot is in full contact with the ground. When using model-based techniques, after acquiring images from one or more cameras, a model is matched with the sequence of images where gait is observed, usually in the form of a human skeleton model or by segmenting the different body parts into independent shapes. After having this information about segment and joint positions it is easy to obtain indicators such as joint angles, stride length etc. In [4] a method is proposed for gait analysis based on the extraction of joint positions and angles from 2D images captured from sagittal and frontal planes. The second part of the study in [2], mentioned before, presents a similar algorithm to obtain an approximated full body 2D skeleton model from the silhouettes of the gait sequences captured with a single camera from the sagittal plane.

The methods based on learned features make use of deep learning algorithms to perform automatic feature extraction and classification. These techniques often require some pre-processing of the images beyond background subtraction. An often seen example of this is the computation of the Gait Energy Image (GEI) [5], which represents the information of full gait cycles in a compact way. These images can then be used as input into a Convolutional Neural Network (CNN) that performs feature extraction [6, 7] and classification can be done using the same network or recurring to more conventional machine learning techniques, that operate on the extracted features. Other deep learning methods involve the use of other types of gait representation, such as skeleton data [8, 9] and optical flow images [10] as well as different types of deep learning architectures, such as Long Short-Term Memory (LSTM) networks [11].

### 1.4. Outline

This report is organized as follows. In Chapter **2**, the acquisition of the GAIT-IST dataset is described. Chapter **3** presents the proposed solution for a system that is able to classify between normal and different types of abnormal gait. In Chapter **4**, the results of the proposed system are presented and discussed. Final conclusions are drawn in Chapter **5** and future work possibilities are presented.

## 2. GAIT-IST Dataset Acquisition

Regarding gait databases, they can belong to two different categories, whether the objective is to perform gait recognition or pathological gait analysis. Regarding gait recognition databases, there is a significant amount of databases, as seen in [12], as their acquisition only requires subjects to walk normally. However, due to the difficulty of obtaining gait sequences from real patients, and also due to privacy and ethical issues, there are not many gait databases available dedicated to the study of pathological gait and the ones that exist are composed of simulated gait impairments and have a reduced number of samples, which is not optimal. The most notable pathological gait datasets are the following:

- The INIT Gait Dataset [13], with samples from 10 subjects, performing gait impairments in the arms and legs;
- The DAI Gait Dataset [14], with samples from 5 subjects that perform normal and random abnormal gait simulations;
- The DAI Gait Dataset 2 (DAI2) [15] that includes sequences of 5 subjects performing normal and four types of abnormal gait simulations.

For the motives mentioned before, it was decided to make a new dataset acquisition, we called GAIT-IST, to be used in this work, in addition to the existing ones.

When deciding about the types of pathological gait to be analyzed and to be featured in the dataset, there is a decision to be made between considering the disorder that causes gait impairment or the type of abnormal gait. For this work, the latter was used and the types of gait chosen, besides normal gait, were the Diplegic, Hemiplegic, Neuropathic and Parkinsonian gaits [1, 16]. This decision was made because each disorder that affects gait, such as Parkinson's, Multiple Sclerosis and Cerebral Palsy, is often composed of a combination of different types of abnormal gait. In turn, each one of these types of gait are defined by a set of symptoms and movements that can be easily simulated. In addition, these four types of pathological gait are also the ones simulated in the DAI2 dataset, making it possible to perform some cross-database tests on the created model.

There were in total 10 subjects (8 males and 2 females) in the age range of 20-50 years old. The participants were voluntary and signed a consent form allowing the use of their images for research purposes. Each one performed 4 sequences for each of the 4 pathological gait types (2 per severity level) and 2 sequences for the normal gait, for a total of 180 sequences. In each one, the subject walks, parallel to the camera plane, one time from left to right and another from right to left, with a minimum of 2 complete gait cycles captured in each direction, which varied between gait types and subjects. For the situation of the freezing of gait, a gait cycle is considered to be the interval involving the step immediately before and after the episode.

The image capture was done using a cellphone camera with a resolution of 720p, fixed on a tripod at a height of about 1,5 meters and at a distance of about 4 meters from the target. The illumination was fairly constant, mostly composed of the artificial light of the room. The background was constant and uniform, consisting of a white wall and gray ground, with some shadows and reflections present.

## 3. Proposed Gait Analysis System

The objective of the proposed gait analysis system consists in, starting with videos of subjects depicting normal gait and abnormal gait representing different pathologies, being able to classify the type of gait into one of those categories. A database of gait sequences will also be constructed to be used and help performing this task.

The development of an image classification system is often divided into different steps, as shown in the block diagram of Figure 1, which represents a general architecture of the system to be developed. The first step is the image pre-processing which firstly aims at obtaining a set of images that can be used as input in the next stages. In this work, two types of input are considered, and so this step will be different for each one. The resulting images are then used as input to a feature extraction module, although they can be used in the classification module directly. Finally, the resulting feature vector is used to perform classification. This chapter starts by describing the acquisition of the GAIT-IST dataset and the implementation of the different steps of the proposed system.

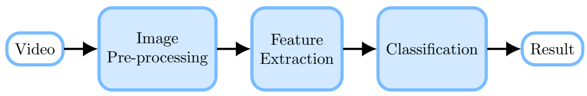


Figure 1: General block diagram of the proposed gait analysis system.

### 3.1. Pre-processing

In the case of the GEIs, the pre-processing stage starts with background subtraction of the original images in order to obtain silhouette images. Then these silhouettes are cropped, resized to a normalized height and the width is padded with zeros in order to obtain 224x224 sized images, which is the input size of the VGG-19. Then the gait cycles are detected, by analysing the variation of the width of the feet of the silhouettes along the frames and the sets of images, $I_t(x, y)$, corresponding to complete gait cycles are averaged according to (1) in order to obtain the final GEIs.

$$GEI(x,y) = \frac{1}{a+N} \sum_{t=a}^{N} I_t(x,y) \tag{1}$$

The SGEIs, the process starts with the use of the OpenPose [17] algorithm, to obtain a set of coordinates corresponding to certain joints of the body. This coordinates are then used to draw an image of the skeleton of the subject in each frame. Finally the images go to a similar resizing and averaging process as described before in the case of the GEI, in order to obtain the final SGEIs.

Examples of obtained GEI and SGEI images are shown in Figure 2.

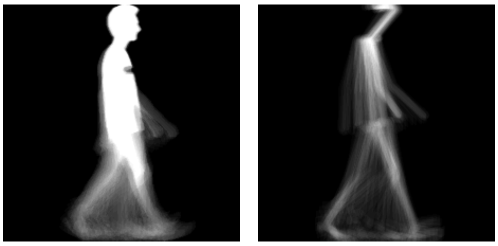

Figure 2: Examples of an obtained GEI (left) and SGEI (right).

### 3.2. CNN Feature extraction

After performing image pre-processing, the next module is the feature extraction, performed by a CNN in order to obtain a feature vector for every image input, which will later be used for classification.

The chosen architecture for this task was the VGG-19, which is a very deep CNN which is easy to implement and train, as it converges quickly, and is generally the best solution for feature extraction. The VGG-19 was chosen over the VGG-16 because it is a deeper network and will have a better performance overall [6]. Its input is a 224x224 image that passes through several blocks of convolutional layers composed of very small sized filters ($3 \times 3$). The stride and padding are both fixed to 1 pixel. Spatial pooling after each convolutional block is done with max-pooling over a $2 \times 2$ pixel window and a stride of 2. The end of the network is composed of three fully connected layers, in which the first two have 4096 channels and the third one, that performs the classification, has originally 1000 channels, and is followed by a soft-max layer. The third fully connected layer is also changed to one with only 5 channels, corresponding to the the 5 classes of gait in our classification problem.

As the available datasets for training are very small, it would be impossible to use them to completely retrain the network, as it would lead to problems such as overfitting. To overcome this issue, transfer learning is performed using a version of the VGG-19 pre-trained using the ImageNet database [18]. The practice of transfer learning is a widely used technique in machine learning, in which a model trained with a given dataset to perform some classification task is partially retrained with a second, usually smaller, dataset to perform a related task. In this case, as the first convolutional layers are responsible for detecting basic generic shapes in the images, they can be frozen during training, maintaining their weights. Succeeding layers are then fine-tuned using the GAIT-IST dataset in order for them to learn features that are more specific to our problem.

In a first stage the VGG-19 was fine-tuned and used to perform classification directly on the GEIs and SGEIs computed from the GAIT-IST dataset. However due to the low amount of data available and the high probability of overfitting when training the network, another classification method was considered. Inspired by the work in [6], the CNN was used as a feature extractor, taking the first fully connected layer as a feature vector for each GEI and SGEI. These vectors, which have a size of 4096, are then used for classification.

### 3.3. Classification

The classification step is done in two different ways. The first option is direct classification using the VGG-19 with the images obtained after preprocessing as input. The second option is, after obtaining feature vectors from the feature extraction module, performing dimensionality reduction using Principal Component Analysis (PCA) and classification with either Linear Discriminant Analysis (LDA) or Support Vector Machine (SVM).

The Matlab Classification Learner app allows us to provide a set of data with labels and apply several types of supervised machine learning methods automatically, in order to evaluate what is the best one to be used. After using some of the extracted feature vectors as input for training all the methods available, the LDA and SVM were observed to be the most promising ones, obtaining a higher training accuracy. The models of these classifiers were then saved and used to perform classification.

In both cases, PCA was applied to the feature vectors for dimensionality reduction, such that the resulting components explained 95% of the original variance.

Regarding the SVM, several models with different kernel functions were tested, namely Linear, Quadratic, Cubic and Gaussian kernels. It was later observed that the Linear kernel function produced the most consistent and generally best results.

In the case of LDA, both full and diagonal covariance structures were tested, of which the regular LDA with full covariance structure performed better.

## 4. Results

In order to test the proposed system in the task of detecting and classifying different types of abnormal gait, results were first obtained using the GAIT-IST. Cross-validation tests were conducted, using the GAIT-IST dataset for training, first using the GEIs and then the SGEIs. After this, to assess the generalization capability of the system, the model trained with the GEIs was tested using the DAI2 dataset [15], which is composed of gait sequences of 5 individuals simulating the same types of gait as our dataset, allowing us to use one dataset for training and the other for testing. The computation of skeletons using the DAI2 dataset was not possible, as the original images were not available, only the silhouettes. This chapter presents results with the different methods proposed.

### 4.1. CNN Hyper-parameter Tuning

The CNN network chosen for feature extraction and classification was the VGG-19 pre-trained with the ImageNet database [18]. The batch size was fixed to 16, based on the memory available on the graphic card used. A higher value could help the model to converge more quickly, as more data would be used in each training epoch to update the network, however it would lead to problems when trying to allocate too much memory. The number of epochs was set to 50, as it was observed that the training accuracy converged close to 100% quickly and further training would lead to overfitting. Also an early stopping strategy was adopted, in which the training always runs for the set number of epochs, but the saved model is the first one to achieve the highest validation accuracy during training.

The chosen loss function was the categorical cross entropy, which is the default loss function to use for multi-class and single label classification, which means that each data instance belongs to only one of the available classes.

After briefly testing some optimizers and observing the accuracies during training, the chosen one was the Stochastic Gradient Descent (SGD), which is a version of Gradient Descent (GD) in which only part of the dataset is used to compute an estimate of the gradient, which achieves faster iterations despite converging more slowly. The learning rate was set to 0.0002.

In order to improve the convergence rate, momentum was introduced, namely a modification called Nesterov Momentum. The use of momentum improves the convergence speed by giving more importance to the directions that decrease the loss function the most. After some experimentation the momentum parameter was set to 0.9.

With these parameters, an example of the evolution of the training and validation accuracy during training is shown in Figure 3.

### 4.2. CNN Fine-tuning and Classification

After the parameters of the CNN were selected as defined in the previous section, another parameter was tested while fine-tuning, which was what layers are actually being retrained with the GAIT-IST dataset. As the convolutional layers are responsible for learning shapes and patterns in the images, it can be advantageous to also fine-tune some of them using data related to the task being performed. Having this in mind, a set of tests were performed to evaluate which configuration of fine-tuning was best, in terms of blocks of layers to be retrained. In each of the tests, like before, 10-fold cross-validation was done, and for each fold a confusion matrix of the classification results for the validation set was obtained. Then a global confusion matrix is calculated, with the mean results from all the validation sets. This process was done first using the GEIs and then using the SGEIs as input. The global confusion matrices for the best experiments are presented in Tables 1 and 2, for the usage of GEIs and GEIs respectively.

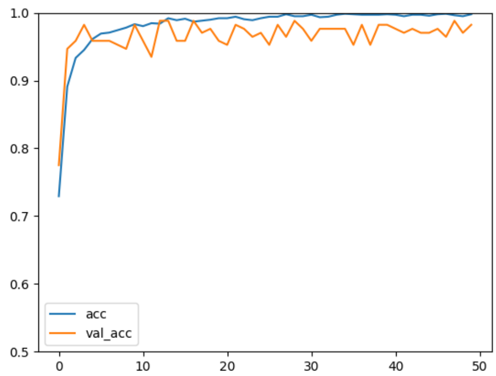


Figure 3: Training (blue) and validation (orange) accuracies (left) and losses (right) during training of one fold of cross-validation with learning rate = 0.0002 and using Nesterov Momentum.

| True Class \ Predicted Class | Diplegic | Hemiplegic | Neuropathic | Parkinsonian | Normal |
|---|---|---|---|---|---|
| Diplegic | 0.95 | 0.02 | 0.00 | 0.03 | 0.00 |
| Hemiplegic | 0.04 | 0.90 | 0.04 | 0.00 | 0.03 |
| Neuropathic | 0.00 | 0.07 | 0.93 | 0.00 | 0.00 |
| Parkinsonian | 0.00 | 0.00 | 0.00 | 0.99 | 0.00 |
| Normal | 0.00 | 0.04 | 0.01 | 0.00 | 0.95 |

Table 1: Confusion matrix with the mean validation accuracy of the CNN GEI classification across all cross-validation folds, for the configuration where the VGG-19 is trained starting at the third convolution block. The overall accuracy was 94.5%.

Regarding the confusion matrices, there are two interesting things to note. First, the Parkinsonian gait is the one most accurately classified, most noticeable when using GEIs. This is probably due to the fact that the GEIs of this class are the most dis-

| | Predicted Class | | | | |
|---|---|---|---|---|---|
| True Class | Diplegic | Hemiplegic | Neuropathic | Parkinsonian | Normal |
| Diplegic | 0.93 | 0.00 | 0.00 | 0.07 | 0.00 |
| Hemiplegic | 0.01 | 0.98 | 0.01 | 0.00 | 0.01 |
| Neuropathic | 0.00 | 0.00 | 0.98 | 0.00 | 0.02 |
| Parkinsonian | 0.01 | 0.00 | 0.00 | 0.99 | 0.00 |
| Normal | 0.00 | 0.01 | 0.01 | 0.00 | 0.98 |

Table 2: Confusion matrix with the mean validation accuracy of the CNN SGEI classification across all cross-validation folds, for the configuration where the VGG-19 is trained starting at the second convolution block. The overall accuracy was 97.4%.

tinguishable from the others due to the posture and very small step size. It may also be because, as the steps are so small, more gait cycles were captured and thus more GEIs were able to be computed in each sequence, which means that there is more data belonging to this class in the dataset and the model is able to learn it better than the others. Also noticeable is the fact that the majority of the times that each gait type is misclassified, the wrongly predicted class is Hemiplegic gait. This occurrence is much less visible when using the SGEI, which suggests that this representation is better suited for this classification task. However, in this case, there is a consistent misclassification of Diplegic gait as Parkinsonian. This may be because of the similar posture that the two abnormal gaits present.

Regarding the global GEI classification accuracy, it can be noted that fine-tuning more convolution blocks generally leads to better results, as the best accuracies are obtained when fine-tuning from block 3 and 2 for GEI and SGEI respectively. Also noticeable are the general better results with the SGEIs experiments, which suggests that it is a more suited representation for this classification task.

### 4.3. CNN Feature Extraction and Classification

In order to avoid some overfitting due to working with a limited amount of data, we can use a classifier trained separately from the CNN, which is now only used for feature extraction.

Using the trained VGG-19 models obtained in the previous section, feature extraction was performed, on every GEI and SGEI, in the form of 4096-sized feature vectors obtained from the first fully connected layer. Like before, 10-fold cross-validation was performed for each fine-tuning experiment. PCA was applied to the feature vectors, for dimensionality reduction, keeping only the most representative components, and they were then used to train LDA and SVM models, to be used as classifiers. Each validation fold always corresponded to the GEIs and SGEIs used to train the CNN model from which those features were extracted, for performance comparison purposes.

The LDA achieved a slightly better overall accuracy and the confusion matrices are shown in Tables 3 and 4, for GEIs and SEGIs respectively. The results are very similar and it does not seem to exist any improvement when comparing to the ones using the CNN classification, obtained from the last fully connect layer of the network. The same observations can be made about the high accuracy on the classification of the Parkinsonian gait, most of the misclassifications corresponding to a wrong prediction of Hemiplegic gait when using GEIs and the confusion between Diplegic and Parkinsonian gaits when using SGEIs. Also like before, the configurations that achieve the highest overall accuracy are the ones where fine-tuning is done starting at the third and second convolution block, for GEIs and SGEIs respectively. Finally, once more, SGEIs produce better results than GEIs.

| | Predicted Class | | | | |
|---|---|---|---|---|---|
| True Class | Diplegic | Hemiplegic | Neuropathic | Parkinsonian | Normal |
| Diplegic | 0.79 | 0.17 | 0.00 | 0.04 | 0.00 |
| Hemiplegic | 0.0 | 0.91 | 0.05 | 0.00 | 0.04 |
| Neuropathic | 0.00 | 0.11 | 0.89 | 0.00 | 0.00 |
| Parkinsonian | 0.01 | 0.02 | 0.00 | 0.97 | 0.00 |
| Normal | 0.00 | 0.05 | 0.05 | 0.00 | 0.90 |

Table 3: Confusion matrix with the mean validation accuracy of the LDA classification using GEIs across all cross-validation folds, for the configuration where the VGG-19, used for feature extraction, is trained starting at the third convolution block. The overall accuracy was 89.6%.

### 4.4. Cross-database Scenario

With the objective of testing the ability of generalization of the proposed system in the classification and feature extraction tasks, tests were made using a different gait database, namely the DAI Gait Dataset 2 (DAI2). This dataset is similar to the GAIT-IST, as it is composed of simulations of the same gait types, performed by 5 subjects. This makes it possible to use it for testing the models already trained with the GAIT-IST dataset. As the original images from the DAI2 dataset were not available, it was not possible to make a cross-database test using the SGEI representation.

The first experiment consisted in doing CNN classification of the GEIs obtained from the DAI2 dataset using the VGG-19 fine-tuned with the GAIT-IST dataset. As previously noted, the con-

| | Diplegic | Hemiplegic | Neuropathic | Parkinsonian | Normal |
|---|---|---|---|---|---|
| Diplegic | 0.91 | 0.01 | 0.00 | 0.08 | 0.00 |
| Hemiplegic | 0.0 | 1.00 | 0.00 | 0.00 | 0.00 |
| Neuropathic | 0.00 | 0.02 | 0.96 | 0.00 | 0.02 |
| Parkinsonian | 0.01 | 0.00 | 0.00 | 0.99 | 0.00 |
| Normal | 0.00 | 0.02 | 0.01 | 0.00 | 0.97 |

Table 4: Confusion matrix with the mean validation accuracy of the LDA classification using SGEIs across all cross-validation folds, for the configuration where the VGG-19, used for feature extraction, is trained starting at the second convolution block. The overall accuracy was 96.4%.

figuration of fine-tuning the network starting at the third convolution block produced the best results for GEIs, so that was the model chosen for the realization of the cross-database testing. The DAI2 dataset was used to test each of the 10 models trained during the previous 10-fold cross-validation training and the mean confusion matrix and accuracy were computed, as seen in Table 5.

As expected, the results were significantly worse than when training and testing on samples of the same dataset, obtaining an overall accuracy of only 43.3%. In this case many factors can contribute to the decrease in accuracy. More specifically, the two datasets may have simulated each gait type emphasizing different characteristics of the pathologies and the GEIs may have been computed in slightly different ways. Besides, the silhouettes available from the DAI2 dataset have some segmentation errors, mainly in the form of holes in the silhouettes, which makes the classification task even harder. Still the same observations as before can be made. Again, the Parkinsonian gait is relatively well classified, second only to the Hemiplegic gait, which may be because of the, now much stronger, tendency of the network to classify every gait type with this category.

In order to improve this result, the alternative method of performing feature extraction using the CNN and then classification using PCA and LDA was tested. The feature vectors were extracted for all GEIs in the DAI2 dataset, using a model fine-tuned with the GAIT-IST dataset. Then the feature vectors were grouped by subject and 5-fold cross-validation was performed using the LDA models, in a way that each fold contains all the features that correspond to each subject. In this experiment an overall accuracy of 76.7% was obtained. The confusion matrix is presented on Table 6. There is a significant improvement when comparing to the CNN classification, which shows that using the extracted features in a different independently trained classifier, improves the generalization ability of the system across different databases.

| | Diplegic | Hemiplegic | Neuropathic | Parkinsonian | Normal |
|---|---|---|---|---|---|
| Diplegic | 0.09 | 0.50 | 0.18 | 0.23 | 0.00 |
| Hemiplegic | 0.0 | 0.97 | 0.02 | 0.00 | 0.00 |
| Neuropathic | 0.00 | 0.86 | 0.09 | 0.01 | 0.04 |
| Parkinsonian | 0.17 | 0.19 | 0.04 | 0.60 | 0.00 |
| Normal | 0.00 | 0.91 | 0.01 | 0.00 | 0.08 |

Table 5: Confusion matrix depicting the mean validation accuracy of the CNN classification across all cross-validation folds, when testing with the DAI2 dataset. The overall accuracy was 43.3%.

| | Diplegic | Hemiplegic | Neuropathic | Parkinsonian | Normal |
|---|---|---|---|---|---|
| Diplegic | 0.53 | 0.16 | 0.01 | 0.30 | 0.00 |
| Hemiplegic | 0.13 | 0.72 | 0.15 | 0.00 | 0.00 |
| Neuropathic | 0.09 | 0.22 | 0.69 | 0.00 | 0.00 |
| Parkinsonian | 0.08 | 0.00 | 0.00 | 0.92 | 0.00 |
| Normal | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |

Table 6: Confusion matrix depicting the mean validation accuracy of the LDA classification across all cross-validation folds, when testing with the DAI2 dataset. The overall accuracy was 76.7%.

### 4.5. CNN Deep Feature Visualisation

Neural networks are currently not very transparent and interpretable, as we do not fully know why and how models make decisions or learn features. As an application based on deep learning can raise questions about its reliability if there is no explanation given by the model, besides the final result, it is important to try to understand how these methods work. The Keras Visualisation Toolkit includes several techniques to give some insight about what is happening inside a CNN during classification. The two visualization techniques tested were the Saliency Maps and the Class Activation Maps, or grad-CAM that result in a visual representation of the attention given by the network to the different areas of the input images. An example of the application of these algorithms on a SGEI corresponding diplegic gait to is presented in Figure 4.

One of the conclusions that was made right away was that the most important pixels of the image

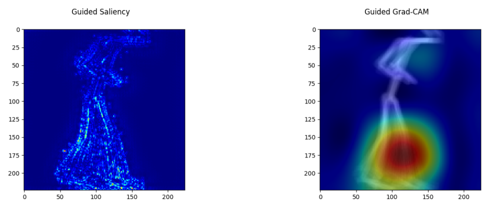


Figure 4: Examples of the application of the saliency map (left) and grad-CAM (right) algorithms on a SGEI corresponding to diplegic gait.

are the ones on the border of the GEI, and on places that correspond to parts of the body that have a larger amplitude of movement. This was expected, as the inside of GEIs is mostly the same across all classes, specially in the upper body region, where there is almost no movement. This once again suggests that the SGEI is a better representation as there is less information that is not used in the image. More evidence to this is the fact that the saliency maps on the SGEIs show a more distributed attention across the entire SGEI, meaning that more information is used by the CNN.

In every type of gait, except Parkinsonian, attention is given to the legs and arms. In the case of Parkinsonian gait this does not happen, probably due to the fact that there is not much movement on the legs and the features learned are more related to the posture and position of the hands.

## 5. Conclusions and Future Work

A new gait dataset, GAIT-IST, is acquired, with sequences from 10 subjects simulating four different types of abnormal gait, to be used in gait analysis and classification systems.

A system is implemented with the objective of performing feature extraction and classification on datasets composed of different types of abnormal gait. The system uses the widely used GEI representation and the new proposed representation SGEI, which performs better in the proposed system. The system is also tested using a different dataset, the DAI2, in order to evaluate its generalization capability.

The results are analysed and discussed, in order to explain the classification output of the system. Furthermore, a deep feature visualisation toolkit is used to create images representing attention maps, indicating which parts of the input images are the focus of the CNN for feature extraction. This helps improving our understanding of the functioning of the network.

There are three main propositions for future developments for this work.

Firstly, the importance of having large and representative datasets is of the utmost importance when dealing with classification tasks. The lack of data in the topic of pathological gait leads to problems when testing this kind of system, for instance, overfitting. Thus one important task is the acquisition of bigger datasets representing abnormal gait.

As the results show, the SGEI representation produces best classification accuracies than the GEI. The next step would be a combination of the two representations. This could be done at different levels. For instance a new image representation combining both of them. Other option would be the combination of the scores or classification output given by the system, regarding the two representations.

The last suggestion is the use of a recurrent neural network such as the LSTM to perform the classification task. Although the GEI and SGEI representations already include some temporal information, the use of an LSTM could further emphasize the time aspect of the gait sequences. Input to the LSTM could be in the form of skeleton coordinates data or feature vectors extracted from the silhouettes composing each gait cycle.

## References


[1] Chris Kirtley. *Clinical Gait Analysis: Theory and Practice*. Churchill Livingstone, $1^{st}$ edition, 2006. ISBN:0-4431-0009-8.

[2] M Nieto-Hidalgo, F J Ferrández-Pastor, R J Valdivieso-Sarabia, J Mora-Pascual, and J M García-Chamizo. A vision based proposal for classification of normal and abnormal gait using rgb camera. *Journal of Biomedical Informatics*, 63:82–89, October 2016. doi:10.1016/j.jbi.2016.08.003.

[3] Tanmay Y. Verlekar, Luís D. Soares, and Paulo L. Correia. Automatic classification of gait impairments using a markerless 2d video-based system. *Sensors*, 18(9):2743, August 2018. doi:10.3390/s18092743.

[4] A. Leu, D. Ristić-Durrant, and A. Gräser. A robust markerless vision-based human gait analysis system. In *2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, Timisoara, Romania, May 2011. IEEE. doi:10.1109/SACI.2011.5873039.

[5] Ju Han and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006.

[6] 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)*, Madrid, Spain, December 2018. IEEE. doi:10.1109/BIBM.2018.8621302.

[7] C Li, X Min, S Sun, W Lin, and Z Tang. Deepgait: A learning deep convolutional representation for view-invariant gait recognition using joint bayesian. *Appl. Sci.*, 7(3):210, February 2017. doi:10.3390/app7030210.

[8] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *CVPR*, 2016.

[9] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

[10] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. *Advances in neural information processing systems*, 1, June 2014.

[11] Rijun Liao, Chunshui Cao, Edel B. Garcia, Shiqi Yu, and Yongzhen Huang. Pose-based temporal-spatial network (ptsn) for gait recognition with carrying and clothing variations. In *CCBR 2017: Biometric Recognition*, Beijing, China, October 2017. Springer, Cham. doi:10.1007/978-3-319-69923-3_51.

[12] Tanmay Tulsidas Verlekar. Gait Analysis in Unconstrained Environments. 2019.

[13] J Ortells, M T Herrero-Ezquerro, and R A Mollineda. Vision-based gait impairment analysis for aided diagnosis. *Medical & Biological Engineering & Computing*, 56(9):1553–1564, September 2018. doi:10.1007/s11517-018-1795-2.

[14] M Nieto-Hidalgo, F J Ferrández-Pastor, R J Valdivieso-Sarabia, J Mora-Pascual, and J M García-Chamizo. Vision based extraction of dynamic gait features focused on feet movement using rgb camera. In *AmIHEALTH 2015: Ambient Intelligence for Health*, pages 155–166, Puerto Varas, Chile, December 2015. Springer.

[15] 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*, Philadelphia, USA, November 2017. Springer, Cham. doi:10.1007/978-3-319-67585-5_28.

[16] Jacquelin Perry. *Gait Analysis: Normal and Pathological Function.* SLACK Incorporated, 1$^{st}$ edition, 1992. ISBN:978-1-55642-192-1.

[17] Openpose. *GitHub*, Accessed in October 2019. https://github.com/CMU-Perceptual-Computing-Lab/openpose.

[18] Imagenet large scale visual recognition competition (ilsvrc). *ImageNet*, Accessed in March 2019. http://image-net.org/challenges/LSVRC/.