Classification of Gait Patterns in Spastic Diplegia

Sofia Martins Camacho Lança
sofia.lanca@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

June 2018

Abstract

The best clinicians excel in their ability to discern the correct diagnosis in different cases. This skill requires an extensive knowledge base, keen interviewing and examination skills and the ability to synthesize correctly all the available information. This level of expertise varies among clinicians and even the most expert can sometimes fail. This is the main justification to implement Clinical Diagnosis Support Systems: reduce the rate of diagnosis error to medicine. This thesis proposes to study the application of these systems to gait pattern analysis of cerebral palsy, namely spastic diplegia. The choice is easily justified by the statistics: cerebral palsy is the leading cause of disability in children. The main idea is to build a multi class supervised model based on the 5 types of spastic diplegia: crouch gait, true equinus, apparent equinus, jump gait and asymmetrical gait. The procedure is mainly based on a careful data processing where scaling and normalization processes are followed by dimensionality reduction and feature extraction. The partition into train and test had to be carefully tackled in order to deal with an unbalanced dataset.

Keywords: Spastic Diplegia, Clinical Diagnosis Support Systems, Gait Cycles, Data Processing,Supervised Classifier

1. Introduction

The best clinicians excel in their ability to discern the correct diagnosis in different cases. This skill requires an extensive knowledge base, keen interviewing and examination skills and the ability to synthesize correctly all the available information. This level of expertise varies among clinicians and even the most expert can sometimes fail. There is also a growing appreciation that diagnostic errors can be just made just as easily in simple cases as in the most complex. Given this dilemma and the fact that diagnostic error rates are not trivial, clinicians are well advised to explore tools that can help them establish correct diagnosis.

Clinical decision support systems (CDSS) are computerized health information systems that use case-based reasoning to provide clinicians assistance in making a diagnosis, selecting appropriate therapy or in making other clinical decisions. CDSS can direct physicians to correct diagnosis and have the potential to reduce the rate of diagnostic error in medicine.

It has been defined as an "active knowledge systems, which use two or more items of patient data to generate case-specific advice." This means that a database with characteristics of individual patients and software algorithms generate patients specific information in the form of assessments or recommendations. It aims to decrease error rates by influencing physician behaviour, improving clinical therapy and improving patient outcome [4].

This means that a CDSS is simply a decision support system that is focused on using knowledge management in such a way so as to achieve clinical advice for patient care based on multiple items of patient data. It has the potential to help clinicians chose the best decision possible.

The aim of the present thesis is to build such a system applied to spastic diplegia patients, so that gait cycles evaluated and the division between the groups, that are about to be defined, are automatically generated and the correct treatment is applied.

An example of a typical CDDS is in the image below.

1.1. State of Art

Cerebral palsy (CP) is the leading cause of disability in children with prevalence rates being in the region of 2 to 2.5 per 1000 live births [10]. CP is a non-progressive medical condition resulting from damage to the developing brain, which depending on the location, can result in difficulties with movement, spasticity, cognition, communication and behaviour [1].

Cerebral palsy is divided into subtypes (spastic, dyskinetic and ataxic), based upon the predominant motor impairment. Spastic cerebral palsy is

also subdivided based upon the number of limbs affected: hemiplegia affects one side of the body, diplegia affects the legs only, while quadriplegia impacts on all four limbs.

Dyskinetic cerebral palsy (athetoid and dystonic) is associated with fluctuating or rigid muscle tone, while ataxic conditions are associated with problems like coordination, muscle tone and balance.

As there is no cure for CP, traditionally interventions are focused on the improvement of physical functioning. Recent advance in medical care, however, have resulted in individuals with CP living longer. Consequently, the need to support has been recognized [2].

Spastic diplegia patients are the study group of the present dissertation and the aim is to distinguish between the subdivision groups that are defined in the section below and serve as a guideline to the correct surgical procedure. It should be again mentioned that those clinical advances can conduct to a lower mortality rate and a better quality of life.

1.2. Spastic Diplegia Types

There is no consensus among the division of spastic diplegia groups but some investigation have been made in order to try to define them.

In 1986, Rang, Silver and De La Garza [6] described a number of gait patterns in spastic diplegia and classified them on a purely observational basis, related to spaticity or contracture of muscles which work in the sagittal plane. They observed associations between contractures of the psoas and lumbar lordosis, of the abductors and scissoring, of the hamstrings and knee flexion, of rectus femoris and of gastrocsoleus and tip-toe gait and linked those patterns to manage the treatment.

In 1993, Sutherland and Davids [9] described four typical abnormalities of gait affecting the knee in children with spastic diplegia, namely jump knee, crouch knee, stiff knee and recurvatum knee. The patterns resort to sagittal plane kinematics data but no quantitative information was given the quantitative assessment of patients.

Following the patterns identified in the two studies mentioned, Miller et al. [5] have further elaborated on the sagittal gait patterns originally described by Rang et al. [6]. The patterns were described as jump, crouch, equinus, jump plus equinus and recurvaturum plus equinus.

In the study by Hullin et al., with a subjective approach that consider pathological angular joint movements the patients were divided in five different groups. [11]

Using visual inspection, kinetic and EMG data, [7], came to establish four distinct groups.

Although there were some studies made before, the one that will be used for gold standard in the present dissertation was the one by [3].

Five main patterns were identified based on sagittal plane kinematics and the definition and characteristics of each one is given below.

- True Equinus
- Jump Gait
- Apparent Equinus
- Crouch Gait
- Asymmetrical Gait

An additional mild pattern was also identified but it will not be analysed due to lack of data and objective quantification.

1.3. Objectives

As mentioned in the previous sections spastic diplegia patterns have been largely studied both subjectively and objectively. There is no agreement regarding the number of groups but [8] will be considered the gold standard because the specialists classification provided, that will be used in this dissertation, follow this approach.

This study will continue the one in the previous dissertation where it was possible to perfectly distinguish spastic diplegia patients from healthy individuals and an unsupervised classifier was designed with no access to clinical classification. Presently there is medical information regarding the groups classification and it will be used to build a supervised classifier to distinguish between the four groups defined: true equinus, jump gait, apparent equinus and crouch gait.

The same data will be used: 26 children with cerebral palsy and 25 healthy children from biomechanical tests provided by Laboratório de Biomecânica e Morfologia Funcional da Faculdade de Motricidade Humana da Universidade de Lisboa as in [3].

The first step will be the scaling of both the kinetic and kinematic data so that the work scale is the same. The kinetic data needs an additional normalization as it depends on the weight of each individual. Different datasets will be created as it will be explained later.

Dimensionality reduction will be used to reduce the amount of features resorting to Principal Component Analysis. In order to select the best combination of features both Sequential Forward Selection and Sequential Backward Selection algorithms will be tested.

A Support Vector Machine classifier based on the group labels provided by specialists will be used for all the datasets created and with the variables chosen by Feature Selection algorithms and tested resorting to different performance metrics.

Additionally unsupervised clustering techniques will be applied to compare with the specialists classification, which relies on subjective methods and is difficult and ambiguous in some cases.

The main objectives are related to the understanding of the gait variables involved in the gait classification.

This study tries to create different datasets to evaluate which data treatment is preferable to this type of classification: looking at data in an absolute way or looking for abnormalities and selecting variables or not.

The idea is to advance in the possibility of designing a reliable and robust classifier that helps in the clinical diagnosis, reducing medical classification error rate. Both supervised and unsupervised learning will have input data that had features extracted and selected.

2. Background

2.1. Data Characterization and Processing

The data consists of 25 healthy children that were considered the control group and 26 patients with spastic diplegia. The patients are individuals with maximum 18 years old that were observed for at least 2 years. The distribution of patients through the four groups is shown below.

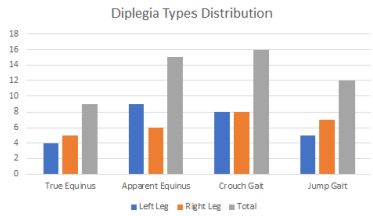

Figure 1: Spastic Diplegia Type

Crouch gait and apparent equinus represent most of the patients that are available which means that the data distribution is quite unbalanced. This bad distribution is justified with the fact that crouch gait and apparent equinus are the most common spastic diplegia groups.

The kinetic and kinematic gait cycles were obtained on the biomechanical laboratory of Faculdade de Motricidade Humana (FMH), using an optoelectronic system with 15 cameras (Qualysis Oqus 300, Qualysis AB, Gothenburg, Sweeden) calibrated with a sampling frequency of 100Hz. The ground reaction force was acquired using two force platforms Kistler (9281B and 9283U014) and an AMTI one. Motion capture was used to reconstruct the 3D diagram of 12 body segments based on 37 reflective markers and 4 gait clusters. Every children performed at least 5 gait cycles, walking through the hall. Of the 12 body segments acquired only 7 were provided: 4 regarding kinematics and 3 regarding kinetics.

Each cycle consists of 101 points for each plane – sagittal, coronal and transversal – and each joint – pelvis, hip, knee and ankle for the kinematic data and only the first three for the kinetic data. Kinematic cycles are calculated as the angular rotation from distal to proximal segment. It is measured in degrees. Kinetic cycles are calculated using Newton-Euler equations of movement and express the internal movements of each joint. It is measured in N/m.

The average of the gait cycles for each individual was calculated in order to use only one gait cycle so that the classification is simplified. This was done using a normal distribution with 101 points and calculating the average and the standard deviation. The diagrams below show each average cycle by group in red and the healthy cycle in black.

The data was normalized to the same scale and kinetic cycles suffered an additional normalization to remove body weight dependence.

$$\text{Normalized Moment} = \frac{Moment}{Gravity \cdot Weight} \tag{1}$$

In order to analyse kinetic and kinematic together, a scaling procedure was necessary. This procedure followed the rescaling method given by:

$$x^{'} = \frac{x - \min x}{100 \times (\max x - \min)} \tag{2}$$

This scales all the data to a range of [0 1] where x is an original value and $x^{'}$ is the normalized value.

The scaling explained above was done to each cycle independently so that it evaluates the maximum and minimum values inside each gait cycle.

In order to posteriorly analyse the cycles two main ways were found: looking at the cycles in an absolute manner or searching for abnormalities, i.e. considering the difference between each patient and the average of the control group.

2.2. Modelling Techniques

In this section the main modelling techniques will be presented. The procedures concerning data processing will be shown, followed by the appropriate data partition and what it involves and, finally the model used to classify the data in a supervised manner.

2.2.1 Principal Component Analysis

Dimensionality reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables. It is used in order to reduce computational time and storage space, to improve the performance of

the model by reducing multi-collinearity and can provide representations that are easier to understand and visualize. Here it will be performed resorting to Principal Component Analysis (PCA). This method is a mathematical tool that converts a given set of possibly related data points into a group of linearly non-related points called the Principal Components, through an orthogonal transformation. It will originate a reduced number of variables projected onto a new space when compared with the original group.

2.2.2 Backward Selection

After feature extraction procedures (PCA for instance) creating new features from functions of the original set, feature selection returns a subset of the best combination of the most relevant ones.

Feature selection techniques are often used in domains where there are many features and comparatively few samples (or data points). The present study is a good example of this case: very few patients (or observations) for several gait cycles. For this reason, it seems logic the usage of such procedure.

The backward search works analogously to the forward selection. However, it starts with the full feature set (of size m) and performs the search until the desired dimension d is reached.

Under this approach, one starts with fitting a model with all the variables of interest or predictors. Then the least significant variables are dropped, as long as it doesn't hit a minimal give by a *a priori* chosen threshold. The algorithm continues by successively re-fitting reduced models and applying the same rule until all remaining variables are statistically significant.

In order to perform backward selection, a situation where observations are more than variables is needed so that n is greater than p. This way, a least squares regression can be applied.

Supposing k features have already been removed from the complete set of measurements $\bar{X}_0 = Y$ to form feature set X_k with the corresponding criterion function $J(\bar{X}_0)$. Furthermore, the values of all supersets) $\bar{X}_i, i = 1, 2, ..., k-1$, are known and stored.

2.2.3 Clustering

The purpose of clustering algorithms is to identify natural groupings of data from a large data set to produce a concise representation of a system's behaviour. There are different techniques and the one used in the present thesis is fuzzy c-means or soft clustering which will be described below. The objective is to use validation indexes in order to choose the number of groups considered and afterwards the membership value to each group will be evaluated.

- **K-Means**
  This algorithm divides the data from a group of n vectors $x_j j = 1, \ldots, n$ in c groups $G_i i = 1, \ldots, c$ and finds the cluster centres or group profile so that the distance is minimized. The Euclidean distance was the metric chosen and the cost function is:

$$J = \sum_{i=1}^{c} \left(\sum_{k, x_k \in G_i} \|x_k - c_i\|^2 \right) \tag{3}$$

  The function J_i is the cost function inside group i. The group partition is typically defined as a binary matrix with a membership function $U_{(c \times n)}$ where $u_{i,j}$ is 1 if x_j belongs to group I and 0 if it doesn't.

- **C-Means**
  Fuzzy C-Means is a data clustering technique wherein each data point belongs to a cluster to some degree that is specified by a membership grade. This technique was originally introduced by Jim Bezdek in 1981 as an improvement on earlier clustering methods. The FCM algorithm aims to partition a finite collection of n elements $X = x_1, \ldots, x_n$ into a collection of clusters with respect to some given criterion. The algorithm returns a list of centres $C = c_1, \ldots, c_c$ and a partition matrix $W = w_{i,j} \in [0, 1], i = 1, \ldots, n; j = 1, \ldots, c$ where each element $w_{i,j}$ represents the degree to which element x_i belongs to cluster c_j.
  Fuzzy C-Means aims to minimize the objective function

$$\underset{C}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{c} w_{i,j}^m \|x_i - c_j\|^2$$

  where,

$$w_{i,j} = \frac{1}{sum_{k=1}^{c} \frac{\|x_i - c_j\|}{\|x_j - c_k\|}^{\frac{2}{m-1}}}$$

2.2.4 Data Partition

One of the main problems of the model creation is to find a dataset that describes the information with precision. The division between training, validation and testing must guarantee that the minimization of the error during training doesn't imply a poorer performance when validating and testing.

A Training dataset is used to create the model whereas the validation and testing datasets are used to evaluate the model.

After creating and evaluating the model in training and validation, a different subset is used to test if the model can describe the reality according to different metrics.

The pursued division will consider a major subset for training and the remaining one for testing.

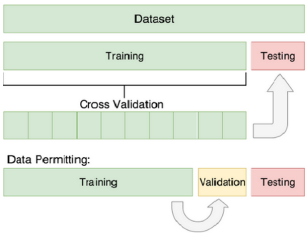

Figure 2: Data Partition

As illustrated in the above figure, the conventional algorithm begins by splitting the dataset into training and test.

Following, the training subset is also divided into training and validation. From a panoply of procedures that perform this desired division, *leave one out cross validation* is the most commonly used for small data sets. As the data set for this present study is indeed small, *leave one out cross validation* become the most obvious procedure to be performed for this task.

In this type of cross validation, the number of folds (subsets) equals to the number of observations that are contained in the dataset. Then all of these folds are averaged and the model is built. It is a computationally expensive method because the number of training sets is large (equal to the number of observations) and, for this reason, should only be used in small datasets. For larger data-sets, K-fold cross validations is the most commonly used.

The figure below illustrates the *leave one out cross validation* method:

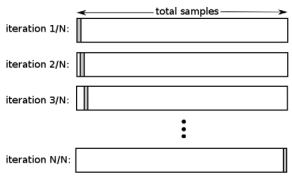

Figure 3: Cross Validation

2.2.5 Classifier

For the classification process, two different approaches were considered, both using Support vector machines (SVM). The first considered binary Support vector machines, while the second multi-class Support vector machines. Support vector machines is a supervised classifier, that creates an hyperplane that separates a group of points into two, while trying to maximize the separation margin between the two. The plane that maximizes this distance is known as the optimal hyperplane.

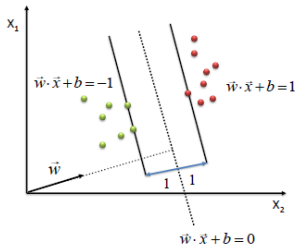

Figure 4: Optimum Hyperplane

Originally the SVM technique was used for binary problems but later it was expanded to deal with multi class problems. There are two methods to accomplish this change.

The first one creates a finite number of classifiers transforming a multi class problem in binary problems. In the end the binary classifiers are combined so that the multi class problem is solved.

The second method considers all the data, formulates and optimized a cost function.

The classification training is different for every pair of different groups and with N classes it creates $\frac{N(N-1)}{2}$ classifiers.

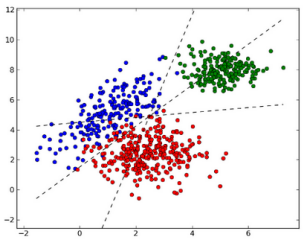

Figure 5: SVM Multi Class problem

2.2.6 Performance Analysis

This step aims to find out how effective is the model based on some metrics using validation and test datasets. Different performance metrics are used to evaluate different algorithms. This thesis is only focusing on the ones used for classification problems.

The metrics choice is important as it allows to measure and compare different ways to evaluate the behaviour of the model created.

The importance of each metric is relative and dependent on the problem that is being studied, i.e the weight given to false positives, false negatives, true positives and true negatives varies.

In the field of machine learning and specifically the problem of statistical classification, a confusion

matrix, also known as an error matrix, is a specific layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is called a matching matrix). Each row of the matrix represents the instances in a predicted class while each column represents the instances the instances in an actual class. The confusion matrix in itself is not a performance measure but almost all performance metrics are based on its information.

The representation for a binary problem is displayed below.

		Predicted class	
		P	*N*
Actual Class	*P*	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Figure 6: Confusion Matrix

Where:
TP: True Positives are the cases when the actual class of the data point was 1 (true) and the predicted is also 1 (true).
FP: False Positives are the cases when the actual class of the data point was 0 (false) and the predicted is 1 (true). **TN**: True Negatives are the cases when the actual class of the data point was 0 (false) and the predicted class is also 0 (false).
FN: False Negatives are the cases when the actual class of the data point was 1 (true) and the predicted is 0 (false).

Thus, the metrics used to evaluate the performance of the model where:

$$ACC = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{4}$$

$$PPV = \frac{TP}{TP + FP} \tag{5}$$

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$TNR = \frac{TN}{TN + FP} \tag{7}$$

$$NPV = \frac{TN}{TN + FN} \tag{8}$$

$$FNR = 1 - TPV \tag{9}$$

$$FPR = 1 - TNR \tag{10}$$

$$FDR = 1 - PPV \tag{11}$$

$$FOR = 1 - NPV \tag{12}$$

$$F1 = \frac{2.TP}{2.TP + FP + FN} \tag{13}$$

3. Implementation

3.1. Data Processing

The data used is composed by 2121 variables and 50 observations for each limb of the control group and 2121 variables and 52 observations for each limb of the patients group. This means that the legs are evaluated independently, which doubles the number of observations. The variables are given by: both the kinematic gait information and the kinetic gait cycles.

The first is composed by gait cycles in 4 joints (pelvis, hip, knee and ankle) and 3 planes (sagittal, transversal and coronal). Each cycle is constituted by 101 points (from 0 to 100%). This translates into a total of 1212 variables.

The remaining 919 variables are contained in the kinetic gait cycles which only consider 3 joints (hip, knee and ankle) and 3 planes (sagittal, transversal and coronal). The cycles are also constituted by 101 points.

For the procedures explained below different datasets were used.

The first approach was to consider the absolute data, that is, 2121 variables and 50 observations that contain kinetic and kinematic information in all planes. The scaling and normalization explained in section 2 was also applied on this data set.

The second dataset looked at data from a relative perspective and searches for abnormality. It is a difference between pathological gait of each patient and the control group gait average. This was done by averaging the 52 observations from the control group and obtaining 1 control cycle of 2121 variables. After, for each one of the 50 pathological observations, this normal cycle was subtracted.

The third dataset was also relative, but instead of considering left and right limb as equal calculates the difference between pathological left leg and normal left leg and the difference between pathological right leg and normal right leg.

Some experiments considering only the stance phase of each cycle (i.e. the first 60% of the cycle), only the kinematic data and only the sagittal plane kinematic data were also performed.

3.2. Dimensionality Reduction and Feature Extraction

Feature reduction and feature extraction, explained in the previous section, were applied to every defined dataset.

Initially, Principal Components Analysis (PCA) was be used in order to transpose the original data set into a new cardinal space, projecting it in new axes according to the directions of higher variance. This procedure also imposes thresholds to the new data, reducing this way the number of variables that explain at lest 95% of the variance.

After, feature selection was used to select the best combination of variables obtained from PCA, minimizing the classification error. As the criterion used to minimize the error cannot be chosen randomly, once the dataset is unbalanced. Bearing this in mind the metric that used weighted a false negative classification more than a false positive. For this reason, the chosen criterion was F1 Score, which is the harmonic mean of sensitivity and precision.

3.3. Supervised Learning

For the supervised learning the data treatment was done according to the previous section and the groups were labelled according to the classification made by specialist.

r A Support Vector Machine (SVM) classifier was used to classify the data into groups and the metrics defined in the last section were applied to evaluate its performance for different experiments.

Several binary SVM classifiers to distinguish between every possible combination of groups, resorting to a one-vs-one (OVO) strategy. Knowing that the multi-class problem has 4 possible classes, the OVO ensemble was composed by 6 binary classifiers.

The data was split considering 75% of the observation for training and 25% for testing. As the dataset is not well balanced the procedure was to try to maintain the representativeness of each group in each subset as similar as possible to its representativeness in the total amount of data.

A flowchart is presented in order to explain the main steps of the implementation.

The flowchart below shows the implementation of the process.

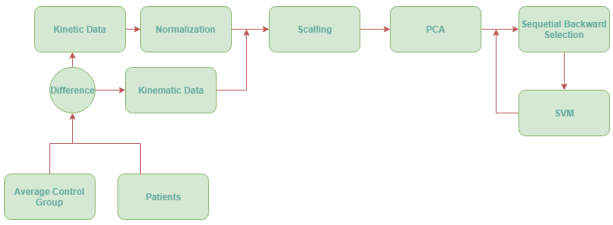

Figure 7: Main method flowchart

3.4. Unsupervised Learning

As explained before the distinction between groups is quite subjective and the patterns are ambiguous. The idea behind this unsupervised section is to classify each patient and make a comparison with the specialist classification.

3.4.1 Kmeans and Cmeans

The data set used to apply both *kmeans* and *cmeans* was the one that achieved the best results in supervised learning with the corresponding dimensionality reduction and feature extraction.

While *kmeans* classifies the patients into only one group *cmeans* presents a degree of membership which is interesting to analyse in the cases where a patient is dubious between two different groups.

The results of this technique will not be described here as they weren't conclusive as is shown in the full thesis.

4. Results

The results achieved, following the methodology already explained are presented in this section. The best model created, corresponding to the dataset where a relative approach with left and right leg considered as equal, will be the one that will have attention here.

4.1. Data Processing

All data, i.e. the complete kinetic and kinematic cycles were used to create the model. This values were scaled and normalized, as explained previously, so that the work scale is the same and body weight dependences are removed.

The difference between patients cycles and control group cycles was calculated so that the model evaluates the data in a relative way and look for abnormalities.

The result is a matrix with with 26 observations, corresponding to the patients. and 2121 variables or features, corresponding to the gait cycles.

4.1.1 Feature Extraction and Feature Selection

In order to avoid overfitting and to transform the dataset so that it contains more observations than variables, dimensionality reduction was needed. Principal Components Analysis was the method chosen to transform the data to a new space, guaranteeing that no relevant information was lost by using a minimum variance threshold.

After reducing the dimension of the data feature selection was used to select the best combination of features to create the model optimally.

4.1.2 Principal Components Analysis

To explain at least 95% of the data variance 28 principal components are needed. After using Principal

Components Analysis, Sequential Backward Selection was used in order to chose the best combination of features.

4.1.3 Sequential Backward Selection

Sequential Backward Selection was used to evaluate the best combination of principal components. It was used in the model creation, considering a error criterion that evaluates F1 Score.

The Support Vector Machine model that is about to be created need good generalization abilities, which means that the classification function needs to work well on data that have not been used during the training. The dataset, as was highlighted previously, is quite unbalanced, with two classes prevalent over the other two. This means that the cost for a false negative classification is significantly higher than the cost for a false positive. Bearing this in mind the metric chosen to evaluate the error was the so-called F1-Score, that is the harmonic mean of sensitivity and precision.

From the 28 components selected by Principal Component Analysis only 16 were used in the model construction.

4.2. Data Partition

The data was split in two subsets, one for training and validation and other for testing the model created. The split is made guaranteeing that the balance of each group in both subsets is maintained the most similar possible to the one of the total dataset.

4.3. Supervised Learning

4.3.1 SVM Classifier

The data was split according to what was explained in the previous subsection and a model was created optimizing the F1-Score.

A subset using the *leave-one-out* method to validate the model was used and the the model created behaved perfectly in this process, i.e. identified all the patients correctly. The next step was to test the model in a different subset and evaluate the generalization abilities using the performance metrics explained previously.

	1	2	3	4
1	1	0	1	0
2	0	2	2	0
3	0	2	2	0
4	0	0	1	2

Table 1: Test Confusion matrix

Only F1 Score values will be displayed here, as it is the main metric used in the process. In general, one can see that group four, jump gait, is the one that achieves best results whereas group three, crouch gait is the one with poorer ones. This is due to the fact of an unbalanced dataset with an higher representativeness of crouch gait. It is important to say that this results come from the fact that when the model misbehaves in the classification of the remaining three classes it always mistakes them with crouch gait.

Group	$F1Score$
1	66.67
2	50
3	40
4	80

Table 2: F1 Score Model Performance (%)

The best result, of 80% is achieved by class 4, followed by class 1 with 66.67%, the ones with lower representativeness. This is due to the fact that besides the model cannot identify all of them correctly, it never mistakes the remaining classes with class 1 and class 4. The poorer results of class 3 are because of the opposite: when the model misbehaves it always mistakes classes 1, 2 and 4 with 3. Class 2 achieves medium results, which can be explained by being the second most representative label. When class 3 is misclassified it is mistaken with class 2.

The problems of unbalanced datasets are clearly shown in this problem. Jump gait, true equinus and apparent equinus when are misclassified are always confused with crouch gait pattern, the most prevalent class whereas crouch gait misclassifications are always as apparent equinus, the second more prevalent class.

The limbs where the model misbehaves are not always limbs where the specialist had doubts however one should bear in mind that the classifier was trained with patterns that were difficult to evaluate and borderline patterns which conducts to a model that can learn from possible errors. This a problem difficult to tackle as 25% of the data, which is the same amount of data that the test data set contains, is based on ambiguous labels. A possible solution is to try to oversample data which is not dubious.

The size of the data set is also small which conducts to less learning and generalization ability and the solution highlighted above can also tackle this problem.

The F1 Score performance in the test phase was compared between the main data sets already defined and this was the one that gave the best results.

4.4. Unsupervised Learning

Clustering methods using the features selected by the best supervised model were used in order to see

if were in agreement with medical classification but no conclusive results were found.

5. Conclusions

Pursuing the work done in [3] , this dissertation aimed to create a classifier that could distinguish patients suffering from spastic diplegia and categorize them according to their gait. It was known a priori the number of existing groups and its description according to the type of gait was made. Medical classification of the patients from the data set was also provided and used as a mean for supervised learning and validation.

It is important to highlight the advancements done relative to the past work:

- The scaling of the data made the integration of kinetic data in the analysis possible;
- The normalization allowed the removal of bodyweight dependence;
- Looking at data and searching for abnormalities led to improvements;
- Backward selection was seen as a better method to select features comparing to Forward Selection;
- The error criterion to select features changed from maximizing accuracy to maximizing F1 Score in order to deal with an unbalanced data set;
- The data partition was more carefully looked at considering the unbalanced data set and a 75%-25% was chosen instead of using one leg for train and the remaining for test;

The procedures mentioned above permitted the overall improvement on the classification results and to understand possible future steps in order to overcome the challenge. To develop this problem in better terms a database with a greater amount of data will be imperative. If possible it should be one that contains more jump gait patients and specially true equinus ones. Patterns which are not dubious will be of great help also, at least until a robust and reliable classifier is built.

A greater amount of gait information proved to be important in improving classifier performance. This means that adding spatio-temporal information and, if possible, force and acceleration can prove essential in the process.

The wavelet network and Shannon entropy approach proved to be non conclusive in the evaluation of gait types but can be used to study the degree of palsy of each individual in order to help in the treatment process.

In what regards feature selection it would be interesting to apply it to each individual binary classifier instead of a multi class feature selection. It would help in understand which features are important to distinguish between each class.

Another method that can be tried if no more data is available is oversampling the least prevalent classes, a typical procedure to unbalanced data sets.

Acknowledgements

I would like to kindly thank both my supervisors, Prof. Susana Vieira and Prof. Filipa João, for guiding me through this project. I am also grateful for the data provided by Faculdade de Motricidade Humana, which made the realization of this work possible.

References

[1] S. Carlon, N. Shields, K. Yong, R. Gilmore, L. Zewski, and R. Boyd. A systematic review of the psychometric properties of quality of life measures for school aged children with cerebral palsy. *BMC Pediatrics*, 2010.

[2] P. M. Evans, S. J. W. Evans, and E. Alberman. Cerebral palsy: why we must plan for survival. *Archieve of disease in childhood*, 1990.

[3] F. Fernandes. Classificação de pacientes com diplegia espástica baseada em padrões de marcha. Master's thesis, Instituto Superior Técnico, Universidade de Lisboa, 2017.

[4] P. L. Jia, P. F. Zhang, H. D. Li, L. H. Zhang, Y. Chen, and M. M. Zhang. Literature review on clinical decision support system reducing medical error. *Journal of evidence based medicine*, 1995.

[5] F. Miller, K. W. Dabney, and M. Rang. Complications in pediatric orthopedic surgery. *Pediatric Orthopaedics*, 1995.

[6] M. Rang, J. D. L. Garza, and R. Silver. Cerebral palsy. *Pediatric Orthopaedics*, 1986.

[7] J. Rodda and H. K. Graham. Classification of gait patterns in spastic hemiplegia and spastic diplegia: a basis for a management algorithm. *European Journal of Neurology*, 2001.

[8] J. Rodda, H. K. Graham, L. Carson, and R. Wolfe. Sagittal gait patterns in spastic diplegia. *The bone and joint journal*, 1995.

[9] D. H. Sutherland and J. R. Davids. Common gait abnormalities of the knee in cerebral palsy. *Clin Orthop Relat Res*, 1993.

[10] S. Winters, A. Autry, C. Boyle, and M. Yeargin. Trends of prevalence of cerebral palsy in a population based study. *Official Journal of the American Accademy of Pediatrics*, 2002.

[11] T. F. Winters, J. R. Gage, and R. Hicks. Gait patterns in spastic hemiplegia in children and young adults. *The Journal of Bone and Joint Surgery*, 2008.