NTNU

Norwegian University of
Science and Technology

Early Detection of Cerebral Palsy Using Two Computer-Based Methods

Hege Gran Lundeby

Preface

This report is a master thesis done at NTNU as a part of the study program cybernetics and robotics. The work was executed in the spring of 2016 as a continuation of an earlier project report done in collaboration with St. Olavs Hospital. The topic in this thesis has been to analyze and compare the detection of fidgety movements in infants, based on results from two computer algorithms. This is done to see if a newly finished method is more accurate than the one that has been in use at St. Olavs Hospital the last few years.

Trondheim, June 5, 2016

Hege Gran Lundeby

ii

Acknowledgment

I would like to thank the following persons for their great help on this master thesis. First and foremost, I would like to thank my supervisor Ole Morten Aamo for his guidance throughout the whole project and the writing of the report. Secondly, I would like to thank Lars Adde for his insight on cerebral palsy and fidgety movements, as well as pointing out ideas I've had to consider. Hodjat Rahmati also deserves a thank you, for using a lot of his time to explain the different codes that are included in his method. Last, but not least, I want to thank Harald Martens for his guidance on the different statistical approaches I have used in my work.

H.G.L.

Abstract

In this master thesis the assignment was to perform a statistical analysis of two different computer-based methods, to see which one that had the highest predictive ability for fidgety movements in infants. One method has been in use at St. Olavs Hospital for a few years, and it is both reliable and effective. However, after a newly finished method performed even better last year, it was in everyone's interest to test both of these methods on a completely new set of infants.

This statistical analysis included two different approaches to test the performance of the chosen features. First, the principal component analysis was done to get an overview of the data and to find possible outliers. While the method from St. Olavs Hospital had no clear outliers, one was found in the data from Hodjat Rahmati. There was also found one clear similarity between the two methods. This was that none of the data sets managed to find enough coherence in the data to separate the different classifications from each other. In the partial least squares regression the similarities between the two methods ended. The method from St. Olavs Hospital reached at best 14,63% explained variance for the calibration and 11,56% for the validation with the features C_{sd}, Q_{mean} and Q_{sd}. This was quite weak compared to earlier results from this method, although these results are from a different form for evaluation. When it came to Hodjat Rahmati's method, the results were quite surprising. For the explained variance the calibration was quite good and reached a value of 93,85% at best. However, for some reason the validation fell downwards to a value of -180,73%. Because of this, the conclusion in the end was that the method from St. Olavs Hospital was marginally better.

Different reasons for these results have also been discussed, such as the data itself, the videos or the classification. Two different approaches showed that the data was most likely correct, while there were no way to check for either of the other two.

Sammendrag

I denne masteroppgaven gikk prosjektet ut på å gjennomføre en statistisk analyse av to ulike databaserte metoder, for å se hvilken som hadde høyest prediksjonsevne for å detektere en type bevegelse som på fagspråket kalles fidgety movements. En metode har vært i bruk på St. Olavs Hospital de siste årene, og den er både pålitelig og effektiv. På en annen side var det en annen metode som oppnådde enda bedre resultater i fjor, og etter dette har det vært i alles interesse å teste begge disse metodene opp mot hverandre på et helt nytt sett med spedbarn.

Denne statistiske analysen inkluderte to ulike tilnærminger for å teste prediksjonsevnen til de valgte kroppsvariablene. Først så ble det gjennomført en prinsipal komponent analyse for å få en oversikt over dataene, og for å finne mulige uteliggere. Mens ingen spedbarn stod frem som klare uteliggere i dataene fra St. Olavs Hospital, ble det funnet en i datasettet til Hodjat Rahmati. Det ble også funnet en klar likhet mellom de to metodene, nemlig at ingen av dem klarte å finne et skille mellom de ulike klassifiseringene til spedbarnene. I minste kvadraters metode sluttet likhetene mellom de to. Metoden fra St. Olavs Hospital oppnådde en høyest forklart varians på 14,63% for kalibreringen og 11,56% for valideringen med variablene C_{sd}, Q_{mean} and Q_{sd}. Dette var ganske svakt sammenlignet med tidligere resultater, selv om disse er oppnådd med en annen form for evaluering. Derimot var resultatene fra Hodjat Rahmati sin metode ganske overraskende. Her var kalibreringen veldig god og hadde en forklart varians på 93,85%. På en annen side oppførte valideringen seg veldig rart, og falt ned til en verdi på -180,73%. Derfor ble konklusjonen til slutt at St. Olavs Hospital sin metode var marginalt bedre.

Ulike grunner for disse resultatene har også blitt diskutert, blant annet dataene, videoene og klassifiseringen. To ulike tilnærminger viste at grunnen mest sannsynlig ikke lå i selve datasettene, mens det ikke var mulig å teste for noen av de andre mulighetene.

Acronyms

CP Cerebral palsy

GM General movement

FM Fidgety movement

PLSR Partial least squares regression

MGT Musical gesture toolbox

GMT General movement toolbox

GMA General movement assessment

OpenCV Open Source Computer Vision

FFT Fast fourier transform

PCA Principal component analysis

PC Principal component

SVD Singular value decomposition

CV Cross validation

EM Expectation maximization

RMSE Root mean square error

x

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Sections 1.1.1 and 1.1.2 are both taken from Lundeby [9]. They are included here to make it easier for the reader, so that the person does not have to go back to the earlier report. After these sections there is a short summary on the work done in the project thesis and the results that were found there.

1.1.1 Cerebral Palsy

Cerebral palsy (CP) is a disability in infants causing reduced motor function (see Bakke and Moster [4]), and it is developed from a brain damage sustained either before, during or after birth. Even though there are many different causes for developing CP, it is shown that especially infants with a gestational age lower than 28 weeks and/or a birthweight lower than 1000 gram are at high risk(see Adde et al. [3]). According to Eriksen [6], 10-15% of preterm infants will develop CP.

1.1.2 Motivation

The process of diagnosing someone with CP today is difficult, and it is usually not done until the child is between 2-5 years old. This is because the symptoms are usually not visible the first few years, and there needs to be clinical signs to set an accurate diagnosis. However, seeing as the brain is not fully developed until the child is 2 years of age, it would be highly beneficial if the diagnosis could be set earlier, so that one can begin

treatment such as physical therapy at an earlier age (see Adde [1]).

General movements (GMs) are spontaneous movements of the infant, and assessment of these have shown to be helpful in earlier detection of CP. By studying video recordings of infants, observers may classify the movements into different categories. This is usually done between the age of 10-18 weeks when certain circular movements called fidgety movements (FMs) generally are present. According to Adde et al. [3], these movements can be classified into either two or four categories. The main categories are normal or abnormal FMs, which again can be divided into continuous or intermittent, sporadic or absent. The results from the observations will then give an indication if further follow-up of the infant is necessary, and has shown to be a reliable method of choice.

However, the assessment of GMs are dependent of skilled personnel, which is not available everywhere around the world. That is why many now are turned toward the computer technology, hoping to find a computer-based algorithm that can be used on a global scale.

Summary of Project Thesis

Since this master thesis is based upon earlier work (see Lundeby [9]) it is important to know which results that have been found so far. In the project thesis the method created by Alexander Refsum Jensenius was analysed and compared against earlier results found by St. Olavs Hospital. When Lars Adde, who is a physiotherapist there, tested the feature C_{sd} on a previous set of babies (see Adde et al. [3]), he got the result of 81,5% for the sensitivity and 70% for the specificity. However, from running the features through a partial least squares regression(PLSR) the predictive ability was found to be at best 57,2%.

There might be different reasons for this, one of them being the validation. In their case, they create a model of the infants on one set of babies before testing this on a completely different set. However, in the PLSR the validation is done on the same set of infants as the one the regression is built from. Another reason might be that while Lars Adde normalized the features in regard to the babies' sizes, this was not done in the project report. If the sizes of the babies are evenly spread out for infants both with and without FMs this will have nothing to say, but if the sizes are poorly distributed the results might be influenced. One last thing to mention is that while he ran the analysis on

only the variable C_{sd}, the feature set in Lundeby [9] included some of the other features as well.

Problem Formulation

In the project thesis the assignment was to collect data from 146 infants using two different computer algorithms. This data would then be statistically analysed to see which bodily features that would result in the best predictive ability. The previous thesis included an analysis of the method implemented by Alexander Refsum Jensenius and St. Olavs Hospital. In this master thesis the same analysis will be executed again for the normalized babies, before the results from the other algorithm will be presented. The results from the two different methods will then be compared against each other so that a final conclusion can be made.

1.2 Objectives

The main objectives of this Master's project are

1. To present the theoretical and mathematical aspects of the two computer-based methods

2. To perform a statistical analysis of the normalized infants from St. Olavs Hospital

3. To perform a statistical analysis of the infants from Hodjat Rahmati's method

4. To compare the two methods before making a conclusion about which method and features that best predict FM in infants

1.3 Limitations

One of the biggest limitations in this project was the running time of some of the files written by Hodjat Rahmati. The file that used the most time was the calculation of the affinity matrix, which describes the similarity between the different trajectories. It had a running time of one video per 24 hours, and with 146 infants it is safe to say that it took a lot of time. Because of this a lot of work had to be delayed, and the features from

this method were not finished before the end of April. As a result the analysis is not as thorough as it could have been.

1.4   Structure of the Report

The rest of the report is structured as following. Chapter 2 presents the different aspects of the two computer-based methods, before the statistical approaches are described in Chapter 3. Chapter 4 gives a brief introduction to the infants that are included in this analysis, and in Chapter 5 are the results presented and discussed. Chapter 6 concludes and summarizes this master thesis.

Chapter 2

Theory

The theoretical descriptions of the two different methods below are taken from Lundeby [9]. In addition to this there is a section on normalization of the infants, as well as the mathematical descriptions.

2.1 Method of Alexander Refsum Jensenius

The first method I have studied is created by Alexander Refsum Jensenius. He is a music researcher who has collaborated with NTNU and St. Olavs Hospital the last few years. He designed a video analysis tool for studying musical gestures in dancers, which converts the movements into different numbers and curves (see NRK [11]). This tool, called the Musical Gesture Toolbox (MGT), has later been customized into the General Movements Toolbox (GMT) (see Adde et al. [3]) used on infants in general movement assessment (GMA), resulting in an effective and reliable method of detecting abnormal FMs in infants.

This method first uses the GMT to crop the image so it will be less pixels to process. After this, a motion image is obtained, which describes in which pixels there are movements, called positive pixels. Furthermore, the motion image is filtered before continuing the analysis. After this process, a motiongram can be obtained. This gives a visual indication of how much the infant moves over time.

In addition to this motiongram, there are also some important variables to be obtained from this method. These variables are the features that have been used in this statistical analysis. In total there are five features, namely $C_{x\_}sd$, $C_{y\_}sd$, C_{sd}, Q_{mean} and Q_{sd}.

Q_{mean} and Q_{sd} stands for the mean and standard deviation of quantity of motion, which describes how much motion there is between one frame and the next. The quantity of motion is calculated by adding all the positive pixels together, and then dividing that on the total number of pixels.

C_{x_sd} and C_{y_sd} are two features that describe the standard deviation of the centroid of motion in x- and y-direction. Those two features have also been collected into one single feature, which is the C_{sd}. This is the standard deviation of centroid of motion. This describes the center of positive pixels in the motion image, which corresponds to the centre point of the infants movements. According to Valle et al. [19], a low C_{sd} usually corresponds to small, circular movements spread evenly throughout the body of the infant, while higher numbers usually are the result from more jerky, stiff movements indicating a possibility of CP. This is the opposite of the values for the quantity of motion, which usually has high values for normal FMs and low for abnormal.

2.1.1 Mathematical description

The quantity of motion describes how much motion there is between one frame and the next. Usually one calculates this by summing the number of pixels that have changed between two frames and divides them on the total number of pixels in the frame. However, according to Jensenius et al. [7], this is only possible when comparing recordings with equal duration. When the time duration is different from one video to the next, it is better to divide the total distance travelled on the duration of the video. This is equal to the average speed that the marker moves with.

The equation used in this method is as following (see Jensenius et al. [7]):

$$Q_{om} \frac{\sum_{n2}^{N} \|p(n) - p(n-1)\|}{T} \tag{2.1}$$

Here p is a position vector, which can be either 2D or 3D based on which measurement tool that is used. The 2D vector then describes the x- and y-position of the pixels from one frame to the next, while the 3D vector is used when the infant is wearing a sensor. The other two variables are N, which is the total number of samples, and T is the total duration of the recording.

OpenCV(Open Source Computer Vision) is a tool used to perform these calculations

and the centroid of motion is obtained by the following approach described in AI Shack(see Sinha [17]).

The centroid of motion describes the center of movement in the frame. The pixels that describes movement are usually white and set equal to one, while the pixels with no movement are black and equal to zero. This is illustrated in Figure 2.1.

Figure 2.1: Binary image (from Sinha [17])
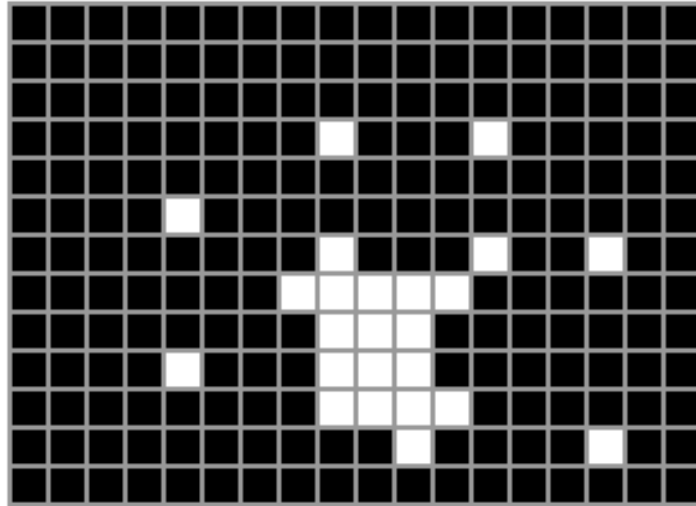
According to Adde [2], the origin is found in the bottom left corner with the x-coordinates to the right and the y-coordinates upwards. To find the centroid of motion one needs to sum up the coordinates for the pixels that describe some sort of movement. This is done for the x- and y-coordinate, as shown below. In these equations are the current pixel described by $f(x,y)$.

$$s_x \sum \sum x f(x,y) \tag{2.2}$$

$$s_y \sum \sum y f(x,y) \tag{2.3}$$

If there is a movement in the current pixel, $f(x,y)$ equals to one. So in these two equations the different x- and y-coordinates where there is movement are summed together. Since the centroid of motion describes the center of movement, this needs to be averaged by the total number of positive pixels. This is the same as the area of the frame, which is described in Equation (2.4).

$$\mu_{0,0} \; \sum_{x0}^{w} \sum_{y0}^{h} f(x,y) \tag{2.4}$$

The w and h corresponds to the width and height of the frame, so that all the pixels are included. Since the frame consists of only zeros and ones, it is seen from the last equation that the number of positive pixels are added up together. This can be used to find the average as following:

$$\mu_{1,0} \; \frac{s_x}{\mu_{0,0}}$$

$$\mu_{0,1} \; \frac{s_y}{\mu_{0,0}}$$

Here the coordinates in both x- and y-direction is divided on the total amount of positive pixels. The centroid of motion is then calculated as in Equation (2.5).

$$c_{motion} \; (\frac{\mu_{1,0}}{\mu_{0,0}}, \frac{\mu_{0,1}}{\mu_{0,0}}) \tag{2.5}$$

As some might notice, the Equations 2.2 and 2.3 are dependent of the babies sizes. According to Adde [2], the mattress the infants are lying on is made so that their placement will be as similar as possible. However, it is correct that if the infant is placed either to the left or the right side of the mattress the data will be affected by this.

2.1.2 Normalization

In the project thesis it was shown that the predictive ability of the method used at St. Olavs Hospital was quite low. It was in fact surprisingly low compared to other results that has been found with earlier sets of infants. In 2009 St. Olavs Hospital got a sensitivity of 81,5% and a specificity of 70% when running it in SPSS (see Adde et al. [3]). In Lundeby [9] the predictive ability at best reached 57.2%. There might be different reasons as to why these numbers differ, such as the different sets of infants, how the results are calculated in SPSS and The Unscrambler or the normalization of the infants.

The evaluation at St. Olavs Hospital is done by building a method upon one set of babies before testing this method upon another set in SPSS. That way there is no

direct evaluation of the performance on one single set of babies. This is different from the evaluation in the Unscrambler, where the predictive ability is tested on the same set as the one the calibration is built upon.

Normalizing the infants may be important in regard to A. Jensenius' method because the features are calculated in regard to bodily measures such as quantity and centroid of motion. A larger infant will naturally have its centroid of motion placed differently than a smaller one, and vice versa. To calculate this, one first need the trunk area. In A. Jensenius' program it is possible to decide the trunk length and width, which is multiplied to find the trunk area. The different features are then divided by the trunk area to obtain the normalized features of each infant. Still, this normalization will only *maybe* improve the results. If the distribution of the infants' sizes are equal between the different FMs, then the normalization will be unnecessary.

2.2 Method of Hodjat Rahmati

This method, which is created by Hodjat Rahmati, is quite different than the one previously mentioned. Instead of basing the calculations on motiongram, this one is based on optical flow and segmentation. Optical flow describes the motion of an object caused by its relative movement to an observer. Using the pattern of how a particle in a pixel move from one frame to the next, one can calculate the displacement of all the particles, ending up with an optical flow vector (see Stahl et al. [18]). These vectors, or trajectories, are analysed pairwise by comparing the distance between their movements. They are then aggregated into an affinity matrix which describes the pairwise similarities. This similarity matrix can then be used to divide the object into multiple motion segments (see Rahmati et al. [14]). A positive result of only comparing trajectories pairwise, is that missing data can be set to zero, or lowest possible affinity.

After the affinity matrix comes the segmentation. In this part all the different segments are labelled manually. Here it is important to make sure that all the segments have trajectories for the total duration of the video. This is because of the tracker that comes afterwards. If any of the segments lacks trajectories, the tracker disappears off to the side of the frame, without being able correct itself. The end result of the tracking is one single trajectory for each segment that is used in the feature extraction.

The important features that are extracted from this approach are the frequency components of the motion trajectories. These components are a good way of studying the motion variability and they are found using Fast Fourier Transform (FFT) (see Rahmati et al. [16]). First, the FFT results in a large magnitude of frequency components that are dependent of the number of trajectories. To reduce the number of components, the log-space equivalent of the frequency range is split into non-overlapping segments. The number of segments is crucial, because too many will lead to redundant information and overfitting, while few segments cause a loss of information. It has been observed that a 100 segments usually lead to a good fit. In the end there are 99 features extracted from both the mean values and related to the standard deviation for each body part and in each dimension. Since there are six segments and two dimensions, the total number of features are $(6*2*99)*2$ 2376 features.

There are many positive sides to this method. In addition to being able to handle occlusions or fast motions, one of the most important aspects must be that there is nothing intrusive to the infant's motion. Earlier methods have been dependent of sensors or other instruments that may hinder important movements, while this uses none.

2.2.1   Mathematical description

After the tracking there is one trajectory for each segment in both x- and y-direction. These are in the time domain and can for an example be written as $x_{arm_{right}}(t)$ and $y_{arm_{right}}(t)$ for the right arm. After this a FFT is done, so that these features are instead represented in the frequency domain. Suppose we have a lower and upper bound for the frequency range w, where w $= [f_i, f_s]$. Here f_i is the greatest lower bound and f_s is the least upper bound. Since the calculated number of frequency components corresponds to the number of trajectories, where half of the components are negative frequencies with equal value to the positive counterparts, the independent components are only half of the total set, ex. $\frac{N}{2}$.

These components can be divided into F bins, with a size L_n $\lfloor nc \rfloor$ L_0. Here L_0 is an initial value of all the trajectories and L_n is the size of bin n. $\lfloor c \rfloor$ describes the integer part of c, a constant which is calculated as shown in (2.6) (see Rahmati et al. [15]).

$$\sum_{n1}^{F}(L_n) \sum_{n1}^{F}(L_0 \ nc) \tag{2.6a}$$

$$FL_0 \ \frac{F^2}{2}c \ \frac{N}{2} \tag{2.6b}$$

$$\Rightarrow c \ \frac{2(\dfrac{N}{2} - FL_0)}{F^2} \tag{2.6c}$$

To find the final features both the mean and standard deviation of the power spectrum is used. According to Rahmati et al. [15] L_0 is set equal to 2 to make sure that each bin has at least two frequency components. The higher the frequency or a higher number of samples in the trajectories also increases the bin.

Chapter 3

Statistical Approach

The principal component analysis (PCA) and PLSR sections in this chapter are taken from Lundeby [9].

3.1 Principal Component Analysis

PCA is a statistical procedure used to get an overview of the data. Running a set of measured samples through a PCA results in certain components called principal components (PCs). According to Martens and Martens [10], these components are a linear combination of the measured samples. The PCs describe where in the data the variance is highest, with the first PC describing the highest variance. Studying plots from a PCA is therefore very useful, because one can easily point out sample outliers and which variables that influence the system most. However, it is not a complete statistical model and is better used along with some kind of regression model.

There are multiple algorithms available for PCA, and three of these are Nipals, stepwise and full singular value decomposition (SVD). Nipals is an iterative procedure that analyses only one PC at a time. It is also possible to add certain criteria, such as smoothing, so that one's own need is met. The stepwise SVD also analyse one PC at a time, however, it does not include all the same possibilities as Nipals. The full SVD is not iterative, and regards the whole matrix at once.

3.2   Partial Least Squares Regression

The PLSR is one of many regression methods. It is used to find a relationship between two matrices X and Y, so that the resulting model can predict the dependent variable Y from X. According to Martens and Martens [10], the prediction model of the PLSR is usually divided into two parts, namely the training or calibration phase and the prediction phase. In the calibration phase the model for predicting Y from X is set with help from a chosen training set. However, in the prediction phase the trained model is used on new X samples so that the unknown values of Y can be predicted.

3.2.1   Cross Validation

Cross validation (CV) is an important part of the PLSR, because it gives a measurement of how well the prediction is. By dividing the data into different training and test sets, one can train the model into making a prediction about the remaining set. There are several ways to do this, such as leave-one-out CV, where all observations in successive rounds are left out as test set. Other division methods are blocked, split-sample or random CV.

3.3   Cluster analysis

According to Dell [5] is cluster analysis a tool used to sort observed data into two or more categories depending on how similar the data is. This way the association between two objects is maximum if they are in the same category, while it is minimum otherwise. A cluster analysis is therefore said to find structure in the data without interpreting the reason behind it.

There are many different algorithms to choose between, such as k-Means, expectation maximization (EM) or Ward's method. The k-Means clustering is very relevant in situations such as the infants' classifications. This is used if one already have an opinion of how many clusters it should be, and these clusters are to be as different as possible. Seeing as the infants either have normal or abnormal FMs, it would be natural to choose two clusters for this analysis.

Chapter 4

Input Data

4.1   Infants

In this project there were in total 146 infants included, all of them recorded in between the age of 10-18 weeks post-term. Since no diagnosis have been set so far, the infants were instead classified by the use of FMs. One set of data was collected from each infant, namely a video capture. This recording features the child laying supine on a mattress while awake, wearing a diaper and a body (see Adde et al. [3]).

However, when all the data was extracted from H. Rahmati's method, a small change had to be made. This is because an error occurred on one of the videos when tracking the segments. The segmentation, in which the different body parts are manually labeled, worked just fine, but for the tracking part something was obviously wrong. In the tracking file all the trajectories from the segmentation are summed and averaged into a single trajectory for each segment, but for one video there was no visible trajectories. With the approval from both supervisor Ole Morten Aamo and Lars Adde at St. Olavs Hospital was this infant taken out of the analysis executed on H. Rahmati's features. This infant had present FMs, so luckily did the sample of those without FMs not decrease.

Chapter 5

Results and Discussion

In this chapter the results from the statistical analysis will be presented. The different features have been analysed by using both PCA and PLSR. These methods are explained in detail in Chapter 3. All results were found using a program called The Unscrambler.

5.1 PCA

5.1.1 Normalized infants from St. Olavs Hospital

When the PCA was run on the infants from St. Olavs Hospital, it was run on two different sets. One of these sets consisted of all the features, while the other set included the variables C_{sd}, Q_{mean} and Q_{sd}. This was done to see if there were any differences in the results, such as the given outliers.

It is also important to mention that in this part of the analysis were all the infants still included. In Section 4 it was mentioned that one infant was left out, although this is only for the analysis done on H. Rahmati's features. However, the analysis on the features from St. Olavs Hospital were also performed without this infant just to check, and it did not change the results in any way. Therefore are these plots not included here.

All features

The first PCA was run with all of the features included. That way the system could be analysed using all the available knowledge. The resulting scores plot and the correlation loadings are shown in Figure 5.1 and Figure 5.2.

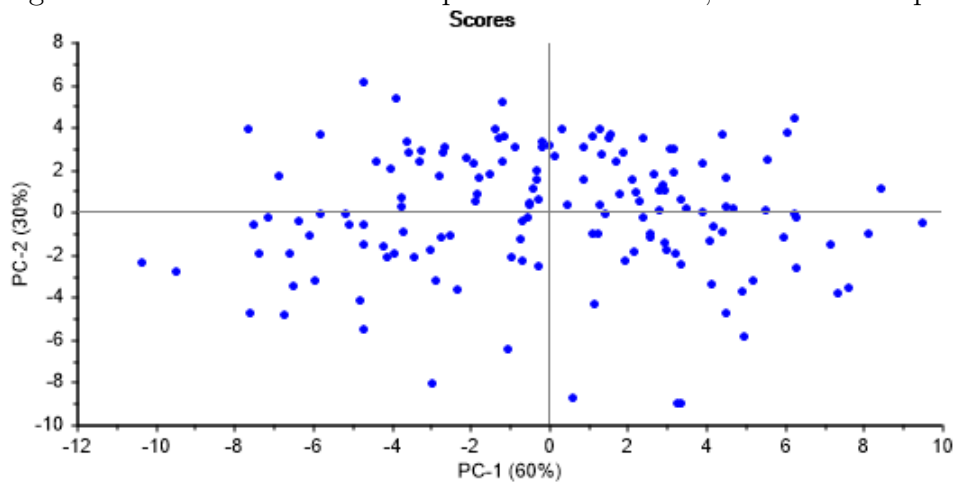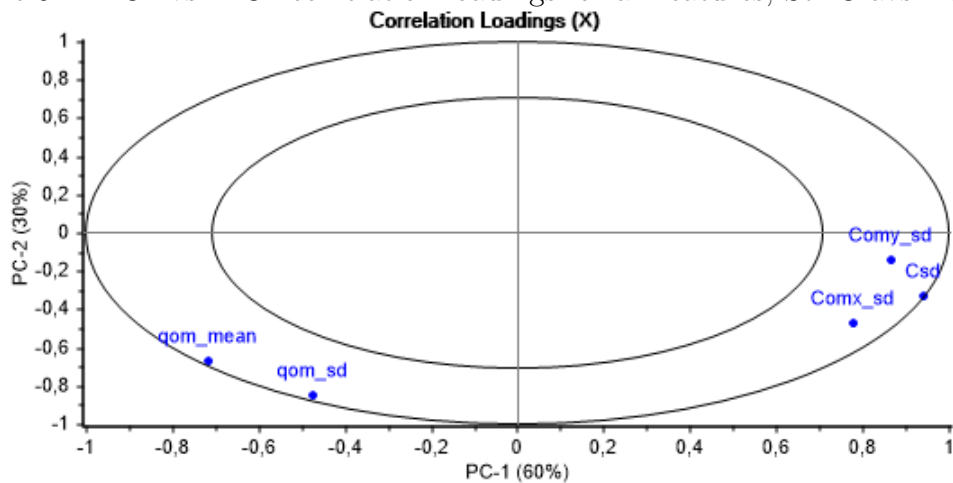Figure 5.1: PC1 vs. PC2 scores plot for all features, St. Olavs Hospital

Figure 5.2: PC1 vs. PC2 correlation loadings for all features, St. Olavs Hospital

The scores plot indicate how similar the different samples are. If the distance between two samples are small, they are similar with respect to the chosen PCs. The larger the distance between two samples becomes, the less similar they are. In this scores plot the two first PCs are chosen. This is done because they are the components that describe where the highest amount of variability in the data is.

In this scores plot one can see that a lot of the samples are quite similar, since they are clustering together in the middle. Still, the distance between the samples on the outskirts are quite large. This is especially true if one studies the horizontal line, which makes sense, since this is described by the first PC. This might be samples that are either outliers, or that strongly describe the condition of the specific infant.

However, there is not much coherence between the samples regarding the two different classifications. The labels have not been included here since it would make the plot too

messy, but a quick check confirms that samples that describe infants with and without FMs are found close together. This proves that the PCA has not managed to distinguish between the different classifications.

The correlation loadings indicate how much of the variance in the different features that are taken into account in the two first PCs. The outer circle describes 100% explained variance, while the inner describes 50%. Here it is shown that almost a 100% of the variance in C_{sd}, Q_{mean} and Q_{sd} are explained in the first two PCs, while the two other features are partially described in later components. The correlation loadings also shows if any of the features are correlated with each other. Since Q_{mean} and Q_{sd} are found in the same quadrant, they are positively correlated with each other. The same is also true about the three other features. One last thing to mention is that all five features explain variance in both PCs, although they seem to have a slight divide.

Another important plot is the influence plot. This can show if there are any of the samples that might be possible outliers, which can affect the model negatively. The influence plot is shown in Figure 5.3.

Figure 5.3: PC1 influence plot for all features, St. Olavs Hospital
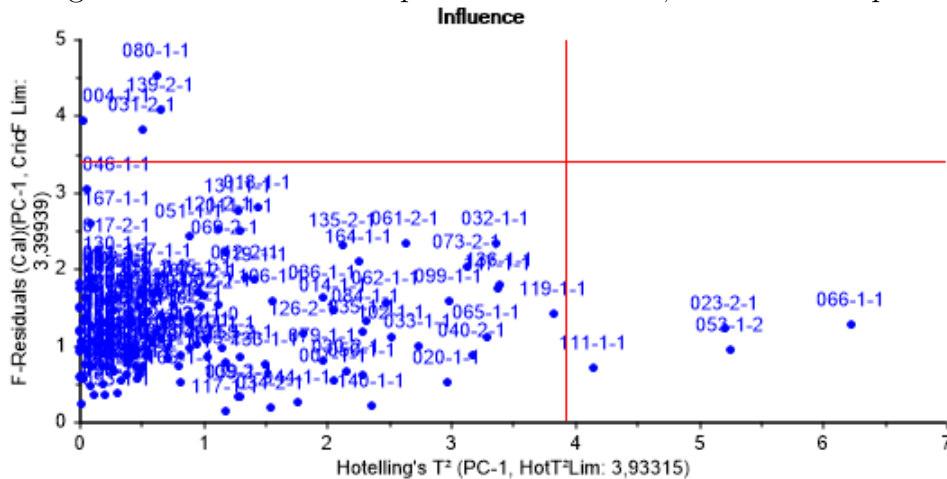
The samples found above the red line have a high residual variance, meaning that they are poorly described by the model. This might be because they are described by another PC than the first, which is a good possibility since the first PC only describe 60%. Still it might be smart to pay attention to these samples, in case they influence the model negatively. The samples on the right side are well described by the model, which is most likely because they are described by the first PC. However, they might be strong influences in building model and need to be treated with care. No samples are found in

the right corner, which is good, because it is there the obvious outliers are found.

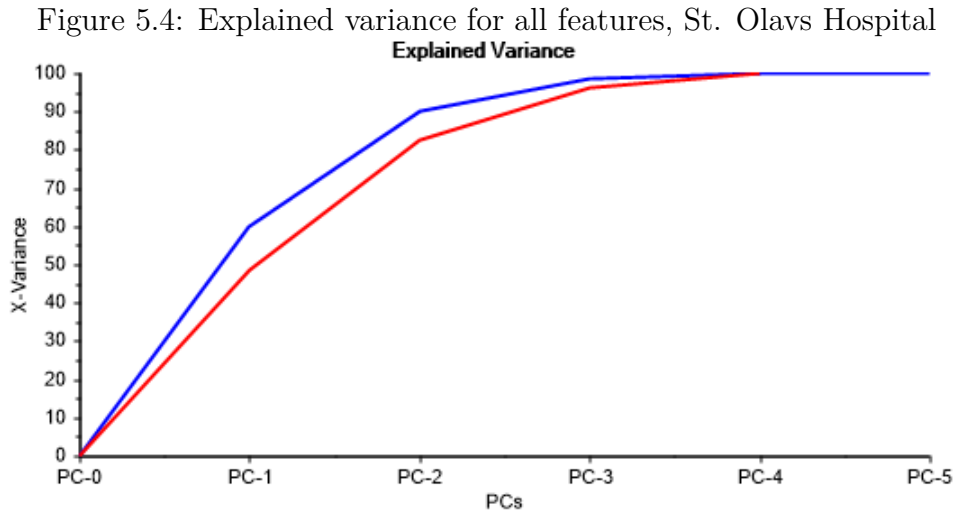Figure 5.4: Explained variance for all features, St. Olavs Hospital

Figure 5.4 shows the explained variance for all the features. This tells how much of the variance that are described by each of the five PCs. While the blue line shows the calibration of the model, the red line shows the validation. Most of the variance is explained by the three first components, leaving little variance left to the two last PCs.

C_{sd}, Q_{mean} **and** Q_{sd}

The other PCA was run with C_{sd}, Q_{mean} and Q_{sd} included. This is because C_{x_sd} and C_{y_sd} are gathered into C_{sd}, and it was seen on the correlation loadings for the full PCA that all three features were found close together in correlation.

Regarding the scores plot and the correlation loadings, the figures will not be included here. This is because the graphs were so similar to the ones in the full PCA. The only difference in the correlation loadings, was that the features Q_{mean} and Q_{sd} were now mostly explained by PC1, while C_{sd} were explained by PC2.

Figure 5.5 shows that there now might be some other possible outliers. This is for an example sample 031, which fits the model well, but is also close to the upper red line. By checking the later PCs, it is clear that the same sample still shows a good fit to the model, indicating that it might be an outlier.

In figure 5.6 one can see that PC1 and PC2 describes almost all of the explained variance.

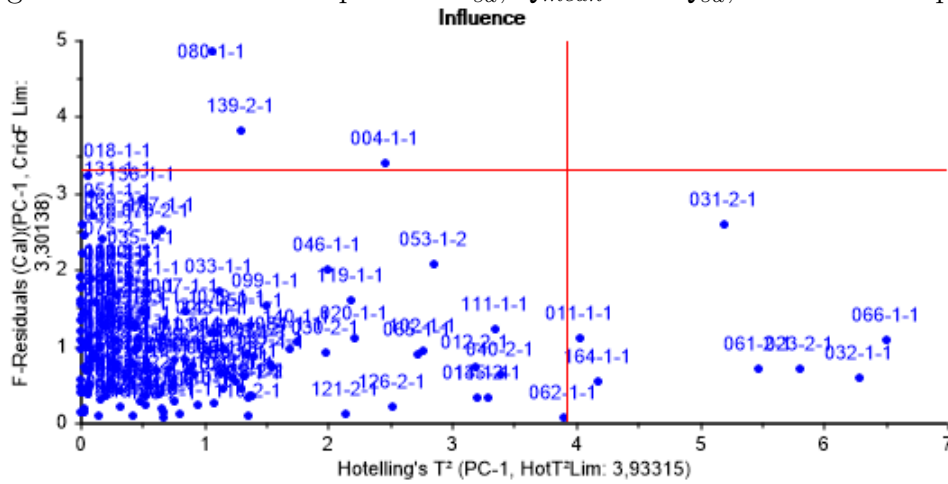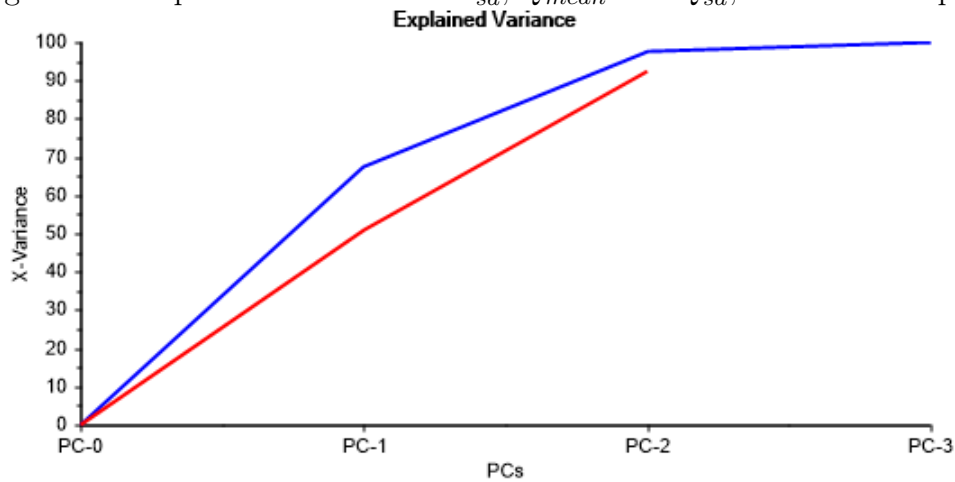Figure 5.5: PC1 influence plot for C_{sd}, Q_{mean} and Q_{sd}, St. Olavs Hospital

Figure 5.6: Explained variance for C_{sd}, Q_{mean} and Q_{sd}, St. Olavs Hospital


5.1.2 Features from Hodjat Rahmati

When the PCA was run on the feature set from H. Rahmati it was only run on the full set, and there are two reasons for this. The first reason is that the results were so similar for each set that it did not seem necessary to perform any further analysis. The other reason was the time limitation, since these features were finished so late in the semester.

All features included

The PCA including features from the method created by H. Rahmati was run with all of them included. Here the number of frequency bins was chosen to be k = 100 - 1 = 99, which resulted in 1188 different features for both the mean and standard deviation. In total there were 2376 features included. In Figure 5.7 is the scores plot shown.

Figure 5.7: PC1 vs. PC2 scores plot for all features, H. Rahmati
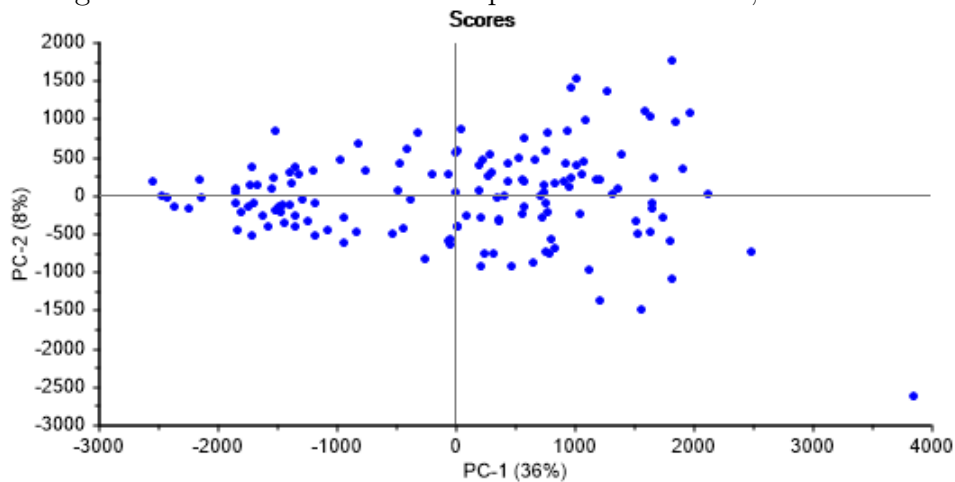
Here it can be seen that the plot is actually quite similar to the one in Figure 5.1, seeing as most of the infants is gathered along the line for PC1. A quick check of the samples also confirms that there is no way to distinguish which of the infants that have normal or abnormal FMs. This is a bit surprising, since one would think from previous results that this would look more promising (see Rahmati et al. [15]). One single infant stand out, which is labeled 082. This might be a possible outlier, and must be treated with care.

An even more surprising plot is the correlation loadings, which can be seen in Figure 5.8. Here most of the features are gathered together, but a slightly troubling point is that almost all the features are found inside the circle representing 50% explained variance. It can be seen that while PC1 describes 36% of the explained variance, PC2 only describes 8%. Most likely is this because there are more PCs included, so the variability becomes more spread out. However, there is a large drop between PC1 and PC2 that might indicate that the data include some sort of noise.

In Figure 5.9 is the influence plot shown. Here almost all the samples are found close together on the left side, while some are more spread out on Hotelling's T^2 or the F-Residuals. In addition to this does all of them seem to be okay except for one single infant, namely 082. This is the one that also stood out on the scores plot, and since it is above both lines on this plot it is most likely an outlier. The smartest thing will probably be to exclude this when the PLSR is run later on.

The explained variance is shown in Figure 5.10. Here it is seen that the highest jump is for PC1, before it slowly moves towards 100% explained variance. The blue line is very

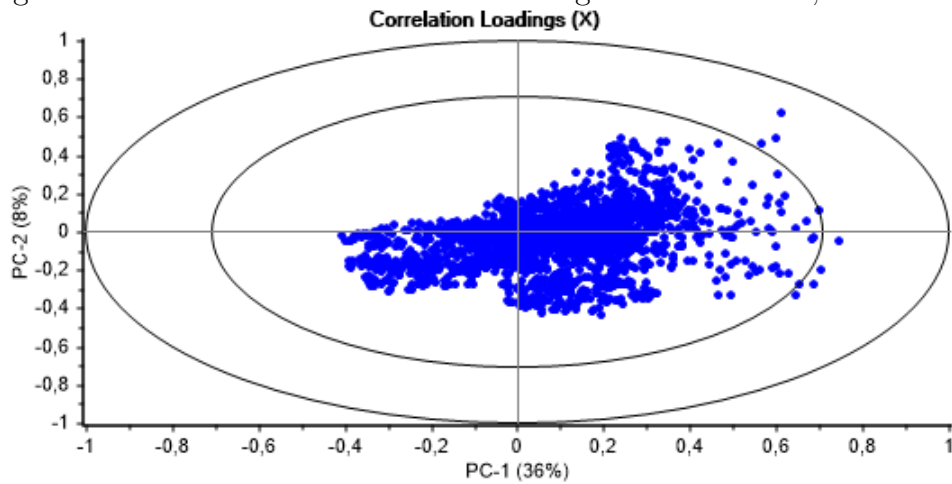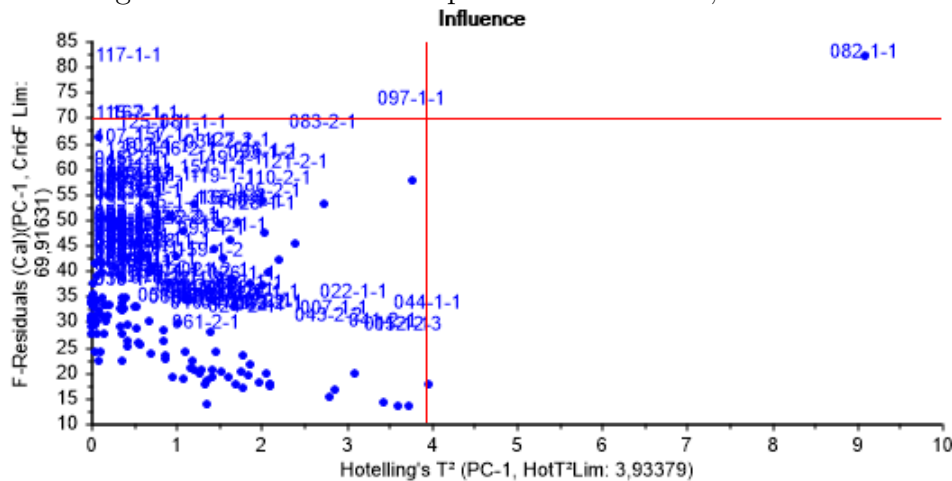Figure 5.8: PC1 vs. PC2 correlation loadings for all features, H. Rahmati

Figure 5.9: PC1 influence plot for all features, H. Rahmati

smooth, while the red line is a little more uneven although it follows the model well. In total there are 25 PCs, but if more were included the lines would probably have continued to rise slowly.

5.2 PLSR

5.2.1 Normalized infants from St. Olavs Hospital

All features, full PLSR

The first PLSR was run on the normalized infants with all the features included. This was done to compare it with the results from the PCA, while also hopefully confirming if the normalization increased the method's predictive ability.

Figure 5.10: Explained variance for all features, H. Rahmati
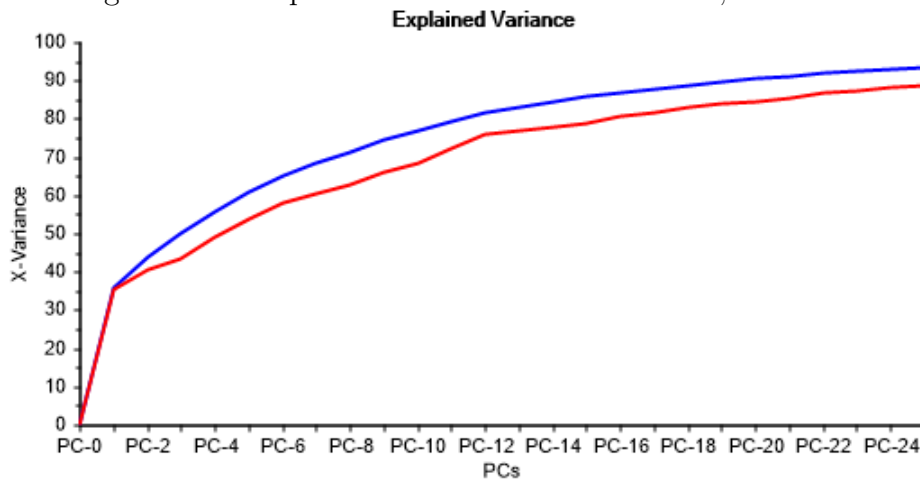

Figure 5.11: Factor 1 vs. Factor 2 scores and validation plot for all features, St. Olavs Hospital
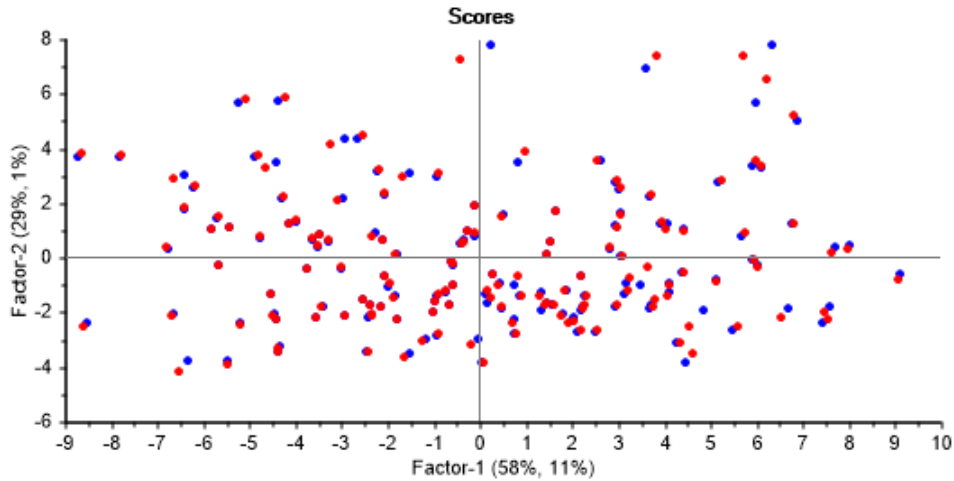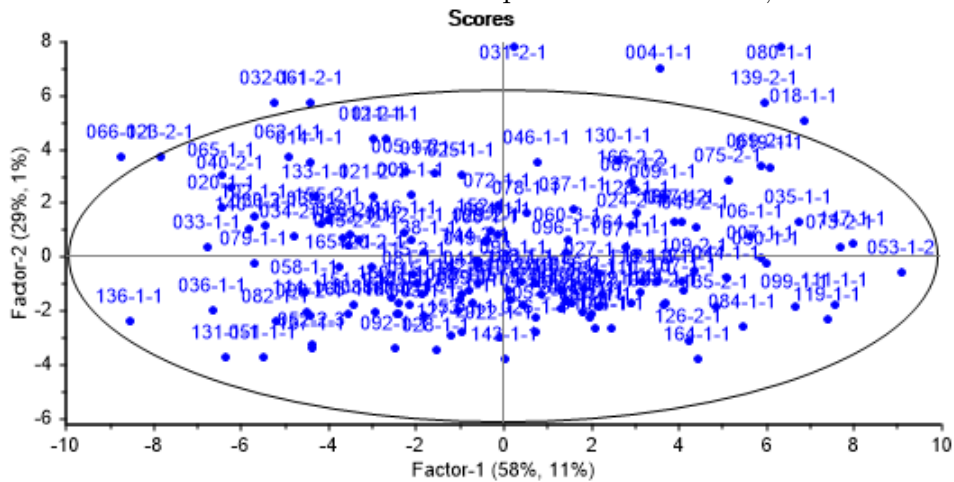

Figure 5.12: Factor 1 vs. Factor 2 scores plot for all features, St. Olavs Hospital

Two different scores plot can be seen in Figure 5.11 and Figure 5.12. The first figure shows how well the validated samples, which is the red ones, correspond to the calibrated samples. Here it is shown that they are quite similar in most cases. The other figure says something about possible outliers, which is the samples found outside of the circle. However, most of these samples are close to the circle and since they seemed fine in the influence plot in Figure 5.3 it is probably safe to assume that they don't affect the model too strongly.

The correlation loadings are shown in Figure 5.13. The first thing one may notice is how the samples have mirrored themselves from the PCA case. This is irrelevant and have nothing to do with the data itself, but it has to do with the configuration of the program used. It can also be seen that some of the samples have small circles around them, indicating that they are important features in building the model. This is all the features except Q_{sd}, which is a little surprising since this feature should have new information compared to the C_x_sd and C_y_sd. Still, this might be because the C_{sd} affects the model so strongly that also it's correlated features count as much on the first and most important factor.

The sample Diag_Arr is also seen, which stands for diagnosis array, or the classification. This is mostly described by the first factor, which is the same as the features describing the centroid of motion. However, the classification barely has any explained variance, seeing as it is far below the inner circle of 50%.

Figure 5.13: Factor 1 vs. Factor 2 correlation loadings for all features, St. Olavs Hospital
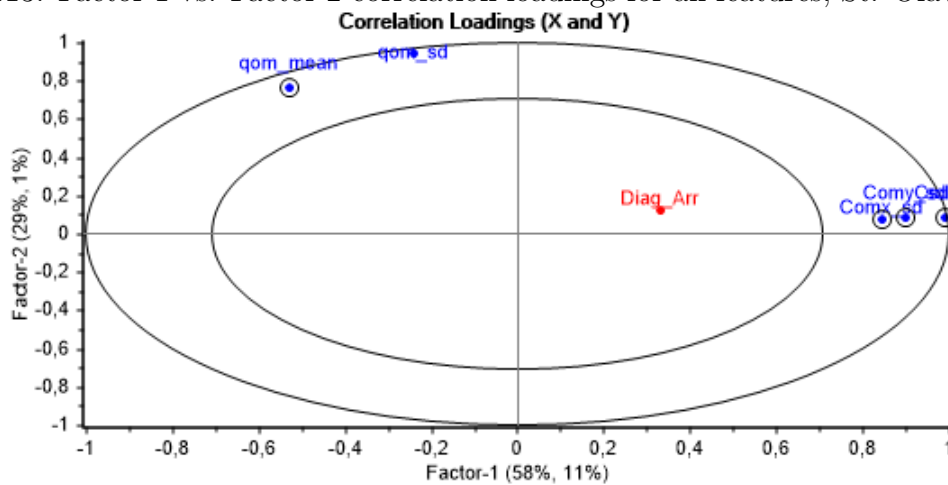

Figure 5.14 shows the explained variance for both the calibration and the validation. The blue line, which is the calibration, shows how well the calibration fits to the model.

This reaches at best 16,63%, which is a very weak performance. The red line shows the validation of samples that were excluded from building the model, which has it's peak at 11%. This has almost no increase after the first factor, and does not manage to follow the line of the calibration.

Figure 5.14: Explained variance for all features, St. Olavs Hospital
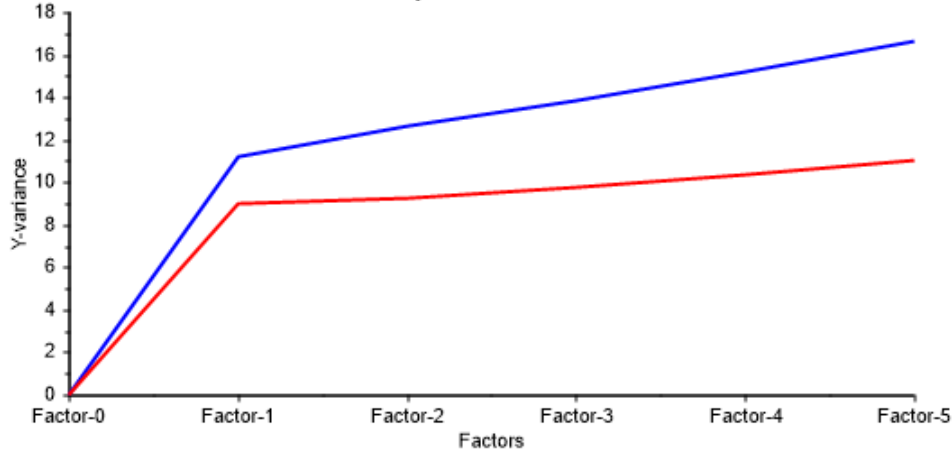
So far the results seem to be quite weak, which is at last confirmed in Figure 5.15. The predicted values of these samples reaches at best 0,63, which is far below the value of 1. This is also the case for infants both with *and* without FMs. Another thing worth mentioning is the root mean square error (RMSE). This explains how far from the desired value the results really are. It is supposed to be as close to 0 as possible, while it here is found to be close to 0,50. This means that the probability of predicting right is slightly larger than it is for predicting wrong, but not by much. Also the slope, which should be close to 1 to indicate good calibration and validation, is quite far below the wanted value.

All features, random PLSR

In the full PLSR, each sample is taken out once while the rest of the samples is used to build the model. However, in the random PLSR, the subjects were divided into 20 segments, where 7 samples were pulled out as unknown for each segment. Usually this have little to say for the scores plot or the correlation loadings, but as it was seen in Lundeby [9], a full or random PLSR can play out differently on the explained variance or the reference vs. predicted plots.

Figures 5.16 and 5.17 show how the two plots are after the random PLSR is run. In the explained variance the change is not too big, although it seems that the validation

Figure 5.15: Predicted vs. Reference for the classification for all features, St. Olavs Hospital
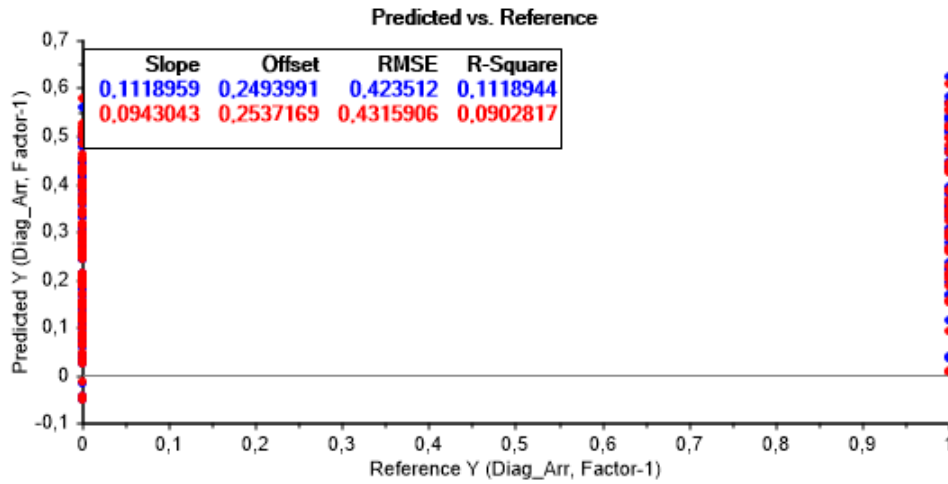

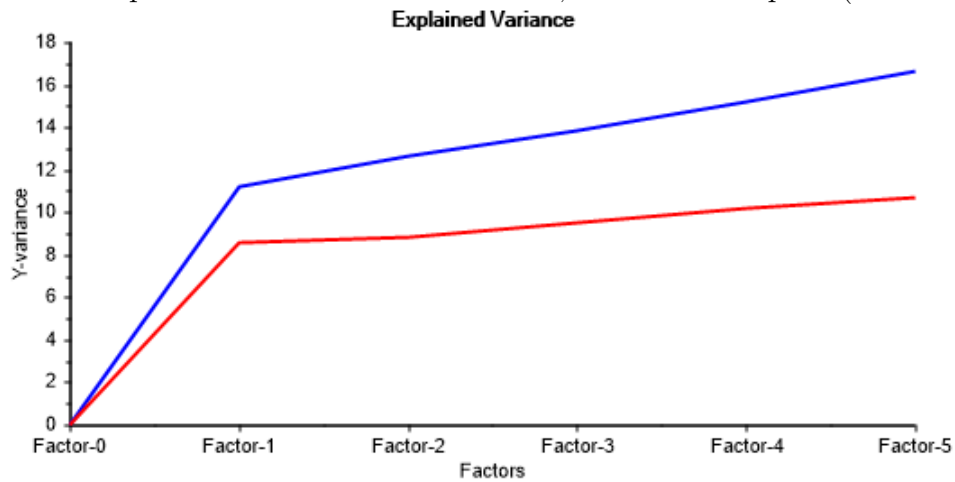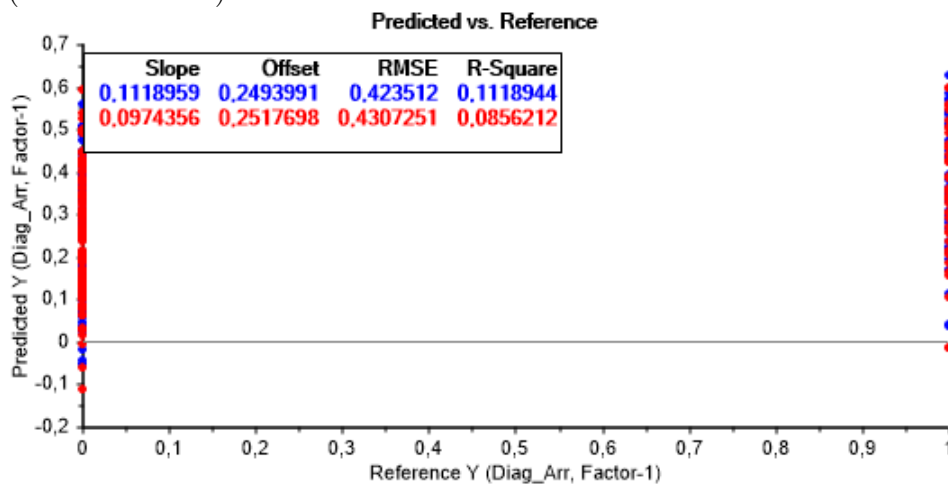Figure 5.16: Explained variance for all features, St. Olavs Hospital (random PLSR)

Figure 5.17: Predicted vs. Reference for the classification for all features, St. Olavs Hospital (random PLSR)

has a slightly larger flat period than the one in the full PLSR. This might be because a possible outlier is pulled out as an unknown, so while the calibration acts similar to before, the validation is weakened by some sample's behaviour. This is also confirmed by the maximum values, which is the same for the calibration, while the validation only reaches 10,67%.

However, if one checks the reference vs. predicted plot, this tells a different story. Although the difference is so small as to be negligible, it can be seen that the validation scores slightly better on the RMSE value. This might indicate that even though the validation had a flat period, it manages to correct itself and actually perform better than in the full PLSR case.

In addition to this, it can be mentioned that the random PLSR was run without some of the possible outliers, such as 004, 031 and 080. The plots are not included here, since the results were so similar to the earlier ones. This may indicate that even though they were placed outside of the circle as possible outliers in the scores plot, they do not affect the model negatively in a too large degree.

C_{sd}, Q_{mean} and Q_{sd}, full PLSR

In this run of the PLSR, the features C_{sd}, Q_{mean} and Q_{sd} were included. This was done so it would be possible to compare it against the PCA run while also keeping most of the known information. The scores plot ended up very similar to the one in Figure 5.12 and is therefore not included. Figure 5.18 shows the correlation loadings, which has had an interesting change from before.

Here it is seen that all the features are explained with approximately 100% explained variance, but now there is a difference in the important features. Earlier it was the C_{sd} and Q_{mean} which were marked as important, while C_{sd} now have been switched out with Q_{sd}. It is difficult to know for sure why this have happened, but one reason might be because while the feature for centroid of motion almost only explain variance in factor 1, the two features for quantity of motion together explain a lot of both factor 1 and 2.

Figure 5.19 shows the explained variance, which is also different from before. It can be seen that the calibration doesn't reach as high as when all the features were included, with a maximum value of 14,63%. However, for the validation the performance has stayed the same with a maximum value right below 11%. An aspect that is positive is that the red

Figure 5.18: Factor 1 vs. Factor 2 correlation loadings for C_{sd}, Q_{mean} and Q_{sd}, St. Olavs Hospital
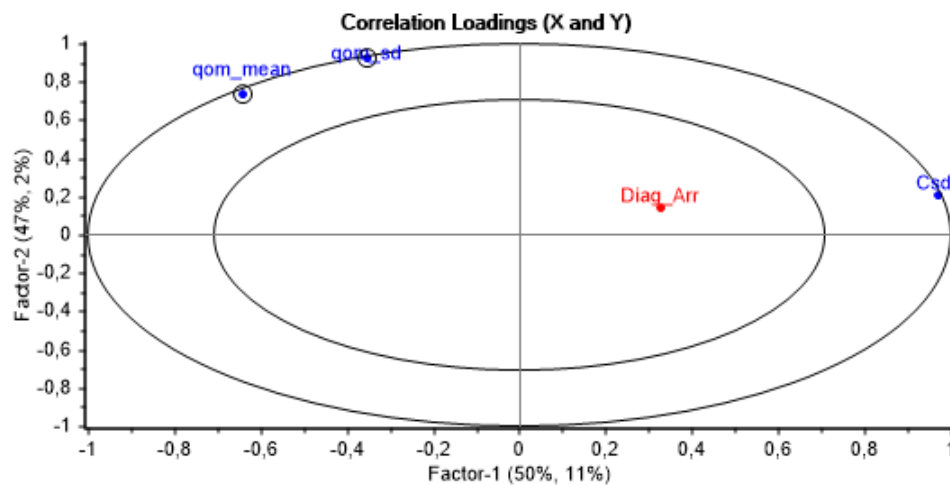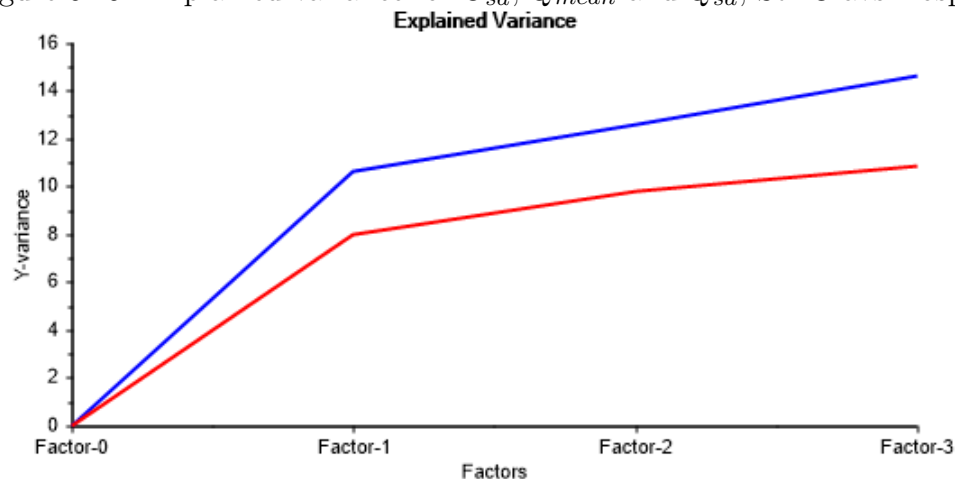

Figure 5.19: Explained variance for C_{sd}, Q_{mean} and Q_{sd}, St. Olavs Hospital

line does not flatten out the same way after factor 1, and the curve follows the calibration much better. On the negative side there is a slight decrease after factor 2. This is probably caused by one of the possible outliers, or a sample that stands slightly out in one of the included features.

Figure 5.20 shows the different results of the reference values and the prediction. Compared to the PLSR where all the features were included, this is a positive development. The RMSE is down to 0,415 for the calibration and 0,427 for the validation, which is the lowest values seen so far. In addition to this, the predicted Y in some cases now reaches the value of 1, although one sample is actually predicted to be 1,37. Again, this sample has received a value which is too high, but there might be two reasons for this. Either the sample is a possible outlier, or the infant is a strong example of how babies with missing

FMs should behave.

Figure 5.20: Predicted vs. Reference for the classification for C_{sd}, Q_{mean} and Q_{sd}, St. Olavs Hospital
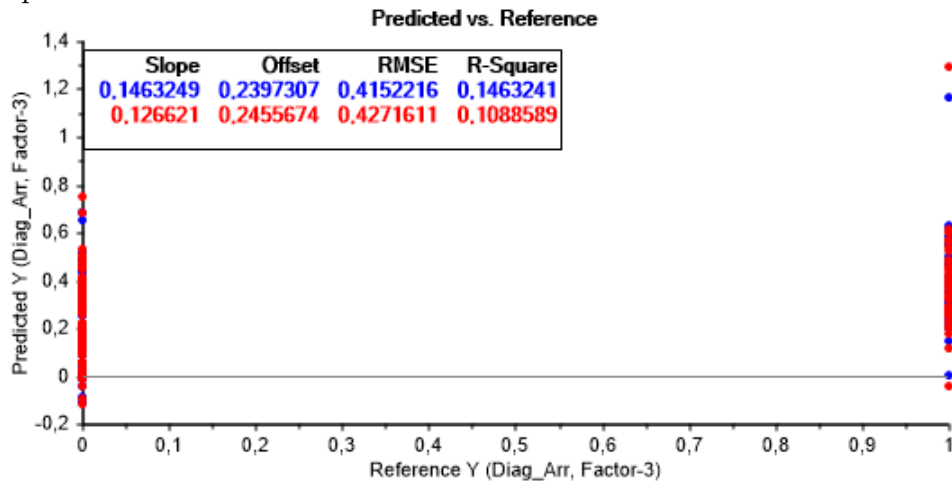
C_{sd}, Q_{mean} and Q_{sd}, random PLSR

In the random PLSR case, the scores plot and correlation loadings are not included because of their similarity to the plots from the full PLSR case of all the features.

Figures 5.21 and 5.22 show the relevant plots for the current PLSR. While the calibration still does not reach the same height as in the PLSR when all the features were included, the similarity in behaviour is much larger between the calibration and the validation. The maximum values are 14,63% for the calibration, just as before, while the validation reaches a new height of 11,56%. The difference is not big, but there is an improvement. This can also be seen by examining the red line, which does not drop as much after factor 2.

The same development is again shown in the reference vs. prediction plot, where the RMSE again has a slight decrease for the validation. This gives a predictive ability of 57,44% at the best.

Also here the PLSR was run without some of the possible outliers. The figures are not included because of their similarity to the others, but the results are worth mentioning. For the explained variance, the calibration had a slight improvement to a value of 15,04%, while the validation stayed the same as before. For the RMSE the results again improved a tiny bit, with a calibration of 0,411 and a validation of 0,423. Still it is important to mention that even though there is a slight improvement from earlier, all of these results

Figure 5.21: Explained variance for C_{sd}, Q_{mean} and Q_{sd}, St. Olavs Hospital (random PLSR)
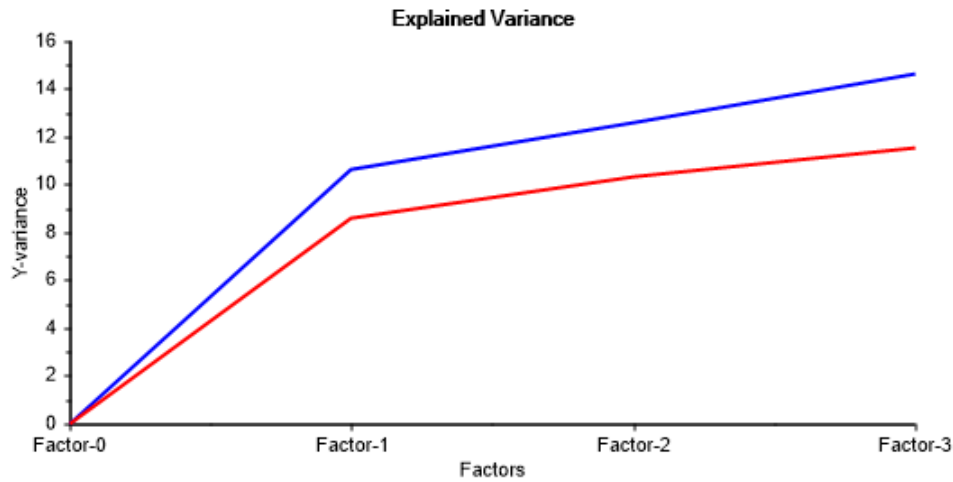
Figure 5.22: Predicted vs. Reference for the classification for C_{sd}, Q_{mean} and Q_{sd}, St. Olavs Hospital (random PLSR)
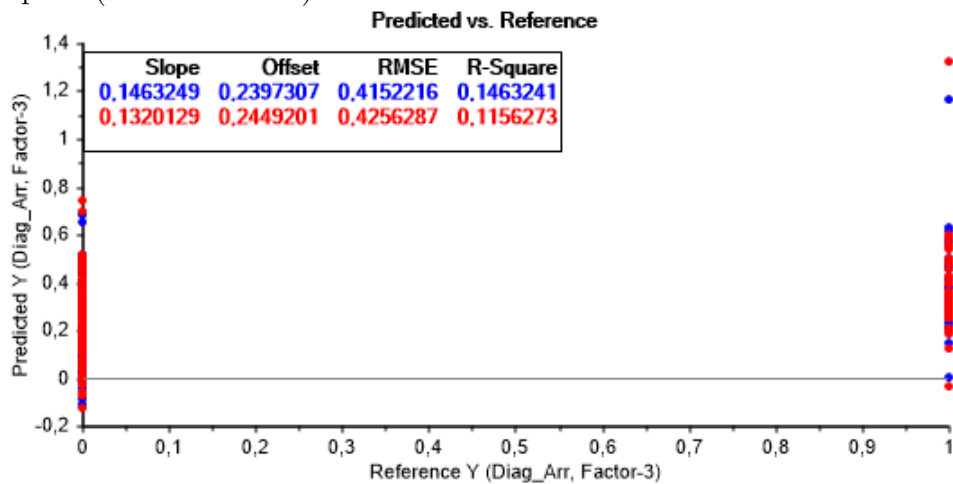
are very weak.

5.2.2 Features from Hodjat Rahmati

All features included, full PLSR

This round of the PLSR was run with all the features included, as in the PCA case. It was also run both with and without the possible outlier 082, and although the change was small it was a slightly better performance without it.

Figures 5.23 and 5.24 show the two different scores plots. In the first plot one can see the comparison between the modelled and the validated samples. In the middle the samples are very close together, while some of the samples found further to the side have

more distance between each other.

Figure 5.23: Factor 1 vs. Factor 2 scores and validation plot for all features, H. Rahmati
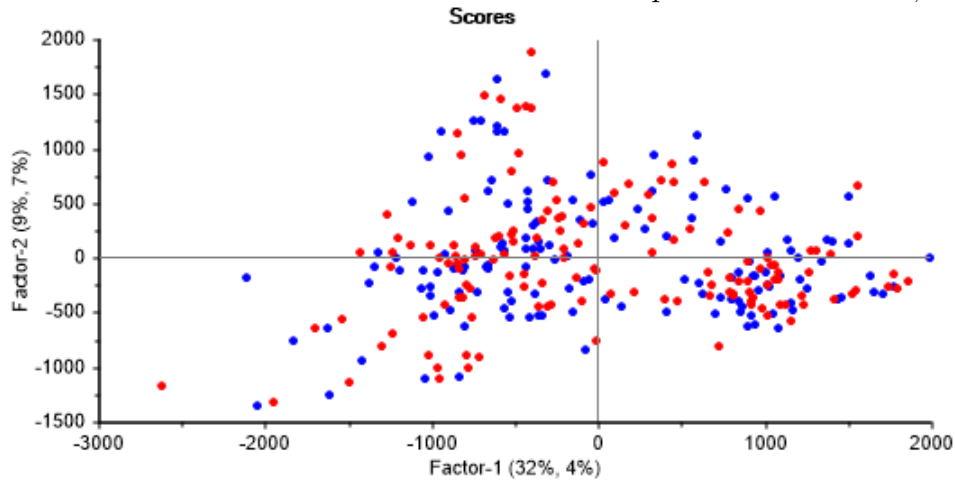
In the other plot are possible outliers found. They must all be treated with care, in case they affect the results strongly in some way. However, since none of them were found as outliers in the influence plot of the PCA, it is probably safe to assume it will be okay.

Figure 5.24: Factor 1 vs. Factor 2 scores plot for all features, H. Rahmati
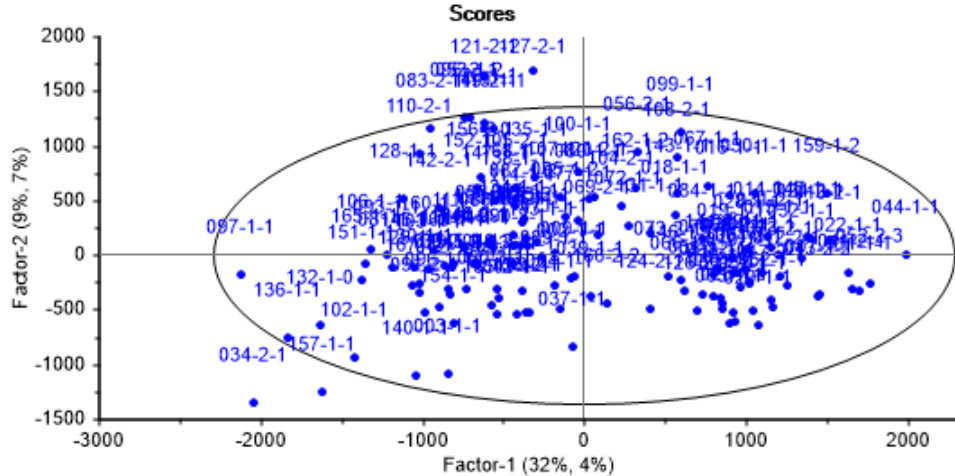
The correlation loadings can be seen in Figure 5.25. As before has the plot now mirrored itself, while no other big changes can be seen. The only difference is the red spot, showing where the classification is placed. This is also found quite far below the line for 50% explained variance, which is not a good sign for further results.

Figure 5.26 shows the explained variance, which is quite different than expected. On one hand, the calibration of the model seems to be quite good, reaching a value of 93,85% explained variance. However, for some reason the validation now has a negative explained

Figure 5.25: Factor 1 vs. Factor 2 correlation loadings for all features, H. Rahmati
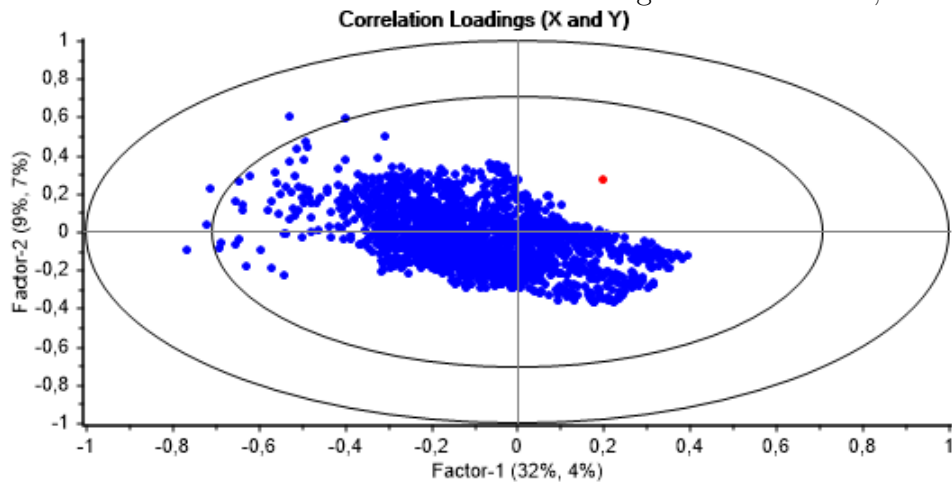
variance of -185,96%.

Figure 5.26: Explained variance for all features, H. Rahmati
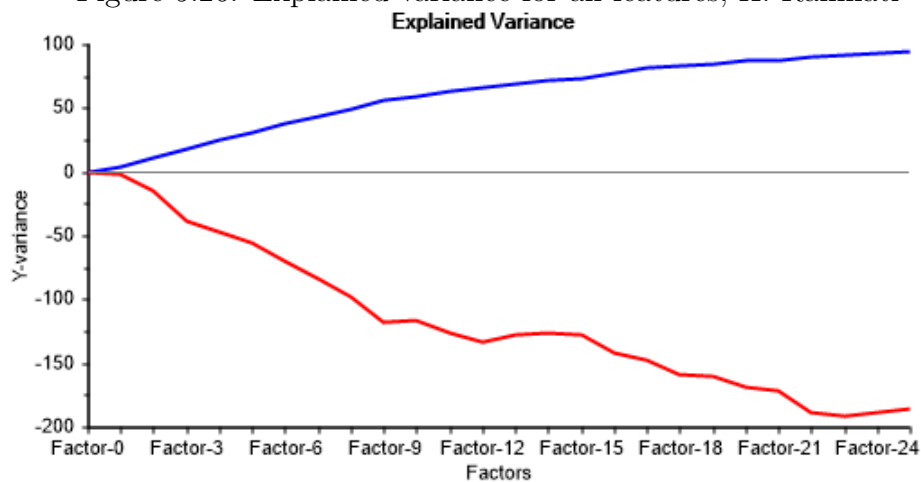
The last plot included in this PLSR is shown in Figure 5.27. This shows that the prediction error is quite large. While the slope has a calibrated value of 0,039 and a validated value of 0,005, the RMSE is 0,442 and 0,458. This is actually even worse than the results produced by the features from St. Olavs Hospital. In addition to this is none of the samples found above the value of 0.45, which is the case for infants both with and without present FMs.

All features included, random PLSR

These features were also also run with a random set, just as it was done for the features from St. Olavs Hospital. Since the scores plot and correlation loadings are just the same,

Figure 5.27: Predicted vs. Reference for the classification for all features, H. Rahmati
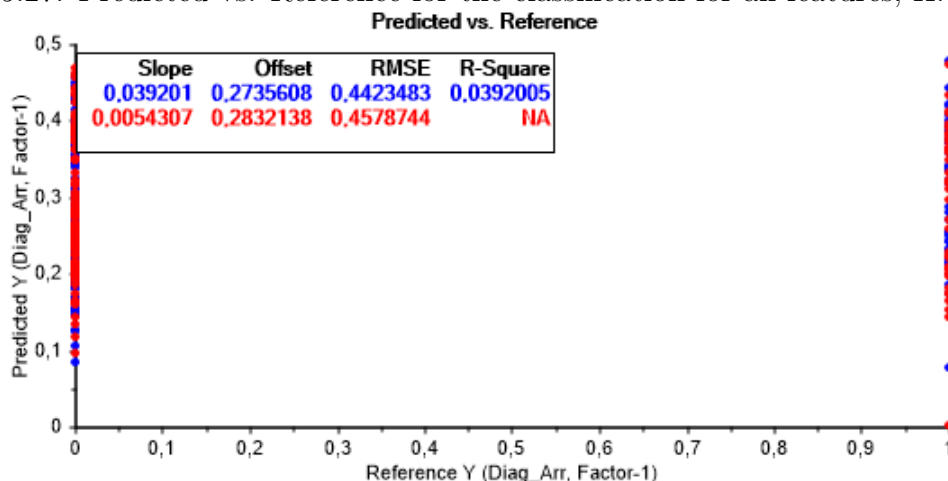
the first plot presented is the explained variance in Figure 5.28. While the calibration seems to be exactly the same as before, reaching a value of 93,85% explained variance, the validation performs slightly better with a value of -180,73%.

Figure 5.28: Explained variance for all features, H. Rahmati (random PLSR)
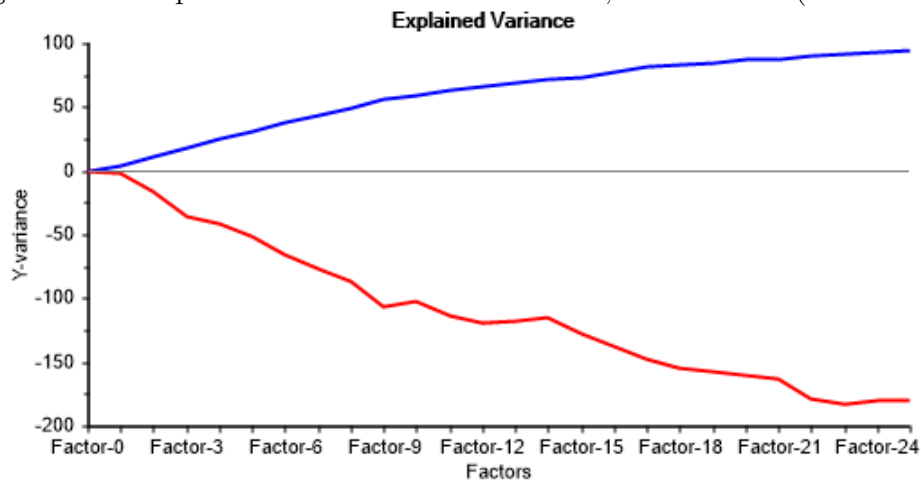
This is also confirmed in Figure 5.29. In this plot are the slope and RMSE values for the calibration the same, while both of the validated values have increased performance. The slope value reaches a new height of 0,008, while the RMSE is equal to 0,457. A positive effect of the random PLSR, is that the samples are predicted with slightly higher values than before.

But why are the results so weak compared to the ones in earlier analyses? In Rahmati et al. [15] the video-based results performed with a sensitivity of 86% and a specificity of 92%, which is even better than the results from St. Olavs Hospital. In this case the

Figure 5.29: Predicted vs. Reference for the classification for all features, H. Rahmati (random PLSR)
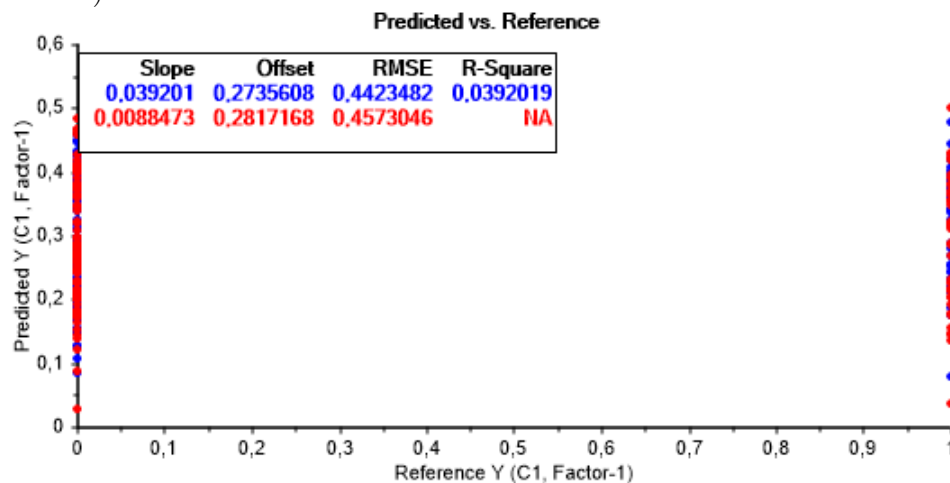
results seem to be more by chance than an actual prediction. One possibility is that the tracking part of the infants might be too slow. According to Adde [2], the infants without FMs have jerky movements to one side, while the infants with FMs have more movements with the whole body. When the infants move in a jerky fashion, such as the ones with abnormal FMs usually do, the trackers have problems following the motion at times. Seeing as these movements are quite important for the detection of unusual behaviour, the results might be weakened. However, in this case it should have had the same effect on H. Rahmati's results.

5.3 Other Approaches

Seeing as the results for both methods are so weak, it is important to try and find out where the problem is placed. Is it in the data, the classification or the videos? While it might be difficult to check this for either the classification or the videos, the data can be examined in different ways. One approach might be to run a PLSR on both data sets together, to see if there is any comparable structure between the two. The other possibility is to run a cluster analysis, to see if the methods actually manage to find some distinguishing classification.

5.3.1   PLSR

In this analysis the PLSR was run with the features from H. Rahmati's method as the explained variables X, while the features from St. Olavs Hospital were the Y variables to be predicted. This was done to see if there would be any correlation between the two data sets. The classification was also included, but downweighted, so that it would not influence the results in any way. Most of the plots are not included here, but by using the correlation loadings it is possible to see if there is any coherence between the two data sets.

Figure 5.30: Factor 1 vs. Factor 2 correlation loadings for H. Rahmati and St. Olavs Hospital
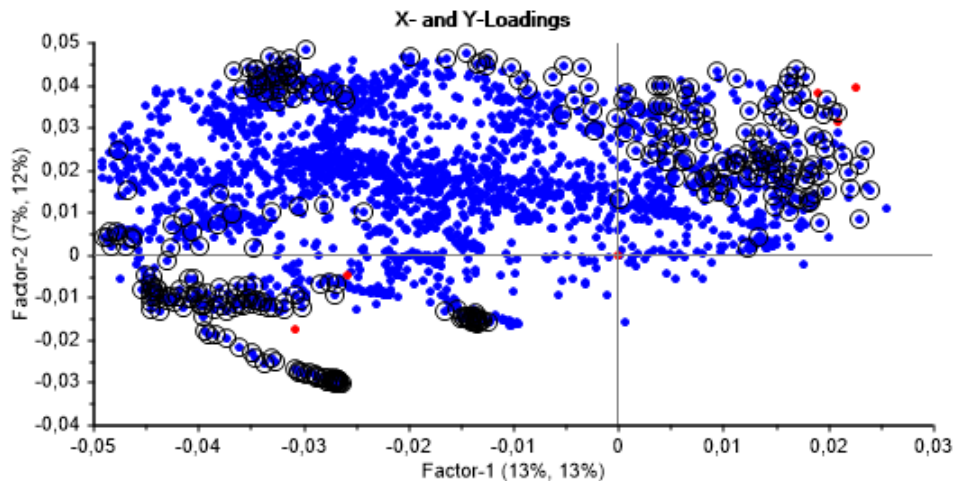
In Figure 5.30 is the plot for Factor 1 versus Factor 2 included. The blue dots are the features from H. Rahmati's method, while the red ones, which might be a little difficult to see, is from St. Olavs Hospital. One red is also found in the middle, which is the classification. To make the features more visible, the circles for explained variance have not been included. In this plot are the red features found in two groupings, just as in the analyses done in Sections 5.1.1 and 5.2.1. The ones in the upper right corner is the features describing the centroid of motion, and it is expected that these will have some correspondence to the head and trunk features from H. Rahmati. By studying the labels it can be concluded that even though a few of the closest features represent an arm or a leg, there is a majority of head or trunk variables. These are then positively correlated with the centroid of motion features. The same is also confirmed for the quantity of motion features, which are strongly correlated with different arms and legs features. As a conclusion one can say that yes, there exists a structure between the two data sets.

5.3.2 Cluster Analysis

Since it has been confirmed that the data has structure, it might be an idea to see if it manages to find some kind of classification itself. This is done by performing a cluster analysis, which is described in Section 3.3.

Figure 5.31: Cluster analysis of H. Rahmati and St. Olavs Hospital
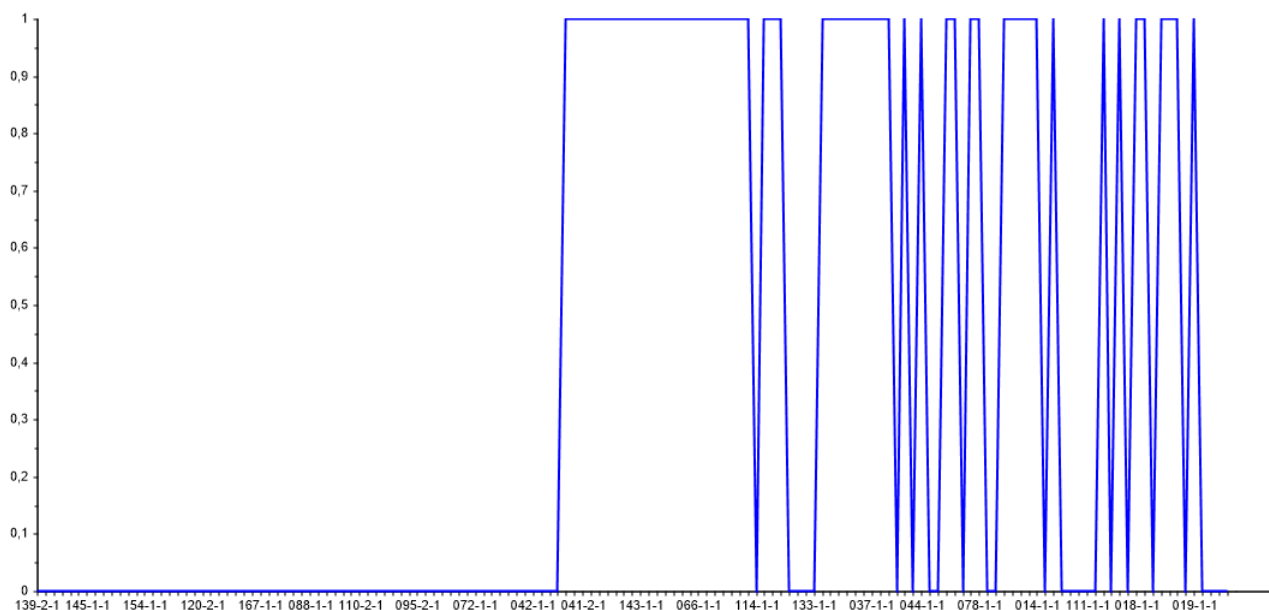
The resulting plot can be seen in Figure 5.31. The left part of the plot contains only healthy infants, which is probably because they have very strong present FMs. On the other hand are the parts with abnormal FMs shorter in interval with healthy babies scattered in between. This might be because some of the infants are questionable in their classification, in addition to some babies being wrongly diagnosed. In reality there are 104 out of 145 infants that are healthy, and it seems to be more than 41 babies with abnormal FMs in this figure. Still, it is obvious that if one uses H. Rahmati and St. Olavs Hospital's data together, they find some sort of classification. One might ask then, if there is something wrong with the classification, or if there is something else that might be easier to find than FMs?

Since it has been shown that the data sets actually have some structure, and it is obvious that they find some kind of classification, the two problems left are the classification from St. Olavs Hospital and the videos. Seeing as the real diagnosis of the infants will be ready this summer, it is not possible to check the classification before this. Neither is

there any easy way to test if the videos affect the results in some way. Therefore it will be relevant to check these possibilities later on.

Chapter 6

Summary and Recommendations for Further Work

6.1 Summary and Conclusions

This report has presented two different methods for detecting cerebral palsy in infants. While one of them focuses on the centroid and quantity of motion, the other focuses on the variability of the movements. These methods are quite different, and this is also shown by the different results given earlier in the report.

While Chapter 2 gave a theoretical and mathematical description of the two methods, Chapter 3 explained the statistical approach used for the analysis. In Chapter 4 the infants were introduced, before the results were presented in Chapter 5. This chapter was divided into three different sections, one containing the PCA, one the PLSR, while the last section presented different possibilities and similarities between the two different computer algorithms.

The first part of the PCA contained the results of the analysis performed on the normalized infants from St. Olavs Hospital. These results were quite similar to the ones found in Lundeby [9], which proved that the normalization of the infants did not affect the results. Although the PCA was run both with all the features included, and only C_{sd}, Q_{mean} and Q_{sd}, the only differences found were in the influence plot. Here it was seen that there was a slight change in which infants that were fitted well to the model, and which infants that stood out. On the other hand did neither set have any clear outliers to worry about.

In the PLSR there was little difference in performance of the predictive ability on the two data sets. The best result was found with C_{sd}, Q_{mean} and Q_{sd}, which gave a predictive ability of 58,86% for the calibration and 57,77% for the validation. This is more than 10% lower than the results found by Lars Adde, and there might be different reasons for this. One of them might be that the validations used are quite different. While they have built a model from one set of babies, they have tested the performance of this on a completely different set. In this report the validation is done on the same set of infants that the model is based upon.

Already in the PCA analysis of H. Rahmati's method could one see some differences. Although the scores plot were quite similar to the one from St. Olavs Hospital, the correlation loadings showed that only a few features had over 50% explained variance. This is expected since the feature set is so much larger, however, the drop from PC1 to PC2 is quite large. In the influence plot there was one clear outlier, but when the PLSR was run without this sample it gave only a slight increase in results. Even though the RMSE values were not that much higher than in the feature set from St. Olavs Hospital, the explained variance for the validation ended up with a value of -180,73% at best.

At the end of Chapter 5 a few of the possible problems were discussed. These could be in either the data, the videos or the classification. By doing a PLSR with H. Rahmati and St. Olavs Hospital's data against each other, in addition to a cluster analysis, it was concluded that the problem is most likely not in the data itself. Therefore it is probably in either the videos or the classification.

It is a little difficult to give a final conclusion as to which method that performs the best, since the results are so weak. However, since the validation in the PLSR from St. Olavs Hospital seems to follow the calibration, even though this is weak, it is concluded that this method is slightly better. Even so are both methods close to the same RMSE value, which means that they have almost the same error in predicting the right FM.

6.2   Discussion

While the problem most likely lies in any of the two points above, there are a few weaknesses in H. Rahmati's method that should be considered. In Rahmati et al. [15] it is said that the number of frequency bins play a large role in the results of the analysis. While

too few bins usually causes one to lose information, too many may lead to overfitting. There, the optimal number was found to be 99 bins, which is also the number used in this report. However, it was also tested with slight changes such as two more or less bins, or drastic ones such as the double number, although this did not influence the results much in any way. How can it be that in one case the number of bins may lead to completely different results, while in another it has nothing to say?

There are a few other limitations to this method as well. The first one is the segmentation, and although this is a small problem it might affect the results. In some cases the trunk had very little movement and the optical flow part would therefore not create any flow vectors. In these cases the segmented trajectories for the trunk usually ended up on the left or right side of the infant, which is not the place they were intended. One would perhaps think that this would not play a role on the results, but in some cases the trunk trajectories would end up close to a leg or arm, causing much more movement to be included in this segment than what it normally would be.

Another weakness is the tracker, which had a tendency to lag behind the largest movements. This was also the case when H. Rahmati ran the tracker (see Rahmati [13]), but who knows how much lag that is to be expected? In some cases the tracker disappeared completely out of the frame, and it was impossible to know the movement of it afterwards. Since there is so much uncertainty at this step, it might be more correct with another solution at the end than the FFT. This is after all used to find the motion variability, but if the strongest variations is lost in the tracking it might not be as suitable as first believed.

6.3 Recommendations for Further Work

There are still many things that can be done in this analysis. One of the main points is to check both the classification and the videos, to see if they have affected the results. As earlier mentioned is the final diagnosis of the infants not ready yet. However, they will be this summer, at which point it would be relevant to check the results once again. In the analysis done by H. Rahmati in Rahmati [12], was the CP diagnosis used instead of FMs. It could also be interesting to test these data with the earlier FM classification, to see how it would perform then.

Another possibility as to why the results were so much better in H. Rahmati's analysis, might be the division of infants. While the data set from before included only 78 infants, this one includes 146. In addition to this were only 14 of the infants classified with abnormal FMs there, while there are 41 in this. Maybe those infants from the earlier analysis had stronger movements with less uncertainties than in this set? Since the data set here is doubled in size, it is a chance that more of these infants are uncertain, neither fitting right into normal or abnormal FMs.

When considering the videos there are one thing that is different from before, namely the resolution. While the videos that H. Rahmati received had been edited repeatedly (see Adde [2]) and had a lower pixel resolution, the new videos have a much higher resolution because of a new camera being used. A higher resolution leads to a higher number of pixels that each contains more detail in each frame (see Leurs [8]). Even though one would think that this is positive, it is also a risk that the tracker does not behave as it should. Because there is much more detail in each pixel, the tracker might get stuck in a local minimum somewhere else than on the intended place. The only negative part with this hypothesis, is that it only considers why H. Rahmati's results were so weakened. Why are also the results from St. Olavs Hospital so much worse than before? There might be different reasons for this, such as the evaluation. However, there is another possibility as well. In the cluster analysis, it was shown that the data managed to find a division between the two classifications. Maybe there is something else than FMs that is easier to use as a sign of CP?

One last possibility that is hopefully not correct is the hypothesis. What if it is not possible to create a computer algorithm that will be able to distinguish between the different FMs? It might be that the only way to do this is by using ones own eyes, the way they do in the GMA. From the results in Chapter 5 it seemed that most infants were predicted with values around 0,04, and maybe that is because most of them are placed in a grey zone between normal and abnormal FMs.

So far the suggestions on further work have dealt with possible ways to check how the results are so bad. However, there is also another way to go. Because of the possibility that the tracking removes some of the most distinct variations in movement, it could also be an idea to come up with some other features than the ones generated from the FFT. After the tracking is done the data includes one trajectory for each of the segments, and the question

is then if there exist any calculations that could be performed on these trajectories that would increase the detection of CP? Since the infants with FMs usually have more similar movements in the different limbs, it might be possible to use the correlation as a tool. As an example would probably the correlation between the right and left arm be much higher for an infant with present FMs, than for an infant without. One might also assume in the same way that the variations between limbs must be higher in infants with abnormal FMs, since their movements usually take place on one side.

No matter what the next step is, there are still many things that can be done. Hopefully, one day there will be a fully working computer algorithm that will make detection of CP much easier.

Bibliography

[1] Adde, L. (2010). *Prediction of cerebral palsy in young infants: Computer-based assessment of general movements.* PhD thesis, Norwegian University of Science and Technology.

[2] Adde, L. (2016). Personal communication.

[3] Adde, L., Helbostad, J. L., Jensenius, A. R., Taraldsen, G., and Støen, R. (2009). Using computer-based video analysis in the study of fidgety movements. *Early Human Development.*

[4] Bakke, T. and Moster, D. (2015). Fakta om cerebral parese (CP). http://www.fhi.no/tema/cerebral-parese/fakta-om-cerebral-parese.

[5] Dell (2015). Cluster analysis. http://documents.software.dell.com/Statistics/Textbook/cluster-analysis.

[6] Eriksen, H. Ø. (2010). Hva er CP. http://www.cp.no/index.asp?id=39228.

[7] Jensenius, A. R., Nymoen, K., Skogstad, S. A., and Voldsund, A. (2012). Study of the noise-level in two infrared marker-based motion capture systems. https://www.duo.uio.no/bitstream/handle/10852/26906/Jensenius_SMC_2012a.pdf?sequence=1&isAllowed=y.

[8] Leurs, L. Image resolution. http://www.prepressure.com/design/basics/resolution.

[9] Lundeby, H. G. (2015). A comparison of features for early diagnosis of cerebral palsy.

[10] Martens, H. and Martens, M. (2000). *Multivariate Analysis of Quality: An Introduction.* Wiley.

[11] NRK (2014). Schrödingers katt 16.10.2014.

[12] Rahmati, H. (2015). *Computer Vision-Based Infant Movement Assessment.* PhD thesis, Norwegian University of Science and Technology.

[13] Rahmati, H. (2016). Personal communication.

[14] Rahmati, H., Dragon, R., Aamo, O. M., Adde, L., Stavdahl, Ø., and Gool, L. V. (2015a). Weakly supervised motion segmentation with particle matching. *Computer Vision and Image Understanding.*

[15] Rahmati, H., Martens, H., Aamo, O. M., Stavdahl, Ø., Støen, R., and Adde, L. (2015b). Frequency analysis and feature reduction method for prediction of cerebral palsy in young infants. *IEEE Transactions on Neural Systems and Rehabilitation Engineering.*

[16] Rahmati, H., Martens, H., Aamo, O. M., Stavdahl, Ø., Støen, R., and Adde, L. (2015c). Frequency-based features for early cerebral palsy prediction. *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE.*

[17] Sinha, U. Image moments. http://aishack.in/tutorials/image-moments/.

[18] Stahl, A., Schellewald, C., Stavdahl, Ø., Aamo, O. M., Adde, L., and Kirkerød, H. (2012). An optical flow-based method to predict infantile cerebral palsy. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6186829.

[19] Valle, S. C., Støen, R., Sæther, R., Jensenius, A. R., and Adde, L. (2015). Test-retest reliability of computer-based video analysis of general movements in healthy term-born infants. *Early Human Development.*