



ARTIFICIAL INTELLIGENCE TO EVALUATE THE SHORT-TERM PROGRESS OF DEVICE ASSISTED SCOLIOSIS THERAPY ON THE EXAMPLE OF FED METHOD

Paula Schumann¹, Andreas Heinke¹, Thurid Jochim¹, Tilman Lieberknecht¹, Jenny Nisser², Steffen Derlien², Zbigniew Śliwiński³, Hagen Malberg¹, Grzegorz Śliwiński^{1,3}

¹Institute of Biomedical Engineering, Technische Universität Dresden, Germany

²Institute of physiotherapy, University Hospital Jena, Germany

³Institute of physiotherapy, UJK Kielce, Poland

Abstract

X-Ray or video raster stereography are used for the progress control of the *FED* therapy but applied only at intervals of months. A short-term evaluation would allow to adjust the therapy parameters based on the individual therapy progression and could also provide a direct feedback for patient. Therefore, this study aims to isolate parameters for a short-term progression monitoring by applying machine learning algorithms on a set of 130 posture characteristics. A measuring procedure using the *DIERS formetric 4D* optical measuring system was developed and validated on six patients. The measuring procedure was repeated eight times (four days, each morning and afternoon). Eight parameters were evaluated. The *Wilcoxon* signed rank test and the *Friedman* test were used to verify the statistical significance. In order to identify small changes in posture correlating with the applied treatment a hierarchical cluster analysis was performed. The evaluation shows that the parameters pelvic tilt, kyphosis angle and lordosis angle changed significantly between the individual measuring points, but not across all eight parameters. The data is highly dependent on the daily form and cooperation of the patient. The cluster classification is not determined on the basis of the four measurement points, but on the basis of patient individuality. Hierarchical clustering can classify new patients to match them with successful treatment plans of similar cases. By further optimizing the setting parameters a better cluster result should be achieved. More measurements will be made to expand the database. In order to obtain a short-term patient monitoring, other methods of artificial intelligence especially neural networks will be considered.

Introduction

Scoliosis is a spine deformity characterized by lateral and rotational curvature of the spine. Most prevalent is idiopathic scoliosis, which emerges during the

pubertal growth affecting approximately 4 % of children under age 16. Scoliosis and its treatment are highly individual. An effective therapy aims to reduce the risk of curve progression that indicates a surgery. As for today, physicians can evaluate the

130 parameters, eight were selected for statistical evaluation, where the causal relationship to posture is given and also plays a part in diagnosing scoliosis. These Table 1.

eight parameters cover all three body planes. The posture parameters and the corresponding plane are described in

Table 1 Description of the eight parameters and their assignment to body plane and body segment

Segment	Plane	Parameter	Description
Pelvic	Frontal	Pelvic elevation	Difference between the height of the dimples (in mm)
	Sagittal	Pelvic tilt	Angle of line connecting right dimple and left dimple and an external horizontal line (in °)
		Pelvic torsion	Degree to which the ilia are rotated against each other (in °)
	Transverse	Pelvic rotation	Degree of deviation of right dimple from left dimple (in °)
Trunk	Frontal	Plumb deviation	Lateral deviation of the vertebra prominens from the center of the dimples (in mm)
		Lateral deviation	The Root Mean Square deviation between the spine line and the calculated line connecting the vertebra prominens and the center of the dimples (in mm)
	Sagittal	Kyphosis angle	Degree of dorsal convex curvature of the spine (in °)
		Lordosis angle	Degree of dorsal concave curvature of the spine (in °)

Data acquisition

The measurements with the DIERS formetric 4D system took place at the rehabilitation clinic in Zgorzelec, Poland. Six female patients age 11 to 17 years were recruited. All patients took part in a

stationary rehabilitation program with a duration of four weeks. The patients had double curve scoliosis. All enrolled patients besides patient 3 wore a brace. The degree of severity was between 17 ° and 60 ° Cobb (Table 2).

Table 2 The detailed description of the patient population

Patient	Age (in years)	Thoracic (left or right convex)	Thoracic Cobb (in °)	Lumbar (left or right convex)	Lumbar Cobb (in °)
1	16	right	32	left	60
2	15	right	35	left	28
3	11	right	17	left	26
4	12	right	37	left	26
5	17	left	34	right	33
6	13	left	42	right	50

Each patient had two therapeutic sessions per day, one in the morning and one in the afternoon. Four postural measurements were taken during each session. The measurement plan is shown in **Error! Reference source not found.**

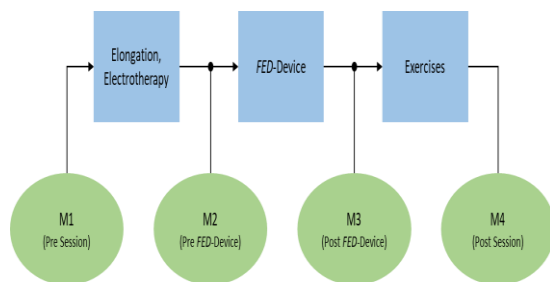


Figure 1 Measuring procedure: M1 to M4 represent the four different measuring points.

Measurements were taken right before muscle warming (M1), right before treatment in the FED device (M2), right after treatment in the FED device (M3) and right after physiotherapeutic exercises (M4). The measurements were repeated over a period of four days at both sessions in eight data sets per patient at each point. The evaluation was carried out on all six patients. Thus, 48 data points emerged for each measuring point (M1 to M4).

4D Average was used as measurement mode. All six patients were familiar with the diagnostic system. Infrared markers were used in case that the anatomical points were not detected by the *DIERS formetric 4D* system. Patient 1 and patient 3 did not need infrared markers. Patient 2 and patient 5 needed markers for the vertebra prominens and dimples, patient 4 and patient 6 only needed them for vertebra prominens.

Statistical methods

For the descriptive statistics box plots visualize the locations and scatter of the data and to mark outliers. Furthermore, the box plots were used to analyze the tendencies. All data was processed with *MATLAB R2018* after exportation from the *DIERS formetric 4D*.

The research question resulted in two research hypotheses. The first research hypothesis claims that there is a difference in the eight parameters between the first and the last (eighth) therapy session. The *Wilcoxon* signed rank test was performed for this purpose. The second research hypothesis claims that the eight parameters change within the four measurement points

M1 to M4. Therefore, the *Friedman* test was performed. The significant results were used to investigate between which measuring points a change of parameters could be observed. Therefore, the *Wilcoxon* signed rank test was applied as post hoc test. Since eight different parameters were examined, the problem with multiple testing arose. To control the familywise error rate, the *Holm* correction was applied. This made it possible to achieve a better power of the tests than with the *Bonferroni* correction [12]. The global significance level α_{global} was set to 5 %.

Hierarchical clustering

In order to check whether the data is suitable for classification, a hierarchical cluster analysis was performed with *MATLAB R2018*. All 130 parameters were used for the analysis. The research hypothesis claims that the data can be divided into four clusters representing the four measuring points M1 to M4. The data was scaled between 0 and 1 to ensure comparability. The *Euclidean distance* was used as distance measure and the *Average Linkage method* as a linkage function. The *Cophenetic correlation coefficient* was applied as a measure of quality for the cluster method, as recommended by

MATLAB R2018 [13]. The coefficient ranges from 0 to 1. A higher value means that the cluster result represents the data more accurately.

Two steps were followed to determine the number of clusters. The first step was to determine the natural cluster classification of the data. For this the *Inconsistency coefficient* was used as a threshold. The coefficient is a measure of the dissimilarity of the links. The highest possible value was taken for the threshold value for which at least 2 clusters were created. The third decimal place of the *Inconsistency coefficients* was looked at. After setting the threshold, the number of clusters and the distribution of objects to the individual clusters were calculated. The second step was to set the number of clusters at four. The resulting cluster distribution was analyzed.

Results

Descriptive evaluation

Boxplots were created for the eight parameters by using the data of the six patients (Figure 2). Any difference of the median level would show a possible therapeutic effect between the individual measuring points (M1 to M4).

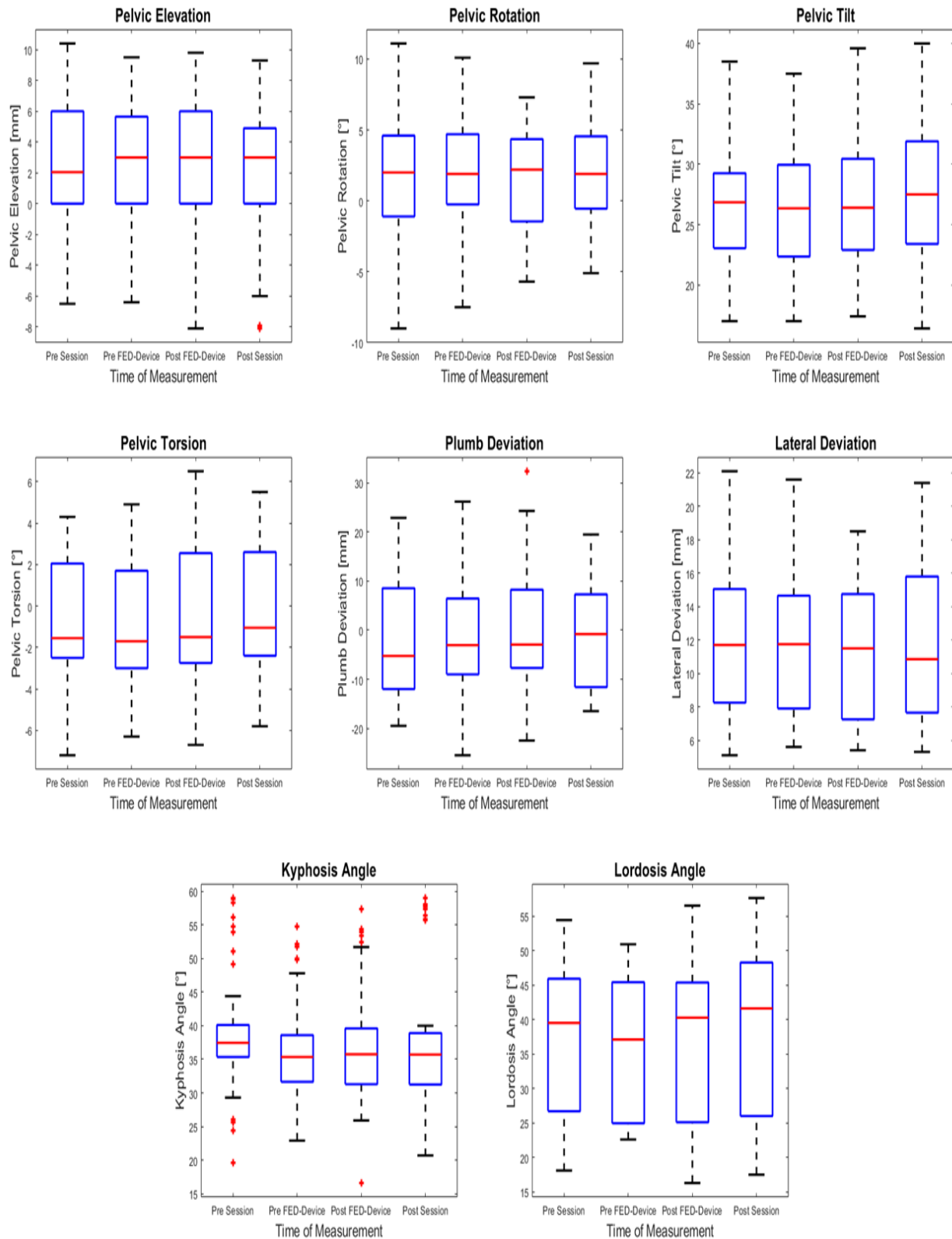


Figure 2 Descriptive statistics: Boxplots of the eight relevant parameters were created. The y-axis represents the absolute values of the parameters and the x-axis represents the four different measurement times (M1 to M4).

The boxplots show a difference in the median values of pelvic elevation, pelvic tilt, plumb deviation, lateral deviation and kyphosis angle between the first and last measurement. However, due to the large

variation of each parameter, no visible difference can be detected. Outliers are present in the parameters pelvic elevation, plumb deviation and especially frequent in the kyphosis angle.

Table 3 Results of the significance test for the eight parameters

Parameter	Wilcoxon signed rank test	Friedman test
	p-value	p-value
Pelvic elevation	1.000	0.331
Pelvic rotation	0.594	0.547
Pelvic tilt	0.438	<0.001*
Pelvic torsion	0.563	0.989
Plumb deviation	0.219	0.519
Lateral deviation	0.438	0.392
Kyphosis angle	0.156	<0.001*
Lordosis angle	0.688	<0.001*

* significant ($p < \alpha_{local}$; $\alpha_{local,i} = \alpha_{global}/(k-i+1)$; number of tests per research hypothesis $k=8$)

The *Wilcoxon* signed rank test indicated that none of the eight parameters changed significantly from the beginning of the measurement to the end of the measurement. The parameters with the lowest p-value were plumb deviation ($p = 0.22$) and kyphosis angle ($p = 0.16$).

Table 4.

Table 4 Results of the *Wilcoxon* signed rank test of the three significant parameters

Comparison of measurement points	Pelvic tilt p-value	Kyphosis angle p-value	Lordosis angle p-value
M1 – M2	0.515	<0.001*	0.012*
M1 – M3	0.754	0.001*	0.114
M1 – M4	<0.001*	0.007*	0.025
M2 – M3	0.825	0.093	0.255
M2 – M4	<0.001*	0.048	<0.001*
M3 – M4	<0.001*	0.508	<0.001*

The *Friedman* test indicated that pelvic tilt ($p = 6 \cdot 10^{-5}$), kyphosis angle ($p = 3 \cdot 10^{-5}$) and lordosis angle ($p = 3 \cdot 10^{-5}$) changed significantly over the four measurement points. In order to verify between which measuring points occurred the difference, the *Wilcoxon* signed rank test was used as a post hoc test. The results are presented in the following

* significant ($p < \alpha_{local}$; $\alpha_{local.i} = \alpha_{global}/(k-i+1)$; number of tests per parameter $k=6$)

There were no two measurement points that differed significantly in all three parameters. Pelvic tilt and kyphosis angle differed significantly between M1 (pre session) and M4 (post session). Pelvic tilt and lordosis angle were significantly different between M2 (pre FED-device) and M4 as well as between M3 (post FED-device) and M4. Kyphosis angle and lordosis angle differed significantly Figure 3 (left side). The data was not divided into the four measurement points. The hierarchical clustering divided the data sets on the basis of the six patients. Patient 1, 2 and 5 were divided into three clusters, Figure 3 (right side). Patient 1 and 5 were divided into two clusters, patients 2 and 6

between M1 and M2. The kyphosis angle was the only one that changed significantly between M1 and M3.

Hierarchical cluster analysis

First, the natural cluster classification was determined. This cluster analysis showed a *Cophenetic correlation coefficient* of 0.93. Values of the *Inconsistency coefficient* ranged from 0 to 1.155. The threshold value was therefore set at 1.154. The natural classification of the data resulted in a cluster number of 10

patients 3 and 4 into one cluster and patient 6 was divided into a total of six clusters.

The second step was to set the number of clusters at four. A similar cluster division occurred

into one cluster and patient 3 and 4 into one cluster.

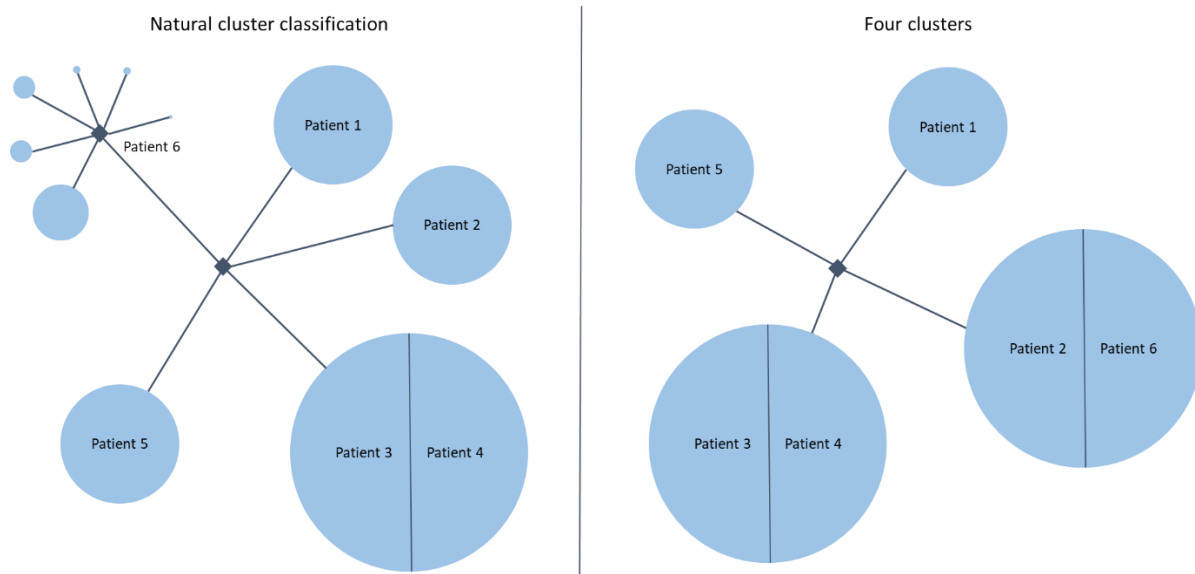


Figure 3 Results of the cluster division for the natural determination of the number of clusters (left) and for the fixed number of four clusters (right). Each circle represents a cluster. The size of the circles reflects the relation of the respective cluster size.

In order to check whether the data is divided into six clusters for all six patients, the cluster analysis was repeated with the aim of forming six clusters. The result is presented in Figure 4. Patient 1, 2 and 6 were divided into three clusters, patient 3 and 4 into one cluster and patient 5 was divided into two clusters.

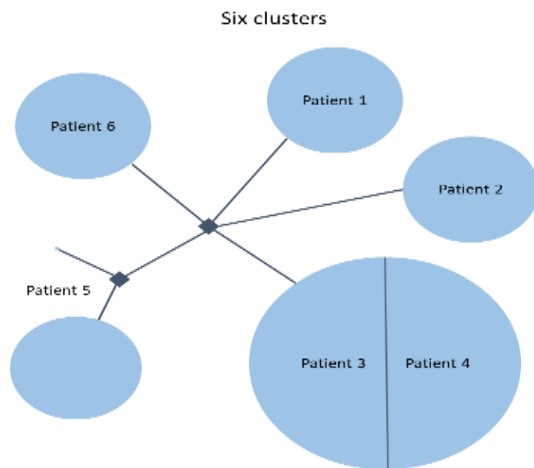


Figure 4 Results of the cluster division for the fixed number of six clusters. Each circle represents a cluster. The size of the circles reflects the relation of the respective cluster size.

Discussion

Four out of six patients were given infrared markers. The influence of the markers was not investigated in this study. However, the effect on the measurement was estimated to be small. The *DIERS formetric 4D* should automatically detect the anatomical points. If markers are used, it can be assumed that the position of the anatomical landmarks is sufficiently accurate.

The descriptive statistics showed that there are no clear differences between the measurement times due to the large variation of the values. The kyphosis angle, differs only in a small margin, but due to the

very large number of outliers, this evaluation is not meaningful. Outliers below the lower whisker were found during measurements with patient 5. Outliers above the upper whisker were found during measurements with patient 6.

Inductive statistics showed that none of the eight parameters had changed significantly between the first and the last therapy session. However, there were three parameters (pelvic tilt, kyphosis angle, lordosis angle) that changed significantly during the therapy session. The variances of the patient bodies were therefore perhaps greater than the influence of the therapy. The variations from person to person were also confirmed in [7]. In addition, the data is highly dependent on the daily form and cooperation of the patient. Therefore, more measurement sessions are planned for the future in order to increase the database.

It is noticeable that the three significant parameters arise from the sagittal plane. This could be an indication that the sagittal plane is especially suitable for therapy monitoring. A further investigation is recommended.

The hierarchical cluster analysis showed a *Cophenetic correlation coefficient* of 0.93. Thus, the cluster analysis already reflects the data very well. There was no clustering on the basis of the four measurement points. Clusters were formed on the basis of patients. The method identified similar patients (patient 3 and patient 4) as to these patients being combined into one cluster on each cluster analysis. Furthermore, it was noticeable that the patient 6 had large variations between the measurements. Therefore, this patient was divided into six clusters

Figure 3 (left side). These variations were similar to the variations between the clusters with the other patients. This phenomenon can be explained with the descriptive statistic. Almost all measurements of patient 6 were identified as outliers. This shows that patient 6 has different characteristics compared to the other patients. After reviewing the data, it can be concluded that the kyphosis angle of this patient is on average 20° greater than that of the other patients.

A similar behaviour was detected in patient 5. This patient was also divided into two clusters (Figure 4). However, only one measurement of patient 5 was divided into a different cluster. Whereas, the descriptive statistics showed that measurements of patient 5 were responsible for the outliers below the lower whisker (Figure 2). Patient 6 and patient 5 with a high severity behaved unstable during the measurements. These variations could have been increased by the therapy. This could be an indication that the measurement system is particularly sensitive to higher severities. A further investigation with more people of higher severity is recommended.

Finally, the clusters are classified according to patient individualities, since these result in larger variations than the variations of the parameters. Further investigations were carried out on the basis of these results. The patient individualities were excluded, by scaling between 0 and 1 of each patient individually. The cluster analysis carried out on the basis of this database had a *Cophenetic correlation coefficient* of 0.57. Thus, the representation of the data by the calculated clusters is much inaccurate than in the first analysis. The cluster classification is no longer based on the

patients. Furthermore, the measurement times M1 to M4 are also not reflected in the clusters.

However, by optimizing the setting parameters for the cluster analysis a better result could be achieved. It would be conceivable to use a different distance measure or similarity measure instead of the *Euclidean distance*. Examples would be the *Minkowski distance* or the *Mahalanobis distance*. Other linkage functions besides the *Average Linkage method* would be *Single Linkage* and *Complete Linkage method*.

After further measurements more methods of artificial intelligence will be investigated. In addition to hierarchical clustering, there are other clustering methods such as the k-means algorithm and spectral clustering [14]. In addition to clustering, Support Vector Machines are also used for pattern recognition [15]. The Support Vector Machines are effective even with high dimensions and simultaneously low observations [16]. Therefore, this method is well suited to the present problem. Deep learning methods are particularly suitable for processing large data. Especially in the field of image classification, Convolutional Neural Networks are the method of choice [17]. In addition to the 130 parameters, the *DIERS formetric 4D* system also generates point clouds and image data. These could be used to assess the progress by using neural networks.

Conclusion

This study investigated which parameters of the video raster stereography are suitable for patient monitoring and whether patterns can be recognized in the data for monitoring

using hierarchical clustering. It could be shown that during a therapy session three parameters changed significantly. It was also detected between which of the four measuring points the change occurred. This makes it possible to develop short-term patient monitoring. In addition to that further measurements must be carried out to observe the progress of therapy over several therapy sessions as well as to increase the database at least tenfold.

It was proven that clustering on the basis of the four measurement points cannot be achieved with the used methods. However, the cluster method is suitable for differentiating between patients. Thus, an application for the classification of patients is conceivable. In case of a new patients, it might be possible to evaluate to which cluster (other patients) he show the greatest similarity and based on that a previous successful treatment plan could be recommended. Therefore, it could be investigated whether this cluster represents a bad treatment progress or a good progress. In addition, different approaches for the optimization of cluster analysis could be evaluated. Further methods of artificial intelligence will be investigated to ensure the best possible short-term patient monitoring.

References

- 1- Z. Śliwiński, W. Kufel, B. Halat, B. Michalak, D. Śliwińska, and G. Śliwiński, “The evaluation of chosen body posture parameters in children with scoliosis – own materials,” *Scoliosis*, vol. 9, no. S1, Dec. 2014.
- 2- Z. Śliwiński, W. Kufel, B. Halat, B. Michalak, D. Śliwińska, and G. Śliwiński, “Radiological progress report of curing scoliosis according to the fed method based on own material,” *Scoliosis*, vol. 9, no. S1, Dec. 2014.
- 3- P. Knott *et al.*, “SOSORT 2012 consensus paper: Reducing x-ray exposure in pediatric patients with scoliosis,” *Scoliosis*, vol. 9, no. 1. BioMed Central Ltd., 25-Apr-2014.
- 4- Z. Śliwiński, W. Kufel, B. Halat, B. Michalak, D. Śliwińska, and G. Śliwiński, “Analysis of body posture of children with idiopathic scoliosis in the image Diers after the application of kinesiology taping application,” *Scoliosis*, vol. 9, no. S1, Dec. 2014.
- 5- J. Nisser *et al.*, “The FED-Method (Fixation, Elongation, Derotation) – a Machine-supported Treatment Approach to Patients with Idiopathic Scoliosis – Systematic Review,” *Z. Orthop. Unfall.*, Aug. 2019.
- 6- J. Nisser *et al.*, “Scoliosis Specific Physiotherapy Approach to Adolescent Idiopathic Scoliosis (AIS) – a narrative review,” *Phys Med Rehab Kuror*, vol. 28(02), pp. 88–102, 2018.
- 7- B. Degenhardt, Z. Starks, S. Bhatia, and G.-A. Franklin, “Appraisal of the DIERS method for calculating postural measurements: an observational study,” *Scoliosis*

- spinal Disord.*, 2017.
- 8- [8] A. Gipsman, L. Rauschert, M. Daneshvar, and P. Knott, "Evaluating the Reproducibility of Motion Analysis Scanning of the Spine during Walking," *Adv. Med.*, vol. 2014, pp. 1–9, 2014.
- 9- W. Frobin and E. Hierholzer, "Rasterstereography: a photogrammetric method for measurement of body surfaces (stereophotogrammetric method)," *Photogramm. Eng. Remote Sensing*, vol. 47, no. 12, pp. 1717–1724, 1981.
- 10- E. Hierholzer, "Objektive Analyse der Rückenform von Skoliosepatienten," *Gustav Fischer Verlag*, 1993.
- 11- B. Drerup and E. Hierholzer, "Back shape measurement using video rasterstereography and three-dimensional reconstruction of spinal shape," *Clin. Biomech.*, vol. 9, no. 1, pp. 28–36, 1994.
- 12- [12] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand J Stat.*, vol. 6, pp. 65–70, 1979.
- 13- "The MathWorks, Inc., Natick, MA, USA. The Statistics and Machine Learning Toolbox™: User's Guide v11.6," 2019. [Online]. Available: https://de.mathworks.com/help/pdf_doc/stats/stats.pdf. [Accessed: 26-Nov-2019].
- 14- A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.
- 15- B. Keshari and S. M. Watt, "Hybrid mathematical symbol recognition using support vector machines," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR, 2007*, vol. 2, pp. 859–863.
- 16- I. Guyon, V. Vapnik, B. Boser, L. Bottou, and S. A. Solla, "Structural Risk Minimization for Character Recognition."
- 17- Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems*, 2010, pp. 253–256.