

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

journal homepage: [www.elsevier.com/locate/bbe](http://www.elsevier.com/locate/bbe)

## Original Research Article

## Gait patterns classification based on cluster and bicluster analysis

J. Pauk<sup>a,b,\*</sup>, K. Minta-Bielecka<sup>a</sup><sup>a</sup> Białystok University of Technology, Białystok, Poland<sup>b</sup> Glenrose Rehabilitation Hospital, 10230 111 Ave NW, Edmonton, AB T5G 0B7, Canada

## ARTICLE INFO

## Article history:

Received 15 August 2015

Received in revised form

1 March 2016

Accepted 7 March 2016

Available online xxx

## Keywords:

Hemiplegia

Gait patterns

Clustering

Biclustering

Joints moments

Spatial-temporal gait parameters

## ABSTRACT

Gait patterns of hemiplegia patients have many potential applications such as assistance in diagnosis or clinical decision-making. Many techniques were developed to classify gait patterns in past years; however, these methods have some limitations. The main goal of the study was to present the performance evaluation results of the new biclustering algorithm called KMB. The second objective was to compare clustering and biclustering methods. The study was performed based on the gait patterns of 41 hemiplegia patients over 12 months post-stroke, at the age of  $48.6 \pm 19.6$  years. Spatial-temporal gait parameters and joint moments were measured using motion capture system and force plates. Clustering and biclustering algorithms were applied for data consisting of joint moments of lower limbs. The obtained results of this study based on joint moments, clustering, and biclustering can be applied to evaluate patient condition and treatment effectiveness. We suggest that the biclustering algorithm compared to clustering algorithms better characterizes the specific traits and abnormalities of the joint moments, especially in case of hemiplegia patients.

© 2016 Published by Elsevier Sp. z o.o. on behalf of Nałęcz Institute of Biocybernetics and Biomedical Engineering.

## 1. Introduction

Hemiplegia is a condition caused by damage to the part of the brain that controls movements of the limbs, trunk, etc. that affects a right or left side of the body. The injury to the left side of the brain causes a right hemiplegia and injury to the right side of the brain results in left hemiplegia [1,2]. While typical gait sequences are repetitive and exhibit periodic behavior, hemiplegia patients usually have problem with gait symmetry [3]. Hemiplegic gait describes altered gait patterns that have been affected by deformity, muscle weakness, impaired motor

control, and pain. Deviations of normal gait patterns are usually observed during both swing and stance phases, and they require systematic evaluation for assessment of functional compensations [3]. Accurate identification of the gait patterns from hemiplegia patients has many potential applications such as assistance in diagnosis or clinical decision-making. The application of intelligent classification systems may enable clinicians to differentiate gait patterns into clinically significant categories. In recent years, many techniques were developed to classify gait patterns, thus making it possible to automatic analysis based on the gait parameters [4–11]. Among the various methods, neural networks models

\* Corresponding author at: Białystok Technical University, Wiejska 45C, 15-351 Białystok, Poland.

E-mail addresses: [j.pauk@pb.edu.pl](mailto:j.pauk@pb.edu.pl) (J. Pauk), [kasia.minta@gmail.com](mailto:kasia.minta@gmail.com) (K. Minta-Bielecka).

<http://dx.doi.org/10.1016/j.bbe.2016.03.002>

0208-5216/© 2016 Published by Elsevier Sp. z o.o. on behalf of Nałęcz Institute of Biocybernetics and Biomedical Engineering.

and support vector machines (SVM) have previously been applied to identify hemiplegia patients. Clustering, a basic tool used in machine learning and patterns classification, is also a very popular technique used to analyze gait patterns. This method involves separating different objects into groups of similar objects called clusters such that the cluster variations are minimized and the variations between clusters are maximized [12,13]. In recent years, various clustering methods have been applied to resolve numerous classification problems [14,15]. In patterns classification, clustering is usually applied to find natural groupings in the data. The main limitation of standard clustering approaches is that they are very sensitive to variations in the gait trials, especially in the hemiplegia patients. The technique that overcomes the limitations of standard clustering approaches may be biclustering, which, in case of the gait patterns analysis, groups both patients and samples simultaneously. Cheng and Church [16] are among the pioneers in introducing the concept of biclustering in data analysis. Their biclusters were based on uniformity criteria, and a greedy algorithm was developed to discover biclusters. Biclustering has been also applied for dimensionality reduction of databases via automatic subspace clustering of high dimensional data, and has been used in a number of other areas [17,18]. The concept of associating biclustering with biological data analysis allows identification of sets of patients sharing consistent expression patterns over the subsets of samples [16–18]. Unfortunately, this method has some limitations. First, it is not easy to present all biclusters simultaneously. Second, a bicluster can show complicated coherent patterns. Finally, the criterion to evaluate a biclustering algorithm depends on the structures of biclusters to be detected [19]. This gave motivation to search for a new efficient algorithm such as proposed in this paper KMB, which can be used to analyze gait patterns. The main goal of this paper was to present the performance evaluation results of the new biclustering algorithm called KMB. The second objective was to compare clustering and biclustering methods.

## 2. Material and methods

### 2.1. Subjects

The study included the gait patterns of 41 hemiplegia patients over 12 months post-stroke. Inclusion criteria for hemiplegia patients characteristics were applied, such as first ever stroke, ischemia or intra-cerebral hemorrhage assessed by computerized tomography or magnetic resonance imaging, occurrence of stroke at least 12 months before enrolment in the study, ability to walk minimum 10 m independently, absence of prior treatments of orthopedic surgery, and cognitive ability to follow the study instructions. Patients were excluded if they had received botulinum injections or a phenol nerve block in the hemiplegic lower limb within the last 6 months before the evaluation, and if they had any other medical disorders, which might have adversely affected their gait patterns. The patient's level of functional independence was assessed by neurologists through the modified Barthel index (BI) [8]. The maximum score of 100 means that the patient is fully independent in physical functioning. The lowest score of 0 corresponds to the

totally dependent, bedridden person. The study was approved by local ethics committee. All subjects received full information about the study before giving signed informed consent.

### 2.2. Measuring spatial-temporal gait parameters and joint moments

Gait parameters were measured using a motion tracking system (Motion Analysis Corp., USA) and two AMTI force platforms (Advanced Mechanical Technology, Inc., USA). Temporal parameters were estimated using sample frequency of 1000 Hz, whereas spatial-temporal gait parameters were estimated using sample frequency of 120 Hz. Each subject was asked to walk barefoot at habitual speed. The measurements were repeated to obtain at least three valid walking trials. Mean values of stride duration (s), velocity (m/s), stance duration (%), stride length (m), and step length (m) of both affected and unaffected side were calculated. The spatial-temporal gait parameters have been summarized in Table 1. The hip, knee, and ankle joint moments were computed using an inverse dynamic approach [3] and were normalized to the body mass. The signals were initially low-pass filtered using a Butterworth filter.

### 2.3. Agglomerative hierarchical clustering

The data set included 177 gait cycles of joint moments for 41 hemiplegia patients. A one-way clustering analysis was considered to be reliable and stable for a data set with continuous variables. The data set is divided into many small sub-clusters, which are then combined into a specified number of clusters using an agglomerative hierarchical clustering algorithm (AHC). In the clustering algorithm, the Ward's method has been applied as agglomeration method, and the Euclidean distance as a dissimilarity measure.

### 2.4. A new biclustering algorithm KMB

In this article we proposed the new algorithm called KMB, which was based on two existing biclustering methods such as node-deletion algorithm [16] and node-addition algorithm [18]. Examination results were represented as a matrix  $A = (R, C)$  with rows  $R$  corresponding to walking trials and columns  $C$

**Table 1 – Spatial-temporal gait parameters for affected and unaffected side in hemiplegia patients.**

Gait parameters	Hemiplegia		p-Value
	Affected side	Unaffected side	
Stride duration (s)	1.48 (.44)	1.51 (.56)	$p = .306$
Mean (SD)			
Velocity (m/s)	.73 (.41)		$p < .05$
Mean (SD)			
Stance duration (%)	63.82 (7.71)	68.13 (8.66)	$p = .479$
Mean (SD)			
Stride length (m)	.93 (.35)	.93 (.37)	$p = .316$
Mean (SD)			
Step length (m)	.48 (.16)	.47 (.19)	
Mean (SD)			

corresponding to captured results. All numbers in the data matrix were transformed by scaling and function  $x \rightarrow x^5$ , and each element  $a_{ij}$  of the matrix  $A$  contained a real number describing the joint moments at  $j$ th measurement point in  $i$ th gait cycle. A submatrix  $A_{IJ}$  containing only the  $a_{ij}$  elements belonging to subsets of rows  $I < R$  and columns  $J < C$  is called a bicluster if it exhibits homogeneous properties. A mean square residue score was used to measure the quality of biclusters, and a parameter  $\delta$  (delta) was used as a threshold that limits the value of mean square residue score. A residue of an element of a bicluster has been defined as the value that differentiates this element from the expected value calculated from its row, column and bicluster means:  $r_{ij} = a_{ij} - (a_{ij} + a_{ij} - a_{IJ})$ , where  $a_{ij}$  is the mean of the  $i$ th row,  $a_{ij}$  – the mean of the  $j$ th column, and  $a_{IJ}$  is the mean of the whole bicluster. The mean square residue score indicated by  $H = (I, J)$  was defined as

$$H(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (r_{ij})^2. \quad (1)$$

Cheng and Church [16] have presented the optimization algorithm, which maximizes the volume of a bicluster while maintaining the upper bound on mean square residue score  $H(I, J) \leq \delta$ . For the purpose of their algorithm, Cheng and Church provided another score, similar to mean square residue  $H = (I, J)$  that they used to qualify rows and columns for removal or addition to the bicluster:

$$d(i) = \frac{1}{J} \sum_{j \in I, j \in J} (a_{ij} - (a_{ij} + a_{ij} - a_{IJ}))^2. \quad (2)$$

The algorithm has been defined into two phases. During the first phase both rows and columns are evaluated in terms of their  $d(i)$  scores and algorithm parameter  $\alpha$  (alpha), which is a threshold describing when the multiple node deletion step is used (a higher  $\alpha$  leads to less multiple node deletion), and every row or column with  $d(i)$  coherency score  $> \alpha H(I, J)$  is removed from the bicluster until  $H(I, J) > \delta$ . In the second phase of the algorithm, the rows and columns are added to the bicluster based on their  $d(i)$  score being less than or equal to  $H(I, J)$ . The node-addition step guarantees the  $H(I, J)$  score will not increase, and for that reason it is not necessary to evaluate it against delta during this phase. The whole process is then repeated after masking the discovered biclusters elements in the original data set with random numbers; without this masking the algorithm would produce the exact same result. A different approach also based on the mean square residue score has been presented by Borah and Bhattachatyaa in [18]. In their algorithm, rows and columns are only added to the bicluster using a random starting point (a single row or column). To include the row or column in the bicluster they also used the coherency score  $d(i)$ , but the qualification criteria were different. Borah and Bhattachatyaa have suggested two passes to choose which rows should be added to the bicluster. During the first pass, all rows with  $d(i)$  score less than  $\theta = \alpha \delta$  are excluded from further consideration. In the second pass, the row is temporarily added to the bicluster, and a new mean square residue score is calculated. The row is permanently added only if the score is increased by not more than half of the  $\delta$ -score difference:

$$newscore \leq \frac{score + \delta}{2}. \quad (3)$$

The method for adding columns is similar, but the first phase is omitted, and condition  $\delta' < \delta$  is used when expected mean square residue scores are qualified. In each iteration of the algorithm, row extension is attempted until no addition takes place, and column extension is done using lower bound for  $\delta$ . In our algorithm KMB, we combined those two approaches by using the initialization, deletion, and masking steps as specified by Cheng and Church [16] and utilizing Borah and Bhattachatyaa's method [18] for addition of rows and columns. This method has the advantage of using the provided  $\delta$  parameter as the upper bound for  $H(I, J)$  and does not require that the  $H(I, J)$  score only decrease. Therefore, it is able to discover biclusters with  $H(I, J)$  close or equal to delta. Its counterpart from Cheng and Church's algorithm is bounded only by the current value of  $H(I, J)$  and is known for being able to greatly decrease the mean square residue score below the requested delta.

## 2.5. Statistical analysis

The data set included gait cycles from 41 hemiplegia patients. Means and standard deviations were calculated for the total subject sample for the data obtained from the gait analysis. For gait variables, an ANOVA for repeated measurements was applied to evaluate differences in quantitative walking parameters. A value of  $p < .05$  was considered significant. Computer software Statistica 10.0 (StatSoft, Tulsa, OK, USA) was used for computations.

## 3. Results

### 3.1. Assessing spatial-temporal gait parameters

Forty-one hemiplegia subjects (mean age 48.6 (SD 19.6) years, BMI 24.3 (SD 4.0), 48.8% of male) were recruited. Twenty two patients had left hemiplegia and 19 patients had right hemiplegia. The average time from stroke was 16.8 (4.6) months. The type of stroke was hemorrhage (32 patients) and ischemia (9 patients). The average value for BI in hemiplegia patients was  $87.5 \pm 13.7$  (min. 50–max. 100). According to the five-grade scale of independence, the most numerous group (15 patients) were patients with a mild degree of dependence, the smallest group (4 patients) were patients with advanced degree of dependence, fourteen patients were moderately dependent, and eight patients were completely independent. Table 1 shows the spatial-temporal gait parameters for affected and unaffected side in hemiplegia patients.

The analysis shows that no significant differences in the stride duration, stride length, and step length ( $p > .05$ ) between affected and unaffected sides in hemiplegia subjects were observed. However, significant differences ( $p < .05$ ) in the stance duration were observed.

### 3.2. Agglomerative hierarchical clustering

One hundred seventy-seven gait cycles of joint moments for 41 patients were subdivided into homogeneous subgroups (clusters). Three clusters of gait patterns were identified as a result of the AHC; cluster 1 included 25 gait cycles (14.1%),

**Table 2 – Cluster properties.**

Cluster	Within-class variance	Minimum distance to centroid	Average distance to centroid	Maximum distance to centroid
1	44,385.977	88.747	196.155	317.398
2	20,419.080	63.667	134.100	260.710
3	33,368.221	55.900	168.350	459.548

cluster 2 included 64 gait cycles (36.2%), and cluster 3 included 88 gait cycles (49.7%). Clusters properties were presented in Table 2.

Average age of the patient in cluster 2 and cluster 3 was 53.5 (12.3) years old and 51.5 (23.8) years old, respectively. Cluster 1 contained the youngest patients, and average age of these patients was 35.3 (14.8) years old. Spatial-temporal gait parameters for patients in individual clusters were presented in Table 3.

The three clusters identified different walking speed. Cluster 2 represents a low walking speed ( $.34 \pm .16$  m/s), cluster 1 – an intermediate walking speed ( $.62 \pm .34$  m/s), and cluster 3 – a fast walking speed ( $1.05 \pm .27$  m/s). Cluster 2 presents the greatest deviations in stance duration compared with the other clusters.

### 3.3. A new biclustering algorithm KMB

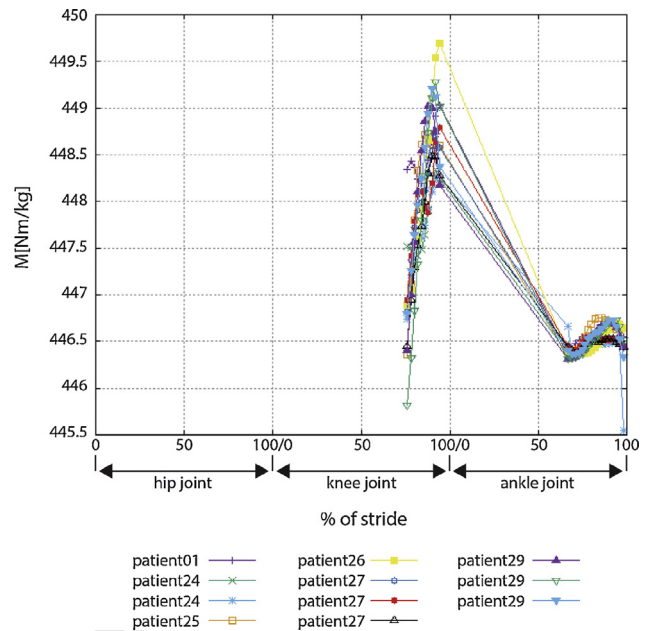
The 177 gait cycles of joint moments data for 41 patients were subdivided into biclusters. The data set have been tested for different values of parameter  $\alpha$  and  $\delta$ . The best results were obtained for the parameter  $\alpha$  equal to 1.2. The higher value of parameter  $\alpha$  caused the lower accuracy of the obtained results – more gait cycles were identified by biclusters. Finally, the three interesting biclusters representing gait patterns in hemiplegia patients were identified for  $\delta = .1$ ,  $\delta = 1$ , and  $\delta = 5$  (Figs. 1–3). Bicluster 1 includes 11 gait cycles (6.2% out of the cycles analyzed), bicluster 2 includes 21 gait cycles (11.9%), bicluster 3 includes 28 gait cycles (15.8%).

Bicluster 1 contains 11 cycles for 6 patients, all corresponding to cluster 2 from AHC.

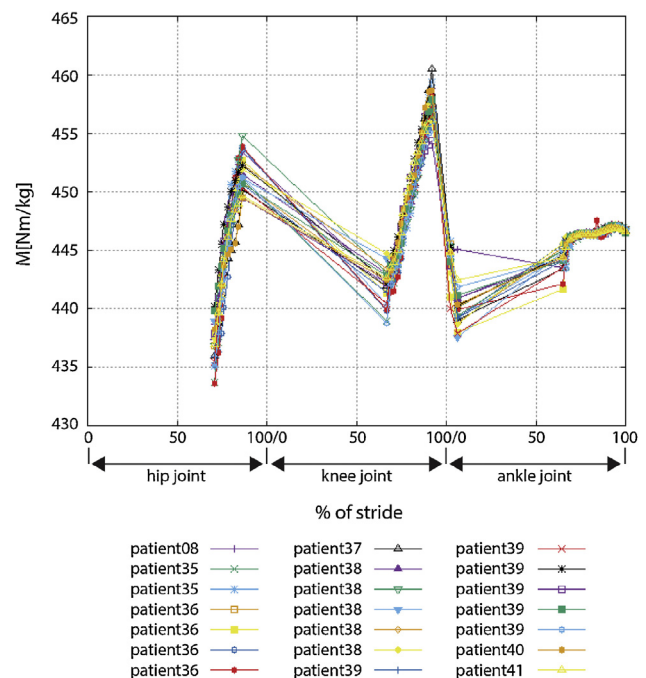
Bicluster 2 contains 21 cycles for 11 patients, all corresponding to cluster 3 from AHC.

**Table 3 – Spatial-temporal gait parameters for hemiplegia patients in each cluster.**

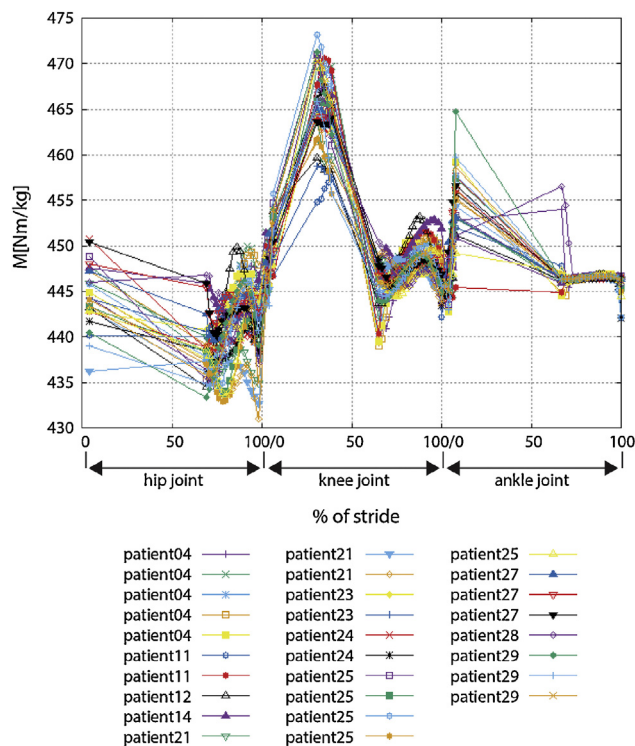
Spatial-temporal gait parameters	Cluster 1	Cluster 2	Cluster 3
Stride duration (s)	1.63 (.39)	1.80 (.43)	1.21 (.25)
Mean (SD)			
Velocity (m/s)	.62 (.34)	.34 (.16)	1.05 (.27)
Mean (SD)			
Stance duration (%)	64.81 (5.80)	63.98 (9.86)	62.74 (3.76)
Mean (SD)			
Stride length (m)	.89 (.22)	.58 (.22)	1.19 (.21)
Mean (SD)			
Step length (m)	.47 (.10)	.33 (.11)	.60 (.09)
Mean (SD)			

**Fig. 1 – Bicluster 1 represents 1st bicluster discovered by algorithm KMB ( $\delta = .1$ ,  $\alpha = 1.2$ ).**

Bicluster 3 contains 28 cycles for 8 patients. The majority of these correspond to cluster 2 from AHC and 2 of them correspond to cluster 3. Biclusters 1 and 3 included the youngest patients from all clusters (50.5 (12.1) years old in the 1st bicluster and 51.5 (18.2) years old in the 3rd bicluster, respectively), while the 2nd bicluster showed the oldest patients (66.0 (17.4) years old). The spatial-temporal parameters for each bicluster were presented in Table 4.

**Fig. 2 – Bicluster 2 represents 2nd bicluster discovered by algorithm KMB ( $\delta = 1$ ,  $\alpha = 1.2$ ).**





**Fig. 3 – Bicluster 3 represents 1st bicluster discovered by algorithm KMB ( $\delta = 5$ ,  $\alpha = 1.2$ ).**

Biclusters 1 and 3 characterize patients with low velocity ( $.32 \pm .12$  m/s and  $.36 \pm .19$  m/s, respectively). Bicluster 2 characterizes patients with fast walking speed ( $1.13 \pm .08$  m/s).

#### 4. Discussion

The categorization of gait pattern using three-dimensional gait analysis provides useful information on the joint dysfunction underlying the gait disability. In our study, the hemiplegic gait was characterized by low values of velocity (.73 m/s), which is in agreement with other studies [20]. The analysis of most spatial-temporal gait parameters did not show significant differences ( $p > .05$ ) between affected and

unaffected sides. However, significant differences ( $p < .05$ ) were observed in the stance duration.

Cluster analysis has been applied to classify groups of stroke patients with similar gait patterns, which are not easy to identify using standard techniques [21]. Amongst the patients examined in the presented study, three clusters have been obtained. Cluster 1 included 14.1% of gait cycles, cluster 2 included 36.2% gait cycles, and cluster 3 included 49.7% gait cycles. The three clusters present patients with different walking speed. Cluster 2 contains patients with a low walking speed (average velocity .34 m/s), cluster 1 represents patients with intermediate walking speed (average velocity .62 m/s), and cluster 3 includes patients with fast walking speed (average velocity 1.05 m/s). According to Perry's classification [22], patients in the presented study characterized a walking speed corresponding to household ambulation ( $< .40$  m/s, clusters 2) or a limited community ambulation (between .40 and .80 m/s, cluster 1), whereas only a few patients had a walking speed useful for full community ambulation ( $> .80$  m/s, cluster 3). These results are different from those considered by Kinsella and Moran [9], where the group of patients with fast velocity characterized a mean walking speed of .42 m/s.

In comparison to clustering, biclustering is a data mining method that simultaneously clusters both rows and columns of a data matrix. Biclustering has been applied to a number of other areas such as gene patterns recognition [16–18]. In proposed approach, missing values were replaced by 0, to not introduce any recognizable data patterns that would interfere with the algorithms' operation. A similar solution based on replacing missing data with random numbers was proposed in [16]. Additionally, in the original implementation [16] the masked data set was used during the node-deletion step, but not during the node-addition step, and the resulting bicluster could include the randomly generated entries. Our algorithm also used the masking technique when deleting rows and columns, but unlike the original it populated the biclusters with the values coming from the unmasked data set. Our approach determines the number of well separated biclusters from the input solutions itself. The 177 gait cycles of joint moments data for 41 patients were subdivided into biclusters. We have tested the data set with different values of parameter  $\alpha$  and  $\delta$ . We have obtained the best results with the  $\alpha$  equal to 1.2. Finally, the three biclusters representing the gait patterns in hemiplegia patients were identified for  $\delta = .1$ ,  $\delta = 1$ , and  $\delta = 5$ . In the study, three of the found biclusters were presented. Biclusters 1 and 3 were characterized by low velocity ( $.32 \pm .12$  m/s and  $.36 \pm .19$  m/s respectively). Bicluster 2 was characterized by fast walking speed ( $1.13 \pm .08$  m/s). The analysis shows, that in the group of patients discovered by the bicluster 1 the average velocity was by 56.2% lower than in the group with all hemiplegia patients (.32 m/s in patients discovered by bicluster vs. .73 m/s in whole group of hemiplegia patients). Furthermore, the average stride length in the group of patients discovered by bicluster 1 was lower by 43% compared to the whole group. The same was observed for the step length. For patients discovered in bicluster 1, the average step length was by 22.9% shorter than the average step length of the whole group. Bicluster 1 contains gait cycles corresponding to cluster 2 from AHC, bicluster 2 contains gait cycles corresponding to cluster 3 from AHC, and bicluster

**Table 4 – Summary of biclusters characteristics.**

Spatial-temporal parameters	Bicluster 1	Bicluster 2	Bicluster 3
Stride duration (s)	1.73 (.38)	1.08 (.04)	1.62 (.43)
Mean (SD)			
Velocity (m/s)	.32 (.12)	1.13 (.08)	.36 (.19)
Mean (SD)			
Stance duration (%)	58.35 (4.49)	61.99 (1.60)	63.13 (5.91)
Mean (SD)			
Stride length (m)	.53 (.16)	1.18 (.17)	.54 (.22)
Mean (SD)			
Step length (m)	.37 (.10)	.61 (.03)	.33 (.10)
Mean (SD)			

3 contains gait cycles the majority of which correspond to cluster 2 from AHC, and 2 of them correspond to cluster 3.

The advantage of the bicluster analysis is the possibility of identification of local patterns in hemiplegia patients, not only for whole gait cycles as it is visible in cluster analysis. It should be noted that while the clustering problem created disjoint clusters that covered all the input set, biclusters overlapped, and they covered only a part of the matrix. Biclustering method gives better characteristic of abnormalities in gait patterns of hemiplegia patients compared to clustering because it discover abnormalities in different parts of gait cycle whereas clusters methods classify patients based on similarity of gait patterns of whole gait cycle. Biclustering took as an input the same conditions activity matrix, and tried to find statistically significant sub-matrices in it, also called biclusters. These structures implied a joint behavior of some data under some conditions. While clustering methods can be applied to either rows or columns of a data matrix separately, biclustering methods perform clustering in two dimensions simultaneously. This is the first study, which presents biclustering application in gait analysis.

## 5. Conclusions

Proposed clustering and biclustering algorithms are relevant to gait analysis because of the complexity of the data. It means that it is possible to extract information about subjects with coherent values in data set based on these algorithms. The obtained results of this study based on joint moments, clustering, and biclustering can be applied to evaluate patient condition and treatment effectiveness. We suggest that the biclustering algorithm compared to clustering algorithms better characterizes the specific traits and abnormalities of the joint moments, especially in cases of hemiplegia.

## Funding

Q2 The research was supported by Polish Ministry of Higher Education.

## REFERENCES

- [1] Olney SJ, Richards C. Hemiparetic gait following stroke. Part I: Characteristics. *Gait Posture* 1996;4(2):136–48.
- [2] Wasay M, Khatri IA, Kaul S. Stroke in South Asian countries. *Nat Rev Neurol* 2014;10(3):135–43.
- [3] Winter DT. Biomechanics and motor control of human gait: normal elderly and pathological. Waterloo University of Waterloo Press; 1991.

- [4] Kramers De Quervain IA, Simon SR, Leurgans S, Pease WS, McAllister D. Gait pattern in the early recovery period after stroke. *J Bone Jt Surg A* 1996;78(10):1506–14.
- [5] Kim CM, Eng JJ. Magnitude and pattern of 3D kinematic and kinetic gait profiles in persons with stroke: relationship to walking speed. *Gait Posture* 2004;20(2):140–6.
- [6] Straudi S, Manca M, Aiello E, Ferraresi G, Cavazza S, Basaglia N. Sagittal plane kinematic analysis of the six-minute walk test: a classification of hemiplegic gait. *Eur J Phys Rehab Med* 2009;45(3):341–7.
- [7] Begg RK, Palaniswami M, Owen B. Support vector machines for automated gait classification. *IEEE Trans Biomed Eng* 2005;52(5):828.
- [8] Kamruzzaman J. Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait. *IEEE Trans Biomed Eng* 2006;53(12):2479–89.
- [9] Kinsella S, Moran K. Gait pattern categorization of stroke participants with equinus deformity of the foot. *Gait Posture* 2008;27(1):144–51.
- [10] Enachescu D, Enachescu C. Clustering and spatial variation in risk. *Biocybern Biomed Eng* 2009;29(2):19–30.
- [11] Salazar AJ, De Castro OC, Bravo RJ. Novel approach for spastic hemiplegia classification through the use of support vector machines. *Conf Proc IEEE Eng Med Biol Soc*, vol. 1. 2004. pp. 466–9.
- [12] Mulroy S, Gronley J, Weiss W, Newsam C, Perry J. Use of cluster analysis for gait pattern classification of patients in the early and late recovery phases following stroke. *Gait Posture* 2003;18(1):114–25.
- [13] Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. *Proc. 1998 ACM SIGMOD International Conference on Management of Data*; 1998.
- [14] Fisher DH. Knowledge acquisition via incremental conceptual clustering. *Mach Learn* 1987;2(2):139–72.
- [15] Jain AK, Dubes RC. Algorithms for clustering data. Prentice Hall; 1988.
- [16] Cheng Y, Church GM. Biclustering of expression data. *American Association for Artificial Intelligence*; 2000.
- [17] Yang J, Wang W, Wang H, Yu P. Enhanced biclustering on expression data. *Proc. 2003 Third IEEE Conference on Bioinformatics and Bioengineering*; 2003.
- [18] Borah B, Bhattachatya DK. Biclustering expression data using node addition algorithm. *Proc. 2007 15th International Conference on Advanced Computing and Communications*; 2007.
- [19] Zhao H, Wee-Chung Liew A, Wang DZ, Yan H. Biclustering analysis for pattern discovery: current techniques, comparative studies and applications. *Curr Bioinform* 2012;7:43–55.
- [20] Wall J, Turnbull G. Gait asymmetries in residual hemiplegia. *Arch Phys Med Rehabil* 1986;67:550–3.
- [21] Manca M, Ferraresi G, Cosma M, et al. Gait patterns in hemiplegic patients with equinus foot deformity. *BioMed Res Int* 2014;2014:1–8.
- [22] Perry J, et al. Classification of walking handicap in the stroke population. *Stroke* 1995;26(6):982–9.

Q3