A new classification of hemiplegia gait patterns based on bicluster analysis of joint moments

JOLANTA PAUK1,2*, KATARZYNA MINTA-BIELECKA1

1 Białystok University of Technology, Białystok, Poland.
2 Glenrose Rehabilitation Hospital, Edmonton, Canada.

Purpose: Hemiplegia is a paralysis on one side of the body resulting from disease or injury to the motor centers of the brain that may lead to difficulty in walking and problems in balance. A new methodology for hemiplegia gait patterns classification based on bicluster analysis, which aims to identify a group of patients with similar gait patterns, and verify if spatial-temporal gait parameters are correlated with the Barthel Index, has been proposed. Methods: Eighteen hemiplegia patients were recruited. Measurements included spatial-temporal gait parameters and joint moments. Gait data were measured using a motion tracking system and two force platforms. Bicluster analysis was used to classify the subjects' gait patterns. The relation between Barthel Index and spatial-temporal gait parameters was determined based on the Spearman correlation. Results: A high correlation between spatial-temporal gait parameters and Barthel Index ($r > .5$, $p < .05$) was observed. Well-separated biclusters presenting similarity among the lower limb joints during the gait cycles were obtained from the data. Conclusions: Bicluster analysis can be useful for identifying patients with similar gait patterns. The relation between the gait patterns and the underlying impairments would allow clinicians to target rehabilitation strategies at the patient’s individual needs.

Key words: classification, bicluster, hemiplegia, joint moments, Barthel Index

1. Introduction

Hemiplegia is a paralysis on one side of the body resulting from disease or injury to the motor centers of the brain. Usually, an injury to the right side of the brain will cause a left-sided hemiplegia, while an injury to the left side of the brain will cause a right-sided hemiplegia. It can occur in all age groups, but most commonly in elderly people over 60 years old [5]. The common cause of hemiplegia is a stroke, which by insufficient blood supply to the brain leads to loss of brain functions. Stroke is one of the main causes of adults’ disability in most countries [13], [23]. In Western European countries, the ratio of hemiplegic patients to the general population is 3.94 per 1000 for men and 2.52 per 1000 for women [11]. Symptoms usually depend on the affected part of the brain. Hemiplegic patients may have difficulty in walking, and problems with balance maintenance [10], [16], [18]. Gait observation, which may provide information about musculoskeletal and neurological conditions in hemiplegia patients, is an important aspect of diagnosis. In the past, hemiplegia patterns were explored in many studies, and a deterioration in spatial-temporal gait parameters (e.g., stride length, stride velocity, stride time, etc.), phasic patterns (e.g., swing, double stance, etc.), and kinetic (e.g., ground reaction force, joint powers, etc.) was reported [2], [9], [12], [14], [15], [17], [19]. Several authors have characterized gait patterns by using a variety of techniques such as mathematical modeling, Support Vector Machine (SVM), or cluster analysis to analyze gait classification of elderly patients [2]–[4], [6], [21], [22]. Carod-Artal et al. [6] found that gait velocity of stroke patients was a strong determinant for group placement. Consideration of several parameters at the same time instead of a single parameter for each pa-
tient is the advantage of this method. However, in con-
trast to clustering, biclustering is a data mining method
that groups both rows and columns of a matrix simulta-
neously. This method has been applied previously to
examine patterns in gene research [7]. We suggest that
bicluster analysis can also be used to identify meaning-
ful but not obvious patterns, in gait data by dividing
data into groups (biclusters). Therefore, the purpose of
this study was to check whether the biclustering
method can effectively group gait patterns in hemiple-
gia population, and to verify if spatial-temporal gait
parameters are correlated with the Barthel Index.

2. Materials and methods

2.1. Subjects

The study has been conducted using 87 gait pat-
terns of 18 hemiplegia patients with a stabilized clinical
condition (12 with left hemiplegia and 6 with right
hemiplegia) over 12 months, post-stroke. Stroke was
defined as an acute event of cerebrovascular origin
causing focal or global neurological dysfunction lasting
more than 24 hours diagnosed by neurologists,
and confirmed by computed tomography or magnetic
resonance imaging. As inclusion criteria for hemiplegia
patients characteristics such as occurrence of stroke
within last 12 months/occurrence of stroke at least
12 months prior to evaluation, equines deformity,
which was evaluated through the observational analysis
and clinical assessment, ability to walk 10 meters inde-
pendently without a walking device, lack of neurologi-
cal and/or orthopedic co-morbidities impairing ambu-
ation, and cognitive ability to understand training
procedures and to follow the study instructions were
applied. Patients were excluded if they had received
a phenol nerve block in the hemiplegic lower limb or
botulinum injections during the six months prior to the
evaluation, and if they had other medical disorders,
which might have adversely affected their gait patterns.
All subjects received full information about the study
before giving signed informed consent. The study was
approved by the local ethics committee.

2.2. Measurement protocol

Assessment of functional level of independence

The patient’s level of functional independence was
assessed by neurologists through the modified Barthel
Index (BI). Application of the BI to determine dis-
ability is considered reliable, valid, and sensitive [8].
The BI evaluates the following ten self-care functions
such as feeding, moving from chair to bed and re-
turning, performing personal toilet, getting on and off
the toilet, bathing, mobility, climbing up and down
stairs, dressing and undressing, bowel control and
urine control. Each activity is scored separately and
the total points are calculated. The maximum score of
100 means that the patient is fully independent in
physical functioning. The lowest score of 0 corre-
sponds to the totally dependent, bedridden person.

Measuring spatial-temporal gait parameters
and joint moments

Gait parameters were measured using a motion
tracking system (Motion Analysis Corp., USA), al-
lowing one to extract the spatial-temporal gait pa-
tamers, and two AMTI force platforms (Advanced
Mechanical Technology, Inc., USA), providing the
ability to determine the exact moments of initial and
terminal contacts, measure temporal parameters.
Temporal parameters were estimated using sample
frequency of 1000 Hz, and spatial-temporal gait pa-
tamers were estimated using sample frequency of
120 Hz. Each patient was asked to walk barefoot at
habitual speed. The measurements were repeated to
obtain at least three valid walking trials, and patients
had to take a five-minute rest between each trial.
Mean values of parameters such as stride duration
(sec), velocity (m/sec), stance duration (sec), stride
length (m), and step length (m) of both affected and
unaffected side were calculated. These gait data were
summarized in Table 1. The joint moments of the
lower extremity were calculated using an inverse dy-
namic approach [24] and were normalized to the body
mass.

Classification of gait patterns
based on the biclustering algorithm

The biclustering algorithm (greedy algorithm),
whose function is to extract a subset of hemiplegia
patients behaving similarly over a subset of joint mo-
ments, was used to classify gait patterns. Data con-
taining ankle, knee and hip joint moments of exam-
ined patients were represented as a matrix $A$ with rows
corresponding to specific gait cycles of joint moments
for hemiplegia patients, and columns corresponding to
subsequent measurements during a specific cycle. All
numbers in the data matrix were transformed by scal-
ing and function $x \rightarrow 1/x^2$, and each element $a_{ij}$ of the
matrix $A$ was a real number describing the result of
the \( j \)-th measurement in the \( i \)-th gait cycle. As proposed in [7], to avoid the formation of additional recognizable patterns in the data, which could interfere with the algorithm, missing values were replaced by 0. The result of proposed method was a set of biclusters, each being a submatrix with sets of rows and columns extracted from the original data. Cheng and Church in [7] proposed a mean square residue score as the measure of similarity of values in a bicluster

\[
H(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (a_{ij} - a_{i..} - a_{.j} + a_{..})^2,
\]

where

\[
a_{ij} = \frac{1}{|J|} \sum_{j \in J} a_{ij}, \quad a_{i..} = \frac{1}{|I|} \sum_{i \in I} a_{i..},
\]

and

\[
a_{.j} = \frac{1}{|I|} \sum_{i \in I} a_{ij}, \quad a_{..} = \frac{1}{|J|} \sum_{j \in J} a_{..}
\]

were defined as mean of the \( i \)-th row of a bicluster, mean of the \( j \)-th column, and mean of all the values in a bicluster, respectively. The goal of presented algorithm was to find a bicluster with the largest possible volume, measured as the number of elements, while keeping the mean square residue score as small as practically possible. It was also desirable to find non-trivial bicluster that is characterized by high row variance. In order to fulfill those requirements, Cheng and Church [7] have formulated a multi-step heuristic that, by successively removing and then adding nodes, produces a bicluster satisfying the constraints given. The starting point of the matrix is having all the rows and columns from the original data set, with its missing elements replaced with zeros. First, they have defined the node deletion method [7]. In that step, the value of function \( d(i) \)

\[
d(i) = \frac{1}{|J|} \sum_{j \in J} (a_{ij} - a_{i..} - a_{.j} + a_{..})^2
\]

is computed for each row and column of the initial bicluster. Then, a row or column having the greatest \( d(i) \) value is removed from initial bicluster, \( d(i) \) values are recomputed for the remaining rows and columns, and the process is repeated until the \( H(I, J) \) is lower or equal to \( \delta \) (delta). Delta is a threshold that limits the value of mean square residue score. Since in each iteration exactly one node of the bicluster is removed, it is guaranteed that the algorithm will proceed by \((n + m)\) iterations at most, where \( n \) is the number of rows and \( m \) is the number of columns of the initial bicluster. The method of deleting one node at a time, which requires recalculation of \( d(i) \) values for each of the remaining rows and columns of a bicluster, can be implemented in \( O(nm) \) time. However, in some cases, it may be more expensive, and therefore, the authors proposed different method of node deletion that can be implemented in \( O(\log n + \log m) \) time. This method [7] also requires calculation of \( d(i) \) values for each row and column, but, unlike the previous one, removes all nodes that have their \( d(i) \) values greater than certain threshold and defined as \( d(i) > \alpha H(I, J) \) for \( \alpha > 1 \) in a single iteration. A parameter \( \alpha \) (alpha) is a threshold describing when the multiple node deletion step is used (a higher \( \alpha \) leads to less multiple node deletion). Similar to the first method, the removal step is repeated until the mean square residue score is less or equal to \( \delta \). By discarding the \( d(i) \) calculation step after each node removal, the efficiency of the algorithm is improved at the cost of possible omission of significant \( \delta \)-biclusters due to rapid convergence of the solution towards smaller biclusters. This behavior can be mitigated by fine-tuning of the alpha parameter. After the node deletion steps are completed, even though the resulting bicluster satisfies the \( H(I, J) < \delta \) requirement, it may not yet be maximal. In other words, there might exist a node or a set of nodes that can be safely added to it without increasing the mean square residue score beyond the defined \( \delta \) threshold. As a solution, the authors suggest a node addition step, which is again based on \( d(i) \) values. Only rows or columns having its mean lower than the mean square residue score of the whole bicluster are considered candidates for inclusion in the resulting bicluster. The node adding step is repeated until no rows or columns that satisfy the addition criteria can be found. Cheng and Church in [7] proved that this selection mechanism guarantees that the bicluster score \( H(I, J) \) will not increase. However, it is still possible that it will decrease greatly below the \( \delta \). Finally, all the steps described above are combined into a single bicluster finding algorithm. After the data conditioning step replaces the missing values with zeros, the bicluster rows and columns are initialized using all of the rows and columns presented in the matrix. Then, the multiple node deletion algorithm is called followed by single node deletion. After the initial bicluster is reduced, the node adding method is used to possibly expand the set of nodes, and possibly reduce the bicluster score even more. Since the node manipulation methods are deterministic, successive
runs of the algorithm would produce the same result. For that reason, a data masking step, where each new discovered bicluster is masked in the original data set by replacing its elements with random numbers, is being introduced. Now, when the finding process is repeated, the node deletion algorithm is able to discover different biclusters. In the original implementation [7], the masked data set is used during node deletion step but not during node adding step, and the resulting bicluster may include the randomly generated entries. Our algorithm also uses the masking technique when deleting rows and columns, but, unlike the original one, it populates the biclusters with the values coming from the unmasked data set.

2.3. Statistical analysis

The data set included gait cycles from 18 hemiplegia patients. Means and standard deviations were calculated for the total subject sample for the data from the gait analysis. An ANOVA for repeated measurements was applied for gait variables to evaluate differences in quantitative walking parameters. The data from trials were averaged together to provide the unique score for each participant. A value of $p < .05$ was considered significant. The spatial-temporal gait parameters of patients were compared with Barthel Index by using the Spearman correlation. Computer software Statistica 10.0 (StatSoft, Tulsa, OK, USA) was used for analysis.

3. Results

3.1. Assessing spatial-temporal gait parameters

Eighteen eligible hemiplegia subjects (mean age 44.4 (SD 17.1), BMI 24.0 (SD 4.2), 66.7% of male) were recruited. Spatial-temporal gait parameters for affected and unaffected sides in hemiplegia patients are shown in Table 1.

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

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

<table>
<thead>
<tr>
<th>Gait parameters</th>
<th>Hemiplegia</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Affected side</td>
<td>Unaffected side</td>
</tr>
<tr>
<td>Stride duration (sec)</td>
<td>1.61 (.39)</td>
<td>1.61 (.40)</td>
</tr>
<tr>
<td>Velocity (m/sec)</td>
<td>.58 (.37)</td>
<td></td>
</tr>
<tr>
<td>Stance duration (%)</td>
<td>65.14 (8.08)</td>
<td>69.25 (7.57)</td>
</tr>
<tr>
<td>Stride length (m)</td>
<td>.84 (.36)</td>
<td>.85 (.37)</td>
</tr>
<tr>
<td>Step length (m)</td>
<td>.45 (.17)</td>
<td>.43 (.20)</td>
</tr>
</tbody>
</table>

3.2. Agreement between Barthel Index and spatial-temporal gait parameters

The distribution of the study subjects according to the five-grade scale of independence is shown in Table 2. The most numerous group (i.e., 7 patients) were patients with a mild degree of dependence, the smallest group (i.e., 1 patient) was patient with advanced degree of dependence, six patients were moderately dependent, and four patients were completely independent.

### Table 2. Barthel Index of hemiplegia patients

<table>
<thead>
<tr>
<th>BI</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–20 Completely dependent</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21–60 Advanced degree dependent</td>
<td>1</td>
<td>5.6</td>
</tr>
<tr>
<td>61–90 Moderate degree dependent</td>
<td>6</td>
<td>33.3</td>
</tr>
<tr>
<td>91–99 Mild degree dependent</td>
<td>7</td>
<td>38.9</td>
</tr>
<tr>
<td>100 Independent</td>
<td>4</td>
<td>22.2</td>
</tr>
</tbody>
</table>

The average value for BI in hemiplegia patients was $88.1 \pm 13.9$ (min. 50–max. 100). Studies have shown that the average BI value of the male patients was $86.0 \pm 16.0$, whereas the average BI value of the female patients was $92.2 \pm 6.4$. No statistically significant difference ($p > .05$) in the BI value between genders was found. When the patients were evaluated in terms of being aged over or under 50 years, it was determined that the BI value of those under 50 years was $88.8 \pm 16.4$, and the BI value of those equal and over 50 years was $87.3 \pm 10.9$. No statistically significant difference ($p > .05$) in the BI value between these two groups of patients was found.

Table 3 presents the agreement between BI and spatial-temporal gait parameters. The correlation between BI and spatial-temporal gait parameters was
A new classification of hemiplegia gait patterns based on bicluster analysis of joint moments

3.3. Classification of gait patterns based on biclustering algorithm

In this section, we present results obtained from the algorithm described in Section 2. The proposed algorithm has been executed using different values of $\delta$ and $\alpha$. The best results with a minimum mean square residue score have been obtained for the parameters $\delta$ equal to 1 and $\alpha$ equal to 1.4, respectively. The proposed algorithm has found most significant results within 10 iterations. Figure 1 shows biclusters, i.e., subsets of gait patterns that exhibit the similarity among the subsets joint moments, identified in hemiplegia patients.

Most patients, who as a result of the algorithm have been selected for clusters, have similar gait patterns for each of the three joints, i.e., hip, knee and ankle. The results for the hip joint show similarity between gait patterns of eight patients (3 female and 5 male). The similarity between gait patterns at the knee joint was found for 5 hemiplegia patients (2 female and 3 male). Additionally, for the ankle joint, the gait patterns were similar for 8 patients (3 female and 5 male). The analysis shows that in the group of patients having similar gait patterns at the knee and ankle joints selected for the biclusters, the average velocity was 37.9% lower than in the whole group of hemiplegia patients (0.36 m/sec in hemiplegia patients selected for the biclusters vs. 0.58 m/sec in whole group of hemiplegia patients, $p < .05$). Furthermore, the average stride length in the group of patients selected for the biclusters was 35.7% lower than in the whole group (0.54 m in hemiplegia patients selected for the biclusters vs. 0.84 m in whole group, $p < .05$). The same was observed for the step length, where the average of the step length was 26.7% lower compared

<table>
<thead>
<tr>
<th>Spatial-temporal gait parameters</th>
<th>Correlation coefficient $r$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected side</td>
<td>Unaffected side</td>
<td></td>
</tr>
<tr>
<td>Stride duration (sec)</td>
<td>.151</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Velocity (m/sec)</td>
<td>-.013</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Stance duration (sec)</td>
<td>.048</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Stride length (m)</td>
<td>0.11</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Step length (m)</td>
<td>0.01</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Affected side</td>
<td>Unaffected side</td>
<td></td>
</tr>
<tr>
<td>Stride duration (sec)</td>
<td>-.22</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Velocity (m/sec)</td>
<td>- .013</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Stance duration (sec)</td>
<td>.178</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Stride length (m)</td>
<td>0.122</td>
<td>$p &lt; .05$</td>
</tr>
<tr>
<td>Step length (m)</td>
<td>0.187</td>
<td>$p &lt; .05$</td>
</tr>
</tbody>
</table>

low ($r < .2$), but statistically significant ($p < .05$) for both affected and unaffected sides.
to the whole group (0.33 m for patients discovered by the biclusters vs. 0.45 m for the whole group, \( p < .05 \)). The difference between spatial-temporal gait parameters in patients selected for bicluster, which present similar gait patterns at the hip joint, and spatial-temporal gait parameters in the whole group are not statistically significant (\( p > .05 \)).

4. Discussion

Gait deviations observed in patients following stroke are remarkably diverse. Over the last few years, numerous methods have been applied to detect gait patterns in stroke patients, i.e., to classify groups of
patients with similar gait patterns [15], [21]. A very well known technique is cluster analysis, which has been used to classify hemiplegia gait patterns based on temporal-distance parameters and sagittal joint kinematics during gait [18]. In this paper, we have proposed the technique biclustering as an alternative to clustering methods. The proposed method is based on a simultaneous grouping of rows and columns of a data matrix. Biclusters were extracted from a set of 87 gait cycles of joint moments for 18 patients. In our study, we decided to group hemiplegia patients according to their gait characteristics instead of dividing them according to gender, BMI, and age. Bicluster analysis was conducted using joint moments only, and as the result of this method, well-separated biclusters were produced from the data. Each bicluster extracted from data contains only a specific subset of hemiplegia patients behaving similarly over a subset of joint moments at the hip, knee, and ankle; as such, it fulfills the primary purpose of this study. The need for this type of data was obvious given the constraints of the other methods that were used to analyze the data in earlier research. It was difficult to assess the differences within joint moment patterns occurring during the gait cycle. If we take into account that most of the studies analyzed gait parameters using cluster analysis, our results could be considered valid and reliable. Although biclustering is potentially very useful, bicluster analysis has never been widely used in clinical biomechanics. This may be caused by lack of knowledge about the new method and biomechanists’ concerns regarding procedural problems related to this technique. One of the main disadvantages of the bicluster analysis method is that it is sensitive to variations in the gait cycles. The various gait cycles of the particular hemiplegia patients may be selected for different biclusters. However, considering that the main objective of this study was to observe the natural grouping of hemiplegia patients, which was aimed to demonstrate the presence of similar gait patterns, this type of issue does not represent a critical source of error.

The spatial-temporal gait parameter values including velocity, stride duration, stance duration, stride length, and step length during walking were different between the affected and unaffected sides of hemiplegia patients. In our study, a hemiplegic patient gait was characterized by low velocity (0.58 m/sec), which is in agreement with other studies [20]. The differences in stance duration and step length between both affected and unaffected sides in hemiplegia patients were significant (p < .05). The goal of this research was also to determine if spatial-temporal gait parameters are correlated with the Barthel Index (BI).

The Barthel Index test was selected as the useful instrument for the evaluation of daily living activities because its validity and reliability have been proven in many studies [8]. Our results show that seven patients (38.9%) were mildly dependent, six patients (33.3%) were moderately dependent, four patients (22.2%) were completely independent, and one patient (5.6%) had an advanced degree of dependence. Although many studies of gait parameters in hemiplegia patients have been presented [2], [9], [12], [14], the relation between the Barthel Index and spatio-temporal parameters during gait observation had been unknown. As a result of our analysis, we found that the degree of independence correlates with some spatio-temporal parameters. The correlation between BI and all measured spatial-temporal gait parameters was low ($r < .2$), but statistically significant ($p < .05$) for both affected and unaffected sides. The obtained results indicate that the relationship between independence in daily living activities and gait parameters exists. These findings are consistent with results presented by Abel et al. [1], who state that the degree of motor recovery corresponds to improvement in the spatio-temporal variables; the gait efficiency improvement is related to velocity increase. However, we did not find any significant differences in BI values between genders ($p > .05$), which was also reported by Carod-Artal et al. in [6].

5. Conclusions

In this article, we proposed the method of biclustering, which allows automatic identification of groups of hemiplegia patients with similar hip, knee, and ankle joint moments during gait. The results obtained by the proposed algorithm can be crucial for further clinical trials. The relationship between gait patterns and underlying impairments would allow clinicians to develop individual rehabilitation strategies according to a patient’s specific needs.

Acknowledgements

This study was supported by The Polish Ministry for Science and Higher Education.

References


