MARMARA
MEDICAL JOURNAL

Original Article

https://dergipark.org.tr/tr/pub/marumj

Classification of hemiplegia through gait analysis and machine learning methods

Hazal TAS[1], Ahmet YARDIMCI[1], Hilmi UYSAL[2], Ugur BILGE[1]

[1] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Akdeniz University, Antalya, Turkey
[2] Department of Neurology, Faculty of Medicine, Akdeniz University, Antalya, Turkey

Corresponding Author: Ugur BILGE
E-mail: ubilge@akdeniz.edu.tr

Submitted: 03.01.2023 **Accepted:** 13.04.2023

ABSTRACT

Objective: Gait analysis is a method that is used for understanding normal walking and determining the stage of the disease as it affects walking. It is important to objectively determine the stage of the disease in order to decide interventions and treatment strategies. This study aims to determine the Brunnstrom Stage of the hemiplegic patients with an analysis of gait data.

Patients and Methods: In the first part of the study, the gait signal data were taken from 28 post-stroke hemiplegic patients and 7 healthy individuals with three-axis accelerometers. In the second part, new gait data were collected from 15 healthy individuals through an accelerometer on the anteroposterior axis.

First the accelerometer signals were decomposed to Daubechies 5 (Db5) level six wavelets using MATLAB software. Subsequently, these attributes were classified through several classifier and machine learning algorithms on WEKA and MATLAB software packages to predict the stages of hemiplegia.

Results: The highest accuracy rate in the prediction of hemiplegia stage was achieved with the LogitBoost algorithm on WEKA with 91% for 35 samples, and 90% for 50 samples. This performance was followed by the RUSBoosted Trees algorithm on the MATLAB software with an accuracy of 86.1% correct prediction.

Conclusion: The Brunnstrom Stage of hemiplegia can be predicted with machine learning algorithms with a good accuracy, helping physicians to classify hemiplegic patients into correct stages, monitor and manage their rehabilitation.

Keywords: Hemiplegia, Stroke, Gait analysis, Brunnstrom, Machine learning

1. INTRODUCTION

Hemiplegia as a result of a stroke affects lots of people every year. Approximately two-thirds of those who have had a stroke do not become ambulatory without assistance and only half of those who were rehabilitated were able to walk independently [1]. The hemiplegic gait is defined as a movement pattern and body posture with a heavy, challenging or weak coordination that the hemiplegic patient experiences during the gait [2]. Although, there are many methods to evaluate hemiplegic patients, the Brunnstrom Staging, is the preferred one amongst all the methods. Therefore, we preferred to evaluate our patients according to the Brunnstrom Staging. Brunnstrom consists of six stages of improvement. However, only the hemiplegic patients at stages III, IV, V and VI were included in the study as the gait is not in question during the first two stages [3].

Gait analysis is widely used in determining human gait disorders. There are two fundamental approaches developed for the gait analysis to analyse the human gait. The first approach uses the marking systems that include video-based systems, active magnetic trackers, and optical marker systems, to acquire the human gait motion. However, as they depend on an artificially-created source, they cannot be used outside a laboratory setting [4]. Muro-De-La-Herran et al., named them as "unwearable sensor systems". The second approach involves wearable sensors. Wearable sensor systems make gait analysis possible outside laboratories and can gather information about the gait during the daily activities of an individual. Wearable systems use the sensors located in various parts of the body such as feet, knees, thighs, or waist [5].

How to cite this article: Tas H, Yardımcı A, Uysal H, Bilge U. Classification of hemiplegia through gait analysis and machine learning methods. Marmara Med J 2024: 37(1):5-10. doi: 10.5472/marumj.1379890

Through a gait analysis, it is possible to define the gait phases, to determine the kinematic and kinetic parameters of human gait events, and to quantitatively evaluate the musculoskeletal functions. Gait analysis has been used since the 19th century to for sports or safety purposes. For instance, in some types of sports training, the method is applied to recognize the faults in athlete performances so that they can improve. For security purposes, interest may centre on distinguishing and identifying persons based on a general characterization of their silhouette and the movements between the subject's different body segments when walking [5,6].

Today, there is a great variety of signal processing methods, and the Wavelet Transform (WT) technique is one of them. The WT is a mathematical method that gives the time-frequency representation of a signal. The WT is an effective signal processing tool thanks to its features such as the time-frequency localization (obtaining a signal at a specific time and frequency, or extraction of attributes at various locations on different scales) and differential-proportional filtering (distinguishing between the signals with various frequencies) [7,8].

In a study conducted by Lee et al., in 2018, through the use of a wearable system, it was aimed to distinguish the hemiplegic gait by extracting the simple properties of the acceleration signals caused by the asymmetry during the gait. The wearable system designed was equipped with a three-axis accelerometer and a three-axis gyroscope. In the study, which employed a "random forest" algorithm for classification, the accuracy, sensitivity, specificity and positive predictive value were found to be 100% [9].

Another study aimed to determine the symmetry, regularity and stability of the gaits in hemiparetic patients in the post-stroke period by using the stability index that was based on a dynamic time-bending algorithm, a sample entropy method and an empirical mode decomposition. The study was conducted with 15 healthy control subjects and 15 post-stroke hemiparetic patients. A total of four different machine learning methods were used, which included the decision support machine, decision tree, multilayer neural network and k-nearest neighbour (kNN), and the maximum area under curve (AUC) value was given as 0.94 by the kNN classifier [10].

In our study we first used WT to decompose gait data then applied several machine learning algorithms to establish the Brunnstrom Stage of hemiplegic patients accurately.

2. PATIENTS and METHODS

Data collection tools

The signals used in the study were obtained through a three-axis accelerometer from the waist of participants (3031-010, IC-Sensors, USA, size: 4x4x3 mm; weight: 0.3 g; range: ±10 g; frequency reaction: 0-500 HZ). The accelerometer was orthogonally mounted to record the anteroposterior (x), lateral (y) and vertical (z) signals. The accelerometers were calibrated by measuring their outputs under a controlled inclination. Then, they were fixed onto an acrylic plate for the waist belt. An elastic waist belt was put on the lumbosacral area of the vertebral column on the back of the patient in the proximity of the center of gravity while the patient was standing upright. The accelerometer unit was connected to a portable data recorder (Micro 8, Shimadzu, 36 Japan) through an interface circuit. This data recorder consists of one Central Processing Unit (CPU), one 10-bit A/D converter, one Integrated Circuit (IC) card interface and one removable 2-MB IC memory card. The interface circuit contains three amplifiers as an anti-aliasing filter for each direction, as well as three second-degree analog Butterworth low permeable filters. The cut-off frequency is 500 Hz. The accelerometer outputs were digitalized by the data recorder at a sampling rate of 1024 Hz and recorded on an IC memory card. Following the completion of the measurements, the data were transferred to a personal computer through a card reader for analysis purposes [11].

The data set

In the first part of the study, the gait signal data taken from 28 post-stroke hemiplegic patients and 7 healthy individuals were used. Of the patients, 9 were female and 19 were male, while the healthy ones were all female. The gait signal data was taken from post-stroke hemiplegic cases and healthy individuals in 2005-2006 at the University of Chiba, Japan [12]. The approval of the local ethics committee was obtained from the University of Chiba, Japan, and all subjects gave their written informed consent for the data used in the study.

Even though, the Brunnstrom Staging consists of six stages, only the patients at stages III, IV, V, and VI were included in the study as the gait is not in question at the first two stages. Table I gives the distribution of patients and their Brunnstrom Stages.

Table I. *Distribution of patients according to Brunnstrom Staging*

Brunnstrom Stage	Number of Patients	Gender (F/M)
III	12	3/9
IV	9	3/6
V	4	3/1
VI	3	0/3

Table II gives the mean age, weight and height characteristics of the hemiplegic and healthy individuals included in the study.

Table II. *Demographic characteristics of the subjects*

Group	Age	Height	Weight
Patient	67 ± 11	155 ± 8.89	55.27 ± 9.81
Healthy	61 ± 5.1	149±1.41	49.66±1.69

In the second part of the study, the gait signals were collected additionally from 15 healthy individuals through an accelerometer with an anteroposterior axis. The total number of samples was increased to 50 by merging the same attributes obtained through the signals decomposed by the WT method with the attributes of the anteroposterior axis of the data set used in the first part of the study.

Table III. Classification algorithms and their accuracy rates

		Accuracy rates (%) with 35 samples	Accuracy rates (%) with 50 samples
WEKA Classification algorithms	LogitBoost	91.4	90.0
	Iterative Classifier Optimizer	91.4	90.0
	J48	88.5	86.0
	CVR	85.7	82.0
	OneR	87.7	78.0
	Bagging	85.7	78.0
	REPTree	80.0	76.0
	Random Forest	74.2	72.0
	Random SubSpace	74.2	68.0
	Multi Class Classifier	65.7	76.0
	AdaBoost	57.1	52.0
MATLAB Software Classifiers	RUSBoosted Trees	86.1	86.0
	Complex Tree	83.3	82.0
	Subspace Discriminant	75.0	76.0
	Linear Discriminant	69.4	64.0

Data analysis

Within the scope of the study, the gait signal data taken from 28 post-stroke hemiplegic patients and 22 healthy individuals were used. The MATLAB and WEKA software were used to analyse the signal data.

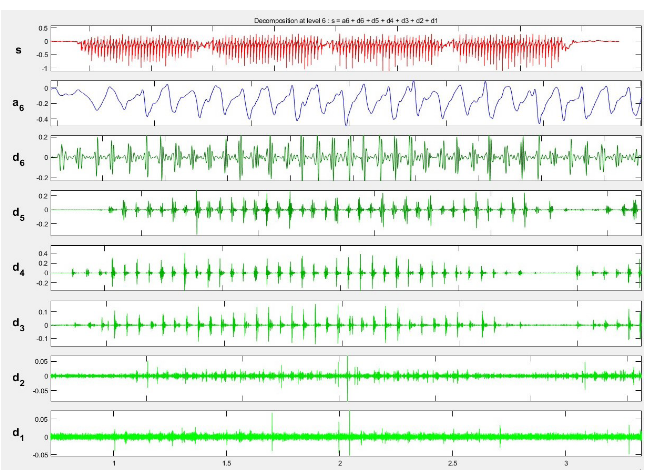

Fig.1. The Db5 main wavelet was decomposed up to level 6, thus each gait signal was decomposed into d1-d6 detail bands and a6 approximate sub-band (s denotes the original state of the signal, d denotes the detail function, and a denotes the approximation function.)

Extraction of attributes from the gait signals through discrete wavelet transform

The selection of the proper wavelet and the number of decomposition levels is of utmost importance for the analysis of the signals through discrete wavelet transform. The dominant frequency components are considered while selecting the number of decomposition levels [13].

Initially, every single axis of the signal taken from each subject was saved as MATLAB file format and thus 105 pieces of signal data were obtained. Subsequently, 105 decomposition procedures were carried out to obtain the coefficients to be used in the classification process for each axis. This process was carried out by using the Db5 decomposed main wavelet, which is the most commonly preferred item in the literature for gait signal analysis [14]. The Db5 main wavelet was decomposed up to level 6, thus each gait signal was decomposed into d1-d6 detail bands and a6 approximate sub-band (Figure 1.)

The attributes of the approximation signal at level 6 were selected to create the classification data following the decomposition of the signals into 6 levels through the use of the Db5 main wavelet (Figures 2-5).

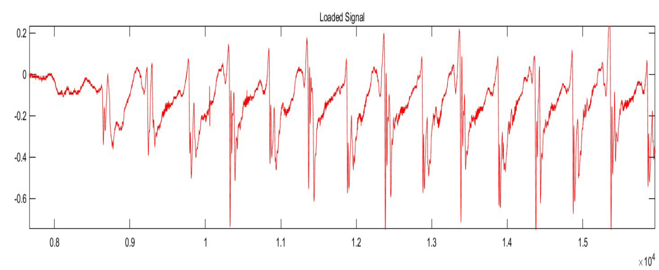

***Fig. 2.** The gait signal's amplitude of the anteroposterior axis taken from the healthy individual*

These attributes are minimum, maximum, mean, median, absolute deviation from the mean, absolute deviation from the median, the first norm of the vector (L1 norm), the second norm of the vector (L2 norm), and maximum norm. Table IV shows some examples of the attributes on the anteroposterior axis for each stage.

Figure 2 shows the gait signal of the anteroposterior axis taken from the healthy individual while Figures 4 shows the gait signal of the anteroposterior axis taken from the patient at the 3rd Brunnstrom Stage. Figure 3 shows the approximation signal at level 6 of the Db5 wavelet of the gait signal of the anteroposterior axis taken from the healthy individual while Figure 5 shows the approximation signal at level 6 of the Db5 wavelet of the gait signal of the anteroposterior axis taken from the patient at the 3rd Brunnstrom Stage.

Table IV. The attributes of the approximation signal at level 6 (anteroposterior axis)

Patient no	Stg	Mean	Median	Max	Min	Med Abs Dev	Mean Abs Aev	L1 Norm	L2 Norm	Max Norm
7	healthy	-0.674	0.735	1.202	-2.346	0.532	0.547	1369	39.63	2.346
8	3	-0.554	-0.554	1.038	-1.769	0.330	0.366	1859	39.7	1.769
23	4	-0.746	-0.731	1.712	-3.047	0.576	0.632	2062	52.2	3.047
29	5	-0.938	-0.955	2.154	-3.466	0.926	0.954	2379	65.82	3.466

Estimation of the Brunnstrom Stage through classification algorithms

The first data set consists of the attributes, which were obtained from the level 6 approximation signal coefficients as a result of the WT of the gait signals taken from the anteroposterior, lateral and vertical axes of 35 individuals (28 patients, 7 healthy), as well as the Brunnstrom stages of those individuals. The second data set, on the other hand, consists of the attributes extracted from the level 6 approximation signal coefficients as a result of the WT of the signals taken from 15 healthy gaits, which were added to the data set that was already present. Nevertheless, this data set only includes the attributes obtained from the signals taken from the anteroposterior axis.

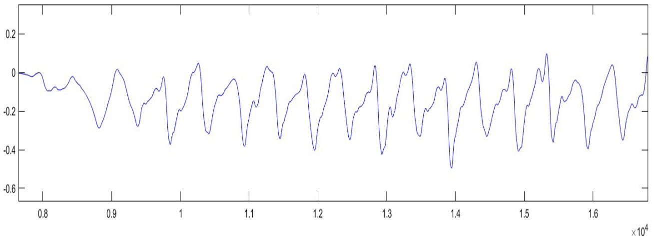

Fig. 3. *The approximation signal's amplitude at level 6 of the Db5 wavelet of the gait signal of the anteroposterior axis taken from the healthy individual*

As specified in the very beginning, the objective of the study was to estimate the Brunnstrom stages based on the gait signal data of individuals. In this context, initially, the classification algorithms on the Weka software were used, which was then followed by those on the MATLAB software, for the classification problem in this study.

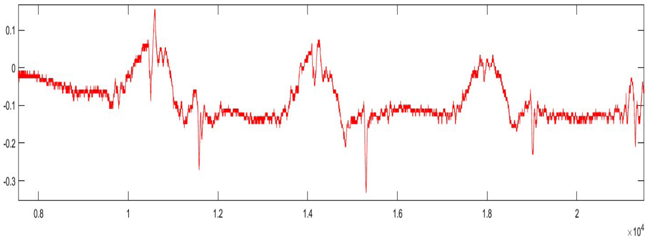

Fig. 4. *The gait signal's amplitude of the anteroposterior axis taken from the patient at the 3rd Brunnstrom Stage*

On the WEKA software the following algorithms were used: The Iterative Classifier Optimizer, AdaBoost, Bagging, Classification via Regression (CVR), LogitBoost, OneR, J48, Random Forest, Random SubSpace, MultiClass Classifier and RepTree classification algorithm.

The WEKA includes various strategies for training and testing. As the dataset is relatively small, the 10-fold cross-validation technique was employed in the study as a test option in all algorithms used for the solution of the problem. As the dataset is relatively small, the 10-fold cross-validation technique was employed in the study as a test option in all algorithms used for the solution of the problem. The cross-validation technique is one of the methods of splitting the data set into parts for training and evaluating the model. In this technique, the dataset is randomly divided into two parts according to a determined k ratio, and the first part is used for both trainings and the second for testing.

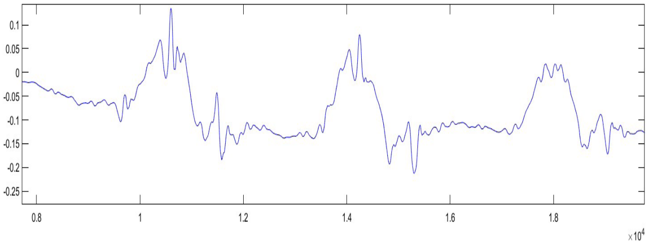

Fig. 5. *The approximation signal's amplitude at level 6 of the Db5 wavelet of the gait signal of the anteroposterior axis taken from the patient at the 3rd Brunnstrom Stage*

3. RESULTS

Results of the classification conducted with a 35-person data set on the WEKA software

The Iterative Classifier Optimizer algorithm yielded the same results as the LogitBoost algorithm, which was caused by the fact that it used the LogitBoost algorithm as an iterative classifier.

Results of the classification conducted with a 50-person data set on the WEKA software

In this section, new data were added to the current data set, and the classification was iterated with the same algorithms. Only the relevant axis of the current data was used as the newly-obtained gait signal covered the gait signal taken from 15 healthy individuals, using an accelerometer with an anteroposterior axis.

In a similar manner to the current signal, the 15 newly-added gait signals were also decomposed at 6 levels through the Db5 main WT. Then again, the attributes of the approximation signals at level 6 were selected and added to the current data set.

As shown in Table III, a decrease was observed in the accuracy rates as a result of the classification process iterated with the new data set.

Results of the classification conducted with a 35-person data set on the MATLAB software

The data set, comprised of the attributes obtained from the gait signals as a result of the WT, was classified by using the classification algorithms available on the MATLAB Software, to estimate the Brunnstrom stages of the individuals. Table III gives the classification algorithms used in the MATLAB software and the accurate classification rates. In a similar manner to the WEKA software, the 10-fold cross-validation technique was also employed in the MATLAB software as a test option in all algorithms used for the solution of the problem.

The RUSBoosted Trees algorithm yielded the highest accuracy rate for the solution to this problem on the MATLAB software, and the accurate classification rate was 86.1%. The true positive rates of the algorithm were found as 0.92 for the 3rd stage, 0.89 for the 4th stage, 0.75 for the 5th stage, 1 for the 6th stage and 0.86 for the healthy individuals. While, the RUSBoosted Trees algorithm accurately classified all of the stage 6 patients, it made one mistake in other groups.

Results of the classification conducted with a 50-person data set on the MATLAB software

The attributes, which were obtained from the gait signals taken from 15 healthy individuals through an accelerometer with an anteroposterior axis, were added to the current data set, and the classification was iterated by using the same algorithms on the MATLAB software.

As a result of the iterative RusBoosted Trees algorithm, the accuracy rate did not change and was found to be 86.1%. Nevertheless, the True Positive rates varied based on the groups. They were found as 1 for the third stage, 1 for the fourth stage, 0 for the 5th stage, 0 for the 6th stage and 0.77 for the healthy individuals. While the Subspace Discriminant algorithm accurately classified all of the patients at stages 3 and 4, it did not accurately classify any of the patients at stages 5 and 6.

4. DISCUSSION

In our study, the WT technique and classification algorithms were used to estimate the Brunnstrom stages of the hemiplegic patients based on their gait signals. On the MATLAB software, the attributes of the approximation signal at level 6 were selected from the gait signals decomposed into 6 levels through the Db5 main wavelet.

It could be argued that the 28 hemiplegic gait and 7 healthy gait samples are insufficient for the study to generalize the success of the classification results due to the insufficient number of samples. The accuracy rates achieved via cross-validation are good and are improved further by the addition of the gait signal data of the healthy individuals due to the difficulty experienced in finding any data that were similar to the gait signal data of the hemiplegic elderly, whose Brunnstrom stages were known and which also constituted the data set of the study. As a result the number of cases reached 50 following the addition of 15 healthy gait signal data. This caused the drop in the accuracy rates and overall performance of the machine learning algorithms. The probable causes of the decrease in the accuracy rates following the reclassification made through the newly-created data set are as follows:

The current data set used the attributes obtained following the WT of the gait signals taken via a three-axis accelerometer. In the new data set, however, only the gait signal data taken from the anteroposterior axis were used. Despite the increase in the number of cases, it is believed that the accuracy rates decreased as a result of the decrease in the number of attributes.

Each analysis can yield different results in machine learning techniques. It is believed that this is caused by the varying data set and the parameters used in the algorithm.

Besides, despite the decrease observed in the classification accuracy rates in all algorithms, an increase of 11% was observed in the Multi Class Classifier algorithm.

The results of the Boosting algorithm were another striking issue in the classification results. Among the accurate classification rates of the AdaBoost and LogitBoost algorithms, a difference was observed by 34% for the first algorithm, which was followed by a 38% difference for the latter. It is believed that this was caused by the fact that the LogitBoost algorithm is designed to solve the excessive conformity problem stemming from the extremely noisy data, which is a problem for the AdaBoost algorithm. The LogitBoost linearly decreases the training error for the solution to this problem [15].

The LogitBoost algorithm and Iterative Classifier Optimizer algorithm yielded the same results. This was caused by the fact that the Iterative Classifier Optimizer algorithm used the LogitBoost algorithm as an iterative classifier.

The accuracy rate did not change as a result of the iterative RuSBoosted Trees algorithm on the MATLAB software. Nevertheless, it cannot be said that the algorithm yielded the same results because the Accurate Positive rates changed.

Using the time-frequency analysis, Ning Wang et al., carried out an accelerometer-based classification of the gait patterns. To determine five different human gait patterns through the data obtained by the use of a three-axis accelerometer attached to the waist over the iliac spinal cord, 33-dimensional time-frequency field properties were developed and evaluated in the study. In the study, 52 subjects were asked to walk on a flat surface along the hallway, go up the stairs and then come down. The time-frequency properties of the acceleration data were developed in the anterior-posterior (AP), mediolateral (ML) and vertical (VT) directions. The acceleration data in each direction were decomposed into three detailed signals on different wavelet scales by using the wavelet packet transform. The Root Mean Square (RMS) values and standard deviations were calculated for the signals decomposed on scales varying from 5 to 2, which correspond to the frequency band of 0.78 – 18.75 Hz. Although, the MV acceleration did not show any significant difference between the gait patterns, the RMS value of the acceleration signal was shown to be a distinguishing property as it was in previous studies. The RMS values were only calculated in the AP and VT directions for the wavelet coefficients at levels 2 to 5, which correspond to 0.78 – 18.74 Hz [15].

ConclusIons

We believe, WT of gait signals, together with machine learning algorithms presented in this study can be used to classify hemiplegic patients into correct Brunnstrom Stages of hemiplegia with high accuracy, helping physicians to monitor and manage the rehabilitation process of their patients.

http://doi.org/10.5472/marumj.1379890
Marmara Med J 2024;37(1): 5-10

Acknowledgement

We would like to thank to Dr. Toshiyo Tamura, Dr. Masaki Sekine and University of Chiba, Japan, for collecting the original data used in this study. We thank Philippa Price for proofreading the article.

Compliance with the Ethical Standards

Ethics Committee approval: The approval of the local ethics committee was obtained from the University of Chiba, Japan, and all subjects gave their written informed consent for the data used in the study.

Conflicts of interest: Authors declare no conflicts of interest.

Funding: Authors received no specific funding for this work.

Authors contribution: HT: Undertook the tasks of processing the walking signal and feature extraction, applying machine learning techniques, and writing the article. AY: Collected and normalized the walking signal data used in the article. HU: Provided support with his clinical expert knowledge. UB: Undertook the design of the study, evaluation of the machine learning results used in the study, and consultancy of the study. All authors approved the final manuscript.

The authors do not report any conflict of interest regarding this manuscript.

REFERENCES

[1] Jørgensen H S, Nakayama H, Raaschou H O, Olsen T S. Recovery of walking function in stroke patients: The copenhagen stroke study. Arch Phys Med Rehabil 2014; 76:27-32. doi:10.1016/S0003-9993(95)80038-7.

[2] Kuan TS, Tsou JY, Su FC. Hemiplegic gait of stroke patients: The effect of using a cane. Arch Phys Med Rehabil 1999; 80: 777-84. doi: 10.1016/s0003-9993(99)90227-7.

[3] Van Sant A F. Movement System Diagnosis. J Neurol Physl Ther 2017; 41: 10-6. doi: 10.1097/NPT.000.000.0000000152.

[4] Guo Y, Wu D, Liu G, Zhao G, Huang B, Wang L. A low-cost body inertial-sensing network for practical gait discrimination of hemiplegia patients. Telemed J E Health 2012; 18: 748-54. doi: 10.1089/tmj.2012.0014.

[5] Muro-de-la-Herran A, Garcia Zapirain B, Mendez-Zorrilla A. Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical Applications. Sensors (Basel) 2014; 14: 3362-94. doi: 10.3390/s140203362.

[6] Tao W, Liu T, Zheng R, Feng H. Gait analysis using wearable sensors. Sensors (Basel) 2012; 12: 2255-83. doi: 10.3390/s120202255.

[7] Meyer Y. Wavelets: Algorithms and applications. Philadelphia: Society for Industrial and Applied Mathematics, 1993: 133.

[8] Samant A. Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis. Comput-Aided Civ Infrastruct Eng 2003; 15: 241-50. doi: 10.1111/0885-9507.00188.

[9] Lee J, Park S, Shin H. Detection of hemiplegic walking using a wearable inertia sensing device. Sensors (Basel*)* 2018; 18: 1736. doi: 10.3390/s18061736.

[10] Li M, Tian S, Sun L, Chen X. Gait analysis for post-stroke hemiparetic patient by Multi-Features Fusion Method. Sensors (Basel) 2019; 19: 1737. doi: 10.3390/s19071737.

[11] Sekine M, Abe Y, Sekimoto M, et al. Assessment of gait parameter in hemiplegic patients by accelerometry. Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No.00CH37143). Chicago 2000; 3: 1879-82. doi: 10.1109/IEMBS.2000.900456.

[12] Yardımcı A. Fuzzy logic based gait classification for hemiplegic patients. Lect Notes Comput Sci 2007; 4723: 344-54.

[13] Toprak IB. Analysis of EEG signals using the wavelet transform and artificial neural network [master's thesis]. University of Süleyman Demirel, Isparta 2007.

[14] Aydın F, Aslan Z. Yapay öğrenme yöntemleri ve dalgacık dönüşümü kullanılarak nörodejeneratif hastalıkların teşhisi. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi 2017; 32: 749-66. doi: 10.17341/gazimmfd.337621

[15] Wang N, Ambikairajah E, Lovell N H, Celler B G. Accelerometry based classification of walking patterns using time-frequency analysis. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Lyon 2007; 36: 4899-902.