INGENIERÍAS USBMed

https://revistas.usb.edu.co/index.php/ingUSBmed

ARTÍCULO DE INVESTIGACIÓN

Uso de algoritmos de aprendizaje automático para la clasificación de la marcha de enfermedades neurodegenerativas

Using Machine Learning Algorithms for Neurodegenerative Disease Gait Classification

Andrés Mauricio Cárdenas[1]
Beatriz Gómez[2]
Luis Carlos Ealo[3]
Juliana Uribe Pérez[4]

[1] *Universidad de San Buenaventura, Medellín, Colombia.*
Email: andresm.cardenas@udea.edu.co

[2] *Universidad de San Buenaventura, Medellín, Colombia.*
Email: beatriz.gomez@usbmed.edu.co

[3] *Universidad de San Buenaventura, Medellín, Colombia.*
Email: luis.ealo@usbmed.edu.co

[4] *Universidad de Antioquia, Medellín, Colombia.*
Email: juliana.uribep@udea.edu.co

Tipo de artículo: Investigación.
Recibido: 01-08-2022.
Revisado: 13-10-2022.
Aprobado: 12-11-2022.
Doi: 10.21500/20275846.6081

Referenciar así:
A. M. Cárdenas, B. Gómez, L. C. Ealo and J. Uribe Pérez, "Uso de algoritmos de aprendizaje automático para la clasificación de la marcha de enfermedades neurodegenerativas," *Ingenierías USBMed*, vol. 14, no. 2, pp. 8–14, 2023.

Disponibilidad de datos:
todos los datos relevantes están dentro del artículo, así como los archivos de soporte de información.

Conflicto de intereses:
los autores han declarado que no hay conflicto de intereses.

Editor: Andrés Felipe Hernández. Universidad de San Buenaventura, Medellín, Colombia.

Resumen. Las enfermedades neurodegenerativas afectan el sistema neuromusculoesquelético generando trastornos del movimiento. La detección de los síntomas suele producirse en las últimas fases de la enfermedad, por lo que una detección temprana ayudaría a introducir terapias para reducir los efectos de las enfermedades y retrasar el deterioro. La base de datos PhysioNet proporciona información sobre la biomecánica de la marcha de voluntarios sanos y de pacientes de Parkinson (PD), esclerosis lateral amiotrófica (ALS) y Huntington (HD). En este trabajo se utilizan datos espacio-temporales para medir el coste energético y la densidad espectral de potencia en esas patologías. Estos parámetros se analizaron estadísticamente para definir descriptores explicativos. Posteriormente, se utilizan la técnica fuzzy c-means, algoritmo de aprendizaje para el análisis de datos multivariados - LAMDA, y redes neuronales para clasificar entre las enfermedades neurodegenerativas y el grupo de control. Se utilizó el método de validación cruzada para evaluar los resultados del algoritmo de clasificación. El análisis estadístico mostró que el coste de la energía aumentaba en la fase de apoyo, la velocidad de la marcha disminuía en condiciones críticas de la enfermedad y la cadencia era diferente según el tipo de enfermedad. Se entrenaron los algoritmos con cuatro clases a priori. Los ajustes de clasificación fueron 92.5% para la red neuronal, 80% para el método LAMDA, y 56.1% para el Fuzzy C-means. Para mejorar los resultados, se entrenaron clasificadores de dos-clases: Ctrl+PD, Ctrl+PD y Ctrl+HD. El emparejamiento mejoró el ajuste de LAMDA a un 98.3%, la red neuronal con un 97.0% y Fuzzy C-means con un 90.2%. El uso potencial de estas técnicas de clasificación permitirá la detección temprana de enfermedades neurodegenerativas, incluyendo nuevos dispositivos que permitan el análisis de la marcha.

Palabras Clave. Enfermedades Neurodegenerativas, Redes Neuronales, Fuzzy C-means, Análisis de datos multivariantes, Aprendizaje automático.

Abstract. Neurodegenerative diseases affect the neuromusculoskeletal system generating movement disorders. The detection of symptoms usually occurs in the late stages of the disease, thus early detection would help to introduce therapies for reducing the effects of the diseases and delay deterioration. PhysioNet database provides information on gait biomechanics of healthy volunteers and patients with Parkinson's disease (PD), Amyotrophic Lateral Sclerosis (ALS) and Huntington's disease (HD). In this work, spatio-temporal data are used to measure the energy cost and power spectral density in these pathologies. These parameters were statistically analyzed to define explanatory descriptors. Subsequently, fuzzy c-means techniques, learning algorithm for multivariate data analysis - LAMDA, and neural networks are used to classify between the neurodegenerative diseases and the control group. Cross-validation method was used to evaluate the results of the classification algorithm. Statistical analysis showed that energy cost increased in the support phase, gait speed decreased in critical disease conditions, and cadence was different according to disease type. Algorithms were trained with four classes. The classification fits were 92.5% for the neural network, 80% for the LAMDA method, and 56.1% for the Fuzzy C-means. In order to improve the results, two-class classifiers were trained: Ctrl+PD, Ctrl+PD and Ctrl+HD. This matching improved the fit of LAMDA to 98.3%, the neural network with 97.0% and Fuzzy C-means with 90.2%. The potential use of these classification techniques will enable early detection of neurodegenerative diseases, including new devices that allow gait analysis outside the laboratory.

Keywords. Neurodegenerative Diseases, Fuzzy C-means, Neural Networks, Multivariate Data Analysis, Machine Learning.

Nomenclature

Fuzzy C-Means (FCM), Neural Networks (NN), Support Vector Machines (SVM), Neurodegenerative Diseases (ND), Parkinson Diseases (PD), Huntington Diseases (HD), Alzheimer Diseases (AD), Amyotrophic Lateral Sclerosis (ALS), Control (Ctrl).

I. Introduction

Neurodegenerative diseases are progressive disorders characterized by an accelerated degeneration of neurons, causing limitations on speech, memory, movement, and balance [1]–[3]. Examples of neurodegenerative diseases include Parkinson's (PD), Huntington's (HD), Alzheimer's (AD) diseases, and Amyotrophic Lateral Sclerosis (ALS).

Symptoms like impaired balance and coordination are common in Parkinson's disease (PD) [4]. Amyotrophic Lateral Sclerosis (ALS) damages the neuromotor system resulting in motor dysfunction and muscle weakness [5]. Huntington's disease (HD) is characterized by fast, sudden and uncontrollable facial movements, spasmodic movements, bradykinesia and an unstable gait [6]. All these symptoms considerably affect the person gait patterns, affecting their quality of life [7].

The human gait has been analyzed to find repetitive patterns caused by neurodegenerative diseases and to provide an early diagnosis. Melo Roiz et al. on [8] analyzed the spatial-temporal gait parameters of subjects with PD and compared them with a healthy group (Ctrl). Significant differences were detected between PD and Ctrl groups for speed and stride length, the initial contact, and the maximum extension in the terminal contact of heel, as well as in the maximum flexion degree in the middle swing phase. A similar statistical study was conducted by Sofuwa et al. on [9]. They found a significant reduction in gait length, gait speed, and ankle and hip force. Delval et al. on [10] discovered akinesia on the gait initiation cycle and a reduction in the range of joint angles movement for Huntington's disease.

The gait characteristics detection can help in the early diagnosis of neurodegenerative diseases, improving the decision making for medication, treatment plans, and adaptation of a new lifestyle [11]. Automatic classification techniques have been used to diagnose neurodegenerative diseases before their onset. Bilgin [12] presents the use of discrete Wavelet transform to extract six frequency bands from foot force signals, then he used the Bayesian Naïve and discriminant analysis technique to classify PD, ALS, HD, and Ctrl groups, achieving a maximum classification fit of 90.93%. On [13] four classification techniques were implemented: the Support Vector Machines (SVM), Random Forest, Multilayer Perceptron and the K Nearest Neighbor. The highest-accuracy classification rate between healthy and neurodegenerative diseases groups was 96.83%, achieved by using SVM. Similarly, in [14] the extra trees method and convolutional neural networks were additionally included, however the best accuracy was achieved with the K Nearest Neighbor method. Another example is presented in [15], which shows that uses of radial function neural networks have been able to classify 93.75%. The star (K*) and Multilayer Perceptron machine learning algorithms were used to classify between PD+ALS+HD achieving accuracies on the order of 96.0% [16]. By applying techniques such as nearest neighbor, decision tree, random forest, adaptive boosting, and naïve Bayes, patients were classified with neuromusculoskeletal gag [17].

Literature reveals multiple algorithms used for classification of neurodegenerative diseases; however, according to our knowledge, this is the first study to use an estimation of energy consumption and power spectral density of spatiotemporal gait data as explanatory descriptors of neurodegenerative gait. In addition, the present study explores the use of two-class classifiers to improve classification accuracy. The objective of this work is to evaluate the effectiveness of machine learning algorithms using two-class classifiers to identify neurodegenerative diseases using clinical information, sociodemographic data, and gait biomechanics.

II. Methods

We used Statgraphics^{TM} software to perform the statistical analysis. Matla^{TM} was used for training the Fuzzy C-Means (FCM) and Neural Networks (NN) classification techniques. The Learning Algorithm for Multivariate Data Analysis (LAMDA) was trained with the SALSA toolbox software [18].

A. Database

Database available on PhysioNet.org [19] was used. Sixty-four subjects were recruited: 15 PD, 20 HD, 13 ALS and 16 Ctrl. Resistive force sensors located in their shoes were sampled at 300 Hz with a resolution of 12 bits [20]. The temporal gait parameters were presented on [21].

The database has 20 variables: seven temporal, one spatial, five variables in percentage and six participants' attributes. Gait parameters were recorded for 5 minutes. Additionally, two new variables were estimated: body mass index (BMI) using weight and height ($BMI = W/H^2$) and cadence (C_d) using stride time (C= $120/T_{st}$). These data were statistically analyzed in section 3.2 to define a descriptor vector $\overline{X}$.

B. Classification Algorithms

The FCM algorithm is based on Picard iteration [22] to minimizes the functional C-means (1) described by Dunn in [23].

$$\overline{J}\left(\overline{X};\overline{U},\overline{V}\right) = \sum_{k=1}^{c}\sum_{i=1}^{N}\mu_{ik}^{m}\|x_i - v_k\|^2 \tag{1}$$

Variable $\overline{X}$ corresponds to the data vector, x_i is a i-th data vector (for each patient i) which has the value of a selected descriptor. $\overline{U}$ is a partition matrix, μ_{ik} is a i-th membership degree of a k-th cluster. $\overline{V} \in \mathbb{R}$ is a prototype cluster matrix where c is the cluster size. The distance $\|x_i - v_k\|^2$ is a data quadratic distances to the center. The m parameter is the fuzziness degree. The algorithm stops when the difference between membership degrees of one iteration and the next one is lower than the tolerance (ε).

The LAMDA method was proposed by Aguilar [18]. This technique is based on fuzzy logic allowing to use quantitative or qualitative descriptors. The classifier admits multiple work modes, the unsupervised and supervised learning, passive recognition, and active learning. The function showed in (2) include a Marginal Adequacy Degree (M) and a Global Adequacy Degree (G) parameter. On unsupervised learning, a Non-Informative Class is generated which could be converted into a new class. The M is the association of descriptors X_n with objects M^N relative to a specific class C^J. The G function evaluates association of objects M^N compared to all classes. Finally, the max values of G are calculated. The α value is the exigence index, which manipulates the quantity of classes.

$$G\left(\overline{X}|k\right) = \alpha\left(M(x_1|k), \ldots, M(x_i|k)\right) + (1-\alpha)\left(M(x_1|k), \ldots, M(x_i|k)\right) \tag{2}$$

The NN uses an interconnected group of neurons and nodes. The machine learning is achieved when the cost function ($C(w,b) = \frac{1}{2N}\sum_x \|y(x) - a\|^2$) is minimized (3). The technique uses several layers (g_i) to describe the neuron interconnection; moreover, a set of weights (w_i) was defined to represent the interconnection strength between neurons. Each node uses an activation function (f) usually, the sigmoid function.

$$y\left(\overline{X}\right) = f\left(\sum_i^N w_i g_i(\overline{X})\right) \tag{3}$$

Parameter w denotes weights, b is the bias, N is the training examples, a represents the activation output vector produced by each input x and y variable represents desired outputs.

The cross-validation was the technique used to evaluate the performance of the machine learning model. A 15-fold cross-validation procedure involved dividing the data into 15 equal parts, using each partition as test set once and the other 14 partitions used as the training set. Repeating this process fifteen times, ensuring that each part is used as the test set exactly once. Having completed this process and average is performed with the evaluation metrics obtained in each of the fifteen iterations in order to obtain an overall measure of the model's performance.

C. Data analysis

1) Statistical analysis

To identify gait characteristics and extract information that could be useful to define descriptors, a statistical analysis of gait data was performed. Four groups or classes were described *a priori*: Ctrl, PD, ALS, and HD. A cluster analysis was conducted by using the Ward's linkage method, while the city-block distance metric [24] was used to determine data cohesion within each group. The statistical analysis confirmed the four classes defined a priori, however the HD and ALS groups overlapped, possibly due to high data variability.

Table 1 shows the mean and standard deviation of some sociodemographic and gait parameters. Differences between groups mean age was found ($p > 0.05$). No statistically significant difference was found between the number of recruited women and men ($p > 0.05$). Using the ANOVA test, a statistically significant difference was found between the cadence standard deviation (SD) within groups (F-ratio $= 6.35$ and $p = 0.0009$), and similarly for swing phase SD (F-ratio $=$ 10.12 with $p = 0.00001$). Using the Kruskal-Wallis test [25], a significant difference of SD median gait speed, stride time, and double support time was found.

The ALS and HD data presented a high dispersion, producing overlaps between the groups. Subjects with differentiating elements such as advanced age, underweight and overweight had similar results to subjects classified with neuromuscular disease. A hypothesis test was performed for the mean value of stance and swing phases in each lower extremity; however, no significant differences were found (p-value ≤ 0.18). No significant difference was observed between the power calculated for each lower extremity in the swing and stance phases (p-value ≤ 0.66). Power spectral density also showed no difference for each limb in the two phases of gait (p-value ≤ 0.68). Statistically significant differences in energy and power used during the stance phase were observed for Control versus ALS, PD and HD (p-value ≤ 0.012).

The variability of gait speed and cadence could mean an increase in the energy expenditure of volunteers during gait (Table 1). Therefore, the energy cost and power spectral density of the temporal signals of the swing and stance phases of gait were calculated to train the classification algorithms.

2) Descriptors selection

The descriptor vector $\left(\overline{X}\right)$ was composed by a geometric mean of cadence (μ_{Cd}), gait speed (μ_{GS}), swing time (μ_{Sw}), stance time (μ_{Se}), and stride time (μ_{St}). The standard deviation was calculated for cadence (σ^2_{Cd}), stride time (σ^2_{St}), double stance (σ^2_{DS}), double support (σ^2_{St}), and Swing-Stance times (σ^2_{SE}).

Considering the effects of BMI on gait speed [26], and dependence between gait speed and energy cost

Table 1. Statistical analysis of the database. [a]Body Mass Index. [b]Measurement of disease severity, normalized between 0 to 1

Classes	Gender		Age (years)	BMI[a]	Severity[b]	Left swing (%)	Right stance (%)	Double Support (%)	Cadence (steps/min)
	F	M							
Ctrl	14	2	39.3±18.5	19.8±2.7	0	36.1±1.7	64.3±1.81	28.2±2.7	110.1±7.9
HD	13	6	47.3±12.5	21.5±4.4	0.2±0.2	34.6±3.45	66.7±3.99	32±6.1	106.4±12.9
PD	5	10	66.8±10.9	21.7±2.79	0.7±0.2	33.2±2.48	67.3±3.71	34.2±5.1	108±10.7
ALS	2	8	55.1±11.9	25.6±5.6	0.2±0.2	32.7±2.74	67.8±2.57	37±9.2	91.8±12.1

[27], the estimation of energy (4) and power spectral density (5) from spatiotemporal signals were considered to identify the signal change rates. These values were obtained from temporal signals $x(i)$ of the stance and swing phases on each lower limb by sample N.

$$E_k = \sum_{i=0}^{N-1} |x(i)|^2 \Delta t \tag{4}$$

$$P_i(\omega) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} R_{xx}(m)\, e^{-j\omega m} \tag{5}$$

The descriptors (E_{Sw}, E_{St}) were estimated using the average Energy produced by both lower limbs. A similar procedure was performed for calculating Power Spectral Density of the Swing phase and Stance phase for both limbs (P_{Sw}, P_{St}). Finally, three more descriptors were used as input information for classifier algorithms: age (A), Body Mass Index (BMI), and gender (G).

III. Results

A. Data pre-processing of volunteers

The database was refined to ensure consistent data. Three people diagnosed with ALS and one person with HD had neither weight nor gait speed records, therefore, these subjects were removed. Some of the temporal signals had outliers, which were also filtered. The final database included 60 volunteers distributed as follows: 15 subjects with PD, 19 with HD, 10 with ALS and 16 Ctrl.

The groups were unbalanced; therefore, five subjects were randomly chosen for each group to train and validate the classification techniques.

B. Fuzzy C-Means Algorithm

The database had groups with an unbalanced number of subjects, i.e., Ctrl (16), PD (15), HD (19), and ALS (10). Generally, the performance of machine learning algorithms decreased with unbalanced classes. Therefore, a random sample of five subjects per group was chosen, resulting in two balanced populations of twenty subjects for training and twenty for validation.

Four clusters were chosen *a priori*. Descriptor vector ($\overline{X}$) was composed by A, G, B, μ_C, μ_{GS}, and μ_{St}. Using a tolerance parameter $\varepsilon = 1 \times 10^{-6}$, a fuzziness weight $m = 2$, and 800 algorithm iterations. The best classification fit was 56.15% distributed as 62.5% for control group, PD and ALS at 60.0%, and 42.1% for HD.

To achieve a better fit, the Fuzzy C-mean algorithm was used with three two-class classifiers: Ctrl+PD, Ctrl+ALS, and Ctrl+HD. To design the classification algorithm a tolerance $\varepsilon = 1 \times 10^{-6}$, fuzziness weight $m = 4$ and one-thousand iterations parameter were used. The results of the cross-validation showed an increase in the total classification fit of 90.2% (Table 2), calculated from the diagonal mean.

C. Neural Network Classification

The classification was performed using four *a priori* chosen classes. Twenty random samples were selected for training, having five samples per class. A matrix including the training data and other twenty random samples was used for validation. All values were normalized including the class number: $[0, 0.25, 0.5, 0.75, 1]$.

Using four classes chosen *a priori*, the neural network was trained with three hidden layers and twenty neurons, classified 92.5% of the data. 100% of volunteers with ALS and HD were correctly classified, however, two persons with PD were misclassified as ALS and one Ctrl subject was classified as PD.

The use of two-class classifiers was evaluated to improve the result. The neural network parameters were chosen according to the size of the database and the results of the cross-validation. Three NNs with 3 hidden layers and 20 neurons were trained to compare between Ctrl and PD, Ctrl and ALS and Ctrl and HD. Table 3 shows the average accuracy of 97.0% between classifier results.

D. Learning Algorithm for Multivariate Data Analysis - LAMDA

The unsupervised LAMDA technique defined two new subclasses to classify the Ctrl and PD groups, three for the ALS, and five new subclasses for HD group. The presence function used was $\rho^x (1-\rho)^{1-x}$, the T-norm & T-conorm was used to calculate the M parameter, and 0.8 as exigence index. Table 4 shows the cross-validation results using one-class classifiers. LAMDA correctly classified sixteen out of twenty volunteers (80.0% accuracy).

Using a stringency index of 0.7 and the same presence function and M operator as in the previous case,

Table 2. Fuzzy C-means technique using two-classes

Classes	**Classes**		
	Ctrl+PD	Ctrl+ALS	Ctrl+HD
Ctrl+PD	**85.7%**	27.3%	27.3%
Ctrl+ALS	60%	**94.1%**	45.5%
Ctrl+HD	36.4%	62.5%	**90.9%**

Table 3. Confusion matrix using neural network technique with two-classes

Classes	**Classes**		
	Ctrl+PD	Ctrl+ALS	Ctrl+HD
Ctrl+PD	**98%**	2%	0%
Ctrl+ALS	0%	**100%**	0%
Ctrl+HD	3%	4%	**93%**

Table 4. Confusion matrix of LAMDA technique using Matlab

Classes	**Classes**			
	Ctrl	PD	ALS	HD
Ctrl	**100%**	0%	0%	0%
PD	20%	**80%**	0%	0%
ALS	10%	30%	**70%**	0%
HD	30%	0%	0%	**70%**

LAMDA in passive recognition classification mode estimated four classes from the descriptor dataset (Table 5). Three two-class LAMDA classifiers were trained: Ctrl+PD, Ctrl+ALS and Ctrl+HD, achieving an average fit close to 98.4%.

IV. Discussion

In this study, we present three methods to classify neurodegenerative diseases from gait information of sixty subjects divided into Control, Parkinson's disease, Amyotrophic Lateral Sclerosis, and Huntington's disease. The use of statistical analysis for descriptor definition improves the performance of the machine learning algorithms. Calculating energy cost and power spectral density from temporal gait parameters decreases the number of sensors used during motion capture. Moreover, these variables could describe gait patterns for the detection of neurodegenerative diseases.

The statistical results were consistent with those already reported in the literature [8], [9], [28]. PD severity and the ALS are age-dependent [29]. There was a positive correlation between the age and the standard deviation of cadence (30.0%), swing phase (25.0%), and support phase (25.0%). There was a 50.0% negative correlation between disease severity and gait speed.

The high variability of the data within the ALS and HD classes resulted in data overlaps within and between classes, which affected the fuzzy membership matrix and the occurrence of local minima. Furthermore, the overlap was verified during LAMDA training, as eight new subclasses were found. These new behavioral patterns were used by LAMDA to avoid misclassification of diseases, improving the results up to 98.33% (Figure 1). When comparing our results, this technique improved performance compared to NN (97.0%), Fuzzy C-means (90.23%) and other machine learning techniques [13], [16], [17].

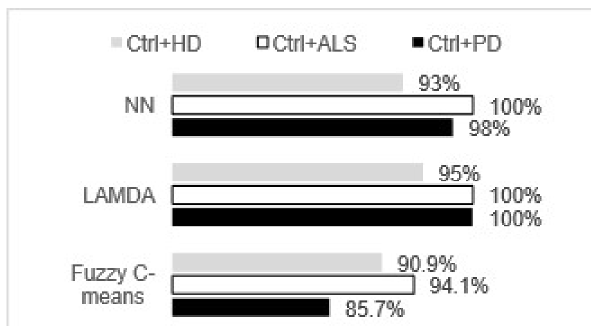

Figure 1. Performance between classification techniques

V. Conclusions

This paper proposes three methods to classify neurodegenerative diseases using descriptors from gait parameters and grouping the disease with the healthy control patients.

The statistical analysis allowed choosing appropriate descriptors for gait parameters, revealing information clinically significant for medical diagnosis.

Describing the energy and power spectral density from spatiotemporal signals might decrease the amount of measured data. Databases with a greater number of patients should give results that are more accurate.

The best classification was achieved using three two-class classifiers. The Fuzzy C-means achieved a fit of 90.23%, the LAMDA fit was 96.66% and the NN achieved 97.0%.

Table 5. LAMDA technique outputs using paired classes: using SALSA software toolbox

Classes	**Classes**		
	Ctrl+PD	Ctrl+ALS	Ctrl+HD
Ctrl+PD	**100%**	10%	40%
Ctrl+ALS	50%	**100%**	30%
Ctrl+HD	40%	40%	**95%**

Descriptors defined by statistical analysis allows to apply lower cost algorithms, however it is necessary to perform a longitudinal study for more complete observations, which will allow the development of more accurate algorithms.

Although results were conclusive, it is not possible to generalize due to the small sample size. It is suggested to increase the database and perform a longitudinal study to conduct a continuously feedback of models and improve the classifications fit.

VI. Future Work

The future work will include the validation of the model in patients with different stages of neurodegenerative diseases, and including the severity of the disease as a model outcome improving the early detection and treatment of the disease. In future work, we propose integrating the degree of disease severity into the modeling process in order to achieve classifiers that give information on disease progression, with the subgroups automatically created by LAMDA being could be useful to solve this problem.

V. Acknowledgment

Andrés Cárdenas expresses his gratitude to Convocatoria 727 – Programa de Becas de Doctorados Nacionales 2015.

Beatriz Liliana Gómez expresses her gratitude to Convocatoria 785 – Programa de Becas de Doctorados Nacionales 2017.

The authors manifest their gratitude to PhysioNet.org for sharing their database and facilitating the further understanding of neurodegenerative diseases.

References

[1] S. Tabrizi, *Neurodegenerative diseases neurobiology pathogenesis and therapeutics,* vol. 77, no. 2, 2006.

[2] D. R. Peterson and J. D. Bronzino, *Biomechanics Principles and Applications,* Second. Boca Raton: CRC Press, 2008.

[3] P. M. Thompson and H. V. Vinters, "Chapter 1 - pathologic lesions in neurodegenerative diseases," in *Molecular Biology of Neurodegenerative Diseases*, ser. Progress in Molecular Biology and Translational Science, D. B. Teplow, Ed., vol. 107, Academic Press, 2012, pp. 1–40.

[4] F. E. Micheli, *Enfermedad de Parkinson y trastornos relacionados*, 2nd ed. Buenos Aires: Editorial Médica Panamericana s.a., 2006.

[5] S. M. Albert, "Amyotrophic Lateral Sclerosis: A Patient Care Guide for Clinicians," *Neurology*, vol. 82, no. 9, pp. 820 LP – 820, Mar. 2014 [Online]. Available: http://n.neurology.org/content/82/9/820.abstract.

[6] V. Medved, *Measurement of Human Locomotion*. Boca Raton: CRC Press, 2000.

[7] S. Mandeep, S. Mooninder, and S. Paramjeet, "Neuro-degenerative disease diagnosis using human gait: A review," *Int. J. Interact. Multimed. Artif. Intell*, vol. 7, no. 1, pp. 16–20, 2013. [Online]. Available: http://www.csjournals.com/IJITKM/PDF.

[8] R. de M. Roiz, E. W. A. Cacho, M. M. Pazinatto, J. G. Reis, A. Cliquet Jr, and E. M. A. Barasnevicius-Quagliato, "Gait analysis comparing parkinson's disease with healthy elderly subjects," *Arq. Neuropsiquiatr*, vol. 68, pp. 81–86, 2010. [Online]. Available: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-282X2010000100018&nrm=iso.

[9] O. Sofuwa, A. Nieuwboer, K. Desloovere, A.-M. Willems, F. Chavret, and I. Jonkers, "Quantitative gait analysis in parkinson's disease: Comparison with a healthy control group," *Arch. Phys. Med. Rehabil*, vol. 86, no. 5, pp. 1007–1013, 2005. DOI: 10.1016/j.apmr.2004.08.012.

[10] A. Delval, "A biomechanical study of gait initiation in huntington's disease," *Gait Posture*, vol. 25, no. 2, pp. 279–288, 2007. DOI: 10.1016/j.gaitpost.2006.04.001.

[11] M. Tinelli, P. Kanavos, and F. Grimaccia, "The value of early diagnosis and treatment in Parkinson's disease: a literature review of the potential clinical and socioeconomic impact of targeting unmet needs in Parkinson's disease," The London School of Economics and Political Science, London, 2016. Accessed: Oct. 17, 2018. [Online]. Available: http://eprints.lse.ac.uk/id/eprint/68479.

[12] S. Bilgin, "The impact of feature extraction for the classification of amyotrophic lateral sclerosis among neurodegenerative diseases and healthy subjects," *Biomed. Signal Process. Control*, vol. 31, no. Supplement C, pp. 288–294, 2017. DOI: 10.1016/j.bspc.2016.08.016.

[13] Y. Xia, Q. Gao, and Q. Ye, "Classification of gait rhythm signals between patients with neurodegenerative diseases and normal subjects: Experiments with statistical features and different classification models," *Biomed. Signal Process. Control*, vol. 18, no. Complete, pp. 254–262, 2015. DOI: 10.1016/j.bspc.2015.02.002.

[14] Ç. B. Erda, E. Sümer, and S. Kibarolu, "Neurodegenerative diseases detection and grading using gait dynamics," *Multimed. Tools Appl*, p. 1, 2023. DOI: 10.1007/S11042-023-14461-7.

[15] W. Zeng and C. Wang, "Classification of neurodegenerative diseases using gait dynamics via deterministic learning," *Inf. Sci*, vol. 317, no. C, pp. 246–258, 2015. DOI: 10.1016/j.ins.2015.04.047.

[16] R. Martins, C. Figueira, H. Gamboa, A. Veloso, and R. Matias, "Automatising gait kinematics classification with fast and accurate machine learning algorithms," en, *Gait Posture*, vol. 42, p. 101, 2015. DOI: 10.1016/j.gaitpost.2015.06.184.

[17] R. Matias, R. Martins, J. Magarreiro, A. L. Gomes, C. Cavaco, and H. Gamboa, "A reliable classification system for neuromusculoskeletal gait disorders," *Gait Posture*, vol. 42, S78, 2015. DOI: 10.1016/j.gaitpost.2015.03.132.

[18] J. Aguilar-Martin and R. Lopez de Mantaras, "The process of classification and learning the meaning of linguistic descriptors of concepts," in *Approximate Reasoning in Decision Analysis*, M. M. Gupta and E. Sánchez, Eds., Amsterdam: North-Holland Pub. Co, 1982, pp. 165–175.

[19] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Gait Dynamics in Neuro-Degenerative Disease Data Base," 2015.

[20] J. M. Hausdorff, Z. Ladin, and J. Y. Wei, "Footswitch system for measurement of the temporal parameters of gait," *J. Biomech*, vol. 28, no. 3, pp. 347–351, 1995. DOI: 10.1016/0021-9290(94)00074-E.

[21] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *J. Appl. Physiol*, vol. 88, no. 6, pp. 2045–2053, 2000. [Online]. Available: http://jap.physiology.org/content/88/6/2045.

[22] S. Bandyopadhyay and S. Saha, *Unsupervised Classification*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[23] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," *J. Cybern*, vol. 3, no. 3, pp. 32–57, Jan. 1973. DOI: 10.1080/01969727308546046.

[24] A. C. Rencher, *Methods of Multivariate Analysis*, 2nd ed. United States of America: Wiley-Interscience, 2002.

[25] S. E. Maxwell, H. D. Delaney, and K. Kelley, *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, 3rd ed. New York, NY, USA: Routledge, 2017.

[26] K. M. Beavers, "Associations between body composition and gait-speed decline: Results from the health, aging, and body composition study," *Am. J. Clin. Nutr*, vol. 97, no. 3, pp. 552–560, Mar. 2013. DOI: 10.3945/ajcn.112.047860.

[27] J. A. Schrack, E. M. Simonsick, and L. Ferrucci, "The relationship of the energetic cost of slow walking and peak energy expenditure to gait speed in mid-to-late life," *Am. J. Phys. Med. Rehabil*, vol. 92, no. 1, pp. 28–35, Jan. 2013. DOI: 10.1097/PHM.0b013e3182644165.

[28] F. Merrikh-Bayat, "Time series analysis of parkinson's disease, huntington's disease and amyotrophic lateral sclerosis," *Procedia Comput. Sci*, vol. 3, no. Supplement C, pp. 210–215, 2011. DOI: 10.1016/j.procs.2010.12.035.

[29] A. Eisen, M. Schulzer, M. MacNeil, B. Pant, and E. Mak, "Duration of amyotrophic lateral sclerosis is age dependent," *Muscle Nerve*, vol. 16, no. 1, pp. 27–32, Jan. 1993. DOI: 10.1002/mus.880160107.