Neurodegenerative disease prediction based on gait analysis signals acquired with force-sensitive resistors

Roger Selzler [1], James R. Green [1], Rafik Goubran[1,2]
[1]*Systems and Computer Engineering*, *Carleton University*, Ottawa, Canada
[2]*Bruyère Research Institute* , Ottawa, Canada
roger.selzler@carleton.ca, jrgreen@sce.carleton.ca, goubran@sce.carleton.ca

***Abstract*—Neurodegenerative diseases such as Parkinson's Disease (PD), Huntington Disease (HD), and Amyotrophic Lateral Sclerosis (ALS) affect the lives of thousands of people around the world. One of the consequences of such diseases occurs in the motor neurons of the patients, resulting in problems in movement, causing a change in gait pattern. Force sensitive resistors can be used to measure the force/pressure between the shoe and the patient's foot, providing information about the gait dynamics when the patient walks. This project uses signals from the Gait Dynamics in Neuro-Degenerative Disease database to extract features for classification of neurodegenerative diseases (NDD). Manually labelled features from the database are used for comparison with previous studies. Time series signals is also used, where algorithms for signal reliability, feature extraction and feature selection are implemented, allowing real-time signal processing and classification. Multiple feature sets are used for classification with algorithms such as K-nearest neighbor, Support Vector Machines, and Decision Trees, and the performance of these algorithms are then reported. This study presents a real-time system with accuracy exceeding 82% for the aforementioned diseases. Finally, a discussion about possible improvements for future studies are presented.**

***Index Terms*—Pressure sensor, pattern classification, measurement, neurodegenerative diseases, gait phase determination.**

I. Introduction

Simple activities such as practicing sports, shopping, going to school, and moving around the city, are easily done for healthy subjects. On the other hand, some medical conditions modify walking patterns making people's lives more difficult. According to Baker *et al.*, clinical gait analysis is a way of recording the biomechanical movement performed when a subject walks, which allow clinicians to assess a patient's state in case of gait dysfunction. When the measured values does not fit a certain pattern, clinicians are able to adapt a certain treatment, and change it as new diagnose are made based on the gait intervals [1].

Usually, the gait fluctuation from healthy subjects are due to specific neurological, muscular or skeletal pathology [1]. The correlation of stride intervals in young and healthy subjects is higher than the correlation of the elderly and for patients with Huntington's disease [2]. In other words, the variability of the stride intervals will be smaller for young and healthy patients. As the disease progresses, it is expected to result in larger variability in the stride intervals. Thus, the gait dynamics analysis can quantitatively monitor the progression of a disease and help with the evaluation of therapeutic evaluations [3].

Different technologies to measure the gait signals have been reported. Force sensors, flexible goniometers, electromagnetic tracking systems, sensing fabrics, electromyography, and accelerometers, gyroscopes, and magnetoresistive sensors are examples of technologies previously reported [4]–[6]. Force sensitive resistors are sensors that measure the force applied across its extremities. As force is applied to the sensor, there is a change in resistance proportionate to the applied force. Thus, by using a force resistive sensor between the shoe and the foot, it is possible to measure the pressure applied over time when a subject is walking. When the subject steps on the ground, there is an increase in pressure and consequently in voltage. When the foot is removed from the ground, less force is applied, resulting in a decrease in voltage. Zeng and Wang used Radial Basis Functions to classify whether the patient has a NDD or not using manually labelled intervals from time series data acquired from pressure sensitive resistors [7]. The present study has similar goals, but avoids the requirement for manually annotating gait intervals. In this way, the solutions developed here are more widely applicable in real-time scenarios.

Figure 1 shows common signals acquired from force sensitive resistors. The blue and red lines represent the left and right foot, respectively. The stance interval is the measured time in which a subject's foot is touching the ground, which starts when the heel touches the ground until the foot is removed from the ground. The duration in which a foot is not touching the ground is called swing interval. Another common measure is the time duration in which both foot are touching the ground, and it is called double support. The double support is shown in the bottom of Fig. 1, and it is highlighted in green.

Gait analysis is often performed in a controlled environment using specialized equipment. With the advancement of wearable devices and the widespread of Internet of Things, more and more signals become available for clinical diagnosis. Force resistive sensors can be embedded to the insole of the

978-1-5386-6709-5/18/$31.00 ©2018 IEEE

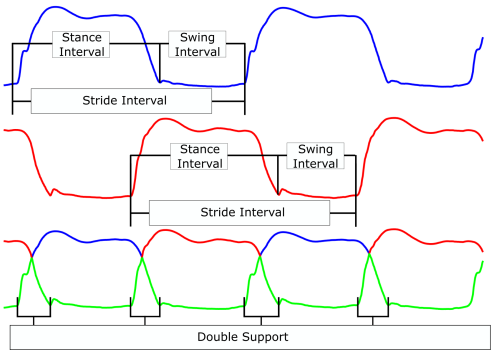


Fig. 1: Gait dynamic phases.

shoes and measure the pressure applied during the gait phases. Thus it is important to implement algorithms to process the signals and provide useful information for clinicians. From the neurodegenerative disease standpoint, it would be interesting if the algorithm could classify whether the patient has a NDD.

This paper presents a way of extracting the gait phases from the time series signals acquired from force resistive sensors, and shows the results of the NDD classification using multiple algorithms. The classification is first performed using the manually labelled intervals provided by [5], which is compared with the work of Zeng and Wang, and also using the gait intervals extracted directly from the time series data with the implemented algorithm.

II. Methodology

A. Data set and signal segmentation

PhysioNet provides a range of physiologic signals broadly used in different research areas. The Gait Dynamics in Neuro-Degenerative Disease database (GAITNDD) consists of signals measured using force-sensitive resistors that roughly represent the forces under the feet of 15 patients with Parkinson's disease (PD), 20 patients with Huntington's disease (HD), 13 patients with amyotrophic lateral sclerosis (ALS) disease and 16 patients with healthy conditions (HC). All signals have a duration of 5 minutes sampled at 300 Hz [5]. The signals were downloaded from [5] and processed using Matlab.

Data for the left and right foot are available in this dataset. Most of the signals are similar with the ones seen in figure 1. However, some signals from different patients have unreliable segments. Such signals do not provide useful information for gait interval extraction. Some values have abrupt changes and periods where the signal is clamped at the minimum or maximum values, which might indicate saturation of the signal or sensor failure. These signal discontinuities indicate a likely sensor failure or bad connection. It is necessary to identify these segments to avoid including erroneous data in the training and evaluation of subsequent NDD classification systems.

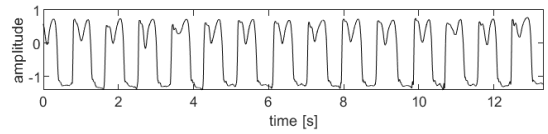
There are also segments that do not represent the usual gait waveform. These signals have a waveform profile with recognizable changes, but usually have a smaller variability on the measured signals. It is believed that in these cases the patient as stopped walking, or that the sensor position is changed, causing unreliable signals for gait analysis. To select the useful segments, variance is computed using a sliding window with duration of 2 seconds or 600 samples, creating a new signal vector. When the variability of the signal is outside the range of the mean of the new vector $\pm 25\%$, the segment is also not considered for analysis. After the signal processing explained in section II-B, only the remaining segments are considered for feature extraction.

B. Signal processing and feature extraction

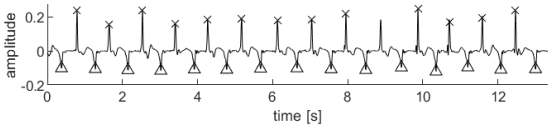
Considering the time series signals, it is necessary to extract the time duration of the stance, swing, and stride intervals. Figure 2 shows the processing steps needed to extract these intervals. Figure 2a shows the signals measured by the pressure sensor from the left foot. To extract the time intervals, the first derivative of the original signal is computed (figure 2b), and the positive and negative peaks, represented by the cross and triangle markers respectively, are found. The positive peaks provide an indication of the beginning of the stance interval, whereas the negative peaks represents the start of the swing interval. The position of these markers is shown in figure 2c. Ideally, these markers would be aligned in the y axis, representing the periodic starting time of each gait interval. To align these markers, ± 10 samples around each marker are used. The closest sample to the mean of all markers is used to update the marker, resulting in a best estimate of the gait intervals.

The difference in time between the cross markers provides the stride intervals. The difference in time between the triangle marker and the previous cross marker provides the stance intervals. The difference in time between the cross marker and the previous triangle marker provides the swing intervals. An example of stride intervals can be seen in figure 2d, in which the triangle markers represents outliers that are eliminated from the signal vector. Figure 2e shows the resulting stance and swing intervals extracted from the aforementioned algorithm. Note that, with the described algorithm, the stance interval will decrease and the swing interval will increase. Further procedures could be implemented to detect the exact time in which the subject touches the ground. However the slope between each step has high variability, causing further process to induce errors. The small delay in time caused by the proposed algorithm will be systematic, and are acceptable for the classification stage.

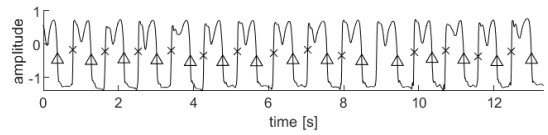
The features are then extracted from the signal vector acquired in the previous steps. The mean (μ) and standard deviation (σ) of the stance, swing and stride intervals are computed and used as features for the classification algorithm. The Fast Fourier Transform (FFT) is applied to the entire time series signal and is divided into linearly distributed bands of 0.0092Hz. The first 40 bands, from 0 to 4.5 Hz, are used and Analysis of Variance is performed over these bands. The bands



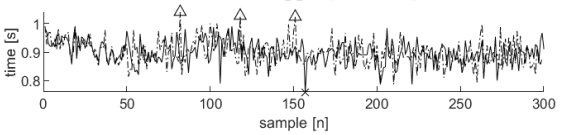
(a) Measured pressure on the left foot.



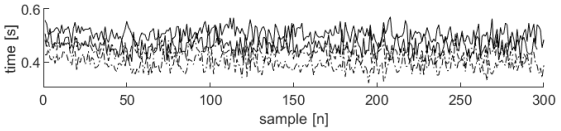
(b) First derivative of fig. 2a and markers for positive and negative peaks.



(c) Markers overlapping with fig 2a.



(d) Stride to stride intervals extracted from 'x' markers in fig. 2b and 2c.



(e) Stance and swing intervals of the left foot after signal processing.

Fig. 2: Signal processing steps required to extract the stride, swing and stance intervals from the time-series signals.

with significance are then used as features in the classification step.

C. Feature sets and classification algorithms

There are four distinct feature sets (FS) used for classification. The first FS (FS1) was used only for performance comparison with Zeng & Wang. FS2 consist of the same features of FS1, but using gait segments extracted from time-series data, making this the ideal solution for real-time feature extraction and consequently real-time classification. FS3 adds the statistically significant FFT bands to the FS2. FS4 uses only the FFT for classification, with an intent of classifying the signals with only the frequency power components of the signals. Table I summarizes the content of each FS.

Multiple classification algorithms were examined in this work. The most important ones and that presented the best results are Fine and Weighted K-nearest neighbor (FKNN, WKNN), Linear and Gaussian Support Vector Machines (SVML, SVMG), and Decision Trees - Fine Tree (FT) and Ensemble Bagged Decision Trees (EBAT). The classification is performed using the leave-one-patient-out cross-validation. In other words, all but one patient are considered training, and the withheld patient is considered the test set. This procedure is repeated until all the patients are classified. The classification is also performed 10 times, and for each instance, the order of the patients is randomly initialized. It appears that only the EBAT algorithm changes the performance with this procedure. This might be due to the small set and the leave-one-patient-out procedure.

TABLE I: Feature sets used for classification

Feature Set	Content
FS1	μ and σ of Stride, Stance, and Swing, all from manually labelled data
FS2	μ and σ of Stride, Stance, and Swing, all from time series data
FS3	FFT bands, μ and σ of Stride, Stance, and Swing, all from time series data
FS4	FFT bands from time series data

III. Results

The classification is performed using the algorithms, features and procedures described in Section II. Tables II-V shows the classification performance such as Accuracy (Acc), Sensitivity (Sen), and Specificity (Spe) using the FS described in Table I, for each individual disease (HC vs. ALS, HC vs. HD, HC vs. PD) and also considering all diseases (NDD vs. HC). For each disease, results are given for the best individual method (IBP) and also for the best overall method (All) which is the method that performs best across all diseases. In this study, EBAT proved to be the best overall method. The IBP is the algorithm that performs best for a specific disease, without considering the results of other disorders. Overall, EBAT performed best considering all NDDs when using the μ and σ of the stance, swing and stride interval, or a conjunction of the time interval features with the FFT bands. Using only the FFT bands, accuracy exceeding 77% was reached, but it was not better than FS1-FS3.

Table II uses FS1, which is the same features used by Zeng and Wang, who used RBF to classify NDD. As can be seen, EBAT performed better than RBF (92.2$\pm$2.2% against 89.7%) for ALS, same for HD, and a little worse for PD (86.3% against 87.1%). This indicates that other algorithms are also effective for NDD classification. Table III shows the results of the classification using the time interval features extracted directly from time series data (FS2). The algorithms had worse accuracy for ALS (82.8% against 89.7%) and PD (80.6% against 87.1%), and same accuracy for HD. When using FS3, the accuracy of ALS and HD remained the same as in FS2, but got worse for the PD (80.6% against 83.9%). Using only the FFT bands as features, the accuracy of ALS increased when compared with FS2 and FS3 (86.2% against 82.8%), remained the same for HD, and decreased for PD (77.4% against 83.9% for FS2 and 80.6% for FS3).

When considering all the NDD, the manually labelled time intervals using FKNN and WKNN classification algorithms had best performance (89.1%). From the features extracted directly from time series data, FS2 presented the best average Accuracy (85.9% using FT).

TABLE II: Classification performance using FS1 and comparison of results with [7]

Disease	Persp.	Acc	Sen	Spe	Alg
ALS	Zeng & Wang	89.7	87.5	92.3	RBF
	IBP	**93.3**	**93.8**	**92.3**	FKNN, WKNN, FT, EBAT
	All	92.2	93.8	92.3	EBAT
HD	Zeng & Wang	83.3	81.3	85	RBF
	IBP	83.3	87.5	80	FKNN, WKNN, EBAT
	All	83.3	87.5	80	EBAT
PD	Zeng & Wang	**87.1**	**87.5**	**86.7**	RBF
	IBP	80.6	81.25	80	EBAT
	All	86.3	81.25	80	EBAT
NDD	Zeng & Wang	NP	NP	NP	RBF
	IBP	89.1	81.3	87.5	FKNN, WKNN
	All	86.3	75	89.6	EBAT

TABLE III: Classification performance using FS2

Disease	Persp.	Acc	Sen	Spe	Algorithm
ALS	IBP	82.8	87.5	76.9	EBAT
	All	82.8	87.5	76.9	EBAT
HD	IBP	83.3	87.5	80.0	EBAT
	All	83.3	87.5	80.0	EBAT
PD	IBP	83.9	81.3	86.7	FT
	All	83.9	81.3	86.7	EBAT
NDD	IBP	85.9	68.8	91.7	FT
	All	84.4	68.8	89.6	EBAT

TABLE IV: Classification performance using FS3

Disease	Persp.	Acc	Sen	Spe	Algorithm
ALS	IBP	82.8	81.3	84.6	EBAT
	All	82.8	81.3	84.6	EBAT
HD	IBP	91.7	93.8	90.0	SVML
	All	83.3	87.5	80.0	EBAT
PD	IBP	83.9	81.3	86.7	SVMG
	All	80.6	81.3	80.0	EBAT
NDD	IBP	82.8	68.8	87.5	FT
	All	81.3	62.5	87.5	EBAT

TABLE V: Classification performance using FS4

Disease	Persp.	Acc	Sen	Spe	Algorithm
ALS	IBP	93.1	87.5	100.0	FKNN, WKNN
	All	86.2	87.5	84.6	SVML
HD	IBP	88.9	93.8	85.0	SVMG
	All	83.3	81.3	85.0	SVML
PD	IBP	77.4	81.3	73.3	EBAT
	All	77.4	87.5	66.7	SVML
NDD	IBP	79.7	62.5	85.4	EBAT
	All	79.7	43.8	91.7	SVML

A. Potential avenues for system improvement

The force measured between the feet and the shoe greatly depends on the positioning of the sensors, and missing sensor measurements may occur when they are not properly positioned. Even when the sensor is correctly positioned at the beginning of the session, it can move around the shoe as the subject walks. Further research could examine the use of multiple force sensors distributed accordingly to the shape of the feet. This reduces the chance of sensor location displacement, and guarantees that a subset of sensors will always have contact with the feet, allowing extra signal processing and signal reliability. In addition, new features might be extracted such as the balance of the subject, as well as the pressure distribution during the walking.

IV. Conclusions

Gate analysis is a powerful tool that can help clinicians make decisions and plan treatments for patients with neurodegenerative diseases. Classification of NDD based on gait analysis can assist in making the diagnosis, thus improving people's live. This paper presents a method to select reliable segments and extract features from time-series signals measured using pressure resistive sensors, and to classify neurodegenerative diseases such as PD, HD, and ALS. Results using multiple machine learning algorithms with the manually labelled data from [5] are compared with previous studies and better accuracy is achieved for ALS classification. Using the proposed algorithm to extract the features and classify NDDs, it was possible to achieve accuracy exceeding 82% for individual disease prediction, and 85.9% for NDD vs. healthy controls. The results shown in this paper indicates that different algorithms will have better results for specific diseases, and as such, a multi-model prediction system could have a better overall performance.

A. Future work

The use of multiple sensors must be considered for next experiments as discussed in section III-A. To improve the performance and have a better generalization of the machine learning algorithms, more data should be collected from a higher number of patients. The training phase did not consider the change in class imbalance introduced by the leave-one-patient-out method, which can be implemented with the increase of the dataset and signals available.

References

[1] R. Baker, A. Esquenazi, M. G. Benedetti, and K. Desloovere, "Gait analysis : clinical facts," *European Journal of Physical And Rehabilitation Medicine*, vol. 52, no. 4, pp. 560–574, 2016.

[2] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C. K. Peng, M. E. Cudkowicz, J. Y. Wei, and a. L. Goldberger, "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease." *Journal of applied physiology*, vol. 82, no. 1, pp. 262–269, 1997.

[3] 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," *Journal of applied physiology*, vol. 88, no. 6, pp. 2045–2053, 2000.

[4] J. Ballesteros, C. Urdiales, A. B. Martinez, and J. H. van Dieën, "On gait analysis estimation errors using force sensors on a smart rollator," *Sensors*, vol. 16, no. 11, 2016. [Online]. Available: http://www.mdpi.com/1424-8220/16/11/1896

[5] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[6] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Sensors*, vol. 12, no. 2, pp. 2255–2283, 2012.

[7] W. Zeng and C. Wang, "Classification of neurodegenerative diseases using gait dynamics via deterministic learning," *Information Sciences*, vol. 317, pp. 246–258, 2015.