وزارة التعليم العالي والبحث العلمي
جامعة كربلاء - كلية الادارة والاقتصاد
المؤتمر العلمي السادس عشر 2023

Classifying Patients with PD using a Human Gait Character and SVM classifier

أ.م.د. كاظم الدفاعي

Assoc. Prof. Dr. Kadhem Al-Daffaie

kadhem.aldaffaie@mu.edu.iq

Mobile: 07827222750

جامعة المثنى – كلية الادارة والاقتصاد –قسم العلوم المالية والمصرفية

Department of Banking and Financial Sciences, College of Administration and Economics, Al-Muthanna University, Samawah 6600, IRAQ

Abstract

Early diagnosis of Parkinson's disease has a significant impact on patient's lives as it is described as a disease in the nervous system of the human body, which leads to disorders in some parts of body. The paper aims at employing a human gait character, i.e. stride time interval, along with the support vector machine (SVM), for classifying patients who suffer Parkinson's disease (PD). To conduct the classification task, the support vector machine classifier was employed on data taken from several human individuals. The data consist of stride time interval records of both normal people and PD patients. The findings of the current study prove the usefulness of using a machine learning technique along with a human gait character to identify Parkinson's disease patients. The current results show that SVM classifier achieved nearly 90% accuracy.

Keywords

Human gait; Stride interval time; Parkinson's disease; Classification; Support Vector Machine

Acronyms

SVM	support vector machine
PD	Parkinson's disease
ABC	artificial bee colony
ELM	Extreme learning machine
RF	random forest

ACC	Accuracy
PPV	Positive Predictive Value
NPV	Negative Predictive Value

1. Introduction

The term "human gait" refers to style of human movement from one point to another, which is done through activities such as walking, running, jumping and etc. (Kadhem Al-Daffaie & Albert K Chong, 2019; Kadhem Al-Daffaie & Albert K. Chong, 2019; Alharthi, Yunas, & Ozanyan, 2019). Human gait is affected by many factors such as diseases, accidents, injuries, and etc. Several characteristics, for instance pressure beneath the foot, force, time to achieve some movements and many more (Al-Daffaie, Chong, & Gharineiat, 2019a), of the human gait have been established. These characteristics are measured and recorded by different techniques and devices.

Experts and researchers use such records to examine and study the status of human gait, where doctors, for example, employ them to diagnose some diseases and then treat them (Al-Daffaie, Chong, & Gharineiat, 2019b).

One of the many diseases that affect human gait is Parkinson's disease (PD) as it affects the nervous system, which leads to abnormal walking (Cicirelli et al., 2021).

Many techniques and methods have been studied and proposed to identify people with abnormal human gait by using machine learning approaches based on data of human gait properties. Proposing such approaches could help doctors and other experts to detect some related illnesses earlier, and then treat them or to prevent reaching severe conditions.

(Patil, Kumar, Gaud, & Semwal, 2019) employed some human gait features to examine the efficiency of several machine learning algorithms to classify people with health condition of multiple sclerosis and stroke. Various machine learning methods were explored. This study revealed that ELM algorithms gave the best results among other algorithms. (Gao, Tian, Yao, Zhang, & Neuroscience, 2021) implemented the classification task by using some features of the so-called surface electromyography. Along with such records, authors of this paper used various classification techniques to reach an accurate gait recognition. In addition, an algorithm called artificial bee colony (ABC) was used to optimize the SVM. The obtained results proved the usefulness of the integrated method (ABC–SVM). (Semwal, Gupta, & Lalwani, 2021) developed different integrated deep learning models based mainly on convolution neural network and long short-term for the purpose of classifying walking movements. In this study, authors used data of different human walking activities. The proposed ensemble based technique performed better than other models. (Gupta & Semwal, 2020) studied recognizing different types of human walking movements, such as fast walk, slow walk and many more, by employing some human gait features. In this study, authors suggest to use ELM model as it gave better performance over SVM and KNN. (Semwal, Lalwani, Mishra, Bijalwan, & Chadha, 2021) explored the classification performance of various techniques to identify different human activities. Applying those algorithms on real human gait dataset revealed that random forest algorithms (RF) had the best accuracy among the other 6 techniques.

This study aims at examining the effectiveness of SVM in classifying patients with PD. To accomplish this aim, the machine learning technique is applied on dataset of normal and patient subjects. The dataset consists of stride time interval records.

The rest of this paper is arranged as follows. Section 2 displays the fundamental theory of the machine learning classifier. Section 3 displays a brief explanation of data and methodology procedure to implement the classification task. Lastly, results are included in Section 4, and conclusions are included Section 5.

2. Theoretical Background

Machine learning can be described as a scientific field that uses computer science and statistics to make computers lean without an explicit programs (Maurya, Hussain, & Singh, 2021). To develop an algorithm that is able to predict, machine learning techniques employs part of the data to develop such algorithms.

In general, machine learning models are categorised in several types. The most popular kinds are supervised learning, unsupervised learning, and etc. (Ayodele, 2010). The supervised learning techniques are usually utilised to perform very common and important tasks which are classification and forecasting, where the first procedure focuses on predicting classes and the second focuses on predicting values.

Now, it has been obvious from the above-mentioned sections that this paper uses one of the machine learning models that works on classification problems, i.e. SVM. Consequently, this section in devoted to present the theoretical background of this algorithm.

SVM has drawn a significant attention during the last years as a sufficient machine learning model that gives promising outcomes for classification tasks in different fields (Sha'Abani, Fuad, Jamal, & Ismail, 2020). The initial idea of its methodology is obtaining a hyperplane between classes, which is then used to determine the class of new inputs. Starting with the simplest case, let us consider the task of binary classification (Al-Daffaie & Al-Ghayab, 2020; Zhang, 2012).

Assume $\{x_i, y_i\}, i = 1,2, \dots, n$ represents the training subset of the data, where n refers to number of samples, and $y_i \in \{-1, +1\}$ere $+1$ refers to 1^{st} class and -1 refers to 2^{nd} class. If these classes are linearly separated, there may be at least one hyperplane to distinguish between the classes:

$$f(x) = w.x + w_0 = 0 \qquad (1)$$

Reaching such a hyperplane requires estimating w and w_0 in a way that

$$y_i(w.x_i + w_0) - 1 \geq 0 \qquad (2)$$

To find the optimal hyperplane that maximize the margin between classes, the so-called support vectors need to be obtained. They are defined as follows:

$$w.x_i + w_0 = \pm 1 \qquad (3)$$

The problem is then takes the following form in order to find the optimal hyperplane:

$$Min \frac{1}{2} ||w||^1 \qquad (4)$$

Subject to $y_i(w.x_i + w_0) - 1 \geq 0 \quad , i = 0, 1, \dots, n$

Which become as follows by using Lagrange method:

$$f(x) = \sum_{i \in S} \lambda_i \, y_i(x_i x) + w_0 \qquad (5)$$

3. Methods and data

Following the purpose of the current paper is examining the classification accuracy of SVM in identifying people with PD condition through a feature of human gait, the experimental aspect consists of:

1- acquiring some stride time interval records of normal subjects and PD patients.
2- implementing SVM classifier., which has been done via some packages of R software.
3- evaluating the classification performance through some measures.

This paper used dataset of human gait feature, i.e. walking stride interval time series, which is publicly available on PhysioNet website (Goldberger et al., 2000). According to (Andrews, Harrelson, & Wilk, 2012), the interval from one heel contact to the ground to the next contact of the same foot called stride time as illustrated in Figure (1). This dataset has been employed by many researchers for classification and other purposes (Alday et al., 2020; Kidger, Morrill, Foster, & Lyons, 2020; Mahmud, Kaiser, McGinnity, & Hussain, 2021).

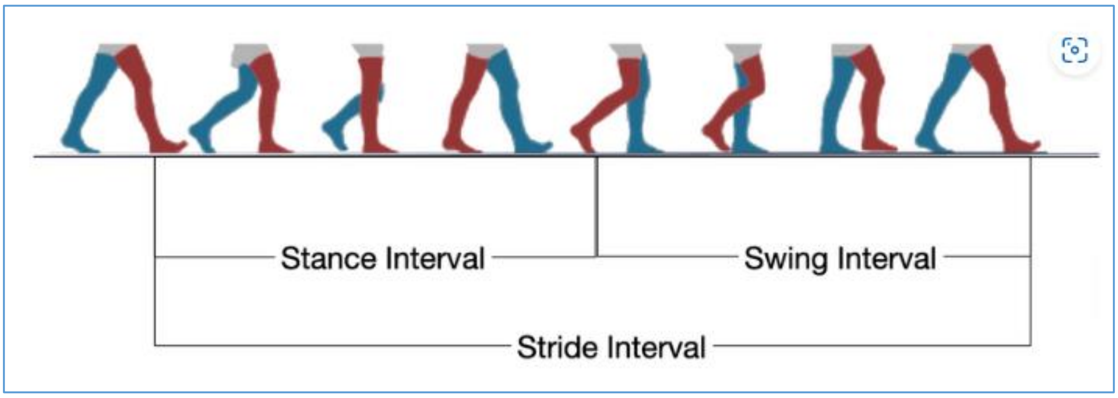

Figure (1): Stride interval

This dataset was collected from different subjects, especially, healthy individuals and PD patients. All individuals were asked to walk on level ground path. To measure the stride interval time, some devices attached to subjects' feet were utilized. These devices included: 1) force sensitive resistors, which was inserted inside the shoe, 2) microcomputer that attached to the ankle. This led to automatic calculations of time between foot strikes (Goldberger et al., 2000).

The following evaluation measures were employed to assess the performance of SVM (Qorri et al., 2022):

$$Accuracy\ (ACC) = \frac{\sum TP + \sum TN}{n} \times 100 \quad (6)$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{\sum TP}{\sum TP + \sum FP} \times 100 \quad (7)$$

$$\text{Negative Predictive Value } (NPV) = \frac{\sum TN}{\sum FN + \sum TN} \times 100 \quad (8)$$

where TP is the true positives, TN is the true negatives, n is the sample number, FP is the false positives, and FN is the false negative.

4. Results

The data was split into two groups, particularly training dataset and testing dataset, by using an R-package called "caTools" (Tuszynski & Khachatryan, 2013). The training subgroup contains 75% of the original data, while the other 25% dedicated to the testing subgroup. Then the machine learning classifier, i.e. SVM, was applied through the "e1071" R package (Meyer et al., 2019). After training the model on the training set of the dataset, the testing group was used to check the ability of the algorithm to detect the class of new entries. The final step was to calculate some indicators to reveal the efficiency of the examined model. The evaluation was implemented by using Equations 6-8.

Values of the performance assessment indicators are shown in Table (1) below:

Table (1): Evaluation metric values of SVM

Metric	Value
ACC	89.58
PPV	74.10
NPV	90.59

The values presented in Table (1) show that the examined machine learning technique in this paper achieved a promising performance, where it had 90.59% of NPV, 89.58% of ACC and 74.10% of PPV.

5. Conclusion

Suggesting an automatic method to identify people with different health conditions has become of a great interest recently as it assists health experts in early diagnosing of various diseases. This paper examines the effectiveness of a machine learning technique, i.e. SVM, in classifying people who suffer from Parkinson's disease. Data of a human gait feature called "stride time interval" were employed in the experimental part of this study. The explored model gave a good performance, where it had a value of NPV od 90.59%. Some possible future works may include adding other human gait features, and applying the same classifier on other problems.

References

Al-Daffaie, K., & Al-Ghayab, H. R. J. I. J. A. S. S. V. (2020). TRANSIENT ARTIFACT REDUCTION AND STATISTICAL METHOD BASED CLASSIFICATION OF NEURODEGENERATIVE DISEASES. *16*(1), 1391-1399.

Al-Daffaie, K., & Chong, A. K. (2019). *Advanced techniques to study foot pressure of healthy young adults.* Paper presented at the Proceedings of the 2019 2nd International Conference on Information Science and Systems.

Al-Daffaie, K., & Chong, A. K. (2019). *Advanced Techniques to Study Foot Pressure of Healthy Young Adults*. Paper presented at the Proceedings of the 2019 2nd International Conference on Information Science and Systems, Tokyo, Japan.

Al-Daffaie, K., Chong, A. K., & Gharineiat, Z. (2019a). *Plantar Pressure Characteristics in Obese Individuals: A Proposed Methodology.* Paper presented at the 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC).

Al-Daffaie, K., Chong, A. K., & Gharineiat, Z. (2019b). *Pressure Beneath the Foot for Older Adults Using an Improved Approach.* Paper presented at the 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE).

Alday, E. A. P., Gu, A., Shah, A. J., Robichaux, C., Wong, A.-K. I., Liu, C., . . . Seyedi, S. J. P. m. (2020). Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020. *41*(12), 124003.

Alharthi, A. S., Yunas, S. U., & Ozanyan, K. B. J. I. S. J. (2019). Deep learning for monitoring of human gait: A review. *19*(21), 9575-9591.

Andrews, J. R., Harrelson, G. L., & Wilk, K. E. (2012). *Physical Rehabilitation of the Injured Athlete: Expert Consult-Online and Print*: Elsevier Health Sciences.

Ayodele, T. O. (2010). *Types of Machine Learning Algorithms*.

Cicirelli, G., Impedovo, D., Dentamaro, V., Marani, R., Pirlo, G., D'Orazio, T. R. J. I. J. o. B., & Informatics, H. (2021). Human gait analysis in neurodegenerative diseases: a review. *26*(1), 229-242.

Gao, F., Tian, T., Yao, T., Zhang, Q. J. C. I., & Neuroscience. (2021). Human gait recognition based on multiple feature combination and parameter optimization algorithms. *2021*.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., . . . Stanley, H. E. J. c. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *101*(23), e215-e220.

Gupta, A., & Semwal, V. B. (2020). Multiple task human gait analysis and identification: ensemble learning approach. In *Emotion and information processing* (pp. 185-197): Springer.

Kidger, P., Morrill, J., Foster, J., & Lyons, T. J. A. i. N. I. P. S. (2020). Neural controlled differential equations for irregular time series. *33*, 6696-6707.

Mahmud, M., Kaiser, M. S., McGinnity, T. M., & Hussain, A. J. C. c. (2021). Deep learning in mining biological data. *13*(1), 1-33.

Maurya, L. S., Hussain, M. S., & Singh, S. J. A. A. I. (2021). Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction Based on Academic Performance. *35*, 403 - 420.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., . . . Meyer, M. D. J. T. R. J. (2019). Package 'e1071'.

Patil, P., Kumar, K. S., Gaud, N., & Semwal, V. B. (2019). *Clinical human gait classification: extreme learning machine approach.* Paper presented at the 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT).

Qorri, E., Takács, B., Gráf, A., Enyedi, M. Z., Pintér, L., Kiss, E., & Haracska, L. J. I. J. o. M. S. (2022). A comprehensive evaluation of the performance of prediction algorithms on clinically relevant missense variants. *23*(14), 7946.

Semwal, V. B., Gupta, A., & Lalwani, P. J. T. J. o. S. (2021). An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition. *77*(11), 12256-12279.

Semwal, V. B., Lalwani, P., Mishra, M. K., Bijalwan, V., & Chadha, J. S. J. C. (2021). An optimized feature selection using bio-geography optimization technique for human walking activities recognition. *103*(12), 2893-2914.

Sha'Abani, M., Fuad, N., Jamal, N., & Ismail, M. J. I. (2020). kNN and SVM classification for EEG: a review. 555-565.

Tuszynski, J., & Khachatryan, M. H. (2013). Package 'caTools'. In: Recuperado.

Zhang, Y. (2012). *Support vector machine classification algorithm and its application.* Paper presented at the International conference on information computing and applications.