

Effect of Machine Learning Techniques for Efficient Classification of EMG Patterns in Gait Disorders

Manoj Kumar¹, Pratiksha Gautam² and Vijay Bhaskar Semwal³

¹Department of Computer Science & Eng., Amity University Gwalior, M.P., India, mannu175@yahoo.com

²Department of Computer Science & Eng., Amity University Gwalior, M.P., India, pgautam@gwa.amity.edu

³Department of Computer Science & Eng., NIT Bhopal, M.P., India, vsemwal@manit.ac.in

*Corresponding Author: Manoj Kumar; E-mail: mannu175@yahoo.com;

ABSTRACT- Gait disorder is very common in neurodegenerative diseases and differentiating among the same kinematic design is a very challenging task. The muscle activity is responsible for the creation of kinematic patterns. Hence, one optimal way to monitor this issue is to analyse the muscle pattern to identify the gait disorders. In this paper, we will investigate the possibility of identifying GAIT disorders using EMG patterns with the help of various machine learning algorithms. Twenty-five normal persons (13 male and 12 females, age around 28 years of age) and 21 persons having GAIT disorders (11 male and 10 females, age around 67 years of age). Four different machine learning algorithms have been used to identify EMG patterns to recognize healthy and unhealthy persons. The results obtained so far have been used to distinguish between GAIT disorders and healthy patients. Our proposed system can also prove that Recurrent Neural Network has achieved the best accuracy with 91.3 % in the case of two classes and 86.95 % in the case of three classes compared to other machine learning algorithms.

Keywords: Machine learning, deep learning, feature extraction, Gait disorders, Kinematic, Random Forest

ARTICLE INFORMATION

Author(s): Manoj Kumar, Dr Pratiksha Gautam and Dr Vijay Bhaskar;

Special Issue Editor: Dr. Vikash Yadav;

Received: 10/03/2022; Accepted: 11/05/2022; Published: 22/05/2022;

E- ISSN: 2347-470X ;

Paper Id: 0422SI-IJEER-2022-07;

Citation: 10.37391/IJEER.100211

Webpage-link:

<https://ijeer.forexjournal.co.in/archive/volume-10/ijeer-100211.html>



This article belongs to the Special Issue on **Recent Developments in Communication Technology using Machine Learning Techniques**.

Publisher's Note: FOREX Publication stays neutral with regard to jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Gait disorder is the common background of different neurological problems [1]. They create a vital problem in fitness as they are a continual source of drop-down, with successive mortality and bitterness. They also decrease the quality of patient's life by spoiling their capability to carry out activities of everyday routine and to engage in conventional daily living [1,2]. Further, they are creating considerably higher costs in the healthcare field [3,4,5].

The different range of pathological units strikes various arrangements of the nervous system might emphasize gait disorganizations. The involvements of various regions of the nervous systems are estimated using different neurological testing. Although, based on phenotypical similarities, recognition of problems with neurologic gait disorganizations is not correct even in the case of expert clinicians. The accurate recognition of etiopathological-based gait disorders is

very necessary. It is also useful for diagnosis, treatment, prognosis, and workup. The increase in the accuracy of recognition of gait disorders would enhance the care quality and able to decrease the cost of therapeutic or diagnosis. In order to inspect gait characteristics, surface electromyography and kinematic recordings are used in so much research in the past years. The EMG is always useful because it is near neuronal control device pathological and normal walking [6,7,8].

In this research, our goal is to scrutinize the assumption that activated muscle patterns in the time of running accommodate maximum details to recognize the gait disorganization, significantly correct to enhance the recognition by clinical evaluation, shown between 50 to 80% accuracy [9,10]. The high dimension and complex nature of EMG data and some uncertainties for the features that are responsible for the healthy or pathological gait relies on the automatic approach using classifiers. We are trying to investigate whether the learning techniques are used in the EMG signals will able to

- (1) Recognize efficiently patients from healthy peoples.
- (2) Recognize efficiently the two classes of gait disorder that is, ataxic and hypokinetic gait from healthy classes and for each other.

The main findings of our paper are:

- (1) We introduced a framework based on machine learning concepts to identify gait disorders.
- (2) Our introduced framework is based on EMG signals.
- (3) We are able to produce our results with various machine learning techniques.
- (4) We are able to recognize efficiently gait patients from healthy people.

- (5) We are able to recognize efficiently the two classes of gait disorder that is, ataxic and hypokinetic gait from healthy classes and for each other.

The rest of the paper is organized as follows: a detailed explanation of the experiment is in *section 2*, results are described and discussed in *section 3*, and finally concluded in *section 4*.

2. EXPERIMENTS

2.1 Participants

Patients with neurological diseases were included in the research. The criteria for including in this research is above 18 years of age having neurological gait disorders categorized into “ataxic gait and hypokinetic gait” patients and healthy patients with no gait disorders. The patients have severe orthopaedic, unable to walk freely, neuropsychiatric or interfering to participate and pregnant lady. Healthy people are obtained from public devices.

2.2 Experimental Methods

Before main experiment, all the participants performed a test. Further, they are instructed to walk down at a normal pace and then asked to move them and conduct tandem gait. They are captured in the time of gait job. We want to investigate the two categories of gait disorders that are, hypokinetic and ataxic gait. All the exercise is done according to the requirement of these categories.

For the experiments, some bipolar electrodes are attached to the leg of the patients like rectus femoris, vastusmedialis, tibialis anterior, biceps femoris, and gastrocnemius lateralis. Further, some additional accelerometers were attached to every foot to obtain the information that is used to identify the gait cycle. The data was recorded by the system with the help of a wireless accelerometer. People are asked to stand up and sit down then walk about 10 meters and then walk and sit again. This process is redone 10 to 12 times in order to get more data. The EMG data are kept where persons freely walk. The electrodes are placed in the same position for all the participants.

2.3 Data Pre-Processing

All accelerometer and EMG data is exported to MATLAB and analysis of the data is done using python. The data is kept only where participants are walking and ignoring the data where participants are standing, sitting, and turning. For data pre-processing, we follow the process adopted by the process of Fricke et al. [11]. We use a high-pass digital filter to delete some noises and DC components that arise due to electrode skin and movement. After that, a low-pass filter is used to remove higher repetition artifacts to process full-wave amendments.

The accelerometer information was further converted in such a way that it represents the gait cycle clearly. The gait cycle was identified with the help of recordings produced by an accelerometer attached to both feet. Since the accelerometer was attached to both feet, the relevant signal was noted from

the attached accelerometer every time when the foot hits the ground when they were walking.

2.4 Classification Techniques

We applied four supervised machine learning techniques to automatically identify gait disorders based on the proposed method using feature extraction techniques. The machine learning techniques used in the given framework are Recurrent Neural Networks, Logistic regression, Support Vector Machines. These techniques are used to classify (1) gait disorder vs healthy (2) ataxic gait vs hypokinetic gait vs healthy.

2.5 Recurrent Neural Networks

RNN is the type of deep learning which has very vast applications in text and speech recognition. The main uses of RNN classifier are its ability to identify features from more from the extracted features [12]. The extracted feature to the RNN is fed from different tasks. The input data is entered into the RNN having various tasks. The spatial layer in the RNN gets the input then filters it having different weights. The given filter is moved horizontal and vertical throughout the layer. Various feature maps have been found based on the number of layers that affect the classification accuracy. Before the training of RNN, the database was first converted into the time-frequency domain with the help of a continuous wavelet transform to improve the accuracy results. In the past research, we have observed that the time-series domain is responsible for the more robust results.

One of the major problems in applying RNN in datasets is the choose a relevant number of layers, weights, and filters to obtained good results while overcoming the problem of under fitting and overfitting because these factors play important role in calculating loss function for the given scenario. To overcome this problem, transfer learning can be a better idea. The past research has shown that transfer learning works well even if the two datasets are obtained from different domains [13]. Transfer learning performs well with small datasets. Further, this is also very useful in medical imaging where the dataset is too small. We have also applied transfer learning to our research due to small datasets.

2.6 Logistic Regression and Support Vector Machine as Classification Technique

SVM and Logistic regression are linear and non-linear types of classifiers used in different, high dimensional features of the database that assign classes to every data point present in data space. Support vector machine is the supervised method applied in many areas like facial expression recognition [14], bioinformatics [16], and face detection [15]. The goal of SVM is to create a distinguished hyperplane in multidimensional space. The hyperplane is further used to separate the entire dataset into two or more labels using tricks [17]. So, it is used to increase the distance between the given data points to the hyperplane to calculate the maximum margin hyperplane. Logistic regression is another machine learning technique that contains no training phase. This technique is used in

situations where no pre-knowledge of data points exists. The logistic regression is a non-linear machine learning algorithm. The logistic regression tries to combine similar objects from the given dataset in some meaningful manner. This is based on some metric that is calculated from the given data points. The LR basically, assigns the class of each data point that is present in the datasets in some manner. The accuracy of results is based on scattered data points present in the given datasets.

To perform logistic regression and SVM, some features have to be extracted from the given dataset. We have followed the technique adopted by Alizadeh et al. [18]. In this research, we have extracted seven features from the dataset from every EMG channel. The three features are calculated in frequency domain whereas four features are calculated in time-domain after Fourier transformation. Every feature is calculated for every muscle resulting 70 features per subject. These features are too big for training purposes to overcome the problem of under fitting and over fitting. Various features of different muscles create redundancy in the data, which is responsible for the recognition results. For removing these redundancies, we used independent component analysis (ICA) and selected only important features to send input to the classifiers.

2.7 Training Classifiers

After pre-processing of the given data, we divided data into validation and training sets contains where validation sets contain the EMG gait cycle information of 17 out of 46 subjects, whereas the training set contains the data of 23 participants, leaving data of 3 from both the training and validations sets. After training all the classifiers, the remaining datasets are used to test the framework. The parameters for the logistic regression and SVM are optimized and leaving out the remaining data for testing. The setup and parameters are kept identical for the RNN. This process is repeated for 46 subjects. In the given investigation, the excluded class was identified as one of two classes ("patient" or "healthy") and in another investigation, the excluded class was identified as one of three classes ("ataxic", "hypokinetic" or "healthy,"). The categorization was processed for every given gait cycle and the categorization outcome for the subject was identified as a class that was achieved most frequently when each and every gait cycle was examined. Incomparably, due to the k-cross-validation method, the classifier was trained without using any trials of the given subject to be identified in the training phase. The k-cross-validation process was repeated by processing every participant as the excluded subject. Finally, we assessed the categorization outcome by computing sensitivity, accuracy, and specificity, for two-class:

Accuracy = (TP + TN)/total

Sensitivity = TP / (TP + FN)

Specificity = TN/ (FP + TN)

FN→false negatives, TP →true positives, FP →false positives, TN→ true negatives

For the three-class problem we used the following methods:

Accuracy = TP class/total number

Sensitivity (atax/hypo) = TP/ (TP + FN)

Specificity (atax/hypo) = TN/ (FP + TN)

3. RESULTS

We simulate RNN, logistic regression, and SVM on our EMG data to recognize pathological gait disorders. RNN is able to identify EMG database with sensitivity, specificity, and accuracy of 92%, 90.47%, and 91.3% respectively (Table 1).

Table 1: Confusion Matrix using RNN for 2 classes

	Healthy	Patients
Healthy	23	2
Patients	2	19

LR categorization obtained the sensitivity, specificity, and accuracy of 80%, 80.95%, and 80.43% respectively (Table 3). SVM performed below other classifiers with sensitivity, specificity, and accuracy of 64%, 66.66%, and 65.21% respectively (Table 5).

Table 2: Confusion Matrix using RNN for 3 classes

	Healthy	Ataxic	Hypokinetic
Healthy	23	1	1
Ataxic	0	13	2
Hypokinetic	0	2	4

All 3 methods were trained to identify EMG design with respect to ataxic, hypokinetic, and healthy. RNN obtained accuracy, sensitivity, and specificity of 86.95%, 66.66%, and 92.5 % respectively for hypokinetic gait, and sensitivity and specificity of 83.33% and 90.32% respectively for ataxic gait (Table 2).

Table 3: Confusion Matrix using LR for 2 classes

	Healthy	Patients
Healthy	20	5
Patients	4	17

The LR model obtained sensitivity and specificity of 50% and 78.04% respectively for hypokinetic gait and sensitivity and specificity of 66.66% and 100% respectively for ataxic gait. (Table 4).

Table 4: Confusion Matrix using LR for 3 classes

	Healthy	Ataxic	Hypokinetic
Healthy	20	2	3
Ataxic	0	10	5
Hypokinetic	0	3	3

For SVM, we calculated sensitivity and specificity of 50% and 74.35% respectively for hypokinetic gait as well as a sensitivity and specificity of 64.28% and 77.41% respectively for ataxic gait (Table 6).

Table 5: Confusion Matrix using SVM for 2 classes

	Healthy	Patients
Healthy	16	9
Patients	7	14

In this research, it has been found that RNN performs best in the present framework to classify EMG gait cycle using pre-processing techniques and machine learning techniques. RNN has also the capability of getting more specificity, sensitivity, and accuracy for 3-classes and 2-classes problem.

Table 6: Confusion Matrix using SVM for 3 classes

	Healthy	Ataxic	Hypokinetic
Healthy	16	4	5
Ataxic	0	9	5
Hypokinetic	0	3	3

The overall accuracy of RNN for 2-classes and 3-classes is found to be 91.3% and 86.95% as compared to the overall accuracy of LR having 80.43% and 71.73% and SVM having 65.21% and 60.86% respectively (Table 6).

Table 7: Comparative Study

Overall Accuracy						
	Fricke et al [11]			Proposed Work		
Type of classier	CNN	KNN	SVM	RNN	LR	SVM
2 class	89.9%	48.7%	67.6%	91.3%	80.43 %	65.21%
3 class	83.8%	32.4%	51.4%	86.95%	71.3%	60.86%

4. CONCLUSION AND FUTURE WORKS

We have investigated the performance of various machine learning techniques to identify gaits disorders using EMG. We observe that the performance of the RNN with respect to other methods used in the given research is best. The performance of logistic regression and support vector machine is significantly not good. Result was consistent and got an accuracy of over 60% throughout the research. This result is better against clinical results found about 50-80% in the real world. We also found that RNN could be tested for the automatic classification of gait disorganizations using data having the accuracy of 80-90%. To further enhance the technique, we need to create a vast normative dataset containing subjects having different and extra specified gait disorganizations, possibly at different phases of their illness

and degrees of indication load that could be used to train more precisely classifiers and neural networks. They are expected to identify disorganization on an etiopathological quantity instead of phenotypic gait disorganizations as experimented here, despite the fact that different pathologies may account for a major obstacle that may be difficult to assess. In the future, neuroscience and robotics in connection with identifying weak muscle, improper walk of robot and proper shoes selection for athletes are few major areas where this work can be majorly used. We have decided to apply more deep learning approaches to improve our results.

REFERENCES

- [1] Baker JM. Gait disorders. Am J Med. (2018) 131:602–7. doi: 10.1016/j.amjmed.2017.11.051
- [2] Gale CR, Cooper C, Aihie Sayer A. Prevalence and risk factors for falls in older men and women: the English longitudinal study of ageing. Age Ageing. (2016) 45:789–94. doi: 10.1093/ageing/afw129
- [3] Thiem U, Klaaßen-Mielke R, Trampisch U, Moschny A, Pientka L, Hinrichs T. Falls and EQ-5D rated quality of life in community-dwelling seniors with concurrent chronic diseases: a cross-sectional study. Health Qual Life Outcomes. (2014) 12:2. doi: 10.1186/1477-7525-12-2
- [4] Van Schooten KS, Pijnappels M, Rispens SM, Elders PJM, Lips P, van Dieën JH. Ambulatory fall-risk assessment: amount and quality of daily-life gait predict falls in older adults. J Gerontol A Biol Sci Med Sci. (2015) 70:608–15. doi: 10.1093/gerona/glu225
- [5] Burns ER, Stevens JA, Lee R. The direct costs of fatal and non-fatal falls among older adults - United States. J Safety Res. (2016) 58:99–103. doi: 10.1016/j.jsr.2016.05.001
- [6] Davis BL, Vaughan CL. Phasic behavior of EMG signals during gait: use of multivariate statistics. J Electromyogr Kinesiol. (1993) 3:51–60. doi: 10.1016/1050-6411(93)90023-P
- [7] Chvatal SA, Ting LH. Common muscle synergies for balance and walking. Front Comput Neurosci. (2013) 7:48. doi: 10.3389/fncom.2013.00048
- [8] Frère J. Spectral properties of multiple myoelectric signals: new insights into the neural origin of muscle synergies. Neuroscience. (2017) 355:22–35. doi: 10.1016/j.neuroscience.2017.04.039
- [9] Litvan I, Goetz CG, Jankovic J, Wenning GK, Booth V, Bartko JJ, et al. What is the accuracy of the clinical diagnosis of multiple system atrophy? A clinicopathologic study. Arch Neurol. (1997) 54:937–44. doi: 10.1001/archneur.1997.00550200007003
- [10] Osaki Y, Ben-Shlomo Y, Lees AJ, Daniel SE, Colosimo C, Wenning G, et al. Accuracy of clinical diagnosis of progressive supranuclear palsy. Mov Disord. (2004) 19:181–9. doi: 10.1002/mds.10680
- [11] Fricke C, Alizadeh J, Zakhary N, Woost TB, Bogdan M and Classen J (2021) Evaluation of Three Machine Learning Algorithms for the Automatic Classification of EMG Patterns in Gait Disorders. Front. Neurol. 12:666458. doi: 10.3389/fneur.2021.666458
- [12] Alaskar H. Deep learning-based model architecture for time-frequency images analysis. IJACSA. (2018) 9:486–94. doi: 10.14569/IJACSA.2018.091268
- [13] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. (2010) 22:1345–59. doi: 10.1109/TKDE.2009.191
- [14] M Rahul, N Kohli, R Agarwal, S Mishra, "Facial expression recognition using geometric features and modified hidden Markov model", International Journal of Grid and Utility Computing 10 (5), 488-496, 2019
- [15] Jee H, Lee K, Pan S. Eye and face detection using SVM. In: Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004: 14-17 Dec. 2004. Melbourne, VIC; Piscataway, NJ: IEEE Operations Center (2004). p. 577–80.
- [16] Melvin I, Ie E, Kuang R, Weston J, Stafford WN, Leslie C. SVM-Fold: a tool for discriminative multi-class protein fold and superfamily recognition. BMC Bioinformatics. (2007) 8(Suppl. 4):S2. doi: 10.1186/1471-2105-8-S4-S2
- [17] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Haussler D, editor. Proceedings of the Fifth Annual Workshop on Computational Learning Theory. New York, NY: ACM (1992). p. 144–52

- [18] Alizadeh J, Vahid A, Bahrami F. Recognizing subjects who are learned how to write with foot from unlearned subjects using EMG signals. In: 23rd Iranian Conference on Biomedical Engineering 2016. Tehran (2019). p. 331–5.



© 2022 by the Manoj Kumar, Dr Pratiksha Gautam and Dr Vijay Bhaskar. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).