The importance of integral time length windows for the classification of activities of daily living based on Machine Learning techniques

A. Ruiz[1], E. Carbone[2], B. Larraga[1], E. Rocon[2], A. Gutiérrez[1]

[1] ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain.

[2] Centre for Automation and Robotics, CSIC-UPM, 28500 Madrid, Spain

Abstract

Pathological tremor, a prevalent movement disorder seen in essential tremor (ET) and Parkinson's disease (PD) patients, is the most common tremor disorder impacting the quality of life of those who suffer from it. This study proposes a method to classify daily life activities using a single wrist-worn IMU for tremor patients. The used dataset involves IMU recordings from the dominant arm during 11 tasks performed by ET and PD patients. Signal features were extracted from different sized windows and used to train Random Forest (RF) and Support Vector Machine (SVM) models, training 10 different models overall. Results shows that although larger window sizes, particularly the 10 seconds window, provided highest average F1-score, certain specific activities were better classified with shorter windows. This approach outperforms prior studies by achieving improved classification outcomes and opens a new line in continuous tremor monitoring. Future research could explore the combination of various window lengths to identify optimal window durations for further accuracy refinement.

1. Introduction

Pathological tremor is the most common movement disorder, characterized by involuntary and rhythmic oscillations of a part of the body, affecting mostly the hands [1]. It is commonly seen in general medical practice; essential tremor (ET) affects approximately 4% of the population above 65 years old [2], and Parkinson's disease (PD) has become the second most common neurodegenerative disorder after Alzheimer's disease [3]. While ET is mostly identified by the presence of tremors that occur during voluntary movements, PD is distinguished by tremors that manifest primarily at rest [1]. However, both frequently exhibit similar symptoms. Other than tremor, patients may also experience bradykinesia, rigidity and balance disorder [4].

As a direct cause of these symptoms, the quality of life of the patients affected by tremor is significantly impacted, gradually undermining their ability to perform their activities of their daily living (ADLs) from early stages and throughout the course of the disease [5].

The most effective treatment of tremor is medication, although drugs are usually prescribed on a trial-and-error basis, which coupled with a wide range of side-effects, represent the major drawback of the actual therapeutic strategies [5]. Motor symptoms caused by pathological tremor are typically assessed based on the mechanical demonstration of tremor and quantified using movement disorder clinical scales such as the UPDRSIII (*Unified Parkinson's Disease Rating Scale*) and the Fahn-Tolosa-Marin scale [6]. However, these evaluation methods are hampered by the bias in the performance of the patients, which may be caused by placebo effects or the "white coat syndrome". In this case patients tend to apply an extra effort due to the presence of a clinician, resulting in a biased reflection of their motor ability. Additionally, this syndrome may also constrain the evaluation of the medication's effect on the patient, which is essential to study the evolution of the symptoms with medication dosages [5], [7].

Nowadays, the combination of wearable sensing technology and data mining algorithms to recognize movement disorders has shown an increasing potential, as they can be used to quantify motor symptoms. In subjects with no tremor, ADLs' classification with IMUs has grown in the past years, approaching recognition of postural movement and activities related to motion [8].

However, the application of these methods to patients, which increases the difficulty of the task due to the tremorous component of their movements, has not yielded conclusive results yet. Previous studies have focused on detecting and classifying ADLs in patients of PD and ET using several IMUs [5,9-11]. Particularly, one of these studies focused on classifying a series of fine and gross movement activities in patients using four IMUs along the arm, with the objective of moving towards an every-day life application that could allow a continuous monitoring [5].

In this paper, a new method which provides an improvement in the classification of activities is proposed. This method uses only one IMU on the wrist and aims to improve the classification performed by a previous study [5], using the same dataset and reducing the number of IMUs used from four to one. Increasing the precision of the classification of activities in PD and ET patients would contribute to the objective of continuous monitoring of tremorous movement and the assessment of the medication's impact throughout the evolution of the disease.

2. Methodology

A dataset which contains records from 4 IMUs, placed over the dominant arm is used. This database contains ET and PD patients' data carrying out different tasks according to a specific protocol [5]. Nonetheless, this work focuses on the IMU placed at the third distal of the forearm, analyzing closest kinematic movements to the wrist. The recorded and analyzed activities were:

- Combing hair (CB)
- Buttoning the buttons of a lab coat (BB)
- Cutting a fake steak (CE)
- Eating the previously cut pieces with a fork (EF)
- Simulate drinking (SD)
- Opening and closing a tupperware container (OT)
- Turning 3 pages in a book/magazine (TB)
- Printing their name/signing a document (SN)
- Simulate tooth brushing (TB)
- Turning doorknob (TD)
- Resting arms on table (RE)

These tasks encompass both fine and proximal movements, portraying two levels of precision.

2.1. Data characteristics

The dataset is composed of acceleration and angular velocity in all three axes (x, y, z) from 16 patients, whose gender and age were not required. Each participant carried out the 11 aforementioned tasks, repeating each of them between three and six times, except for the "RE" task, which was only performed once. The tremorous and voluntary movements were separated by an adaptative algorithm based on frequency separation, and the signals were resampled at 1kHz [5].

2.2 Preprocessing and filtering

As a first step, the data was exported to dataframes to visualize and analyze the IMU signals.

First of all, a data screening was implemented. The signals were then analyzed empirically, considering the time distribution of each task. The signals that were farthest from the distribution center were discarded, as they could potentially represent erroneous recordings and lead to misclassification.

2.3 Feature extraction

Considering that the duration of the signals was different and variable for different tasks and patients, the signals were divided in different sized overlapping windows. The selected windows were (see Figure 1):

- 2.5 seconds with 1 second of overlap
- 5 seconds with 2 second of overlap
- 10 second with 4 second of overlap
- 15 second with 6 second of overlap
- 20 second with 8 second of overlap

Once the signals were segmented, the following features were extracted from each window [5], [11]: mean, standard deviation, median, maximum, minimum, difference between first and last value of window, variance, and RMS.

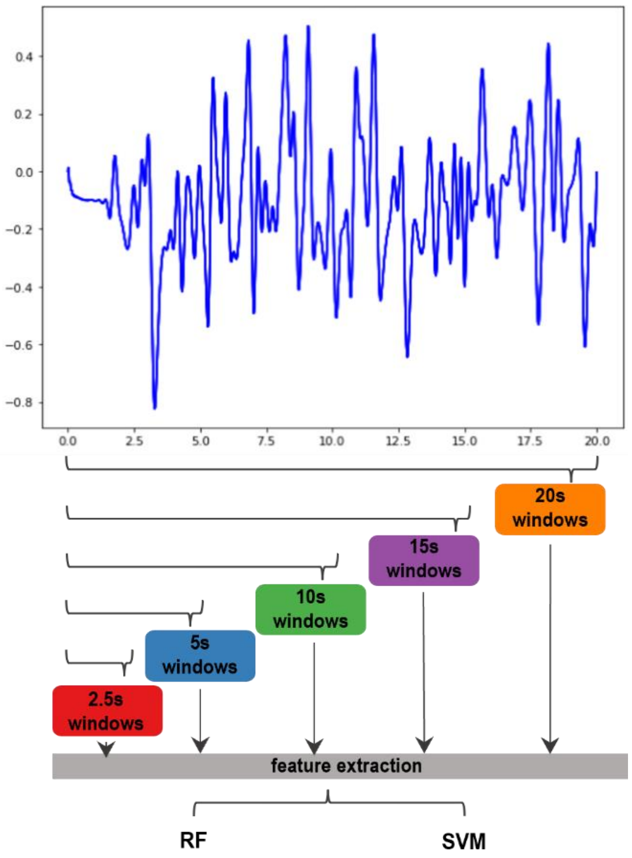


Figure 1: *Methodology followed to train and test different models based on segmentation of different sized windows.*

2.4 Classification approach

The features for each window were extracted and collected in five different dataframes corresponding to the different windows sizes. The dataset was divided into test and train groups with a 30-70% proportion respectively. It is worth noting that this division was done within patients, making sure that the patients within the train group were not in the same test group to avoid bias and overfitting problems. In this way, 11 patients were used for training and 5 for testing. All the samples were normalized using MaxMinScaler from sklearn's library in Python.

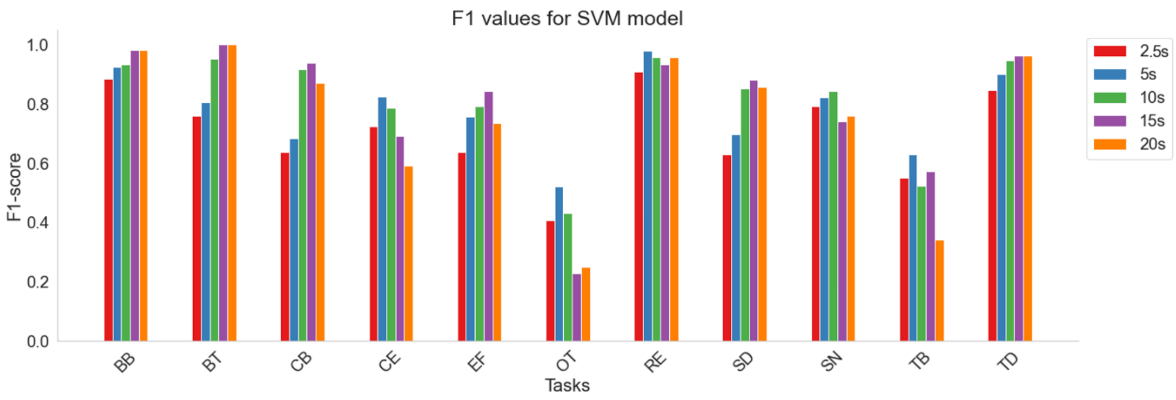
Then, two models were used: a Random Forest (RF) and a Support Vector Machine (SVM) classifier. Both models were trained with the five different window sizes proposed, hence ten different models were trained and tested.

3. Results

The F1-score was the selected metric to assess the performance of the classification models. Table 1 shows the average F1-score calculated for the ten different models. Although the SVM model provides slightly better results,

F1 score	2.5s	5s	10s	15s	20s
SVM	70.65%	77.64%	81.22%	79.76%	75.48%
RF	67.73%	74.11%	80.05%	77.93%	77.34%

Table 1: *Average F1-score of different classifiers*



(a)

(b)

Figure 2: *Bar graph showing the F1-score obtained in the classification using different time windows, in both SVM (a) and RF (b) models.*

there is no significant difference, so it cannot be determined that one model is better than the other.

The trained and tested model with the 10 seconds window performed the best classification in both models, followed by the 15 seconds one. This shows that larger windows provided better results than the shorter ones.

However, to fully understand how each individual task was being classified by the different models, Figure 2 shows the F1-score obtained by different windows for both the SVM (Figure 2a) and RF (Figure 2b).

Despite the fact that the highest average F1-scores were given by larger windows, these Figures show that, depending on the task, some of them were better classified with models trained with shorter windows.

4. Discussion

The assessment and evaluation of tremor and its evolution with medication dosages remains a challenging problem. The existing clinical evaluation methods are limited by the bias in the performance of the patients and the subjectivity of the evaluator. This could be improved by implementing a continuous monitoring in an every-day life application which could allow to correlate the tremor to the activity carried out. However, the recent developed state-of-art methods in ADL classification in tremor patients have not yielded conclusive evidence yet. They use fixed size windows to train machine learning models, but there is no clear consensus on which window size should be preferably used. Most designs are based on randomly chosen values from past successful cases, which may not necessarily be the optimal fit for the particular problem being addressed.

In this paper, a different methodology to classify ADLs considering the segmentation of the signals in different time-sized overlapping windows is presented. Although it was noted that the best results were generally given by the 10 seconds window, the performance of the classifiers was analyzed task by task.

On the one hand, some of the tasks were found to be better classified with shorter windows, such as OT and TB tasks. For these tasks, the F1-score was higher with the 5 seconds windows trained model. None of the tasks showed to be better classified with 2.5 seconds windows.

On the other hand, most of the tasks were best classified with windows between 10 and 20 seconds. For example, for CB, SD, and TD tasks the larger windows, 15 seconds, and 20 seconds windows, provided a higher F1-score. Regarding BB, BT, and EF tasks, it was seen that windows equal to 10 seconds or above provided better results, taking into consideration both SVM and RF models. Finally, RE and SN tasks showed smaller differences between the results provided by different windows, reaching similar F1-score values for each of them.

Table 2 shows the different results obtained from the tested models for every different task. The first column shows the F1-scores achieved in [5]. It is worth noting that this previous study did not include RE's classification results, although the data was present and labeled in the dataset. For this reason, the value doesn't appear in the first column, but was obtained for the following ones. The following two columns show the results obtained from both classifiers, SVM and RF, trained and tested with the segmentation of 10 seconds windows, as it gave the best overall results. Finally, the last two columns show the results obtained considering the optimal window which gave the highest F1-score for each task (see Figure 2).

These last two columns provided the best results, increasing notably the performance in the classification from the previous studies [5]. Therefore, a classification based on different time-sized windows seems to improve the prediction of ADLs.

Tasks	Previous results	10s SVM	10s RF	max SVM	max RF
BB	86,55%	93,15%	97,06%	98,18%	97,06%
BT	85,00%	95,24%	97,78%	100%	97,78%
CB	82,80%	91,53%	88,24%	93,88%	98,04%
CE	52,31%	78,72%	65,96%	82,54%	65,96%
EF	46,40%	79,25%	74,51%	84,44%	78,26%
OT	32,43%	43,24%	36,36%	52,00%	42,31%
RE	-	95,65%	100%	97,92%	100%
SD	83,34%	85,25%	82,14%	88,14%	92,31%
SN	71,43%	84,44%	81,32%	84,44%	81,32%
TB	33,33%	52,46%	62,50%	62,90%	64,00%
TD	92,21%	94,55%	94,74%	96,15%	98,11%

Table 2: *F1-score obtained for each task and model.*

5. Conclusion and future development

In this study, a different methodology to classify ADLs based on different time-sized windows was presented, which brings flexibility in the recognition of different tasks and movements carried out by the ET and PD patients. In general, it can be stated that considering the 10 seconds window models, the methodology presented gives better results than the ones shown in previous studies, considering that only one of the IMUs (the one on the wrist) was taken into consideration, instead of four in the arm.

Results show that although the segmentation in 10 seconds windows provided a better overall classification, some tasks are better classified depending on the window size, providing even better results than the ones obtained considering only the 10 seconds window model.

Consequently, this supposes a step forward in the objective of real-time activity classification for the study and evaluation of medication dosage and its effects on tremor.

In the light of the results presented in this paper, a classification based on the combination of different time windows which segments the signals in different sizes should be considered as a future development. Moreover, a new algorithm to evaluate the optimal window length should be developed.

References

[1] A. Anouti, W. C. Koller, and K. City, "Articles Tremor Disorders Diagnosis and Management."

[2] B. Thanvi, N. Lo, and T. Robinson, "Essential tremor - The most common movement disorder in older people," *Age and Ageing*, vol. 35, no. 4. pp. 344–349, Jul. 2006. doi: 10.1093/ageing/afj072.

[3] F. Demrozi, R. Bacchin, S. Tamburin, M. Cristani, and G. Pravadelli, "Toward a Wearable System for Predicting Freezing of Gait in People Affected by Parkinson's Disease," *IEEE J Biomed Health Inform*, vol. 24, no. 9, pp. 2444–2451, Sep. 2020, doi: 10.1109/JBHI.2019.2952618.

[4] M. A. Thenganatt and J. Jankovic, "The relationship between essential tremor and Parkinson's disease," *Parkinsonism Relat Disord*, vol. 22, pp. S162–S165, Jan. 2016, doi: 10.1016/j.parkreldis.2015.09.032.

[5] J. I. Serrano, S. Lambrecht, M. D. del Castillo, J. P. Romero, J. Benito-León, and E. Rocon, "Identification of activities of daily living in tremorous patients using inertial sensors," *Expert Syst Appl*, vol. 83, pp. 40–48, Oct. 2017, doi: 10.1016/j.eswa.2017.04.032.

[6] S. Fahn, "Classification of movement disorders," *Movement Disorders*, vol. 26, no. 6. pp. 947–957, May 2011. doi: 10.1002/mds.23759.

[7] T. Iluz *et al.*, "Automated detection of missteps during community ambulation in patients with Parkinson's disease: A new approach for quantifying fall risk in the community setting," *J Neuroeng Rehabil*, vol. 11, no. 1, Apr. 2014, doi: 10.1186/1743-0003-11-48.

[8] K. Frank, M. Josefa, V. Nadales, P. Robertson, and M. Angermann, "Reliable Real-Time Recognition of Motion Related Human Activities Using MEMS Inertial Sensors."

[9] L. Sigcha *et al.*, "Deep learning and wearable sensors for the diagnosis and monitoring of Parkinson's disease: A systematic review," *Expert Syst Appl*, vol. 229, p. 120541, Nov. 2023, doi: 10.1016/j.eswa.2023.120541.

[10] H. Nguyen, K. Lebel, S. Bogard, E. Goubault, P. Boissy, and C. Duval, "Using Inertial Sensors to Automatically Detect and Segment Activities of Daily Living in People with Parkinson's Disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 1, pp. 197–204, Jan. 2018, doi: 10.1109/TNSRE.2017.2745418.

[11] B. Jiang, J. J. Han, and J. Kim, "A Wearable In-home Tremor Assessment System via Virtual Reality Environment for the Activities in Daily Lives (ADLs)," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, Jul. 2022, pp. 1117–1120. doi: 10.1109/EMBC48229.2022.9871008.

[12] M. G. Martín, "Universidad Politécnica De Madrid Escuela Técnica Superior De Ingenieros De Telecomunicación Contributions To Human Motion Modeling And Recognition Using Non-Intrusive Wearable Sensors Tesis Doctoral," 2022.