The 12th International Conference on E-Health and Bioengineering - EHB 2024
Grigore T. Popa University of Medicine and Pharmacy Iasi, November 14-15, 2024, Iasi, ROMANIA

Deep Learning for Continuous Recognition of Activities of Daily Living

Ainhoa Ruiz-Vitte[1], Alberto Comesaña[2], Blanca Larraga-García[1], Eduardo Rocón[2], Álvaro Gutiérrez[1]
Affiliation 1: ETSI Telecomunicación, Universidad Politécnica de Madrid, Complutense 30, 28040 Madrid, Spain, a.rvitte@upm.es, blanca.larraga@upm.es, a.gutierrez@upm.es
Affiliation 2: Centro de Automática y Robótica, CSIC-UPM, Carr. Campo Real, km 0, 200, 28500 Madrid, Spain, alberto.comesana@cscic.es, e.rocon@csic.es

***Abstract*— Pathological tremor, which impairs the ability to perform daily activities, presents challenges for current therapeutic strategies based on subjective clinical assessments and symptomatic management. This paper introduces an innovative system for classifying activities in patients with tremor, using models trained on time series data segmented into different window sizes. The system achieved an overall accuracy of 80%, although performance varied across tasks. This represents a significant advancement in the current state of continuous activity classification systems, highlighting the potential to enhance tremor monitoring and optimize medication dosage.**

Keywords—tremor; deep learning; time series; activity classification.

I. INTRODUCTION

Pathological tremor, characterized by involuntary and rhythmic oscillations primarily affecting the hands, is the most prevalent movement disorder seen in medical practice [1]. Essential tremor (ET) impacts around 4% of the population over 65, while Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease [2,3]. These symptoms deeply affect patients' quality of life, impairing their ability to perform activities of daily living (ADLs) from the early stages of the disease and throughout its progression [4].

While medication is the most effective treatment for tremor, it is often prescribed on a trial-and-error basis, accompanied by a broad range of side effects, presenting a significant drawback in current therapeutic strategies [5]. Motor symptoms resulting from pathological tremor are usually assessed through mechanical demonstrations of tremor and quantified using clinical scales. However, these methods are limited by patient performance bias, influenced by placebo effects or the “white coat syndrome” [4,6], where patients exert extra effort in the clinician's presence, leading to an inaccurate reflection of their motor abilities. This syndrome also complicates evaluating the medication's effectiveness, essential for studying symptom evolution with dosage adjustments [6].

In an attempt to minimize the subjectiveness involved in assessing movement disorders and alleviate the burden of current therapeutic strategies, motion analysis has been widely developed, approaching recognition of postural movement. Although vision-based sensing has been the focus of research studies, recently, the use of Inertial Measurement Units (IMUs) have been further investigated [7]. The application of deep learning models for ADLs recognition using wearable-based sensors has recently gained significant attention, with new techniques being continually developed [8,9,10]. Providing information on the tasks performed during the monitored tremor occurrences could enhance the understanding of the patient’s clinical condition. However, applying these task identification methods to patients who face additional challenges due to the tremor in their movements, has not yet provided clear results [6,11,12]. Often, studies involving patients do not cover the identification of complex everyday tasks, focusing instead on postural and transitional activities, and those that do tackle complex task identification use multiple IMUs, which can be burdensome for patients.

Regarding task identification, accurate selection of window size is crucial for effective activity classification. However, it remains highly ambiguous and undefined yet [13,14,15], with most designs relying on estimates from previous studies without substantial empirical support.

The main contribution of this study is the classification of ADLs from continuous sequences of time series data. This is achieved by using various windowing techniques and applying Long Short-Term Memory (LSTMs) models. A smartwatch was used to capture real-time wrist motion data. This approach will enable neurologists to improve the monitoring of tremor progression and assess the impact of medication on patients’ quality of life by measuring tremors during ADLs. The objective is to facilitate continuous monitoring of tremorous movements and evaluate the medication’s effect throughout the disease’s progression.

II. METHODOLOGY

A. Data Collection

Based on previous studies [16], a population of 70 subjects was segmented: 63 of them (5 with ET, 5 with PD, 37 control participants, and an additional dataset of 16 patients with indeterminate ET and PD) were used to train and validate the

2024 E-Health and Bioengineering Conference (EHB) | 979-8-3315-3214-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/EHB64556.2024.10805639

979-8-3315-3214-7/24/$31.00 ©2024 IEEE

ADL classifier, while the remaining 7 control participants were used to extract continuous signals to test the application of the classifier in a daily home environment, where they wore the wristband for prolonged periods of time.

In the first step, a smartwatch (Fitbit Sense, [17]) was placed on the dominant wrist to collect data, recording linear acceleration and angular velocity along three axes (x, y, z), as shown in Figure 1. The motion data were sampled at 30 Hz. Subsequently, all the data were sent to an Android device, which transferred them to a database for storage. During the tests for training the model, each participant performed nine daily life tasks at a self-selected pace, in the same way they would normally perform them. The tasks included: comb hair (CH), button buttons (BB), cut with knife and eat with fork (EF), simulate drinking (SD), open/close box (OB), turn pages (TP), write (WW), brush teeth (BT), and turn doorknob to open/close door (TD).

Each participant from the training group performed these tasks, repeating each one between three and six times. The movements between repetitions and during task performance were also recorded and labeled as ‘no-task’ (NT).

In a second phase, to obtain continuous signals from the validation group in the home environment, subjects wore the smartwatch for extended time periods, marking on the Android device when they started or finished each task, without any predetermined order.

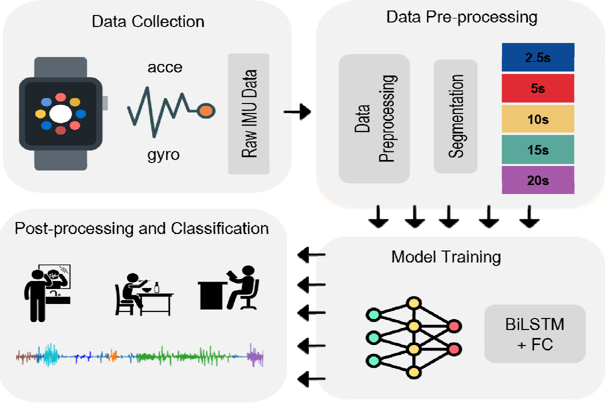


Fig. 1. Schematic representation of the data flow.

B. *Preprocessing and filtering*

First, data screening and cleaning were performed to discard signals outside the central time distribution and to avoid defective recordings. A Butterworth filter with a cutoff frequency of 4 Hz and a 6th order was applied to separate voluntary movements from tremor, ensuring that tremor was filtered out without affecting voluntary movement [18].

C. *Classification strategy*

We applied different window sizes to classify continuous, longer-duration signals from subjects without tremor. This approach is based on the idea that varying window sizes can optimize task classification and provides a holistic perspective for signal analysis [16]. A set of models based on LSTMs was trained and validated using data from the 63-subject group and tested on the continuous longer signals collected from the 7 control participants.

First, the data were segmented into training, validation, and testing sets, comprising 70%, 20%, and 10% respectively. Four different window sizes—5, 10, 15, and 20 seconds—were applied, with overlaps of 2, 4, 6, and 8 seconds. Subsequently, four distinct LSTM models were trained on these window segments. Parameters such as learning rate and batch size remained consistent across all four models. The models featured a Bidirectional LSTM layer, to capture temporal dependencies. Dropout layers were included to reduce overfitting as well as L2 regularization, followed by two dense layers.

Once trained, each model produced a classification vector, assigning a task label to each window on the test set. By using interpolation, the labels for each window were extended to each timestamp, resulting in four different vectors. Each vector corresponded to a different windowed LSTM and contained a task label for each timestamp in the test group.

These four vectors were then combined to determine the most appropriate label for each timestamp. The mode of the labels was selected, and in cases of ambiguity, neighboring values were considered. Finally, a postprocessing stage was conducted in which labeled sequences shorter than a specified threshold were merged with neighboring sequences, only if they shared the same label and exceeded a certain length threshold. This postprocessing step helps to mitigate incoherent misclassifications.

III. RESULTS

Table 1 compares the metrics provided by the ensemble model with those from each individual classifier, demonstrating that the combined approach yields improved results across all evaluated metrics.

TABLE I. WINDOW SIZE ON PERFORMANCE COMPARISON

LSTM model	Precision	Recall	F1-score
Combined	79.81%	80.07%	79.53%
5-seconds	78.77%	77.43%	77.73%
10-seconds	78.50%	73.48%	74.98%
15-seconds	76.62%	73.72%	74.51%
20-seconds	66.66%	63.09%	63.46%

To demonstrate the model’s performance, graphs like the one in Figure 2 were constructed for each patient, where distinct colors represent various tasks over time, including ‘no-task’, which indicates the absence of any specific task. The bottom image shows the tasks performed, while the top image shows the tasks classified by the model.

Table 2 presents the metrics used to evaluate the model’s performance by task. The overall accuracy is 80.07%. However, there is a notable variability between tasks that will be further discussed.

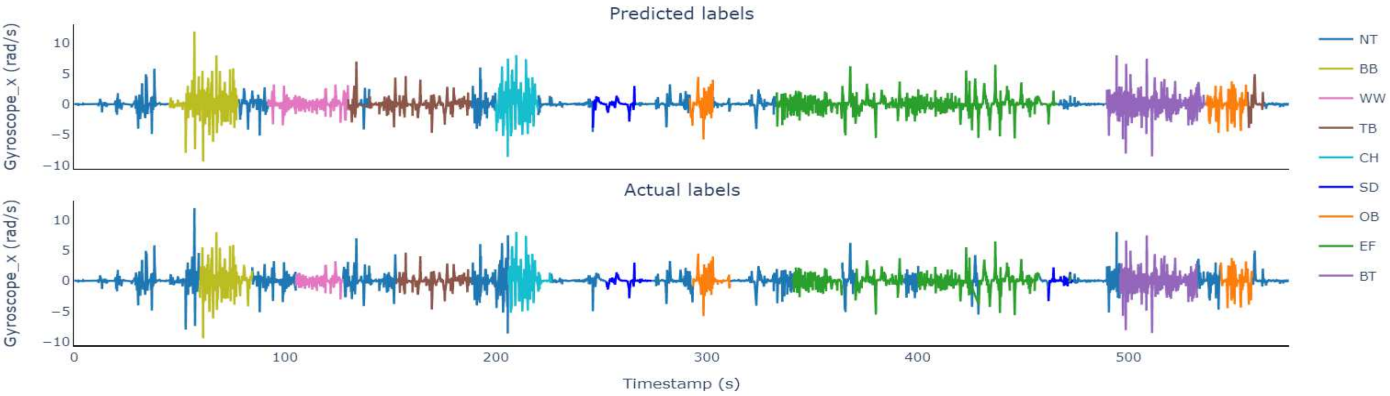


Fig. 2. Graphs comparing the predicted labels with the true ones in a continuous sequence of ADLs.

TABLE II. CLASSIFICATION RESULTS

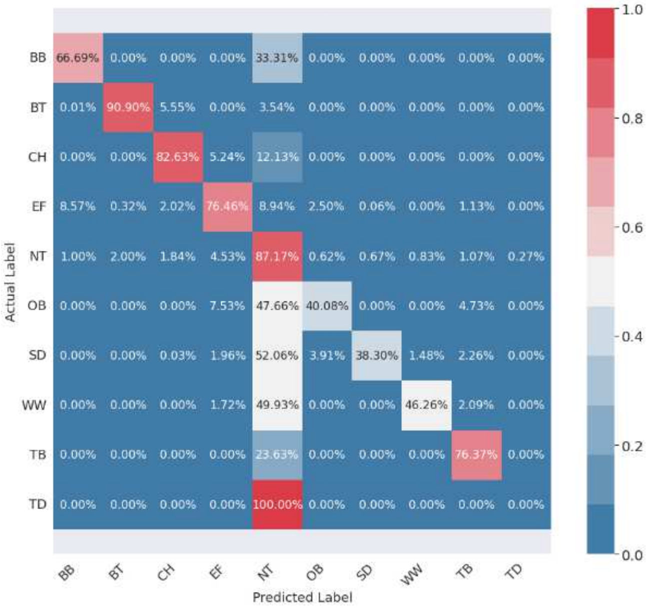
Tasks	Precision	Recall	F1-score
BB	39.96%	66.69%	49.98%
BT	75.56%	90.90%	82.52%
CH	54.22%	82.63%	65.47%
EF	79.24%	76.46%	77.82%
NT	86.12%	87.17%	86.64%
OB	43.33%	40.08%	41.64%
SD	71.25%	38.30%	49.82%
WW	73.21%	46.26%	56.69%
TB	76.49%	76.37%	76.43%
TD	00.00%	00.00%	00.00%
Accuracy	-	-	80.07%

In Table 3, more detailed information is provided on the classification of TD regarding the performance of the individual models.

TABLE III. F1-SCORE ON TD'S CLASSIFICATION

LSTM-5	LSTM-10	LSTM-15	LSTM-20
26.77%	10.48%	0.00%	5.48%

The confusion matrix in Figure 3 illustrates the model's performance across different classes. Although the matrix shows a strong diagonal, indicating accurate predictions for many tasks, it also reveals significant classification errors, particularly false positives with the task NT. This suggests that, while the model is generally effective, there are specific areas where its ability to differentiate between tasks could be improved.



Actual Label	BB	BT	CH	EF	NT	OB	SD	WW	TB	TD
BB	66.69%	0.00%	0.00%	0.00%	33.31%	0.00%	0.00%	0.00%	0.00%	0.00%
BT	0.01%	90.90%	5.55%	0.00%	3.54%	0.00%	0.00%	0.00%	0.00%	0.00%
CH	0.00%	0.00%	82.63%	5.24%	12.13%	0.00%	0.00%	0.00%	0.00%	0.00%
EF	8.57%	0.32%	2.02%	76.46%	8.94%	2.50%	0.06%	0.00%	1.13%	0.00%
NT	1.00%	2.00%	1.84%	4.53%	87.17%	0.62%	0.67%	0.83%	1.07%	0.27%
OB	0.00%	0.00%	0.00%	7.53%	47.66%	40.08%	0.00%	0.00%	4.73%	0.00%
SD	0.00%	0.00%	0.03%	1.96%	52.06%	3.91%	38.30%	1.48%	2.26%	0.00%
WW	0.00%	0.00%	0.00%	1.72%	49.93%	0.00%	0.00%	46.26%	2.09%	0.00%
TB	0.00%	0.00%	0.00%	0.00%	23.63%	0.00%	0.00%	0.00%	76.37%	0.00%
TD	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Fig. 3. Confusion matrix of continuous ADL classification.

IV. DISCUSSION

The model evaluation reveals that, despite an overall accuracy of 80%, significant challenges persist in the classification of ADLs in prolonged continuous signals.

Upon analyzing Figure 2, it is evident that, although the system correctly identifies most tasks, some classification errors persist. For instance, early cases of WW and very short signals of TB are detected, which are inconsistent with the expected behavior. This inconsistency suggests that post-processing mechanisms could be improved to correct potentially erroneous classifications based on contextual information.

This lack of precision is reflected in the metrics presented in Table 2. For tasks such as BB and OB, which are potentially more complex due to their bimanual nature, the accuracy is low (< 50%), despite a high recall in the case of BB, indicating the presence of false positives, as observed in Figure 3. In contrast, tasks such as BT, EF or TB show balanced performance metrics, suggesting more reliable classification. The task with the highest metrics is NT, due to its greater representation in the dataset. The confusion matrix supports this observation, revealing that many tasks are incorrectly classified as NT, due to class imbalance. Another significant observation is that TD is successfully classified by the individual LSTM-5,10, and 20 models, but not by the ensemble model. This suggests that, while combining different window sizes improves overall accuracy compared to using individual windows, as shown in Table 1, the refinement of postprocessing and merging methods could avoid the discrepancies observed in Table 3.

These results suggest that, while the system is promising, especially for certain tasks, it needs refinement to address issues related to class imbalance and improve the model's ability to distinguish between closely related or bimanual tasks. Enhancing post-processing techniques and implementing strategies to mitigate class imbalance could significantly boost the system's overall performance, making it a more reliable solution for continuous activity monitoring in home settings. Additionally, incorporating convolutional layers alongside the LSTM for feature extraction may lead to better outcomes and further improve the model's accuracy and generalization ability, addressing those tasks with lower performance.

V. CONCLUSIONS AND FUTURE WORKS

This study presents an innovative system for the continuous classification of daily activities in patients with tremor, using LSTMs trained with data segmented into different window sizes. With an overall accuracy of 80%, the system shows potential for classifying tasks in the home environment, covering sequences from 5 to 70 minutes, with the aim of extending this capability to analyze longer continuous signals. Although the system is promising, it faces challenges related to class imbalance and the accurate identification of bimanual tasks. Future work should include expanding the dataset and testing with patients in their homes. Including a larger and more diverse patient group would enhance the generalizability and real-world applicability of the results. Additionally, refining post-processing techniques to smooth sequences of classified activities could minimize misclassifications. Addressing class imbalance is essential, and exploring data augmentation or oversampling techniques could provide effective solutions.

ACKNOWLEDGMENTS

This work has been partially funded by the National Program for Scientific, Technical and Innovation Research 2021-2023 within the context of the project NETremor (TED2021-130174B-C32) and by Universidad Politécnica de Madrid under the Técnicas de Control y Análisis de Variables Cinemáticas Aplicadas a la Extracción de Patrones de Movimiento project (RP2109560107).

REFERENCES

[1] A. Anouti, W. C. Koller, and K. City, "Articles Tremor Disorders Diagnosis and Management," Jun. 1995.

[2] 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 Journal of Biomedical and Health Informatics*, vol. 24, no. 9, pp. 2444–2451, Sep. 2020, doi: 10.1109/JBHI.2019.2952618.

[3] 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 Systems with Applications*, vol. 83, pp. 40–48, Oct. 2017, doi: 10.1016/j.eswa.2017.04.032.

[4] E. Rocon, J. M. Belda-Lois, J. J. Sanchez-Lacuesta, and J. L. Pons, "Pathological tremor management: Modelling, compensatory technology and evaluation," *Technology and Disability*, vol. 16, no. 1, pp. 3–18, Jul. 2004, doi: 10.3233/TAD-2004-16102.

[5] Gironell A, Marín-Lahoz J, Póveda S. Essential tremor: Update of therapeutic strategies. June 2024;162(12):599-605.

[6] I. Vanmechelen *et al.*, "Assessment of movement disorders using wearable sensors during upper limb tasks: A scoping review," *Frontiers in Robotics and AI*, vol. 9, Jan. 2023, doi: 10.3389/frobt.2022.1068413.

[7] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, "Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey," in *23rd International Conference on Architecture of Computing Systems 2010*, Feb. 2010, pp. 1–10.

[8] K. Al-Majdi, R. S. H. AL-Musawi, A. H. Ali, S. A. Abudlmoniem, and Y. S. Mezaal, "Real-Time classification of various types of falls and activities of daily livings based on CNN LSTM network," *PEN*, vol. 9, no. 3, p. 958, Sep. 2021, doi: 10.21533/pen.v9i3.2599.

[9] V. Nunavath *et al.*, "Deep Learning for Classifying Physical Activities from Accelerometer Data," *Sensors*, vol. 21, no. 16, p. 5564, Aug. 2021, doi: 10.3390/s21165564.

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

[12] 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)," presented at the 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.

[13] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window Size Impact in Human Activity Recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, Apr. 2014, doi: 10.3390/s140406474.

[14] N. Yala, B. Fergani, and A. Fleury, "Towards improving feature extraction and classification for activity recognition on streaming data," *J Ambient Intell Human Comput*, vol. 8, no. 2, pp. 177–189, Apr. 2017, doi: 10.1007/s12652-016-0412-1.

[15] Tao Gu, Zhanqing Wu, Xianping Tao, H. K. Pung, and Jian Lu, "epSICAR: An Emerging Patterns based approach to sequential, interleaved and Concurrent Activity Recognition," presented at the 2009 IEEE International Conference on Pervasive Computing and Communications, IEEE, Mar. 2009, pp. 1–9. doi: 10.1109/PERCOM.2009.4912776.

[16] A. Ruiz-Vitte, E. Carbone, B. Larraga, E. Rocon, and Á. Gutiérrez, "The importance of integral time length windows for the classification of activities of daily living based on Machine Learning techniques," presented at the XLI Congreso Anual de la Sociedad Española de Ingeniería Biomédica, Cartagena, Nov. 2023.

[17] Google Fitbit, "Advanced Health Smartwatch | Fitbit Sense."

[18] J. Gallego, E. Rocon, J. O. Roa, J. Moreno, and J. L. Pons, "Real-Time Estimation of Pathological Tremor Parameters from Gyroscope Data," *Sensors*, vol. 10, no. 3, pp. 2129–2149, Mar. 2010, doi: 10.3390/s10030