Computers & Graphics 131 (2025) 104276

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on SIBGRAPI 2023 Tutorials

Hyperkinetic movement disorder analysis using multidimensional projections☆

Andressa Silva da Silva [a], Eduardo F. Ribeiro [b], Jelle R. Dalenberg [c], Alexandru C. Telea [d], Marina A.J. Tijssen [c], João Luiz Dihl Comba [a,*]

[a] Instituto de Informática, UFRGS, Brazil
[b] Universidade Federal de Tocantins, Palmas, Brazil
[c] Expertise Center Movement Disorders Groningen, Department of Neurology, University of Groningen, Groningen, The Netherlands
[d] University of Utrecht, Utrecht, The Netherlands

ARTICLE INFO

Keywords:
Computers and graphics
Formatting
Guidelines

ABSTRACT

Hyperkinetic movement disorders are a group of conditions characterized by involuntary movements such as tremors, sudden/uncontrollable jerks, abnormal postures, and random movements, which may have major impacts on the quality of life of individuals. The diagnosis of these disorders is often dependent on subjective clinical assessments, and there is a need for automatic methods that can support this diagnosis. Established clinical neurophysiological approaches use motion sensors to collect motion data from patients performing postural, action, or resting tasks to analyze and classify the types of disorders that affect patients. However, making sense of the high-dimensional space formed by patients, tasks, sensors, and disorders is challenging and time-consuming. In this paper, we propose a workflow to explore this space to select appropriate subsets of its data, transform it, and analyze it using multidimensional projections. We show how our workflow can lead to insights into the design of automated pipelines that automatically separate individuals with disorders from healthy individuals.

1. Introduction

Movement disorders refer to illnesses related to either a reduction in motor activity or abnormal involuntary movements. Among these disorders, *hyperkinetic movement disorders*, including myoclonus, dystonia, tremor, chorea, and tics, manifest by excessive involuntary movements [1]. Accurate classification of movement disorder phenotypes is crucial for guiding clinical decisions in subsequent diagnostic steps, such as conducting additional tests and determining appropriate treatment strategies. However, the clinical diagnosis of each movement disorder phenotype can be complex and may share overlapping clinical features of other disorders, such as ataxia, spasticity, and functional movement disorders [2]. In addition, the lack of evident anatomical brain abnormalities leads to a diagnosis based on a subjective visual assessment by experts. An additional challenge relates to the relatively high cost of traditional clinical diagnosis: medical specialists trained in movement disorder diagnosis typically study each patient individually as they perform several live tests. This costly procedure requires direct access to a small pool of trained specialists.

To aid diagnosis, data collection from patients performing specific tasks using motion sensors (i.e., accelerometry, inertial measurement units), electromyography (EMG), and three-dimensional video has been recently proposed in the Next Move in Movement Disorders (NEMO) project [2] as a way to improve and automate the movement disorder classification problem. The project aims to develop a movement disorder classification tool that uses the abovementioned data to aid specialists in posing diagnoses based on standard tasks taken by patients.

Developing such a classification tool is challenging for two reasons. First, having a large number of motion disorder patients who can undergo the aforementioned tests, have their data collected, and be labeled as having a given disorder by a specialist is difficult due to the relative rarity of these patients; the cost of the procedure; and the difficulty in posing unambiguous diagnoses. Secondly, and more importantly, the complexity of the generated data space from the aforementioned testing procedure is extremely high. This complexity arises from multiple patients performing various test types, each monitored by

☆ This article was recommended for publication by A. Ynnerman.

* Corresponding author.
E-mail addresses: ssandressas@gmail.com (A. Silva da Silva), uft.eduardo@mail.uft.edu.br (E.F. Ribeiro), j.r.dalenberg@umcg.nl (J.R. Dalenberg), a.c.telea@uu.nl (A.C. Telea), m.a.j.de.koning-tijssen@umcg.nl (M.A.J. Tijssen), comba@inf.ufrgs.br (J.L.D. Comba).

https://doi.org/10.1016/j.cag.2025.104276
Received 22 July 2024; Received in revised form 16 May 2025; Accepted 1 June 2025
Available online 10 July 2025

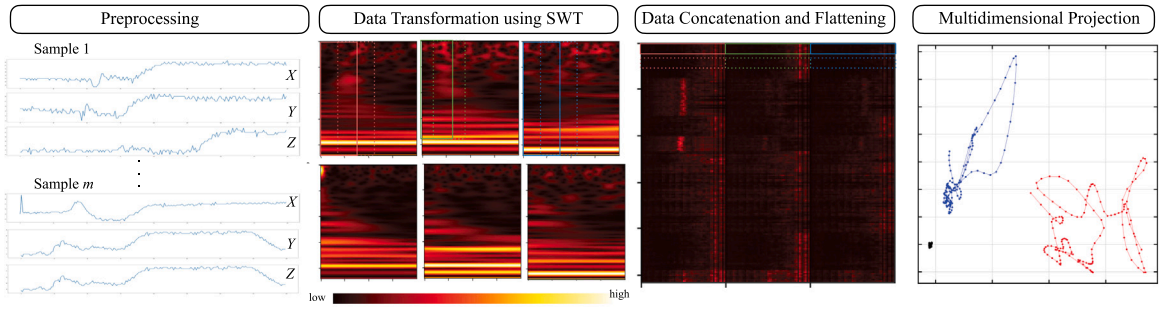

Fig. 1. Overview of the proposed workflow: Raw sensor data $x_i^j(t)$ is collected from various locations on a patient's body while performing specific tasks. The data undergoes preprocessing, including normalization, filtering, and standardization. Subsequently, spectrograms are generated using the Stationary Wavelet Transform (SWT). These spectrogram images are then concatenated and flattened to form a high-dimensional, time-dependent dataset. Finally, the dataset is projected onto a 2D space, where smooth trajectories representing patient behavior are constructed using Akima spline interpolation.

several sensors collecting multiple signals at different sampling rates, all subject to various types of noise. Consequently, identifying the most relevant data subset for predicting a given disorder is a highly non-trivial task. Nevertheless, a proof-of-concept classification approach has surfaced recently [3].

In this paper, we propose a workflow to analyze and process the multidimensional data collected by the NEMO project to select a (small) subset that best predicts a given movement disorder. Fig. 1 illustrates the steps in the proposed workflow. We use dimensionality reduction (DR) methods to transform the high-dimensional data into lower dimensions. DR techniques are widely used to simplify the feature space while preserving the data's essential structure and relevant features. However, although well-known in machine learning and information visualization, to our knowledge, DR methods have not been widely used to analyze dynamic (time-dependent) data.

In our work, we focus on two movement disorders: essential tremor (ET) and cortical myoclonus (CM), comparing them against health volunteers (HV). Distinguishing between these hyperkinetic movement disorders is clinically challenging due to their overlapping symptoms. This overlap leads to significant variability in diagnosis both between different observers and by the same observer at different times. Although our goal does not directly address the clinical problem of phenotypic classification, our approach offers a step-by-step workflow for processing the available data to make it further amenable to analysis and selecting relevant data that can lead to a good separation of patient groups having different conditions. Our workflow (and its underlying methodology) can further reduce the effort of practitioners in the movement disorders field when designing tasks, deploying additional sensors, and collecting supplementary data towards constructing an automatic classification tool.

In summary, the main contributions of this work are:

- We propose a pipeline for exploring and evaluating non-stationary signals using dynamic projections;
- We validate our pipeline with a dataset collected from individuals with hyperkinetic disorders performing specific tasks while wearing sensors at different body locations;
- We present experimental results that identify key factors for effectively distinguishing different types of conditions;
- We provide visual and clustering analysis results showing how projections aid in separating different conditions.

2. Related work

We next review previous work relevant to our research, focusing on four key areas. First, we discuss studies on the classification of hyperkinetic disorders. Next, we examine signal analysis techniques applied to movement disorder data. We then explore multidimensional projection methods for analyzing time-dependent high-dimensional data. Finally, we review techniques for visual clustering analysis, which can reveal patterns or anomalies in high-dimensional datasets.

2.1. Classification of hyperkinetic disorder studies

Hyperkinetic disorders are characterized by involuntary, uncontrollable, and undesirable movements, standing in contrast to hypokinetic movement disorders, such as Parkinson's Disease, characterized by the loss of voluntary and automatic movements, accompanied by slowness and an increase in muscular tone or stiffness. The NEMO dataset [2] was collected as part of a broader initiative to study sensor data to classify hyperkinetic disorders automatically. The study comprised participants with single and mixed phenotype movement disorders (ET and CM), as well as HV subjects. Participants performed distinct motor tasks in a movement registration setup, along with one motor task and three non-motor tasks in neuroimaging settings. Recent advances in the NEMO project [4] include the addition of neuroimaging data acquisition (such as fMRI and PET scans) to assess more complex and mixed movement disorders, such as functional and movement disorders.

The NEMO dataset was recently used in a proof of concept for classifying movement disorders using explainable machine learning applied to power spectra from accelerometry recordings [3]. They applied Generalized Matrix Learning Vector Quantization (GMLVQ) to classify ET and CM disorders, demonstrating the potential of machine learning in aiding clinical diagnosis and improving the accuracy and efficiency of movement disorder classification. However, while classification results were good, this work used only a hand-picked subset of eight from the 16 available accelerometry sensors; did not consider the 10 available gyroscope sensors; and did not consider the HV disorder (more details on the sensor data are given in Section 4.1. Considering all these variables creates a significantly larger data space to explore. How to identify good variable combinations that further assist in predicting conditions is precisely the focus of our work in this paper.

A related study [5] uses the NEMO dataset to examine differences in brain glucose metabolism among hyperkinetic movement disorders using PET scans. Significant metabolic differences were observed, with ET patients showing increased glucose uptake in the right cerebellum and reduced uptake in parietal areas. Comparisons across disorders revealed distinct metabolic patterns, suggesting that glucose metabolism differences could aid in classifying hyperkinetic movement disorders. In contrast to this work, which aims to classify hyperkinetic movement disorders based on differences in brain metabolism, we aim to support the same classification using arguably simpler accelerometer sensor data.

2.2. Time–frequency analysis of movement disorders signals

In clinical practice, combined Electromyography (EMG) and accelerometer are the main electrophysiological measurements to support patient diagnosis and movement analysis [6]. Typically, these

measurements are conducted during movement tasks or at rest and clinically assessed using the patient's movement pattern characteristics, such as muscle activation patterns, movement burst duration, and frequency [7]. Additionally, Inertial Measurement Units (IMUs) have gained increasing popularity in movement analysis due to their affordability and accuracy [8]; IMUs are characterized by their incorporation of tri-axial sensors, including accelerometers, gyroscopes, and magnetometers, which provide information about linear acceleration, angular velocity, and rotation [9].

Time-domain analysis of motion sensor signals provides insights into energy distribution over time but falls short in revealing the underlying physiological processes driving movement disorders. Better approaches involve decomposing signals into their constituent frequencies. Yet, traditional methods like the Fourier Transform (FT) assume signal stationarity, making them unsuitable for capturing the dynamic nature of our data.

To address this, time–frequency representations describe signals in terms of time, frequency, and amplitude [10], which are more suitable for non-stationary data. The wavelet transform is a popular and effective alternative to FT, especially for extracting meaningful information from EMG signals [11].

Several studies have used time–frequency analysis to examine movement disorders. Phinyomark et al. [11] explored EMG feature extraction through multi-level wavelet decomposition. Nijmeijer et al. [12] examined EMG coherence and spectral analysis to find dystonic muscles in cervical dystonia. Kramer et al. [13] used wavelet coherence analysis to differentiate between tremor types, focusing on distinguishing functional from organic tremors. Go et al. [14] used frequency domain analysis for the quantitative diagnosis of lower extremity dystonia. By quantifying the spectral content of these movement-related signals, characteristic frequency bands and their temporal variations have been found leading to a deeper understanding of the neurophysiological basis of each disorder.

In contrast to such studies, we focus on applying *projection techniques* to motion data derived from movement disorders, diverging from the main focus in previous research on analyzing EMG and accelerometer signals using FT and wavelet transform methods. The use of dimensionality reduction to this problem has shown potential in [15] in a small subset of the NEMO data. We expand on this work and propose a workflow that can tackle the full complexity of the NEMO dataset.

Given the extensive data collected from movement registration, we focused our investigation only on the accelerometer and gyroscope data of the motion sensors. We decided not to include EMG and video data because of the additional complexity associated with non-standard preprocessing procedures.

2.3. Multidimensional projections

Data generation in medicine, business, biology, and social media [16] has led to datasets with millions of samples (also called observations) each with hundreds of dimensions (also called variables or attributes). Such high-dimensional datasets present significant challenges for analysis and visualization.

Dimensionality Reduction (DR) techniques, also called projections, are an established solution for such analysis and exploration tasks [17]. DR techniques transform high-dimensional data into low-dimensional data while preserving intrinsic relationships among samples in the original space [18]. This simplifies and accelerates analysis and exploration tasks. More formally, given a dataset $D = \{\mathbf{x}^i\}_{i=0}^{n}$ with samples $\mathbf{x}^i \in \mathbb{R}^d$, a projection can be denoted as

$$P : \mathbb{R}^d \rightarrow \mathbb{R}^q \tag{1}$$

where $q \ll d$. For $q \in \{2, 3\}$, the projection $P(D) = \{P(\mathbf{x}^i) | \mathbf{x}^i \in D\}$ is a scatterplot which can be directly visually explored.

Tens of different projection techniques have been proposed for static (time-independent) data, most notably Principal Component Analysis (PCA) [19], t-distributed Stochastic Neighbor Embedding (t-SNE) [20], and UMAP [21]. Extensive studies have shown how projection methods compare in terms of quality and accuracy [22] and suitability for exploration tasks [23].

Projecting *dynamic* (time-dependent) high-dimensional data has been studied significantly less. Early applications include time series analysis and clustering [24–29]. Fujiwara et al. [30] proposed a PCA-based method for visualizing streaming multidimensional data, demonstrating its effectiveness in handling temporal data. Ali et al. [31] presented a visual analytics system that supports PCA, t-SNE, and UMAP to analyze multivariate time series, facilitating the identification of patterns, outliers, and trends in data from diverse sectors such as medicine, finance, and earth sciences. Abdullah et al. [32] proposed a visual analytics system that uses projections and cluster analysis to process high-dimensional data from electronic health records. More recently, Vernier et al. [33] surveyed the field of dynamic projections and highlighted how traditional projection techniques, designed for static data, fare when projecting dynamic data. Importantly, they observed that a trade-off seems to exist between accuracy (the ability of a projection to keep the data structure) and stability (the ability of a projection to change only when the underlying data changes). Following this observation, Vernier [15] next introduced two new dynamic projection algorithms called Landmark Dynamic t-SNE (LD-tSNE) and Principal Component Dynamic t-SNE (PCD-tSNE). PCD-tSNE, in particular, represents a hybrid approach that combines the stability of PCA with the accuracy (measured by neighborhood preservation) of t-SNE. Munz-Körner and Weiskopf [34] addresses projection errors in multidimensional time series visualization, proposing quality metrics and uncertainty visualization techniques to mitigate misinterpretations caused by dimensionality reduction.

Our work applies dynamic DR techniques to the domain of movement disorders analysis. Our workflow addresses the challenges of understanding the high dimensionality of data collected from patients, tasks, sensors, and disorders.

2.4. Visual clustering analysis

Comparing multiple clustering results is essential for evaluating different methods and parameter choices. Visual cluster analysis integrates clustering techniques with data visualization to enhance interpretability and facilitate result assessment. Several approaches have been developed to address this challenge. Sawada et al. [35] introduced a technique for clustering multidimensional time series subsequences, enabling visual feature extraction from Blazar observation datasets. Their interactive framework supports dynamic exploration of clustering results, allowing users to identify patterns in high-dimensional astronomical data. Similarly, Mohammed Ali et al. [31] proposed a system for visual analysis of temporal data, leveraging dimensionality reduction to cluster similar patterns while detecting and exploring outliers and repeated trends in large time-series datasets. Other visualization systems developed for clustering comparison include XCluSim [36], Clustervision [37], and Clustrophile 2 [38] provide interactive tools to visually compare clustering outcomes, helping users assess different approaches effectively. These techniques offer valuable insights in the development of a visual analytics tool for analyzing our clustering results, which we plan to explore as future work.

3. Proposed workflow

Our proposed pipeline for analyzing hyperkinetic movement disorders using dynamic multidimensional projections is shown in Fig. 1. Our pipeline consists of three key stages: data preprocessing, transformation, and projection – described next.

3.1. Data preprocessing

We next introduce the notations used to describe the dataset collected in the NEMO project which we further explore. A *sensor* is a measurement device attached to a location on a subject's body (Fig. 1). A sensor measures several *channels*, *e.g.*, the three x, y, and z components for a 3D accelerometer. We denote by $x_i^j(t)$ the signal, sampled over time t, produced by the jth channel of sensor i. For a task k, $1 \leq k \leq T$, let $D_k = \{(x_i^j(t), t_k^{start}, t_k^{end})\}$ be the all measurements produced by all sensors. Here, t_k^{start} and t_k^{end} give the time interval over which $x_i^j(t)$ were recorded, *i.e.*, $t \in [t_k^{start}, t_k^{end}]$. For each patient $1 \leq p \leq P$, we have a dataset $P_p = \{D_1, \ldots, D_T\}$ containing all measurements performed by all sensors over all tasks.

The data recorded for all tasks and patients exhibits slight variations mainly in the duration of the recordings of movement execution, *i.e.*, t_k^{start} and t_k^{end} are not the same for all tasks k and patients p. To further compare data over multiple tasks and patients, we normalize the data by (1) cropping the recorded sequences to uniform durations t_k^{start} and t_k^{end} and aligning the start and end movement executions across patients.

Next, we apply a signal processing filter to the signals $x_i^j(t)$ to remove unwanted frequencies that might distort these signals. Specifically, we employ a 6th-order Butterworth bandpass filter with a cutoff frequency set at 1.0 Hz. The choice for Butterworth was based on its capability to maintain a maximally flat frequency response in the passband. This "maximally-flat" characteristic ensures a smooth and uniform frequency response within the specified range, without ripples or variations in amplitude, allowing for consistent signal integrity.

Finally, we scale the filtered signals to unit variance to achieve normal distributed data (Gaussian with a mean of 0 and a variance of 1). A sample standard score $z(t)$ is computed as $z(t) = (x(t) - u)/s$, where u and s represent the sample mean and standard deviation of signal $x(t)$, respectively.

3.2. Data transformation via wavelet transform

When exploring time-series data, capturing frequency components and understanding their temporal evolution becomes essential. Several methods can be considered for this task, as follows. Traditional methods like the FT operate under the foundational assumption of signal stationarity, which is incompatible with the intrinsically dynamic nature of our experiments — that is, the fact that motions exhibited by patients when undergoing tests have a high variability over the time, as well as over different tests and/or different patients. The Short-Time FT (STFT) enables the division of the signal into shorter, equal-length segments, with the FT computed separately on each segment. However, using a fixed resolution is a limitation of STFT since a wide window gives better frequency resolution but poor time resolution (and vice-versa). This emphasizes the necessity for a method that can achieve a more balanced resolution between frequency and time. The Continuous Wavelet Transform (CWT) introduced by Daubechies et al. [39] provides such a balance. Specifically, using wavelets of different scales and positions offers a flexible depiction of signal dynamics with improved time–frequency resolution. Formally put, the CWT of a signal $x(t)$ at scale s and translation u is given by

$$CWT_x(s,u) = \int_{-\infty}^{\infty} x(t)\overline{\psi}\left(\frac{t-u}{s}\right) dt, \tag{2}$$

where $\overline{\psi}$ denotes the complex conjugate of the mother wavelet.

In this work, we used the Synchrosqueezing Wavelet Transform (SWT), a time–frequency reassignment (TFR) method for instantaneous frequency estimation [40]. SWT enhances the Continuous Wavelet Transform (CWT) by adjusting its coefficients based on instantaneous frequencies, producing a sharper, sparser, and more noise-robust time–frequency representation. This reassignment technique concentrates signal energy into specific time–frequency regions, making the resulting features clearer and easier to interpret. Fig. 2 shows a visual representation of the SWT, also known as a spectrogram, where the x-axis represents time, the y-axis represents frequency, and colors indicate frequency magnitude (black = low, red = high).

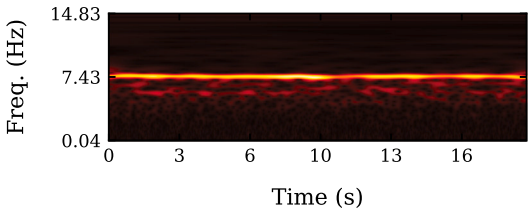

Fig. 2. Spectrogram of the accelerometer X-axis channel of a patient diagnosed with ET.

3.3. Data concatenation and flattening

Spectrograms capture the dynamics of a sensor signal $x_i^j(t)$ over a given time interval $[t_k^{start}, t_k^{end}]$. Analyzing such data in spectrogram form is difficult. Consider, for example, comparing two spectrograms, *e.g.*, from different patients. Small differences in the patients' speeds of performing the (same) task can lead to complex stretching and compression of the time axis. Also, a patient's behavior is described by *multiple* sensors. We need a way to compare the aggregated measurements over patients in terms of their dynamic patterns while allowing for (small) local differences.

We address the above issues as follows. First, we reduce a spectrogram to a set of measurements using a sliding window approach. Each window w_i starts at time $t_k^{start} + i \cdot t_s$ and has a width of t_w seconds. Low t_w values yield a higher temporal resolution, thus allow detecting rapid signal changes; larger t_w values highlight broader frequency components but may sacrifice small temporal details. The stride t_s gives the overlap of successive windows, acting as a low-pass filter. We set $t_s = t_w = 1$ second. We 'flatten' all values in a window w_i to a vector $\mathbf{v}_i$ with dimensions that vary according to the tasks and sensors selected. This way, the entire spectrogram is reduced to a set of (consecutive) high-dimensional vectors $\mathbf{v}_i$, one per window w_i.

As our data has multiple signals x_i^j (from various sensors and/or sensor channels), we concatenate the high-dimensional vectors $\mathbf{v}_{i,k}^j$ resulting from the sliding windows w_k of each signal in a single output vector $\mathbf{v}_k = (\mathbf{v}_{i,k}^j)$. The vectors $\mathbf{v}_k$ are the input of our projection-based analysis described next.

3.4. Projecting time-dependent data

We analyze the high-dimensional vectors $\mathbf{v}_k$ resulting from the sliding-window spectrograms (Section 3.3) using dimensionality reduction. For this, we use PCA, which was shown to balance well visual quality and temporal stability [33]. In detail, given the sets P_p containing all measurements done for a patient p for a set of tasks (Section 3.1), we compute for each such set the projections of the vectors $\mathbf{v}_k$ for all the measurements in P_p. Next, we connect these projections using Akima spline interpolation [41]. This way, the behavior of each patient p yields a 'trajectory' in the 2D projection space. Comparing such trajectories across patients (or, if desired, tasks) allows further analysis of the factors differentiating disorders.

Fig. 3 shows the projected trajectories of a HV, an ET, and a CM patient, all performing the same task, color-coded by condition. We clearly see three different movement patterns: HV (black) has very low variability, essentially a dot — that is, there was almost no change over time of the monitored signals. In contrast, ET (red) and CM (blue) show more variability, in line with the small-scale involuntary movements associated with these disorders. Notably, the red and blue trajectories do not intersect. This means that the information captured by the projections can be used to *discriminate* between these two disorders. We show next that this discrimination is indeed present.

Table 1
Task descriptions.

Task ID	Task description
1	Arms at rest, palms upward
2	Arms stretched in front, wrist straight, palms upward
5	Draw a spiral with dominant hand*
11	Arms stretched in front, wrist flexion, palms downward
12	Arms stretched in front, wrist straight, palms downward
15	Arms stretched in front, wrist straight, palms downward while counting backward
17	Arms stretched in front, wrist extension, palms downward
[19–20]	Arms stretched in front, subsequently touching every finger with the thumb, palms downward*
21	Arms sideways, elbow flexion, hands pointing towards each other in front of the chest
[22–23]	Arms stretched to the side, move fingertip to the nose (tip-to-nose test)*
[24–25]	Holding one finger in front of the nose*
[30–31]	Drinking from a cup*
[35–36]	Arms stretched in front, wrist straight, palms downward, close and open a hand on frequency of metronome *
	*Tasks performed with right and left hand.

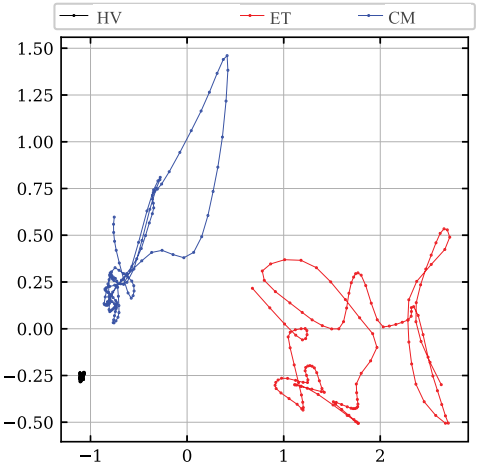

Fig. 3. Projection results for patients in the HV, ET, and CM classes.

Table 2
Device placements.

Task ID	Task description
[1–2]	Triceps brachii*
[3–4]	Biceps brachii*
[5–6]	Extensors forearm*
[7–8]	Flexors forearm*
[9–10]	Proximal interphalangeal joint index finger*
[11–12]	Inside of the forearm near the wrist*
[13–14]	Back of the hand*
[15–16]	Jaw*
	*right and left respectively

Table 3
Number of sensors and their combinations.

Sensor name	Number of devices	Axis combinations	Total combinations
ACC	6	7	42
IMU_ACC	10	7	70
IMU_ORR	10	7	70
Grand total			182

4. Evaluating the proposed workflow

We used the NEMO dataset [2] to validate our proposed workflow. We next describe the dataset and experimental analysis, followed by visual and cluster analysis results.

4.1. The NEMO dataset

The NEMO dataset consists of data collected from subjects performing specific tasks while wearing sensors placed at multiple body locations. The selection of clinical tasks was curated with movement disorder specialists to identify movements where disorder-related symptoms are most likely to appear. The 18 most relevant tasks are detailed in Table 1.

Data acquisition included electromyography (EMG), accelerometry (ACC and IMU_ACC), gyroscope (IMU_ORR) and 2D/3D video recording. Sensors were positioned bilaterally on the arms and neck to capture muscle activity and 3D accelerometry, recording EMG signals from specific muscles and accelerometry data from the forearm, hand, and jaw regions. For our analysis next, we excluded EMG and 2D/3D video data due to (a) inconsistent availability across subjects and tasks; and (b) the increased complexity associated with their analysis.

Our analyzed data consists of three types of *sensors* (IMU_ACC, IMU_ORR, and ACC), each with 10, 10, and 6 devices at different *placements* on the subject (Table 2). Each sensor collects data along the X, Y, and Z axes, leading to 7 combinations of axes (X, Y, Z, XY, XZ, YZ, and XYZ).

The NEMO dataset covers $P = 46$ participants (24 HV, 12 CM, and 12 ET) performing $T = 18$ tasks. For a single task, we have a total of 182 combinations of data from different sensors and devices (see Table 3). Multiplying these 182 combinations by 18 tasks yields a total of $C = 3276$ task-and-sensor combinations. Running each such combination over P participants, we get a dataset D having a total of $PC \times P = 150696$ trajectories projected in 2D space.

4.2. Research questions and experimental analysis

Recall our aim to find which of our measurement combinations (tasks, sensor types, sensor axis combinations, and sensor placements) are most effective in discriminating the three conditions ET, CM, and HV. More specifically, we aim to answer the following research questions (RQs):

- **RQ1:** Which **task** provides the most effective class separation?
- **RQ2:** Which **sensor** type (IMU_ACC, IMU_ORR, and ACC) provides the most effective class separation across tasks?
- **RQ3:** What combination of sensor **axes** (X, Y, Z, XY, XZ, YZ and XYZ) provides the most effective class separation?
- **RQ4:** Which device **placements** provide the most effective class separation across tasks?

To answer these questions, we proceed as follows. We consider all three class pairs HV *vs.* ET, HV *vs.* CM, and ET *vs.* CM. We next denote such a pair as (A, B). For each pair, we consider the two trajectory sets T_A and T_B from the full dataset D. These can be, in turn, split into C subsets $T_A^1 \dots T_A^C$ and $T_B^1 \dots T_B^C$ based on the above-mentioned task-and-sensor combinations. For each pair (T_A^i, T_B^i), we measure how *separable*

Table 4
Top 20 ARI results for the three class comparisons.

	HV vs. ET				HV vs. CM				ET vs. CM			
	Task	Sensor	Device_axes	ARI	Task	Sensor	Device_axes	ARI	Task	Sensor	Device_axes	ARI
1	31	IMU_ACC	['8_y']	0.7823	2	IMU_ACC	['4_z']	1.0000	23	IMU_ACC	['1_y']	0.6529
2	31	ACC	['12_x']	0.6853	2	IMU_ORR	['2_x', '2_z']	1.0000	2	IMU_ORR	['4_x', '4_y', '4_z']	0.5072
3	24	IMU_ORR	['7_x', '7_z']	0.6800	2	IMU_ACC	['4_x']	0.8889	2	IMU_ORR	['4_x', '4_z']	0.5072
4	24	ACC	['11_x']	0.6800	2	IMU_ACC	['3_x', '3_z']	0.8889	2	IMU_ORR	['4_y', '4_z']	0.5072
5	24	IMU_ORR	['7_x']	0.6800	2	IMU_ORR	['2_y', '2_z']	0.8889	2	IMU_ORR	['6_y', '6_z']	0.5072
6	31	IMU_ORR	['8_y', '8_z']	0.6800	2	IMU_ORR	['2_x']	0.8889	2	IMU_ORR	['8_x', '8_y', '8_z']	0.5072
7	31	IMU_ACC	['8_x']	0.6800	11	IMU_ACC	['3_x', '3_y', '3_z']	0.8889	2	IMU_ORR	['8_y', '8_z']	0.5072
8	35	IMU_ACC	['3_z']	0.6000	2	IMU_ACC	['2_x', '2_y']	0.7855	23	ACC	['16_x']	0.5056
9	30	IMU_ACC	['1_z']	0.6000	2	IMU_ACC	['2_y']	0.7855	36	IMU_ACC	['10_x', '10_y', '10_z']	0.3793
10	30	IMU_ACC	['3_y', '3_z']	0.5895	11	IMU_ACC	['4_z']	0.7823	2	IMU_ORR	['10_x', '10_y', '10_z']	0.3793
11	2	IMU_ACC	['3_y', '3_z']	0.5895	11	IMU_ORR	['4_x']	0.7823	2	IMU_ORR	['10_x', '10_y']	0.3793
12	31	IMU_ACC	['6_x']	0.5895	11	IMU_ACC	['6_x']	0.7823	2	IMU_ORR	['10_y', '10_z']	0.3793
13	36	IMU_ORR	['5_x', '5_z']	0.5818	11	IMU_ACC	['4_y', '4_z']	0.7823	2	IMU_ORR	['10_y']	0.3793
14	31	IMU_ACC	['4_x', '4_z']	0.5818	11	IMU_ACC	['2_x', '2_y', '2_z']	0.7823	2	ACC	['12_x', '12_z']	0.3793
15	24	IMU_ORR	['1_x', '1_z']	0.5818	11	IMU_ORR	['4_x', '4_y', '4_z']	0.7823	2	ACC	['12_z']	0.3793
16	31	IMU_ACC	['8_z']	0.5818	11	IMU_ORR	['4_x', '4_z']	0.7823	22	IMU_ORR	['2_x', '2_y', '2_z']	0.3793
17	24	IMU_ORR	['5_z']	0.5818	2	IMU_ORR	['5_y']	0.7823	22	IMU_ORR	['2_x', '2_y']	0.3793
18	31	IMU_ACC	['4_x', '4_y', '4_z']	0.5818	11	IMU_ORR	['4_x', '4_y']	0.7823	22	IMU_ACC	['2_y', '2_z']	0.3793
19	31	IMU_ORR	['8_y']	0.5818	11	IMU_ORR	['4_y', '4_z']	0.7823	2	IMU_ORR	['4_x', '4_y']	0.3793
20	31	IMU_ACC	['4_y', '4_z']	0.5818	24	IMU_ACC	['2_x', '2_y']	0.7823	2	IMU_ORR	['4_y']	0.3793

Table 5
Quantitative results for the proposed pipeline. We examine several clustering metrics to assess projection quality. For all metrics, higher values indicate better results. Overall best results are highlighted.

Data dimensions	ARI	Silhouette	Homogeneity	Completeness	V-measure
Experiment 1: Gyroscope, CM VS. HV, Task 2					
$\{X, Y, Z\}$	0.78	0.61	0.68	0.74	0.71
$\{X, Y\}$	0.78	0.58	0.68	0.74	0.71
$\{X, Z\}$	1.00	0.61	1.00	1.00	1.00
$\{Y, Z\}$	0.88	0.64	0.81	0.84	0.82
Experiment 2: Gyroscope, CM VS. HV, Task 11					
$\{X, Y, Z\}$	0.78	0.59	0.68	0.74	0.71
$\{X, Y\}$	0.78	0.58	0.68	0.74	0.71
$\{X, Z\}$	0.78	0.59	0.68	0.74	0.71
$\{Y, Z\}$	0.78	0.57	0.68	0.74	0.71
Experiment 3: Gyroscope, ET VS. HV, Task 2					
$\{X, Y, Z\}$	0.39	0.57	0.31	0.48	0.38
$\{X, Y\}$	0.22	0.56	0.17	0.38	0.24
$\{X, Z\}$	0.67	0.58	0.57	0.66	0.61
$\{Y, Z\}$	0.39	0.56	0.31	0.48	0.38

the trajectories T_A^i are from the trajectories T_B^i if we do *not* know the actual labels A and B, *i.e.*, using only the projected trajectory data. If separability is high, the tasks-and-sensor combination i provides an effective way to discriminate condition A from B.

We measure this separability by K-means clustering (K = 2) the *joint* set of trajectories $T_A^i \cup T_B^i$ – that is, only the measured trajectories without any labels A or B – and comparing the obtained two clusters with the ground truth cluster pair (T_A^i, T_B^i) via the Adjusted Rand Index (ARI) metric [42]. ARI measures how similar two clusterings are by comparing how sample pairs are grouped in both predicted and true clusterings. ARI ranges from −1 to 1, where 1 means perfect agreement, while values near 0 or negative suggest random or poor clustering. In other words, if ARI is close to 1, the set $T_A^i \cup T_B^i$ can be *easily* separated into the sets T_A^i and T_B^i, *i.e.*, the respective trajectories can be used to discriminate condition A from condition B. In general, ARI does not have a universal threshold for defining "good" cluster assignments because it depends on the data, but values above 0.25, 0.5, 0.65, and 0.8, respectively, suggest weak, moderate, good, and excellent clustering.

We compute all $C = 3276$ ARI values for all task-and-sensor combinations. We present in Fig. 4 a series of 12 violin plots that illustrate the statistical distribution for all combinations with ARI above 0.25. We divide the plots into three parts using the class comparisons (HV *vs.* ET, HV *vs.* CM, and ET *vs.* CM). Each group displays statistical distributions for task, sensor, axis, and device, sorted from left to right by the mean ARI. From the total of 3276 combinations analyzed, the pairwise comparisons yielded 357 (HV vs. ET), 1273 (HV vs. CM), and 72 (ET vs. CM) combinations with ARI above 0.25.

We next focus on only the highest 20 ones since these are the *best* options (from all existing combinations) to predict underlying conditions based on the separability of the measured trajectories. Importantly, we note that a poor trajectory separability does *not* formally imply that the underlying measurements cannot be used, in some other way, to predict said conditions, *e.g.* by more complex manipulations such as machine learning methods [3]. Yet, the key hypothesis we aim to study in this paper is that such predictions can be given *only by trajectories computed by dimensionality reduction* – as such, we focus solely on using just trajectory data.

Fig. 5 shows histograms highlighting the 20 top-scoring tasks, sensors, devices, and axes for each comparison (HV *vs.* ET, HV *vs.* CM, and ET *vs.* CM). Each row corresponds to one of these three different comparison pairs. Columns show the tasks, sensors, devices, and axes combinations (from left to right). To complement the analysis, we detail the top 20 highest ARI results for each of the three class comparisons in Table 4 using the same color mapping as in Fig. 5. We analyze these results next.

Fig. 5 (first column) highlights the **most effective tasks for distinguishing between classes (RQ2)**. These are tasks 31 and 24 for HV *vs.* ET, tasks 11 and 2 for HV *vs.* CM, and tasks 2, 22, and 23 for ET *vs.* CM. A closer examination of Table 4 reveals that task 2 (arms stretched

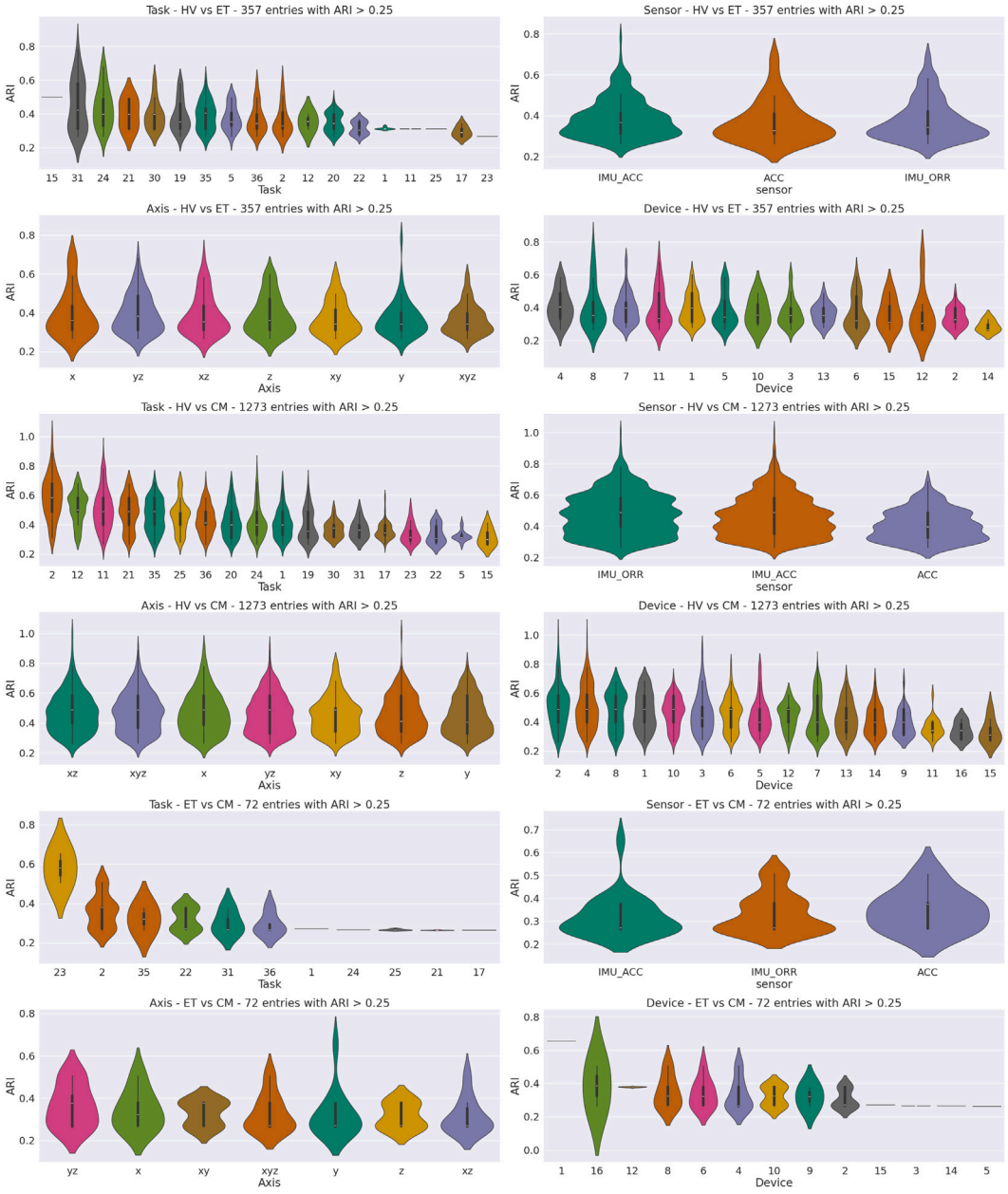

Fig. 4. Statistical distributions of all combinations with ARI above 0.25 were analyzed, organized by task, sensor, device, and axis for each class comparison. Each plot is sorted from left to right by mean ARI value. Out of the initial 3276 combinations, the comparisons HV vs. ET, HV vs. CM, and ET vs. CM yielded 357, 1273, and 72 combinations with ARI exceeding 0.25, respectively.

in front, wrists straight, palms upward) ranks highest for both HV-CM and ET-CM, indicating its strong relevance in distinguishing CM from the other two classes. Separately, task 23 appears twice in the top 10 for ET-CM, suggesting it may be more significant than task 22, which occurs three times but in lower-ranking positions.

Fig. 5 (second column) shows that the IMU_ACC sensor **better separates HV from other classes**, achieving the top results across all three comparisons with 11 out of the 20 best outcomes. In contrast, the gyroscope (IMU_ORR) has 14 out of 20 top results in the **ET *vs.* CM discrimination**. These findings answer our RQ3. Interestingly, the gyroscope was not used in earlier work [3] that aimed to construct a machine learning model for predicting ET *vs.* CM. Our analysis shows the added value of this additional sensor.

Fig. 5 (third column) shows the **top-scoring device placements** and answers RQ5. Device 4 (positioned on the left biceps brachii) ranks highest in HV-CM and ET-CM comparisons and ties for second in HV-ET, highlighting its significance in distinguishing CM from other classes. Similarly, device 2 (placed on the left triceps brachii) is also important in separating CM from other groups. Meanwhile, device 8 (located on the left flexor forearm) ranks highest in separating ET from other classes (first in HV-ET and second in ET-CM). The most informative devices are on the left side of the subject — an interesting finding which may warrant further investigation.

Finally, Fig. 5 (fourth column) shows the **top-scoring axes combinations** and thereby answers RQ4. There is no clear winner here — there is no sensor axis combination that consistently achieves better separation of the trajectories of the studied classes (HV-ET, HV-CM, and ET-CM). Yet, we observe that the yz axes are the most informative in separating the ET-CM classes, having yz, xyz, y, and z at the top four locations in the histogram. Looking at Table 4, we observe that these combinations also correspond to using the IMU_ORR gyroscope sensor.

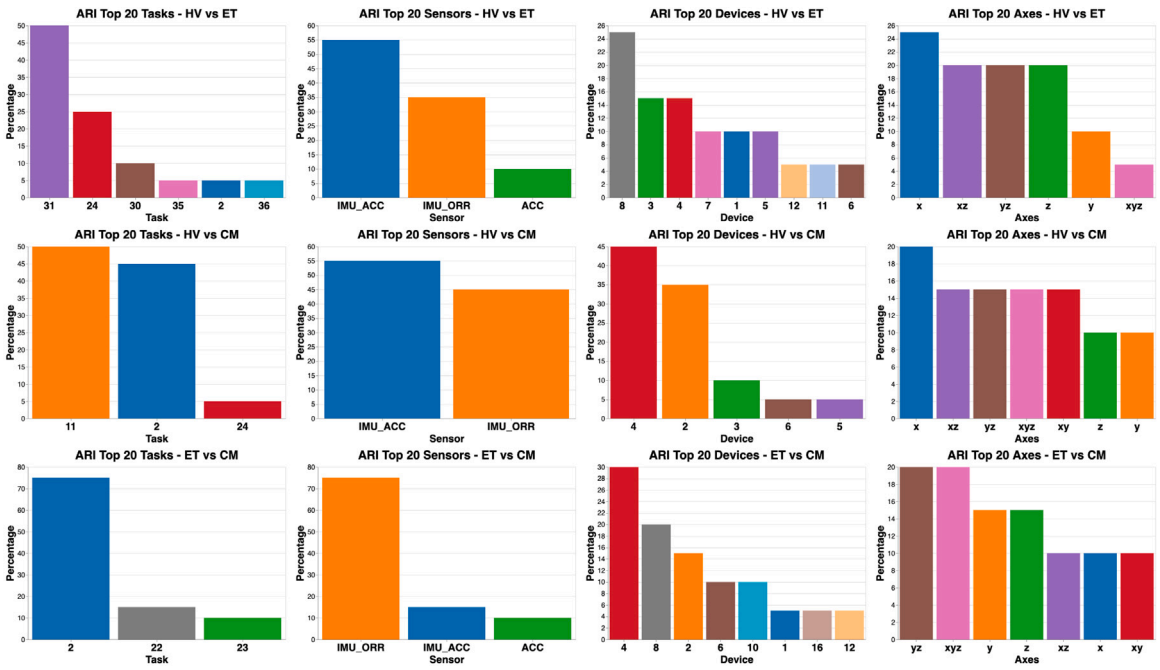

Fig. 5. Histograms for the top 20 tasks, sensors, devices, and axes using the ARI results for each class comparison.

4.3. Visual analysis

The results of our analysis in Section 4.2 effectively narrow down the space of task-and-sensor combinations to evaluate from an initial set of $C = 3276$ combinations to the handful of settings shown in Fig. 5. We next select a subset of these combinations to visually analyze the corresponding projected trajectories. The goal of these visual analyses is to refine the aggregated findings based on ARI scores presented so far to better understand *how* trajectories given by top-scoring task-and-sensor combinations actually differ between conditions.

Case 1: CM *vs.* HV, acceleration sensor: Our first analysis focuses on CM, a disorder characterized by recurrent, often non-rhythmic, abrupt jerks resembling short shock-like movements resulting from involuntary muscle contractions or relaxations [43]. We consider task 2 (arms stretched in front, wrist straight, palms upward) and the readings from the Z-axis accelerometer sensor on the subject's left biceps brachii. Fig. 6 shows the behavioral patterns of 12 individuals diagnosed with CM compared to 24 HV for the above task and sensor. Two distinct clusters appear in the projection:

- HV subjects (black) show very similar trajectories with low variability — see the cluster left in the projection.
- CM patients (blue) show a broader spread to the right of the projection. Their trajectories have a wider dispersion, indicating a diverse range of behaviors compared to HVs.

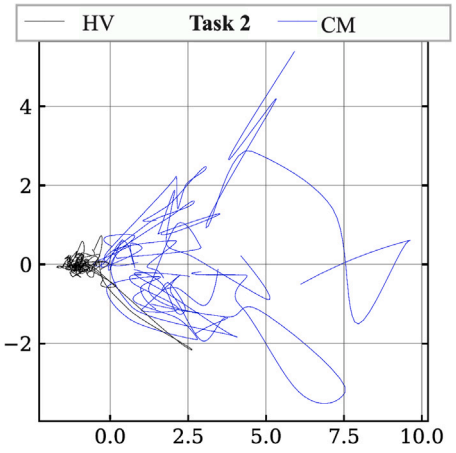

Fig. 6. Projection results for 12 CM, 24 HV, Z-axis accelerometer, task 2 (arms stretched in front, wrist straight, palms upward).

Case 2: CM *vs.* HV, sensor combinations: Fig. 7 performs a similar analysis, but this time using four different combinations of the X, Y, and Z axes from the IMU_ORR on the subject's left triceps during the execution of tasks 2 and 11. On top of Fig. 7 (task 2), all sensor combinations show the HV subjects (black) tightly grouped, indicating a consistent and controlled motion. In contrast, the CM patient trajectories (blue) are wider, indicating more erratic and varied movements.

The bottom row of Fig. 7 repeats this analysis but now using task 11 (arms stretched in front, wrist flexion, palms downward) and a gyroscope sensor on the subject's left biceps brachii. We now see that the trajectories of HV subjects are more dispersed and occasionally overlap with those of CM patients. Still, CM patients continue to display a broader dispersion, highlighting the distinct movement patterns associated with this disorder. As a result, we infer that task 2 and the left triceps gyroscope sensor perform better in separating CM from HV than task 11 and the left biceps gyroscope sensor.

Case 3: ET *vs.* HV, gyroscope: ET manifests as involuntary, rhythmic oscillations within a body region around a joint axis [43]. Here, we consider the same 24 HV and 12 ET patients as before, as well as task 24 (holding one finger in front of the nose, right hand). We use readings from the Z-axis gyroscope sensor on the subject's right flexor forearm. The projected data (Fig. 8) shows two separated clusters, much like Fig. 6:

- Most HV subjects exhibit closely aligned movements, with their trajectories near other HV subjects. Yet, a few subjects show trajectories that are close to ET trajectories.
- ET patients show a broader spread across the projection, indicating a more varied range of trajectories than HV.

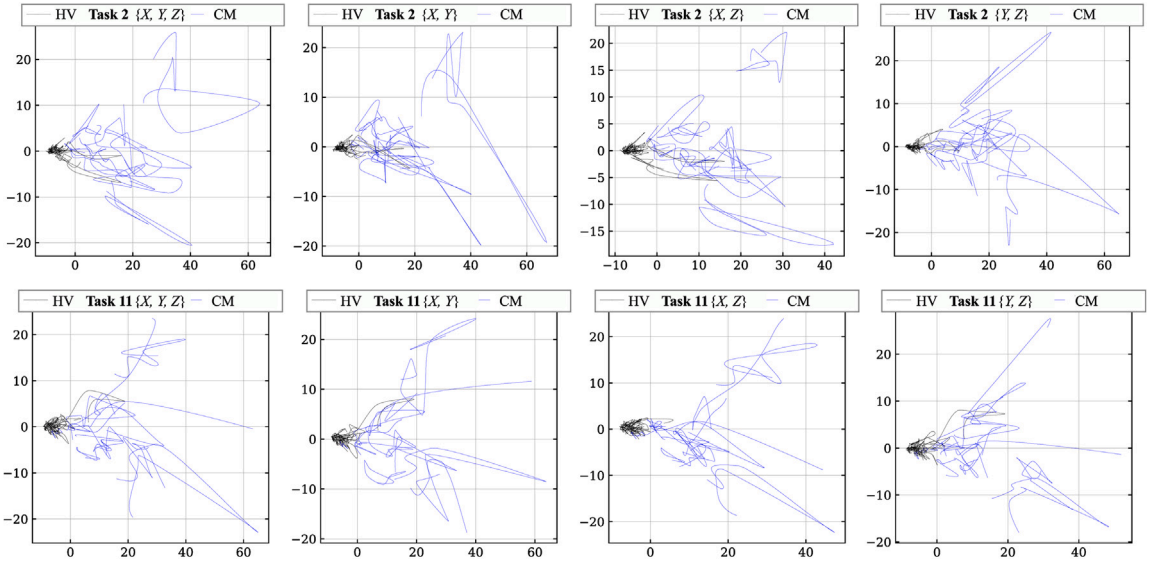

Fig. 7. HV vs. MC comparison: projection results for different combinations of gyroscope sensor data for task 2 (above) and task 11 (below).

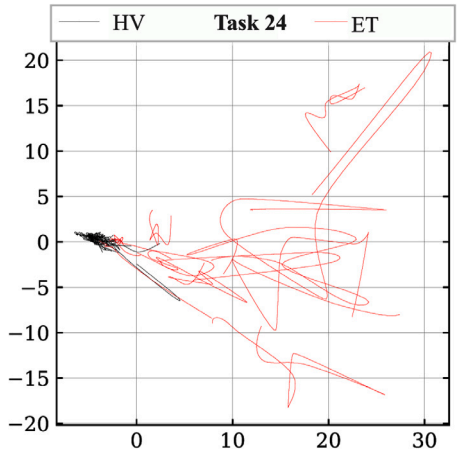

Fig. 8. Projection results for 12 ET and 24 HV, Z-axis gyroscope for task 24 (holding one finger in front of the nose, right hand).

Case 4: ET *vs.* HV, sensor combinations: Fig. 9 studies different X, Y, and Z channel combinations of the gyroscope sensor used earlier in Fig. 8. As in our earlier analysis of sensor combinations (Fig. 7), we see that all the studied combinations lead to HV subjects maintain closely aligned trajectories, while ET patient trajectories exhibit more spread over the projection.

Case 5: ET *vs.* HV, acceleration sensor: Fig. 10 compares the HV and ET subjects using the Z-axis accelerometer sensor for task 12 (arms stretched in front, wrist straight, palms downward). As in the other examples so far, HV subject trajectories cluster well and are largely separated from the ET patient trajectories. An initial overlap in movement trajectories between ET patients and HV individuals is noted. As the movement progresses, ET patients exhibit a more comprehensive range of movement trajectories. This pattern suggests that although the initial movements appear similar in both groups, patients with ETs present more significant variability and dispersion.

Case 6: CM *vs.* ET: Our final example considers the more challenging separation of the two disorders. Fig. 11 shows the projection computed using the gyroscope sensor data (all X, Y, and Z axes) from the left biceps brachii of 11 patients with each condition. CM patients (blue) show a greater diversity of movement patterns, suggesting higher variability. In contrast, ET patients show more similar behavior. Despite some overlap, the trajectories for both classes still show noticeable differences.

Task and sensor suitability for disorder prediction: The NEMO dataset considers a wide set of tasks designed to assess the severity and characteristics of different motor disorders. Yet, as the above examples show, not all tasks and sensors are identical in how well they separate trajectories of subjects having different conditions.

After having analyzed all tasks, we conclude that *static* tasks – that ask subjects to place hands in a given fixed position – can identify CM patients better than tasks involving *dynamic* movements – that ask subjects to perform a specific motion. More precisely, tasks 2 and 11 yielded better results when using the gyroscope sensor placed on subjects' left triceps, left biceps, left extensors forearm, and left flexors forearm along the Z, X, and combined X-Z axes. This refines our current answers to questions **RQ1-4** in the context of CM *vs.* HV.

The better separation of CM patients observed for static tasks can be attributed to several factors. Firstly, static tasks minimize confounding variables associated with dynamic movements, such as speed, amplitude, and coordination variations. By focusing on maintaining a specific posture or position, static tasks provide a controlled environment for assessing subtle differences in motor function and coordination. Moreover, static tasks have the potential to accentuate underlying motor abnormalities and impairments, particularly in individuals who struggle to maintain the prescribed posture or exhibit involuntary movements. The sustained nature of static tasks allows for prolonged observation of motor behavior, facilitating the identification of subtle nuances and abnormalities that may not be apparent during dynamic movements.

For ET patients, our findings showed that the *dynamic* tasks 24, 31 (drinking with left hand), and 20 (finger tapping with right hand) yielded better results when using both gyroscope and accelerometer sensors placed on subjects' flexors forearm right, triceps brachii right, extensors forearm right, flexors forearm left, and biceps brachii left along the Z, X, and combined X-Z axes. This refined our current answers to questions **RQ1-4** in the context of ET *vs.* HV.

Our findings align with the known fact that ET primarily affects the forearms and hands, becoming more pronounced during movement. This consistency with the symptomatic profile of ET supports the relevance of our approach.

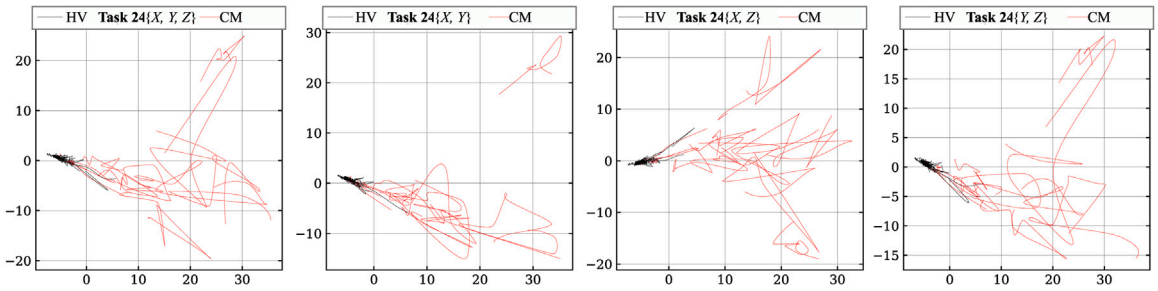

Fig. 9. Projection results for different combinations of gyroscope sensor data for the task 24 ("holding one finger in front of the nose, right hand"). All visualizations use the same subset of subjects (11 Essential Tremor, 24 Healthy control).

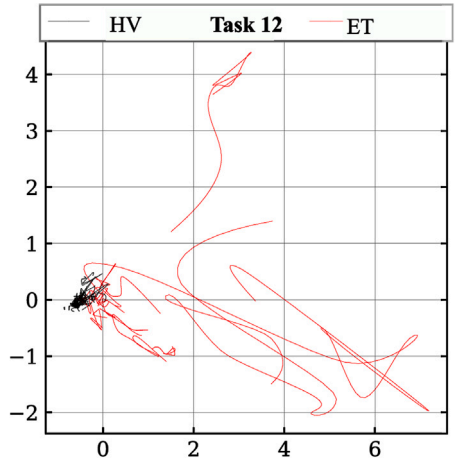

Fig. 10. Projection results for 12 ET, 24 HV, Z-axis accelerometer, task 12 ("pronated outstretched arms and hands").

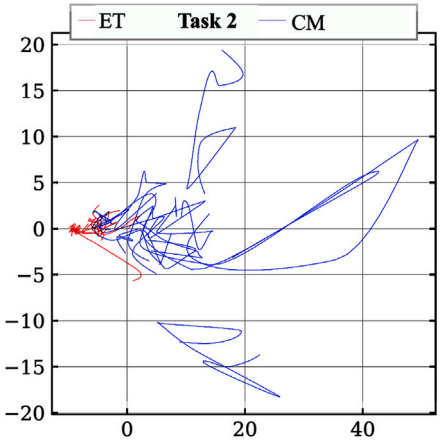

Fig. 11. Projection results for $\{X, Y, Z\}$- axes gyroscope, 12 ET and 12 CM patients, task 2 ("pronated outstretched arms and extended wrists").

4.4. Cluster analysis

The visual analysis presented in Section 4.3 showed how we can use projections to methodically examine tasks and sensor combinations and find the best ones for visually separating subjects having a specific condition. While this helps in 'culling down' the large amount of available data to a smaller, more manageable subset, it does not answer whether information present in the projections can be used to directly *classify* subjects according to their condition.

A first step towards such an automatic classification is to study the characteristics of trajectories of a given class in the projection. If we can find objective metrics that separate classes, we have evidence that projections contain the needed information to perform automatic classification further. To study this, we clustered the trajectories in a projection using K-means (with $K = 2$) and next computed, in addition to the ARI index, the following quality metrics of the resulting clusters:

- Silhouette Coefficient [44]: Measures how similar an object is to its cluster compared to others. This metric has a $[-1, 1]$ range, with values near 1 indicating dense, well-separated clusters.
- Homogeneity, completeness, and V-measure [45]: Homogeneity evaluates if a cluster exclusively contains members of a single class. In contrast, completeness assesses if all class members are assigned to the same cluster. The V-measure represents the harmonic mean of homogeneity and completeness. Scores range from $[0, 1]$, with 1 indicating optimal clustering performance.

We next present our cluster quality results in the same order as the visual analyses described in Section 4.3 – see Figs. 6–11.

Case 1: CM *vs.* HV: Fig. 6 already showed two distinct clusters containing the CM (blue) and HV (black) subjects. The clustering metrics confirm this: A Silhouette score of 0.59 indicates a strong cluster separation. The ARI, homogeneity, completeness, and V-measure achieved a perfect score of 1.0. This performance across metrics tells that each computed cluster corresponds uniquely to a single class, with all class members accurately grouped without errors or misplacements.

Case 2: CM *vs.* HV, sensor combinations: Table 5 (Experiment 1) summarizes the clustering quality metrics for the different sensor data combinations shown in Fig. 7. These results demonstrate that the combination of dimensions X and Z achieved the best results, with perfect scores in ARI, Homogeneity, Completeness, and V-measure, reinforcing the initial visual observations noted in the projection.

Fig. 7 (bottom row) shows good class separation but also some overlaps between the trajectories of HV and CM subjects. This highlights the intricate challenge of distinguishing their movement patterns effectively. The presence of overlap accentuates the importance of choosing optimal sensor data combinations. These findings are confirmed by the clustering metrics in Table 5 (Experiment 2).

Case 3: ET *vs.* HV: Fig. 8 shows a clear visual separation of these two classes. Yet, a few HV trajectories are close to ET ones, suggesting less perfect separation. Clustering analysis confirms these findings, with a Silhouette score of 0.58 telling a good separation between clusters. The ARI of 0.68 implies a high consistency level between the formed clusters and ground-truth ones. A homogeneity score of 0.57 indicates relatively pure clusters, mainly containing data points from a single class. In contrast, the completeness score of 0.66 reveals that most data points from a single true class are assigned to the same cluster. A V-measure of 0.61 reflects a well-balanced trade-off between homogeneity and completeness, emphasizing the discernibility of movement data between HV and ET patients, although with some overlap observed during visual inspection.

Case 4: ET *vs.* HV, sensor combinations: Table 5 (Experiment 3) shows the clustering metrics for the sensor combinations presented

earlier in Fig. 9. The values indicate that it is harder to separate ET from HV as compared to CM from HV.

Case 5: CM *vs.* ET: The projection in Fig. 11 yielded a Silhouette score of 0.36, telling moderate separation between clusters. This is further evidenced by an ARI of approximately 0.51, and moderate scores for homogeneity (0.52) and completeness (0.55), which suggest that clusters, although relatively uniform, do not perfectly represent all instances within a class. A V-measure of about 0.54 reinforces these observations.

General findings: Our analysis used accelerometer and gyroscope data from motion sensors across various tasks. We observed that the gyroscope outperforms the accelerometer in terms of clustering quality – a finding which goes beyond state-of-the-art work in the same area using only accelerometer data [3]. Within both sensor types, the Z and X axes improved clustering quality. Combining data from these axes increased differentiation between classes. We conclude that using both these sensor channels in the analysis adds clear value.

Separately, we observed that K-means clustering struggled to effectively group trajectories with varying sizes, densities, and outliers. This difficulty becomes more pronounced when clusters share overlapping characteristics. In our case, this issue may stem from trajectory overlaps between different classes, particularly during early stages of certain tasks.

5. Conclusion

We presented a workflow for analyzing and exploring hyperkinetic movement disorders based on the application of dimensionality reduction methods on multidimensional sensor data obtained from the NEMO study. Our findings highlight the effectiveness of projection methods in facilitating motion data analysis, providing valuable insights into the complexity of these conditions. In particular, our workflow showed how analysts could methodically go through the large space of combinations of sensors, sensor channels, and tasks to select subsets thereof that perform best in terms of separating subjects having one of the three analyzed conditions (CM, ET, and HV).

While our results show that the available motion data shows significant separation of the studied conditions, our current results also indicate that performing a *fully automatic* classification is still hard. Our findings indicate areas requiring further refinement, particularly in distinguishing between patients with distinct diagnoses, where achieving clear separation remains challenging. For this purpose, in future work, we intend to explore other dynamic projection techniques, other extractors from the trajectories, and add EMG data in the workflow.

Another point of future work is to develop a visual analytics interface that supports interactive exploration of the analysis results we developed. The current limitation to this approach lies on the confidentiality requirements of the NEMO dataset that pose limitations to the deployment of such tool. Once these issues are resolved, we plan to conduct a user study with medical experts to further validate our proposal.

CRediT authorship contribution statement

Andressa Silva da Silva: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Eduardo F. Ribeiro:** Software, Methodology, Investigation, Conceptualization. **Jelle R. Dalenberg:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Alexandru C. Telea:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Marina A.J. Tijssen:** Supervision, Methodology, Investigation, Conceptualization. **João Luiz Dihl Comba:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT in order to check for grammar mistakes. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors report no declarations of interest.

Acknowledgments

This work was partially financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and CNPq.

Data availability

The data that has been used is confidential.

References

[1] Gershanik OS. Movement disorders. In: Della Sala S, editor. Encyclopedia of behavioral neuroscience (second edition). second ed.. Oxford: Elsevier; 2022, p. 148–59.
[2] van der Stouwe AM, Tuitert I, Giotis I, Calon J, Gannamani R, Dalenberg JR, van der Veen S, Klamer MR, Telea AC, Tijssen MA. Next move in movement disorders (NEMO): Developing a computer-aided classification tool for hyperkinetic movement disorders. BMJ Open 2021;11(10):e055068.
[3] van den Brandhof EL, Tuitert I, Madelein van der Stouwe A, Elting JW, Dalenberg JR, Svorenova T, Klamer MR, Marapin RS, Biehl M, Tijssen MA. Explainable machine learning for movement disorders - classification of tremor and myoclonus. Comput Biol Med 2025;192:110180.
[4] Dalenberg JR, Peretti DE, Marapin LR, van der Stouwe AMM, Renken RJ, Tijssen MAJ. Next move in movement disorders: Neuroimaging protocols for hyperkinetic movement disorders. Front Hum Neurosci 2024;18.
[5] Dalenberg J, Peretti D, Marapin L, Jonker S, Timmers E, Stouwe AV, Tijssen M. The next move in movement disorders: Differentiating between hyperkinetic movement disorders using [18F] FDG PET. 2024, https://www.mdsabstracts.org/abstract/the-next-move-in-movement-disorders-differentiating-between-hyperkinetic-movement-disorders-using-18f-fdg-pet/. (Accessed 27 February 2025).
[6] Chen K-HS, Chen R. Principles of electrophysiological assessments for movement disorders. J Mov Disord 2020;13(1):27.
[7] van der Veen S, Klamer M, Elting J, Koelman J, Van der Stouwe A, Tijssen M. The diagnostic value of clinical neurophysiology in hyperkinetic movement disorders: A systematic review. Parkinsonism Rel Disord 2021;89:176–85.
[8] Ahmad N, Ghazilla RAR, Khairi NM, Kasi V. Reviews on various inertial measurement unit (IMU) sensor applications. Int J Signal Process Syst 2013;1(2):256–62.
[9] Timmer J, Lauk M, Deuschl G. Quantitative analysis of tremor time series. Electroencephalogr Clin Neurophysiol/ Electromyogr Mot Control 1996;101(5):461–8.
[10] Issartel J, Marin L, Gaillot P, Bardainne T, Cadopi M. A practical guide to time—frequency analysis in the study of human motor behavior: The contribution of wavelet transform. J Mot Behav 2006;38(2):139–59.
[11] Phinyomark A, Limsakul C, Phukpattaranont P. Application of wavelet analysis in EMG feature extraction for pattern classification. Meas Sci Rev 2011;11(2):45–52.
[12] Nijmeijer S, De Bruijn E, Forbes P, Kamphuis D, Happee R, Koelman J, Tijssen M. EMG coherence and spectral analysis in cervical dystonia: discriminative tools to identify dystonic muscles? J Neurol Sci 2014;347(1–2):167–73.
[13] Kramer G, Van der Stouwe A, Maurits N, Tijssen M, Elting J. Wavelet coherence analysis: A new approach to distinguish organic and functional tremor types. Clin Neurophysiol 2018;129(1):13–20.
[14] Go SA, Coleman-Wood K, Kaufman KR. Frequency analysis of lower extremity electromyography signals for the quantitative diagnosis of dystonia. J Electromyography Kinesiol 2014;24(1):31–6.
[15] Vernier EF. visualization of dynamic multidimensional and hierarchical datasets [Ph.D. thesis], Double Doctorate (University of Groningen and Universidade Federal do Rio Grande do Sul); 2022.
[16] Gao L, Song J, Liu X, Shao J, Liu J, Shao J. Learning in high-dimensional multimedia data: The state of the art. Multimedia Syst 2017;23:303–13.

[17] Liu S, Maljovec D, Wang B, Bremer P-T, Pascucci V. Visualizing high-dimensional data: Advances in the past decade. IEEE TVCG 2016;23(3):1249–68.
[18] Marcilio-Jr WE, Eler DM. Explaining dimensionality reduction results using Shapley values. Expert Syst Appl 2021;178:115020.
[19] Jolliffe IT. Principal component analysis and factor analysis. In: Principal component analysis. Springer; 1986, p. 115–28.
[20] Maaten Lvd, Hinton G. Visualizing data using t-SNE. JMLR 2008;9:2579–605.
[21] McInnes L, Healy J. UMAP: Uniform manifold approximation and projection for dimension reduction. 2018, arXiv:1802.03426.
[22] Espadoto M, Martins RM, Kerren A, Hirata NS, Telea AC. Towards a quantitative survey of dimension reduction techniques. IEEE TVCG 2019.
[23] Nonato L, Aupetit M. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. IEEE TVCG 2018.
[24] Singhal A, Seborg DE. Clustering multivariate time-series data. J Chemom: A J Chemom Soc 2005;19(8):427–38.
[25] Lewandowski M. Advanced non linear dimensionality reduction methods for multidimensional time series : applications to human motion analysis [Ph.D. thesis], Kingston Universit; 2011, Available at https://eprints.kingston.ac.uk/id/eprint/20313/.
[26] Ali M, Alqahtani A, Jones MW, Xie X. Clustering and classification for time series data in visual analytics: A survey. IEEE Access 2019;7:181314–38.
[27] Pealat C, Bouleux G, Cheutet V. Improved time-series clustering with UMAP dimension reduction method. In: 2020 25th international conference on pattern recognition. ICPR, 2021, p. 5658–65.
[28] Brich N, Schulz C, Peter J, Klingert W, Schenk M, Weiskopf D, Krone M. Visual analytics of multivariate intensive care time series data. Comput Graph Forum 2022;41(6):273–86.
[29] Rodriguez-Fernandez V, Montalvo-Garcia D, Piccialli F, Nalepa GJ, Camacho D. DeepVATS: Deep visual analytics for time series. Knowl-Based Syst 2023;277:110793.
[30] Fujiwara T, Chou J-K, Shilpika S, Xu P, Ren L, Ma K-L. An incremental dimensionality reduction method for visualizing streaming multidimensional data. IEEE TVCG 2019;26(1):418–28.
[31] Ali M, Jones MW, Xie X, Williams M. TimeCluster: Dimension reduction applied to temporal data for visual analytics. Vis Comput 2019;35(6–8):1013–26.
[32] Abdullah SS, Rostamzadeh N, Sedig K, Garg AX, McArthur E. Visual analytics for dimension reduction and cluster analysis of high dimensional electronic health records. In: Informatics. vol. 7, MDPI; 2020, p. 17, 2.
[33] Vernier E, Garcia R, Silva Id, Comba J, Telea A. Quantitative evaluation of time-dependent multidimensional projection techniques. In: Proc. euroVis. 2020.
[34] Munz-Körner T, Weiskopf D. Exploring visual quality of multidimensional time series projections. Vis Inform 2024;8(2):27–42.
[35] Sawada N, Uemura M, Fujishiro I. Multi-dimensional time-series subsequence clustering for visual feature analysis of blazar observation datasets. Astron Comput 2022;41:100663.
[36] L'Yi S, Ko B, Shin D, Cho Y-J, Lee J, Kim B, Seo J. XCluSim: A visual analytics tool for interactively comparing multiple clustering results of bioinformatics data. BMC Bioinformatics 2015;16(11):S5.
[37] Kwon BC, Eysenbach B, Verma J, Ng K, De Filippi C, Stewart WF, Perer A. Clustervision: Visual supervision of unsupervised clustering. IEEE TVCG 2018;24(1):142–51.
[38] Cavallo M, Demiralp C. Clustrophile 2: Guided visual clustering analysis. IEEE Trans Vis Comput Graphics 2019;25(1):267–76.
[39] Daubechies I. The wavelet transform, time-frequency localization and signal analysis. IEEE Trans Inform Theory 1990;36(5):961–1005.
[40] Thakur G, Brevdo E, Fučkar NS, Wu H-T. The synchrosqueezing algorithm for time-varying spectral analysis: Robustness properties and new paleoclimate applications. Signal Process 2013;93(5):1079–94.
[41] Akima H. A new method of interpolation and smooth curve fitting based on local procedures. J ACM 1970;17(4):589–602.
[42] Milligan GW, Cooper MC. A study of the comparability of external criteria for hierarchical cluster analysis. Multivar Behav Res 1986;21(4):441–58.
[43] Sanger TD, Chen D, Fehlings DL, Hallett M, Lang AE, Mink JW, Singer HS, Alter K, Ben-Pazi H, Butler EE, et al. Definition and classification of hyperkinetic movements in childhood. Mov Disorders 2010;25(11):1538–49.
[44] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
[45] Rosenberg A, Hirschberg J. V-measure: A conditional entropy-based external cluster evaluation measure. In: Proceedings of EMNLP-coNLL. 2007, p. 410–20.