International Journal of Medical Informatics 114 (2018) 6–17

Contents lists available at ScienceDirect

International Journal of Medical Informatics

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

Rest tremor quantification based on fuzzy inference systems and wearable sensors

Luis A. Sanchez-Perez[a,b,*], Luis P. Sanchez-Fernandez[b], Adnan Shaout[a], Juan M. Martinez-Hernandez[c], Maria J. Alvarez-Noriega[c]

[a] *Department of Electrical and Computer Engineering, University of Michigan – Dearborn, MI, USA*
[b] *Instituto Politecnico Nacional, Centro de Investigacion en Computacion, Mexico City, Mexico*
[c] *Instituto Politecnico Nacional Escuela Nacional de Medicina y Homeopatia, Mexico City, Mexico*

A R T I C L E I N F O

Keywords:
Rest tremor
Wearable sensors
Fuzzy inference
Tremor quantification
Continuous scale

A B S T R A C T

Background: Currently the most consistent, widely accepted and detailed instrument to rate Parkinson's disease (PD) is the *Movement Disorder Society sponsored Unified Parkinson Disease Rating Scale* (MDS-UPDRS). However, the motor examination is based upon subjective human interpretation trying to capture a snapshot of PD status. Wearable sensors and machine learning have been broadly used to analyze PD motor disorder, but still most ratings and examinations lay outside MDS-UPDRS standards. Moreover, logical connections between features and output ratings are not clear and complex to derive from the model, thus limiting the understanding of the structure in the data.

Methods: Fifty-seven PD patients underwent a full motor examination in accordance to the MDS-UPDRS on twelve different sessions, gathering 123 measurements. Overall, 446 different combinations of limb features correlated to rest tremors amplitude are extracted from gyroscopes, accelerometers, and magnetometers and feed into a fuzzy inference system to yield severity estimations.

Results: A method to perform rest tremor quantification fully adhered to the MDS-UPDRS based on wearable sensors and fuzzy inference system is proposed, which enables a reliable and repeatable assessment while still computing features suggested by clinicians in the scale. This quantification is straightforward and scalable allowing clinicians to improve inference by means of new linguistic statements. In addition, the method is immediately accessible to clinical environments and provides rest tremor amplitude data with respect to the timeline. A better resolution is also achieved in tremors rating by adding a continuous range.

1. Introduction

Tremor is the most frequent initial motor disorder associated with Parkinson's disease (PD) [1]. The Movement Disorder Society (MDS) sponsored Unified Parkinson Disease Rating Scale (UPDRS), henceforward referred as the *scale*, contains several items to rate different tremor categories [2,3]. One of these items evaluates tremors that may appear at any time during the entire examination when some body parts are moving but others are at rest. The MDS has sponsored various revisions [2–5] of the original UPDRS [6] in order to incorporate current scientific knowledge, thenceforward becoming the most widely used PD clinical rating scale [7–14]. However, currently all motor examination items from the *scale* are quantified using low resolution ranges based on subjective observations gathered by the examiner [15]. Recently, wearable sensors have been widely used not only to detect, but also to objectively quantify motor signs in PD patients [16,17]. Most common approaches rely on inertial sensors (accelerometers and gyroscopes) [18–26], digitography [27,28], surface electromyographic [29], force detection surfaces [28], and more sophisticated, complex and costly setups such as video motion analysis systems [30].

Tremors in PD patients have been widely studied using wearable sensors [19,29,31–35] and recent review articles have focused on highlighting advantages and disadvantages given the current progress and limitations [16,36,37]. In this sense, tremors quantification is being performed in two steps: detection and severity rating. Detection has mainly relied on frequency threshold [19,31–33,35] and dynamic classifiers [29,34]. In spite of the high accuracy reported in the latter models some caution must be taken when trying to describe tremors as a time-dependent variable, since human behavior is adaptable to circumstances, flexible, changeable, and tremors may disappear suddenly

* Corresponding author at: University of Michigan – Dearborn, 215G ELB, 4901 Evergreen Rd, Dearborn, MI 48128, USA.
E-mail addresses: alejand@umich.edu (L.A. Sanchez-Perez), lsanchez@cic.ipn.mx (L.P. Sanchez-Fernandez), shaout@umich.edu (A. Shaout), jmmartinezh@ipn.mx (J.M. Martinez-Hernandez), majoalvarez@qimed.mx (M.J. Alvarez-Noriega).

https://doi.org/10.1016/j.ijmedinf.2018.03.002
Received 10 March 2017; Received in revised form 27 January 2018; Accepted 8 March 2018
1386-5056/

even in severe PD patients [16,35]. The severity rating estimation has been performed mainly through linear regression and classification using raw and frequency domain features from accelerometers, gyroscopes and surface electromyographic signals [19,29,31–35]. Despite all previous features are related to tremors amplitude, their relationship inside these models with respect to the output ratings are not clear and complicated to derive, thus limiting the understanding of the structure in the data. Moreover, if the model performs poorly there is no meaningful feedback in terms of what relationships are not accurate. This is currently a significant issue in PD motor disorder analysis based on wearable sensors and machine learning [16]. Additionally, trying to quantify tremors using a linear combination of the inputs must be further analyzed, particularly when all tremors severity ranges used in the *scale* exhibit a logarithmic behavior [38], which can also be inferred by examining limits established in tremor items defined in the same *scale*. On the other hand, tremors quantification using classification requires large datasets to avoid overfitting problems and to represent a reasonable large combination of the inputs from different patients [16]. What is more, although tremor ratings on previous works are expressed using similar categories to those proposed in the *scale* for short time windows, an overall rest tremor rating as that required in the *scale* is never performed due to models constraints. In this respect, both the classification models and the *scale* use a discrete range, thus producing a floor/ceil effect, i.e., even when a particular combination of inputs represents a stage between two severity ratings, still, one of the two must be chosen [39].

Despite the current limitations of the works previously mentioned, tremors severity estimation using wearable sensors remains appealing since it enables objective and accurate quantification of signs. To overcome current limitations, in this work the same motor examination procedure described in the *scale* is performed to collect data from wearable sensors. Also, a useful approach to compute tremors amplitude at any time is provided. Consequently, all features extracted are closely related to the highest amplitude tremors recorded during the entire session as proposed in the *scale*. The main purpose is to quantify exactly what the examiner is looking at when performing a subjective rating by means of observations. In this respect, clinicians knowledge is modelled using a rule-based fuzzy inference system in the same way the entire *scale* is built in, but avoiding sharp boundaries between severity ratings and the floor/ceil effect by adding a continuous range. This approach is primarily focused to route these quantification models towards a clinimetric validation for regulatory approval using the *scale* as the guideline.

2. Review

The motivation of this work is to provide an objective and accurate quantification of rest tremors severity covered by motor examination items in the *scale,* but using wearable sensors, advanced digital signal processing techniques, and fuzzy inference systems. In general, several improvements over the current rating process are behind this paper proposal:

– Avoid the lengthy review of video material used by examiners as the main tool for applying the *scale*, which often discourages its use. Instead, computer ratings can be delivered in real-time.
– Provide consistent features quantification including tremors amplitude using wearable sensors and advanced computer tools, to make any given set of inputs always yield the same corresponding outputs. This is difficult to achieve by subjective human observations.
– Computationally represent existing clinician knowledge gathered in the *scale* using a fuzzy inference system with *if-them* rules, which is a similar idea to the rating process in the *scale*.
– Avoid using discrete rating values (0, 1, 2, 3, or 4) to prevent the floor/ceil effect [39].

Several works have used wearable sensors and computer tools to quantitatively evaluate tremor manifestations in PD. These quantifications include differentiation between essential and parkinsonian tremors [40], detection and monitoring during daily life activities and motor exercises [19,31] and severity estimation [29,32–35]. However, all previous works have any of the following fundamental limitations:

1. Features extracted from physical properties measured by the sensors are often not linked directly to the rating process defined in the *scale*. For instance, the main feature used to rate rest tremors including re-emergent rest tremors is the amplitude measured in centimeters. However, in previous works [19,29,31–35,40], all features are statistical representation of inertial signals energy trying to represent tremors intensity. Therefore, clinicians find hard to reason about the PD patient status in those terms. According to the *scale*, when the maximum tremor observed is between 1 and 3 cm, the rating should be *Mild*. Nonetheless, finding the equivalent rating in terms of inertial signals energy is not straightforward and required supervised machine learning algorithms.

- In this paper, advanced digital signal processing techniques are used to acquire features that allow directly applying the existing knowledge already modeled in the *scale*.

2. Computers output severity ratings are not related to the *scale* terms and guidelines [19,31], so the reasoning is difficult and discourages clinical applicability.

- In this paper, examiners and clinicians can make sense on any quantified input/output since it strictly follows *scale* definitions.

3. Since all relationships between features from inertial signals energy and the final rating are not clear, machine learning algorithms must be used to associate inputs to outputs. The most common and advanced approaches so far are multivariate linear regression [33,35] and classification [19,29,34], where the main goal is to treat each severity rating as a class and determine their mathematical relationships using supervised learning to associate a group of features to a severity rating or class. Three major drawbacks arise from this:

a These techniques do not generally supply explanatory power. In other words, they are “black boxes” that ingest data to produce generalized outputs [16]. Mathematical relationships between inputs and outputs are difficult to understand. Therefore, if the algorithm goes wrong, it is difficult to find out why.
b If the dataset is not large enough, i.e., the selected PD patients do not represent properly all severity ratings, then overfitting problems can occur, and classification techniques will fail.
c All classes used during supervised learning correspond to severity ratings given by examiners still under subjective quantification and reasoning. Every patient was subjectively rated as *Normal*, *Slight*, *Mild*, *Moderate* or *Severe* before applying any supervised machine learning algorithms, thus adding uncertainty to the model.

– In this paper, the existing knowledge is modelled by means of fuzzy inference systems imitating how clinicians perform the rating process through if-them rules but with a consistent, quantitative and adaptable model. For instance, in the *scale*, tremors between 1 and 3 cm correspond to a *Mild* level while tremors between 3 and 10 cm correspond to a *Severe* level. However, given an accurate procedure to quantify tremors amplitude A, it is uncertain how and why 3 cm is the proper threshold to differentiate between *Mild* and *Severe*. This uncertainty can be computationally modelled using fuzzy logic rules such as:

IF A is *Medium* **THEN** *Rating* is *Mild*.
IF A is *High* **THEN** *Rating* is *Severe*.