

Article

Validation and Discussion of Severity Evaluation and Disease Classification Using Tremor Video

Takafumi Hayashida ¹, Takashi Sugiyama ², Katsuya Sakai ², Nobuyuki Ishii ^{2,3}, Hitoshi Mochizuki ^{2,4}
and Thi Thi Zin ^{1,*}

¹ Graduate School of Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki 889-2192, Japan; hl17031@student.miyazaki-u.ac.jp

² Department of Neurology, Faculty of Medicine, University of Miyazaki Hospital, Miyazaki 889-1692, Japan; t_sugiyama@med.miyazaki-u.ac.jp (T.S.); katsuya_sakai@med.miyazaki-u.ac.jp (K.S.); nobuyuki_ishii@med.miyazaki-u.ac.jp (N.I.); hitoshi.mochizuki@gmail.com (H.M.)

³ Department of Neurology, Chiyoda Hospital, Miyazaki 883-0064, Japan

⁴ Department of Neurology, Ebihara General Hospital, Miyazaki 884-0006, Japan

* Correspondence: thithi@cc.miyazaki-u.ac.jp; Tel.: +81-985-58-7411

Abstract: A tremor is a significant symptom of Parkinson's disease, but it can also be a characteristic of essential tremor, thereby hampering even specialists' ability to differentiate between the two. This study proposes a system that leverages a single RGB camera to evaluate tremor severity and support the differential diagnosis of Parkinson's disease and essential tremor. The system captures motor symptoms, performs time–frequency analysis using wavelet transforms, and classifies severity and disease using linear classification models. The results showed an accuracy rate of 0.56 for disease classification and 0.50 for severity classification (with an acceptable accuracy rate of 0.96). The analysis indicated that there was a low level of correlation between disease and each feature and a moderate correlation (about 0.6) between severity and each feature. Based on these results, this study recommends classifying severity with a linear model and disease with a nonlinear model to obtain improved accuracy.

Keywords: Parkinson's disease; essential tremor; tremor quantification; differential diagnosis support; wavelet transformation; linear classification models



Citation: Hayashida, T.; Sugiyama, T.; Sakai, K.; Ishii, N.; Mochizuki, H.; Zin, T.T. Validation and Discussion of Severity Evaluation and Disease Classification Using Tremor Video. *Electronics* **2023**, *12*, 1674. <https://doi.org/10.3390/electronics12071674>

Academic Editor: Maria Evelina Fantacci

Received: 19 February 2023

Revised: 16 March 2023

Accepted: 30 March 2023

Published: 1 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tremor is a disorder that can interfere with daily life, and its incidence is known to be correlated with aging [1]. Parkinson's disease and essential tremor are two conditions that can cause tremors, with each having a disease-specific tremor frequency distribution [2,3]. The severity and progression of Parkinson's disease are typically assessed by a neurologist using the MDS-unified Parkinson's disease rating scale (MDS-UPDRS) [4]. This scale consists of 65 items divided into 4 parts, for which some items are completed by the patient alone (Part I and Part II), while others are interactive evaluation items between the patient and the evaluator (Part I and Part IV). Part III is particularly important, as it allows the evaluator to objectively assess the actual motor symptoms. While the MDS-UPDRS is widely used as a clinical evaluation scale for Parkinson's disease, it can be challenging to interpret at the stage when a disease cannot yet be differentiated. Therefore, neurologists often rely on empirical methods to differentiate between diseases based on neurological examinations and interviews. However, this can be difficult in the early stages of Parkinson's disease, when motor symptoms may not be noticeable and differentiating it from essential tremor can be challenging. Thus, there is a need to develop an evidence-based system that can efficiently differentiate Parkinson's disease from essential tremor and estimate the severity of tremor in the early stages of the onset of these conditions.

The purpose of this study is to develop a support system for estimating the severity of tremor and differentiating between Parkinson's disease and essential tremor using a smartphone application. To achieve this goal, we propose an analysis of a time series obtained by the non-contact acquisition method of motor symptoms using a single RGB camera [5,6]. This system may significantly contribute to the accuracy and robustness of diagnoses made by neurologists.

2. Related Work

Many studies have been reported using accelerometers (including gyroscopes), electromyography (EMG), and cameras (RGB cameras and RGB depth cameras) to observe and analyze tremor.

In a research report conducted by a research team at the Nagaoka University of Technology around the year 2000, for which an accelerometer was used for the time series conversion of motor symptoms, Kunimoto Shindo et al. aimed to estimate the severity and differentiation of a disease similar to Parkinson's syndrome and essential tremor [7]; Yukiyo Gondaira et al. (1998) aimed to perform a differential diagnosis and severity estimation of Parkinson's disease and essential tremor [8]; and Yoshinobu Matsumoto et al. (2001) aimed to conduct a frequency analysis of tremor acceleration information measured in the upper extremities of patients with Parkinson's disease and essential tremor, for which effective parameters designed for use in a biofeedback system were extracted [9]. These works reported that the time series of tremor intensity correlated with clinical scores and that it was difficult to discriminate disease using tremor frequency.

Additionally, the objective of a study by Roberto López-Blanco et al. (2019) was the quantitative assessment of resting tremor in patients with Parkinson's disease using the gyroscope of a smartwatch worn on the wrist of patients with Parkinson's disease [10], wherein the authors calculated the average tremor intensity and the clinical score in the three-axis time series of observations using Spearman's rank correlation coefficient and reported that the average tremor intensity and the clinical score were correlated. According to the findings presented in [7–10] wherein accelerometers were used (including gyroscopes) were used, the differential diagnosis of Parkinson's disease and essential tremor is similar in terms of the distribution frequency bands of each disease; it is a common view that identification using frequency information is difficult [7–9]. However, since we can confirm that there is a correlation between the clinical score (severity) and the time series of tremor intensity [8,10], we hypothesize that tremor intensity can be used to estimate severity.

In [11], which was reported by Jie Zhang et al. (2017), the authors aimed to investigate the value of electromyographic tremor analysis for the differential diagnosis of Parkinson's disease and essential and physiological tremor using electromyography, for which the tremor frequency in Parkinson's disease was about 4–6 Hz, the tremor frequency in essential tremor was about 5–8 Hz, and the frequency in physiological tremor was 6–12 Hz [11] (p. 2). Additionally, it was reported that Parkinson's disease and essential tremor are difficult to distinguish according to tremor frequency [11] (p. 3).

In addition, in Zixiang Hu et al.'s study, in which the authors aimed to analyze the interference effects of tremor [12] using a motion sensor and an electromyograph, the authors instructed patients to perform a specified trajectory of motion to analyze tremors during exercise and compared the differences in motion between Parkinson's disease subjects and control subjects. The observation of motor symptoms using accelerometers and electromyographs [11,12] can be used to directly measure time series because the sensors are in contact with the patient, and they seem to encounter fewer measurement errors due to noise, which may cause the patient to become aware that they are being measured and cause a situation wherein motor symptoms do not appear. Therefore, in recent research, many studies have reported the measurement and analysis of this motor symptom using a camera, which is a type of non-contact measurement method.

There were three camera-based methods for observing motor symptoms in the hands of Parkinson's disease. (1) using accelerometer and RGB camera [13], (2) using depth camera [14,15], (3) using RGB camera [5,6,16].

Stefan Williams et al. (2020) estimated tremor frequency via the optical flow of pixels in a video of motor symptoms using an accelerometer and an RGB camera, and an accelerometer was attached to a patient's hand to allow for the comparison of measurements [13]. The time series from each approach were analyzed using fast Fourier transform (FFT) and Bland–Altman analysis. The Altman analysis showed excellent agreement, with 36 of 37 videos (97%) showing a difference of less than 0.5 Hz between the computer vision and accelerometer readings and presenting levels of agreement and tremor amplitude. It was reported that there is no significant correlation between these two measures [13] (pp. 71–74). Accordingly, we hypothesize that the estimated time series from a video shot by a smartphone can reproduce the time series obtained by a contact type device.

In a study by Ferraris Claudia et al. (2018), an RGB depth camera was used to acquire time series of motor symptoms occurring in a patient's hand during an assessment of the three items of the MDS-UPDRS (3.4: finger tapping <FT>, 3.5: opening and closing <OC>, and 3.6: pronation and supination <PS>). The reported accuracy values (clinical score prediction) obtained via the cross-validation of five classifiers [14] (p. 16, Table 6) after extracting the features [14] (p. 11) from each time series were 76%, 65%, and 58% for FT, OC, and PS, respectively [14] (p. 17).

Similarly, in a study [15] conducted by Yan Pang et al. (2020), three-axis time series were obtained by stereoscopically capturing hand motor symptoms using two RGB cameras. To analyze the time series, the authors used the discrete wavelet transform to extract six features [15] (p. 5, Table 2) from each of the three input time series decomposed into five levels. The mean separability between severity in the control subjects and those with Parkinson's disease was determined. The results showed that the mean separability of the five Parkinson's disease patients was 0.411 with a standard deviation of 0.106. The correlation coefficient between the mean separability and UPDRS score was 0.9115, and the P value was 0.0312. Furthermore, they reported a mean separability of 0.179 with a standard deviation of 0.024 for two control subjects aged ≥ 60 years, whose features were clearly distinguishable from those of the Parkinson's disease patients [15] (p. 8).

Ferraris Claudia et al. (2018) and Yan Pang et al. (2020) believe that the three-dimensional analysis of motor symptoms is an effective method for the detailed analysis of motor symptoms. However, it is not a consumer-oriented method because the measurement equipment is bulky and special devices such as an RGB depth camera are used. We hypothesize that the RGB camera standardly installed in smartphones, which is widely used by consumers and does not require a large measurement device, can contribute to solving the above-mentioned problems. We also confirmed the results of a study that attempted an approach using a single camera.

In [5] conducted by T. Hayashida et al. (2021), using one RGB camera, velocity and acceleration waveforms were obtained by differentiating the time series obtained by one RGB camera. Seven statistical features were obtained from each waveform. As a result, the authors reported that the average validation accuracy obtained using the five classifiers was 76.8% [5].

In addition, in [6] conducted by Takafumi Hayashida et al. (2022), which was a continuation of Ref. [5], a time series was decomposed into five layers of wavelet coefficients, each feature value was obtained, and time series analysis was performed. As a result of analyzing the correlation using Pearson's product correlation coefficient for each feature value and four classes of severity, the maximum correlation with severity was $|r| = 0.65$, while most of these values were reported to be around $|r| = 0.60$. They also reported that the maximum correlation with disease was $|r| = 0.34$. However, this method uses Pearson's product correlation coefficient, which is used when combining quantitative and quantitative variables, even though the objects of analysis were quantitative and qualitative variables.

Finally, in a study by Rui Guo et al. (2022) concerning the quantitative assessment of tremor severity in patients with Parkinson's disease [16], a skeletal hand sequence was obtained from a video recorded by a single camera using a pre-trained hand pose estimation model. Clinical scores were estimated by the proposed tree-structure-guided GCN (TSG-GCN) [16] (pp. 3–6). As a result, the authors achieved an accuracy rate of 73.71% and an acceptable accuracy rate of 99.20% in severity estimation (i.e., the probability that a true positive (TP) value includes a false negative (FN) value before and after determining the true value) [16] (p. 7). In addition, they reported that the accuracy rate was 71.88% when tested with a dataset that was independent of the dataset at the time of learning [16] (p. 7).

As described above, the camera-based time-series acquisition of motor symptoms [5,6,16] is a more consumer-oriented approach than the accelerometer-and-electromyography-based approach described above. In the former, the patients themselves can determine the severity of their motor symptoms without travelling to a hospital or ascertain the disease they are likely afflicted with. Moreover, most of the studies presented focused on severity estimation and differential diagnosis between control subjects and those with Parkinson's disease. A comprehensive diagnostic support system, including the necessary differentiation between Parkinson's disease and essential tremor, has not yet been proposed.

3. Acquisition of Clinical Data and Time Series

Time series analysis involves the use of statistical methods to analyze patterns and trends in data that are collected over time. This can be particularly useful for the analysis of clinical data, where changes in a patient's condition may be tracked over time using various types of measurements.

An algorithm flowchart for the proposed system is described in Figure 1, including the following processes:

1. Collect clinical data using a smartphone's built-in camera;
2. Pre-process the data by segmenting foreground and background images;
3. Execute labelling process and time series data transformation;
4. Extract statistical features from the time series data;
5. Visualize the importance of the extracted features;
6. Perform data classification using logistic regression model;
7. Estimate posterior probabilities using the SoftMax function.

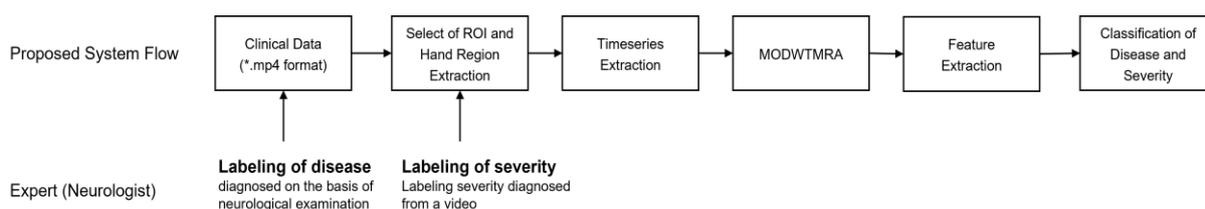


Figure 1. The overall proposed system for extracting clinical data (* refers to the file name)".

The main contribution of this paper is the proposal of a system that can collect clinical data using a smartphone's built-in camera, pre-process the data, extract statistical features, and classify the data using a logistic regression model. The proposed system can be used for medical diagnosis and monitoring. The broader contribution of this paper is the development of a system that can enable the remote monitoring of patients and provide timely feedback to medical professionals.

3.1. Acquisition of Clinical Data

The subjects of this study were 25 outpatients at the University of Miyazaki Hospital, 19 of whom were diagnosed with Parkinson's disease (7 males/12 females), while the remaining 6 were diagnosed with essential tremor (4 males/2 females). Table 1 shows the subjects' information. In addition, the subjects were given an oral explanation of the purpose of this study, how the data would be obtained, and how they would be used,

and their written consent was obtained. In consideration of the patients' privacy, the data were processed so that individuals could not be identified. This study was approved by the Ethics Committee of the UNIVERSITY OF MIYAZAKI DEPARTMENT OF MEDICINE (protocol code: O-0577, 20 September 2019). The data collection period spanned from October 2020 to October 2022. Motor symptoms were recorded with MDS-UPDRS 3.15. Postural tremor of hands was assessed under the supervision of a neurologist designated as an intractable disease doctor.

Table 1. Subjects' information.

	Parkinson's Disease	Essential Tremor
Age		
(mean \pm sd)	75.3 \pm 8.4	64.7 \pm 13.4
(range)	(63–91)	(42–84)
Participants (male-to-female ratio)	19 (7:12)	6 (2:1)
Number of data	23	9
MDS-UPDRS (3.15) equivalent		
(mean \pm sd)	1.73 \pm 1.07	1.88 \pm 0.31
(range)	(0–3)	(1–2)

The recording conditions are as follows. Videos were recorded for about 15 s at a resolution of 1920 \times 1080 and a sampling frequency of 30 fps or 60 fps using a GoPro HERO5 or smartphone's built-in camera. At this time, the camera was fixed, and the distance from the lens to the hand was about 30 cm. At the time of recording, a green, felt fabric was placed behind the hand to facilitate the extraction of the hand region.

3.2. Labeling Clinical Data

Each captured video was objectively evaluated by a neurologist designated as a disease-specific specialist with a score equivalent to the MDS-UPDRS Part III: Motor Examination (3.15: Postural tremor of the hand). MDS-UPDRS 3.15 is rated on a five-level scale, but the score used for correct labeling is a four-level scale ranging from zero to three. A four-level scale was used to eliminate the following ambiguous ratings in MDS-UPDRS 3.15.

3.15		POSTUAL TREMOR OF HANDS
0:	Normal:	No Tremor.
1:	Slight:	Tremor is present but less than 1 cm in amplitude.
2:	Mild:	Tremor is at least 1 cm but less than 3 cm in amplitude.
3:	Moderate:	Tremor is at least 3 cm but less than 10 cm in amplitude.
4:	Severe:	Tremor is at least 10 cm in amplitude.

As shown above, the difference between MDS-UPDRS 3.15 scores of 0 and 1 is ambiguous, and it is possible that different evaluators will yield different evaluations. To avoid this, the four-level evaluation index used in the actual evaluation was a combination of the 0 and 1 scores from MDS-UPDRS 3.15. In addition, the correct labels for Parkinson's disease and essential tremor were findings from a neurological examination conducted by a neurologist designated as an intractable disease doctor.

3.3. Acquisition of Time Series

The selection of the region of interest (ROI) for each captured video was performed by an expert in the field who had extensive clinical experience with the motor symptoms being evaluated. The expert manually identified the specific regions on the video frames that were most relevant to the motor symptoms under study, and these regions were marked as the regions of interest. To ensure the accuracy and consistency of the ROI selection procedure, a second expert was also involved in the process. The second expert reviewed and verified the ROI selections made by the first expert, and any discrepancies were discussed and resolved through consensus. The ROI was then used to trim the captured

video data to remove any extraneous frames outside of the selected region. These trimmed data were then used as the input video to the system. Overall, the ROI selection process was performed with great care and attention to detail in order to ensure the accuracy and reliability of the data's analysis. We believe that this rigorous approach to ROI selection helps strengthen the validity of our findings.

A region of interest (ROI) was manually created for each captured video, and trimming was performed according to the ROI. The data after trimming were used as the input video to the system. In addition, this research was processed using the technical computing language environment MATLAB (The MathWorks, Inc., Natick, Massachusetts, USA; MATLAB Version 9.13.0.2105380 (R2022b), Update 2, The MathWorks, Inc., 2022. Computer Software. www.mathworks.com/ accessed on 1 January 2022). Ref. [5] was used as a reference for the method of acquiring motor symptoms as a time series from the input video.

To extract the hand region for each image, we divided the input videos into image sequences. As an example, Figure 2 shows the image sequences after processing.



Figure 2. Image sequence example for each input video (10 videos).

We would like to clarify that the processing procedure referred to in Figure 2 is the segmentation process, which is a method used to separate the foreground (hand) from the background in an input video. The image sequence in Figure 2 shows the result of this segmentation process, where the hand is displayed as the foreground on a green background.

As can be seen in Figure 2, the image sequence after the segmentation process is an RGB image in which the hand is reflected as the foreground on the green background. To extract the hand region, the hand region was obtained by converting the image to the HSV color space, which can handle color expression more intuitively than the RGB color space, and manually setting the hue range of the background, as shown in Figure 3.

$$0.00 \leq \text{Hue} < 0.20 \cup 0.80 \leq \text{Hue} < 1.00 \quad (1)$$

Therefore, we set the thresholds for the hue range and saturation range of the background, for which we assumed that the foreground would not be included, as follows.

$$0.25 \leq \text{Hue} < 0.75 \quad (2)$$

$$0.20 \leq \text{Saturation} \quad (3)$$

Using these threshold processing steps, the background area aside from the hand area is output. The hand region is extracted by logical negation (NOT). At this time, when a region other than the hand region was extracted, processing was performed so that only the largest region with a logical value of 1 remained. The exclusive OR was calculated for continuous binary images of the hand region. When the exclusive OR is calculated, an

image of displacement (displacement amount) of the hand region between adjacent images is output.

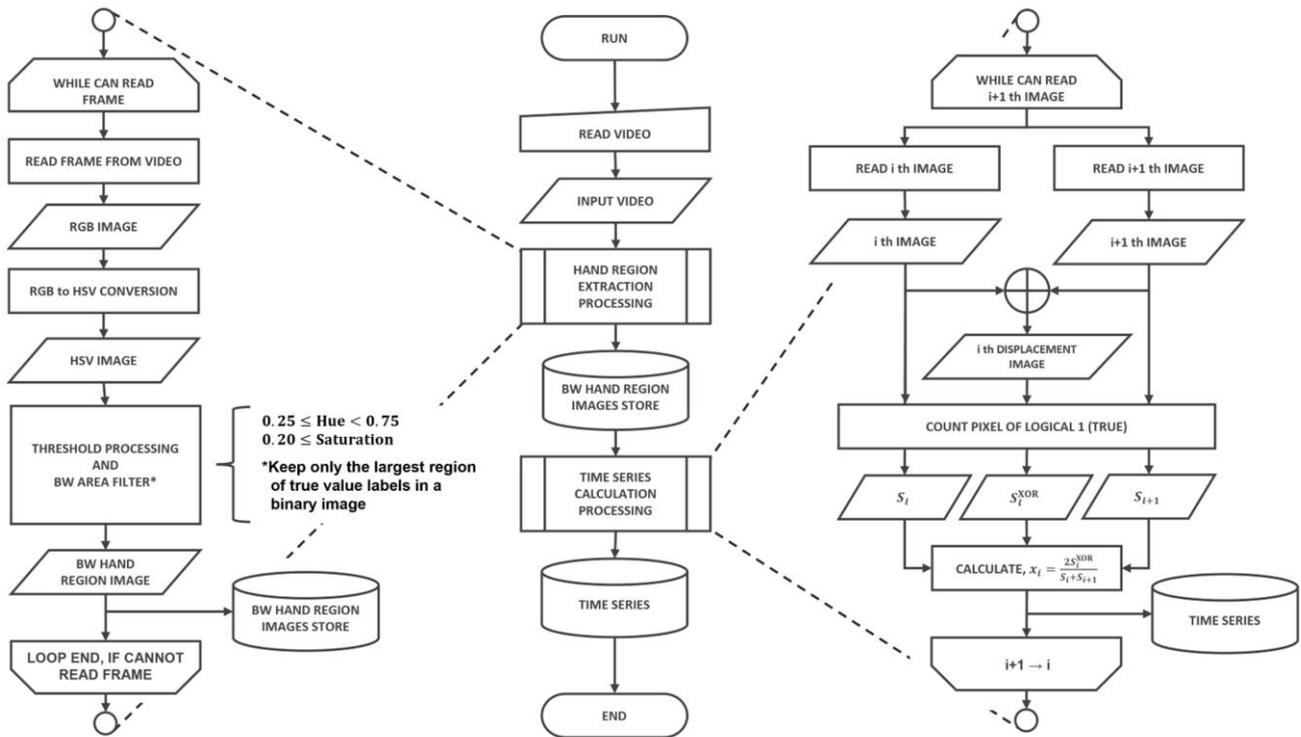


Figure 3. The flowchart of time series extraction from selected video input.

The number of white pixels per frame of the binary image of the hand region and the number of white pixels per frame of the exclusive OR image are defined as follows:

$$\{S_1, S_2, \dots, S_{n-2}, S_{n-1}, S_n\}, \tag{4}$$

$$\{S_1^{EXOR}, S_2^{EXOR}, \dots, S_{n-2}^{EXOR}, S_{n-1}^{EXOR}\}. \tag{5}$$

To suppress the effect of hand size, the time series of motor symptoms was obtained by dividing the number of white pixels in the displaced image by the average number of white pixels in the binary image of the two-handed region. This process can be defined as follows:

$$\{x_1, x_2, \dots, x_{n-2}, x_{n-1}\}, \tag{6}$$

$$x_n := \frac{2S_n^{XOR}}{S_n + S_{n+1}}, \tag{7}$$

where n is the frame number.

When acquiring binary images of the hand region from the image sequence, the hand region could not be adequately extracted, and spike noise (values clearly different from other observed values) occurred in the time series. To remove this spike noise, we detected and removed abnormal values (outliers) using the Hampel classifier for the time series of motor symptoms. The Hampel identifier is a type of parametric method, specifically, an outlier detection method that assumes that the frequency distribution of sample values follows a normal distribution.

Most biological signals are non-stationary signals whose frequencies fluctuate over time. Therefore, the discrete Fourier transform (DFT), which assumes that the target of analysis is a stationary signal, is a frequency transform method that removes time

information from the time series and is thus unsuitable for nonstationary signal analysis. Therefore, in this study, we decomposed the time series of motor symptoms into five levels using undecimated discrete wavelet transform, which is a kind of time–frequency transform. The decomposition level (resolution) of the time series used in this study was set at five due to the tremor frequency distribution and time series length. The mother wavelet used is the Daubechies (Db) orthogonal wavelet Db4, which was reported to have the highest accuracy in [15]. A total of six time series were calculated from each video by combining the input time series and the wavelet coefficients of each level after multiresolution analysis (detailed coefficients—cD1, cD2, cD3, cD4, cD5; approximation coefficient—cA5).

While we realized that AI-based methods can be more efficient and faster, we chose to use the manual method of hue and saturation extraction in our research for several reasons. First, our focus was on evaluating the effectiveness of the overall system in detecting motor symptoms rather than the specific method of hue and saturation extraction. Second, we had limited resources and time available for this project and, therefore, opted for a simpler and more accessible method that could be performed using the tools and expertise available to us.

However, we will consider the use of AI-based methods in future studies. We acknowledge that these methods have the potential to improve the efficiency and accuracy of motor symptom detection, and we agree that they constitute a promising area of research in this field.

4. Experimental Methods

4.1. Feature Extraction

In this study, the statistical features shown in Table 2 were obtained from each time series, where x_n is the n th element of the time series. A total of 31 features were extracted, 3 of which (MAD, RMS, and IQR) from the time series of motor symptoms and approximation coefficients, and 5 features (MAD, RMS, ZCR, IQR, and DTW) were extracted from the time series of each level after wavelet analysis.

Table 2. Statistical features.

Feature	Formula
Median Absolute Deviation (MAD)	$MAD_i = \frac{1}{N} \sum_{n=1}^N x_n - median(x) $ (8)
Root Mean Square (RMS)	$RMS_i = \sqrt{\frac{1}{N} \sum_{n=1}^N x_n - \bar{x} ^2}$ (9)
Zero Cross Ratio (ZCR)	$ZCR_i = \frac{count(x_n x_{n+1} < 0)}{N}$ (10)
Interquartile Range <IQR>	$IQR_i = Q_3 - Q_1$ (11)

Since the tremor frequency band for each disease is known [2,3], the dynamic time warping (DTW) method [17] is used to measure the distance between a sinusoidal signal (stationary signal) of the same length of 5 Hz as the acquired motor symptom time series and the detail coefficients for each level, where the amplitude of the sine wave is the absolute value of the amplitude of each detailed coefficient. The formula for the DTW method is shown below:

$$t_m := \max(|S|) \sin\left(\frac{\pi}{6}(m - 1)\right) \in T, \tag{12}$$

$$\delta(i, j) = |s_i - t_j|, \tag{13}$$

$$DTW(S, T) = \min_W \sum_{k=1}^{m^2} \delta(w_k), \tag{14}$$

where S is each time series $\{s_1, s_2, \dots, s_{m-2}, s_{m-1}, s_m\}$, T is a sinusoidal discrete value $\{t_1, t_2, \dots, t_{m-2}, t_{m-1}, t_m\}$, and W is the set of brute force Euclidean distances (w_k) between each element of and each element of.

4.2. Visualization of Feature Importance

Spearman's rank correlation coefficient is an index of correlation (rank correlation coefficient) obtained from an ordinal scale in statistics that was proposed by Charles Spearman (1904) [18]. In this study, Spearman's rank correlation coefficient was used to visualize the importance of each feature and the correct label.

4.3. Data Classification

Based on the obtained feature values, the severity of tremor and disease were label-encoded, and class classification, which is a form of supervised learning, was performed. Many algorithms have been proposed for classification. In this study, two-class and four-class classification were performed using logistic regression. Logistic regression is a statistical regression model that can explain and predict the probability of an outcome occurring based on several factors for a qualitative variable. It is a generalized linear model (GLM) that uses logit as the coupling function.

In this study, we adopted k-fold cross-validation ($k = 4$) as the verification method of the learning model because the dataset obtained was small and the quantity of each objective variable was imbalanced.

5. Results

Table 3 shows the rank correlation coefficient between each feature value and the objective variable.

Table 3. Rank correlation coefficient between each feature value and target variable.

	Max	Min	U-var	U-std	Iqr	Avg	Med
disease	0.38	0.00	0.01	0.10	0.19	0.13	0.13
level	0.73	0.02	0.03	0.18	0.08	0.58	0.65

The rank correlation coefficients between disease and each feature value were maximum—0.38 (vs. ZCR2); minimum—0.00; unbiased variance—0.01; unchanged standard deviation—0.10; interquartile range—0.19; mean—0.13; and median—0.13. Thus, it can be surmised that there is almost no correlation between the disease and each feature, including the 5 Hz sine wave used in this study and the time series similarity DTW of each time series. The rank correlation coefficients between severity and each feature were maximum—0.73 (vs. MAD and IQR); minimum—0.02 (vs. DISEASE); unbiased variance—0.03; unchanged standard deviation—0.18; interquartile range—0.08; mean—0.58; and median—0.65. As a result, it was confirmed that there is a correlation of around 0.6 in terms of the absolute value between the degree of severity and each characteristic. In addition, the correlation between time series amplitude and severity is consistent with the reports of related studies.

Table 4 shows the performance of the learning model, and Figure 4 shows the confusion matrix of the learning model. The accuracy was 0.56 for two-class classification (disease) and 0.50 for four-class classification (severity) (an acceptable accuracy rate is 0.96).

Table 4 reveals the following characteristics of the performance of the learning model with respect to illness and severity:

- Class 0 (Parkinson's disease) and class 1 (essential tremor) cannot be distinguished;
- Regarding severity, there are few false positives (predictions other than before and after the correct label) because there is a correlation between the objective variable and the explanatory variable;
- In terms of severity, the classification accuracy is 0.80 for class 0, around 0.55 for classes 1 and 2, and 0.22 for class 3, which is the lowest.

Table 4. Learning model’s performance.

	Precision	Recall	F1-Score		Precision	Recall	F1-Score
class 0	0.67	0.78	0.72	class 0	0.80	1.00	0.90
class 1	0.00	0.00	0.00	class 1	0.57	0.67	0.62
				class 2	0.55	0.40	0.47
				class 3	0.22	0.29	0.25
macro avg	0.33	0.39	0.36	macro avg	0.53	0.59	0.56

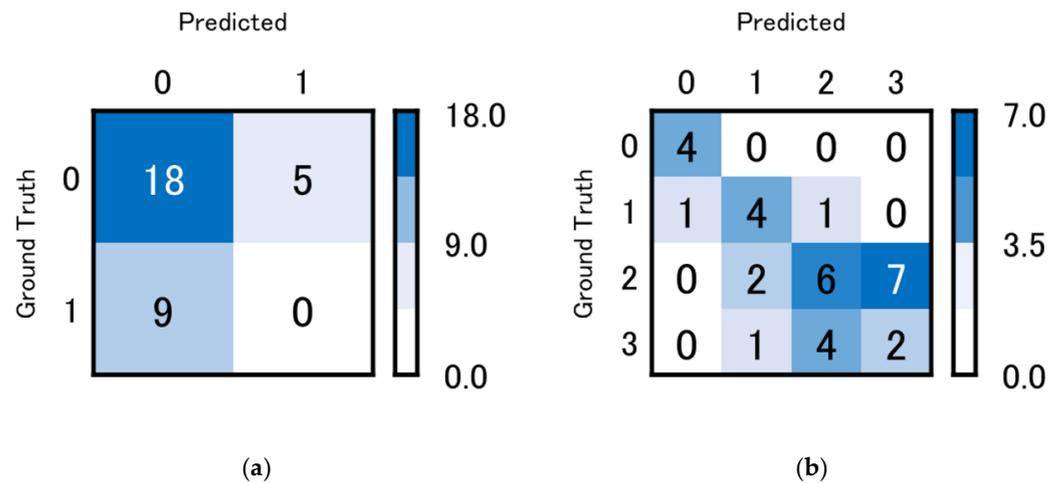


Figure 4. Confusion matrix for training model: (a) 2-class classification result (disease); (b) results of 4-class classification (severity).

Figure 5 shows the ROC curve and AUC of the learning model. Figure 5a shows that the prediction performance of the learning model is lower than that of the random case and that a completely biased prediction was made using the two-class classification result (disease). In addition, of the four-class classification results (severity) presented in Figure 5b, severity class 0 has a high F1 scale and area = 0.96, indicating that classification was remarkably achieved. Similarly, severity class 1 has a high F1 scale (ibid, class 1 F1 score = 0.62) with area = 0.78. On the other hand, for severity classes 2 and 3, the classification prediction is random and becomes more difficult as the severity increases.

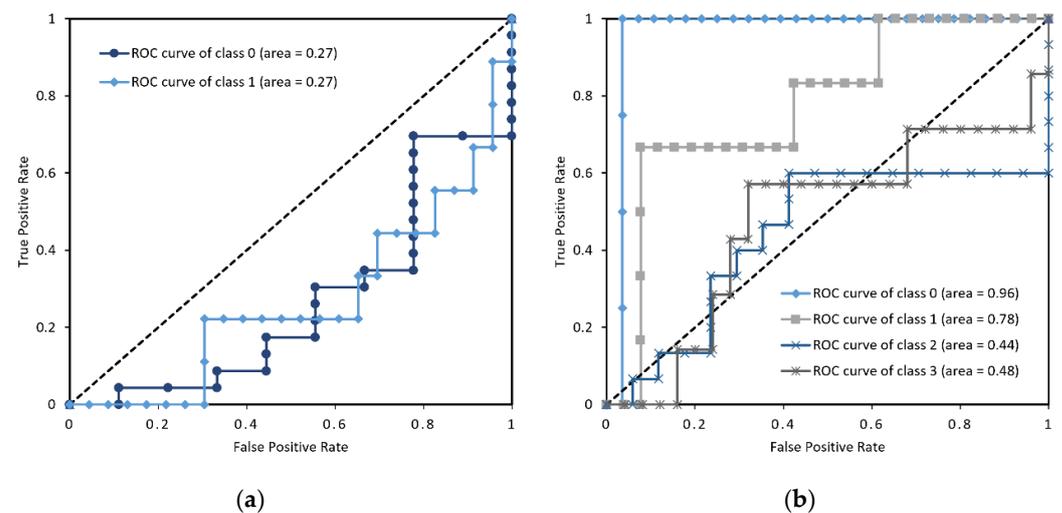


Figure 5. ROC curve and AUC of learning model: (a) 2-class classification result (disease); (b) 4-class classification result (severity).

6. Discussion

The purpose of this study was to develop a support system for estimating the severity of tremor and achieve a differential diagnosis between Parkinson's disease and essential tremor via the implementation of a smartphone application. Therefore, in this study, by analyzing the time series obtained by the non-contact acquisition method of motor symptoms using a single RGB camera [14], we estimated the severity of tremor and examined the possibility of differentiating Parkinson's disease from essential tremor. The elements required to achieve the research objectives are shown below.

1. A practical method for acquiring motor symptoms should be employed;
2. Do not interfere with a medical examination (regardless of location or whether short measurement times are used);
3. Find features (explanatory variables) that are highly significant in terms of severity prediction;
4. Find features (explanatory variables) that are highly significant in terms of disease prediction;
5. Create a learning model and output fuzziness;
6. Ensure reliability (high accuracy) of the classification results.

The degree to which each of these elements was achieved in this research is described below.

Regarding 1. In this study, an RGB camera was used to acquire motor symptoms. An RGB camera is a common device and a standard piece of equipment included in smartphones possessed by numerous individuals. Accordingly, we believe that the non-contact diagnostic support system proposed in this study has achieved this requirement.

Regarding 2. In relation to 1, the timeseries acquisition method of motor symptoms in this study is an original, non-contact, RGB-camera-employing method. By making the system proposed in this research an app that can be installed on smartphones, we believe that it may be possible to overcome the problem of being "anywhere". There is a constraint that we have identified: we believe that a device cannot be capable of being used "anywhere" without overcoming the constraints of this background condition. In addition, the measurement time employed in this study was about 15 s. We believe that this does not interfere with examination, but we did not investigate the effect of the time series length of 15 s on the classification results. Therefore, it is necessary to investigate the effect of the recording time on the classification results in future studies. Therefore, we believe that element 2 has not been achieved.

Regarding 3. We found that there is a high degree of correlation (maximum $r_s = 0.73$) between the correct labeling of severity and each explanatory variable, especially the feature values derived from the amplitude of the time series. Regarding the degree of importance, we believe that it is highly possible to classify disease and severity using the feature value derived from the amplitude of the time series. Therefore, we believe that element 3 has been achieved.

Regarding 4. No correlation was found between the correct labelling of disease and each explanatory variable. In addition, the prediction of the learning model was worse than random (class 0, 1; area = 0.27). There are three possible reasons for this.

First, the features do not adequately represent the characteristics of the disease.

Second, the parameters of the learning model are not optimal for discrimination.

Third, the learning model used for disease prediction is not a suitable method for disease prediction.

The first reason falls into the category of feature quantity engineering and could not be tackled in this study. We also think that it is necessary to reduce the dimensions of the features. This is because, as can be seen from Table 3, for the rank correlation coefficient between each feature value and the objective variable, multicollinearity can be confirmed between many features. In this case, it is necessary to perform dimensionality reduction and feature selection. Specifically, there is a method called VIF (variance inflation factor) that excludes whichever variable has the lowest correlation coefficient and does not use the feature with multicollinearity as an explanatory variable. It is necessary to select features

using an index for detecting multicollinearity between variables. We would like to address this point in our future work.

In addition, we were unable to express the periodic fluctuation characteristic of Parkinson's disease. In addition to the periodic increase and decrease in amplitude, we were able to confirm the phenomenon of the temporary cessation of movement from the numerical data on Parkinson's disease. As a future area of research, we believe that whether or not the shape features that appear in this time series can be reduced to feature values will be the key to disease discrimination.

Regarding the second reason for the poor prediction of the learning model, it must be noted that we could not optimize the parameters of the learning model in this study, and it cannot be said that the experiment was performed with the optimal learning model. Therefore, in future research, it will be necessary to optimize the parameters of the learning model and compare it with several methods (neural network, SVM (support vector machine), decision tree, etc.). Therefore, we believe that element 4 has not been achieved.

The third reason for the learning model's poor outcome is that logistic regression, which is the classifier used in this experiment, is not suitable for disease prediction. As mentioned above, there is almost no correlation between the disease (objective variable) and each explanatory variable, and as the severity increases, the value of the explanatory variable increases as well (No proportional relationship). Logistic regression is a linear classification model that performs well when there is a proportional relationship between the objective function and the explanatory variables. To express objective variables and explanatory variables that cannot be expressed in a proportional relationship, it is necessary to select nonlinear classification models such as SVM, neural networks, and Naive Bayes. Therefore, we plan to use a linear classification model for severity prediction and a nonlinear model for disease prediction.

Regarding 5. In this study, the learning model was created using logistic regression. In addition, since the posterior probability can be output by inserting the SoftMax function in the output, we find that this element has been achieved.

Regarding 6. We acknowledge that our method did not achieve a high level of accuracy in distinguishing between Parkinson's disease and essential tremor. One possible explanation for this failure is that the motor symptoms of Parkinson's disease and essential tremor can be similar and, therefore, difficult to distinguish (even for expert doctors). Another possibility is that the limitations of our method, such as the manual extraction of hue and saturation and the potential impact of lighting conditions and camera positioning, may have contributed to the lower accuracy.

Specifically, it would be useful to evaluate the ability of expert doctors to distinguish between Parkinson's disease and essential tremor based on the videos used in our study. However, we did not conduct such an analysis in our study. We will consider this idea in future research.

In this paper, we investigated whether two diseases can be distinguished by analyzing posture-related tremor features. However, our analysis did not reach a level of differentiation that could be applied clinically. In this study, depending on the interpretation of the degree of correctness, it was shown that the tremor quantification process was consistent with an actual qualitative evaluation, and the validity of the quantification method was verified. We suggest that changes over time in tremors, resting tremors, and tremors due to oral medication can be measured to quantitatively assess improvement in motor symptom severity.

On the other hand, deep learning has the potential to improve the accuracy of motor symptom detection by automatically memorizing complex patterns and relationships in large quantities of data. With deep learning, it is possible to train neural networks to recognize specific motor symptoms, such as tremor or bradykinesia, by feeding them a large dataset of labelled examples. By using deep learning techniques, it may be possible to overcome some of the limitations of traditional computer vision techniques, such as the need for manual feature extraction and the potential impact of lighting conditions and

camera positioning. Deep learning can also improve the generalization of the system to new patients and conditions. Additionally, deep learning can potentially reduce the amount of data storage required for motor symptom detection by only storing the learned parameters of the neural network rather than the entire dataset. This attribute can be particularly useful for mobile applications and remote-monitoring systems wherein storage and bandwidth are limited.

However, deep learning also has its own limitations and challenges, such as the need for large quantities of labelled data and computational resources and potential issues regarding overfitting and interpretability. Nevertheless, it is an exciting area of research whose findings can be used to improve the accuracy and efficiency of motor symptom detection.

Not achieved in relation to 4. The learning model of this research does not produce accuracy sufficient for use at the practical level, and there is a need for improvement. As for the prospective addressal of these issues, we suggest the following:

1. Perform feature quantity engineering after time series extraction, and create optimal feature quantities for each objective variable;
2. Perform feature dimension reduction (feature selection) and select a classifier suitable for the objective variable;
3. Optimize classifiers' hyperparameters;
4. Eliminate background condition constraints when extracting hand regions.

By addressing these issues in the above order, we would like to improve the accuracy of the classifier and develop a discrimination support system.

7. Conclusions

The purpose of this study was to develop a support system for estimating the severity of tremor and to perform a differential diagnosis between Parkinson's disease and essential tremor through the implementation of a smartphone application. Therefore, in this study, by analyzing the time series obtained by the non-contact acquisition method of motor symptoms using a single RGB camera [14], we estimated the severity of tremor and examined the possibility of differentiating Parkinson's disease from essential tremor. It was found that there is almost no correlation between the disease and each feature, including the 5 Hz sine wave used in this study and the time series similarity DTW of each time series. For the rank correlation coefficient of severity and each feature, the maximum value was 0.73 (vs. MAD, IQR), the minimum value was 0.02 (vs. DISEASE), the unbiased minimum value was 0.03, the unchanged standard deviation was 0.18, the interquartile range was 0.08, the mean was 0.58, and the median was 0.65. As a result, it was confirmed that there is a correlation of around 0.6 with respect to the absolute value between the degree of severity and each characteristic. We also confirmed that the correlation between the timeseries amplitude and severity is consistent with the reports of related studies. In addition, many features with correlation coefficients greater than 0.70 can be confirmed. Thus, we can see that multicollinearity may exist in many features. In other words, the features used in this research include features that have little effect on classification, so it is necessary to select them. Furthermore, according to the two-class classification result (disease) presented in Figure 3 the prediction performance of the learning model (for both class 0 and class 1) was lower than the random case (area = 0.50), and a completely biased prediction was performed. In Figure 5b, the four-class classification results (severity) are presented, where severity class 0 has a high F1 scale (according to Table 4, which concerns the learning model's performance, the class 0 f1 score = 0.90) and an area = 0.96; thus, the features were accurately classified. Similarly, severity class 1 has a high F1 scale (ibid, class 1 f1 score = 0.62) with an area = 0.78. On the other hand, for severity classes 2 and 3, the classification prediction was random, and it was found that the classification prediction became more difficult as the severity increased. In upcoming research, we will conduct feature quantity engineering after time series extraction, create the optimal feature quantity for each target variable, reduce the dimensions of the feature quantities (feature quantity selection), select the optimal classifier, optimize the hyperparameters of

the classifier, and eliminate the background condition constraints when extracting hand regions. We would like to implement these affordances so that patients can self-diagnose and provide diagnostic support to neurologists.

Author Contributions: Conceptualization, T.H., T.T.Z. and T.S.; methodology, T.H.; software, T.H.; validation, T.H., T.T.Z. and T.S.; formal analysis, T.H.; investigation, T.H., T.T.Z., K.S., N.I., H.M. and T.S.; resources, T.T.Z., T.S., K.S., N.I. and H.M.; data curation, T.S., K.S., N.I. and H.M.; writing—original draft preparation, T.H.; writing—review and editing, T.H., T.T.Z. and T.S.; visualization, T.H.; supervision, T.T.Z.; project administration, T.T.Z.; funding acquisition, T.T.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the UNIVERSITY OF MIYAZAKI DEPARTMENT OF MEDICINE (protocol code: O-0577, 20 September 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kazushi, T. Clinical management of elderly patients with Parkinson disease in a super-aging society: Current status and issues. *Neurotherapeutics* **2022**, *39*, 51–55. [[CrossRef](#)]
2. Deuschl, G.; Bain, P.; Brin, M. Consensus Statement of the Movement Disorder Society on Tremor. *Mov. Disord.* **1998**, *13* (Suppl. S3), 2–23. [[CrossRef](#)] [[PubMed](#)]
3. Vaillancourt, D.E.; Newell, K.M. The dynamics of resting and postural tremor in Parkinson's disease. *Clin. Neurophysiol.* **2000**, *111*, 2046–2056. [[CrossRef](#)] [[PubMed](#)]
4. Goetz, C.G.; Tilley, B.C.; Shaftman, S.R.; Stebbins, G.T.; Fahn, S.; Martinez-Martin, P.; Poewe, W.; Sampaio, C.; Stern, M.B.; Dodel, R.; et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Mov. Disord.* **2008**, *23*, 2129–2170. [[CrossRef](#)] [[PubMed](#)]
5. Hayashida, T.; Zin, T.T.; Sakai, K.; Mochizuki, H. Evaluation of the severity of tremor based on each signal acquired from the displacement of the hand movements. In Proceedings of the 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), Kyoto, Japan, 12–15 October 2021; pp. 760–761. [[CrossRef](#)]
6. Hayashida, T.; Zin, T.T.; Sugiyama, T.; Sakai, K.; Ishii, N.; Mochizuki, H. A Study on the Possibility of Distinguishing between Parkinson's disease and Essential Tremor using Motor Symptoms Observed by an RGB camera. In Proceedings of the 35th Annual Conference of Biomedical Fuzzy Systems Society (BMFSA2022), Himeji, Japan, 17–18 December 2022.
7. Shindo, K.; Gondaira, Y.; Sakakibara, M.; Matsumoto, Y.; Tamura, M.; Fukumoto, I. Study of automatic analysing system for diagnosis and treatment of parkinsonian disease by main frequency. *Bull. Nagaoka Univ. Technol.* **1996**, *18*, 39–44.
8. Gondaira, Y.; Sakakibara, M.; Shindo, K.; Shindo, I.; Matsumoto, Y. Physiological and Pathological Analysis of Tremor in Human Beings by Main Frequency. *Soc. Biomech. Jpn.* **1998**, *14*, 109–117. [[CrossRef](#)]
9. Matsumoto, Y.; Koyama, R.; Takai, S.; Shiota, H.; Shinbo, S.; Fukumoto, I. Studies of evaluation of tremor diseases by frequency analysis on tremor acceleration waveforms. *Bull. Nagaoka Univ. Technol.* **2001**, *23*, 93–97.
10. López-Blanco, R.; Velasco, M.A.; Méndez-Guerrero, A.; Romero, J.P.; del Castillo, M.D.; Serrano, J.I.; Rocon, E.; Benito-León, J. Smartwatch for the analysis of rest tremor in patients with Parkinson's disease. *J. Neurol. Sci.* **2019**, *401*, 37–42. [[CrossRef](#)] [[PubMed](#)]
11. Zhang, J.; Xing, Y.; Ma, X.; Feng, L. Differential Diagnosis of Parkinson Disease, Essential Tremor, and Enhanced Physiological Tremor with the Tremor Analysis of EMG. *Park. Dis.* **2017**, *2017*, 4. [[CrossRef](#)] [[PubMed](#)]
12. Hu, Z.; Hao, M.; Xu, S.; Xiao, Q.; Lan, N. Evaluation of tremor interference with control of voluntary reaching movements in patients with Parkinson's disease. *J. NeuroEngineering Rehabil.* **2019**, *16*, 1–13. [[CrossRef](#)] [[PubMed](#)]
13. Williams, S.; Fang, H.; Relton, S.D.; Wong, D.C.; Alam, T.; Alty, J.E. Accuracy of Smartphone Video for Contactless Measurement of Hand Tremor Frequency. *Mov. Disord. Clin. Pract.* **2020**, *8*, 69–75. [[CrossRef](#)] [[PubMed](#)]
14. Ferraris, C.; Nerino, R.; Chimienti, A.; Pettiti, G.; Cau, N.; Cimolin, V.; Azzaro, C.; Albani, G.; Priano, L.; Mauro, A. A Self-Managed System for Automated Assessment of UPDRS Upper Limb Tasks in Parkinson's Disease. *Sensors* **2018**, *18*, 3523. [[CrossRef](#)] [[PubMed](#)]
15. Pang, Y.; Christenson, J.; Jiang, F.; Lei, T.; Rhoades, R.; Kern, D.; Thompson, J.A.; Liu, C. Automatic detection and quantification of hand movements toward development of an objective assessment of tremor and bradykinesia in Parkinson's disease. *J. Neurosci. Methods* **2020**, *333*, 108576. [[CrossRef](#)] [[PubMed](#)]

16. Guo, R.; Li, H.; Zhang, C.; Qian, X. A tree-structure-guided graph convolutional network with contrastive learning for the assessment of parkinsonian hand movements. *Med. Image Anal.* **2022**, *81*, 102560. [[CrossRef](#)] [[PubMed](#)]
17. Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [[CrossRef](#)]
18. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.