Enhancing Parkinson's Disease Recognition through Multimodal Analysis of Archimedean Spiral Drawings

Attila Zoltán Jenei[1], Dávid Sztahó[1], and István Valálik[2]

***Abstract*—Parkinson's disease is one of the most common neurodegenerative diseases, which is incurable according to recent clinical knowledge. Evaluating motor symptoms across diverse modalities such as speech, handwriting, and movement composes a conventional diagnostic approach. However, concurrently utilizing multimodal datasets encompassing drawing and acceleration data remains an underexplored field. Our investigation involved examining drawing and movement data of 45 Parkinson's disease (PD) patients and 47 healthy individuals (HC). The PD group presented mild symptoms in the right hand. We transformed drawing data into spiral images and used visual representations of motion data, employing pre-trained models for feature extraction and classifiers. While motion representations exhibited superior performance compared to drawing images, a comprehensive evaluation with the Mann-Whitney U test at a significance level of 0.05 revealed no statistically significant difference between the efficacy of movement and drawing data in all classification scenarios. Significant improvements were made by combining the drawing data predictions with the motion data predictions. The key finding of the research is that the recognition of the disease can be improved by connecting (post-model) the two modalities. Furthermore, it can be concluded that with the present approach, neither the drawing nor the movement data produced lower results on average.**

***Index Terms*—Acceleration Data, Classification, Parkinson's disease, Pre-trained Models, Mann-Whitney U Test**

I. INTRODUCTION

Parkinson's disease (PD) is one of the most common neurological disorders, which affects mainly the aging population. According to current clinical knowledge, the disease is incurable, promoting this area as an extensive research field. The goal is typically to recognize the disease early enough to alleviate symptoms, slow disease progression, and maintain quality of life.

Its prevalence is 1% in people over 60 and 3% in people over 80 [1]. These values tend to increase due to aging societies, environmental factors, and accessibility of health care (more people get recognized). The tendency to the disease is increased by the male sex, certain chemicals, and genetic factors [2].

[1] Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest, Hungary (e-mail: jenei.attila.zoltan@vik.bme.hu, sztaho.david@vik.bme.hu; orcid id: 0000-0003-1007-9907, 0000-0002-7361-4260).

[2] Department of Neurosurgery, St. John's Hospital, Budapest, Hungary (e-mail: valalik@parkinson.hu).

DOI: 10.36244/ICJ.2024.1.8

The destruction of dopaminergic neurons with subsequent depigmentation of the susbtantia nigra pars compacta (SNpc) and the appearance of Lewy bodies can be observed in the development of PD [1]. The importance of early detection of PD is shown by the fact that with the current diagnostic procedure, 60% of the dopamine-producing cells are already dead and cause problematic symptoms.

Non-motor symptoms appear earlier on the onset of PD. These include, for example, loss of smell, memory loss, digestive problems, and difficulty with sleeping [3]. These symptoms can appear even years earlier than motor symptoms. However, using these symptoms is difficult because they can indicate other illnesses, and not everyone develops the same symptoms.

In addition, the motor symptoms appear later in time, of which the three main are slowed movement (bradykinesia), muscle stiffness, and limb tremors at rest (resting tremor) [4]. These are vital symptoms taken into account by the neurologist to a large extent when establishing the diagnosis. It is important to emphasize that the diagnostic procedure relies heavily on the visual assessment of symptoms, imaging procedures, and drug tests. Still, there is currently no objective test for PD. Furthermore, the assessment can be influenced by the physician's subjectivity [5].

Because of the former, many researchers use AI and different modalities of data to recognize the disease using motor symptoms that help to increase the diagnosis accuracy and objectivity. Moreover, it allows the possibility of personalized care. Speech can be such a modality since 70% of PD patients develop dysphonic speech [6]. In addition, drawings/handwriting [7][8] and different forms of movement [9][10] are often used to analyze limb symptoms and help the diagnosis process.

In the present research, we investigate whether drawing (more precisely, drawing a spiral pattern) as pictorial information or the acceleration data measured during drawing provides a better recognition of the disease. Furthermore, we attempt to use the two modalities jointly to see whether recognition performance improves.

The paper's main contributions are 1) *comparing the spiral drawing and the simultaneously recorded movement data as image representation with the same processing approach to detect PD* and 2) *attempting to improve PD detection with the joint usage of the drawing and the simultaneously recorded motion data*.

The following section presents the **literature** on the problem statement, highlighting the key points. Then, in the **methods** section, we discuss the applied procedure. In the **results** and **discussion** section, we present and consider the possible outcomes. Finally, in the **summary and conclusion** section, we recap our work and highlight the essential findings and concerns.

II. LITERATURE STUDY

A requirement for technology to support diagnosis is that it does not overtax the patient. Ideally, this means short-in-time, simple, non-invasive tests. The machine learning algorithm makes these automatic evaluations possible, which can provide an objective output. This can significantly contribute to the physician's opinion and provide a universal measurement procedure.

One such test could be the recording and analysis of drawing or handwriting, which tries to capture the motor symptoms in PD patients' hands. Currently, this is not part of the criteria for diagnosis. However, McLennan et al. pointed out that 5% of patients have handwriting difficulties before the onset of motor symptoms, and 30% of those patients report deteriorating handwriting later [11].

Changes in fine motor skills, such as writing/drawing speed, continuity, and text or shape size, can be seen in PD patients. The reduction of the writing size is called micrography, of which two categories are consistent and progressive micrography [12]. Presumably, one and the other develop depending on the involvement of different brain areas and respond differently to dopaminergic drug treatment [13]. Similar changes appear in the drawings as well.

Many patterns are common, such as spirals, waves, or lines. They are mostly made on a tablet device, while it was more common to use traditional pen and paper in the past. The drawing should be simple so that it does not require special drawing skills but complex so that fine motor changes can be detected.

Kotsavasiloglou and his colleagues [14] conducted experiments involving 20 healthy (Healthy Control - HC) and 24 PD individuals by drawing a horizontal line. They extracted a number of features related to the speed of the pen tip and vertical deviations (deviations from a straight line). With multiple classifiers (Naïve Bayes, AdaBoost, Log. Regression, Support Vector Machine [SVM], Random Forest [RF], J48), 79.4-88.5% accuracy was achieved.

Sharma et al. [15] investigated the recognizability of PD by involving several databases using the Modified Gray Wolf Optimization algorithm on classifiers. Their drawing database was the Hand PD database created by Botucatu Medical School, São Paulo State University, which included 105 PD patients and 53 HC individuals (mean age 44.2 and 58.8, respectively). The classification algorithms were k-Nearest Neighbors (k-NN), RF, and Decision Tree (DT). With the help of predetermined features, 73.4-92.4% accuracy was achieved on the spiral drawings and 72.8-93.0% accuracy on the meander drawings.

In addition, further research deals with monitoring the drawing task with acceleration sensors and examining the usability of the data generated in this way for recognition.

Ali et al. [16] investigated Essential Tremor (ET) with acceleration data acquired while drawing spiral patterns. 17 ET patients and 18 HC individuals were included in the databases. Three sensors were placed at three points: on the dorsum of the hand, on the posterior forearm, and the posterior upper arm. SVM was used to classify the power spectral density (calculated from the acceleration data after creating a single vector magnitude). 74.3-85.7% accuracy was achieved.

Pereira et al. [17] used time series data from sensors mounted in a pen to detect PD. Their research used the HandPD database with 14 PD patients and 21 HC. Images were created from the sensor data (sound, finger grip, axial pressure of ink, acceleration and tilt data in X, Y, and Z directions). These were classified using image processing algorithms (ImageNet, CIRAF-10, LeNet) and a baseline model (Optimum-Path Forest [OPF]) on the raw data. 85.0-87.1% accuracy was achieved with ImageNet on meander data, and 77.9-83.8% accuracy with OPF on spiral drawings as the best performance.

Savalia and his colleagues [18] similarly used digital pen-provided time signals as images with image processing algorithms. They named their database newHandPD, which included 35 HC and 31 PD patients. With the help of Convolutional Neural Network (CNN) and EffNet-based classifiers, they achieved 84.8-88.8% accuracy.

Taleb et al. [19] experimented with several approaches: classification of raw signals from a pen and classification with a spectrogram (a 2D representation of the raw signals). They pointed out that combining several writing tasks improves the classifier's performance. Their best result with data augmentation was 97.6% accuracy.

Cascarano and his colleagues [20] examined the drawing patterns of 21 PD and 11 HC individuals. Descriptive characteristics were calculated from geometric, dynamic, and muscle activity data. 90.8% accuracy was achieved with spiral drawing without feature selection, while 93.8% accuracy was achieved with selection (with Multi-Objective Genetic Algorithm).

The literature study shows that the performance of the classifiers varies widely and may result from the database, feature extraction, and classification algorithms. Paper-based drawing also created a need to record dynamic data, which nowadays is easier to do with the help of tablets and digital pens. With this, the drawn pattern can be used as an image (the drawing itself) and as a set of time series. We also saw an approach where the researchers created images (2D representations) from time series, taking advantage of the performance of image processing algorithms. We could also find examples where combining several time signals/drawing tasks improves recognition. We wish to contribute to this with our present research, in which the participants draw a spiral pattern. During the process, a sensor attached to the wrist records the acceleration data. We examine these two modalities separately and together.

III. METHODOLOGY

A. Database

The database contains drawing and drawing-related sensor data of 45 PD patients and 47 HC individuals. All participants were informed beforehand about the research details and gave their consent by signing a consent form. Participants volunteered to participate in the research and were informed that their participation could be withdrawn without explanation.

There are 37 men and eight women in the PD class, whose average age is 66.0 years with a standard deviation of 14.2 years. Twenty-one people were recorded with drug onset, 18 people with Deep Brain Stimulation (DBS) switched on (10 people belonged to both categories). The severity score of PD patients was defined based on the Unified Parkinson's Disease Rating Scale (UPDRS) scale. The resting tremor (3.17), postural tremor (3.15), rigidity (3.3), and finger tapping (3.4) tasks were evaluated by the neurologist. The average severity was 1.10, where 0 means a healthy (normal) stage, and 4 means the severe stage. Severity scores for the right hand are presented since data from the right hand were examined.

There were 20 men and 27 women in the HC class, with an average age of 56.7 years and a standard deviation of 15.2. According to their admission, the members of the HC class did not have Parkinson's disease or any other disease that affects their movement.

The drawing and motion recordings were made using an application developed for Android tablets. In the case of the drawing, the participants followed the pattern of an Archimedean spiral by moving between the lines from the inside to the outside of the spiral template. The drawing data was sampled at a maximum of 110 Hz. A sensor attached to the wrist provided acceleration data in three dimensions (X, Y, and Z) with a sampling frequency of 50 Hz to record the movement.

Recordings were anonymized before use. Their metadata, such as gender and age, were used only to describe the classes.

B. Preprocessing

The drawings were plotted along X and Y coordinates and resized to 224x224 pixels with 24-bit depth. The average along the corresponding coordinate was subtracted from the acceleration data as standardization (remove the bias of the measurement device). Furthermore, a 4-order bandpass Butterworth filter was applied to the signals between 2 and 15 Hz. Finally, one data vector was created from the three coordinates, which contained the length of the vector pointing to the point described by the three coordinates at each instant of time. Two-dimensional representations were made, as shown in Fig. 1, such as *MarkovTransitionField*, *RecurrencePlot*, and *GramianAngularField*. For this data representation, the *pyts* (v0.12.0) Python package was used with default parameters [21]. These images were also resized to 224x224 pixels with 24-bit depth.

The feature vectors from the resulted input images were derived from the feature extraction part of pre-trained deep learning algorithms. For this, the Keras API was used with the *Tensorflow* (v2.0.0) machine learning platform. Among the available models, *Xception* [22], *ResNet50* [23], and *MobileNet* [24] were used. We removed the classification (*Dense*) layer from the end of each model and assigned a *GlobalAveragePooling2D* layer to get one-dimensional feature vectors. Finally, the feature vectors were standardized using the *StandardScaler* function of sklearn (v0.0.post7).

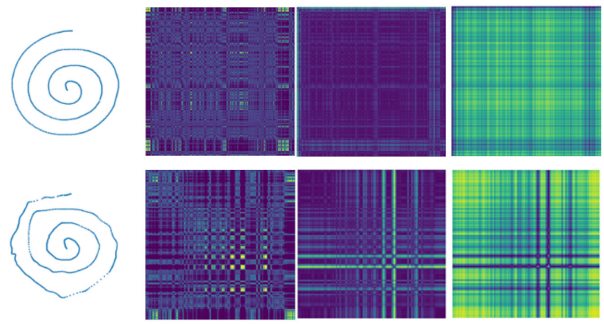

Fig. 1. The spiral drawing and the 2D movement representations for an HC person (upper images) and a PD patient (lower images).

C. Classification and Evaluation

To classify the feature vectors, SVM, RF, and k-NN classifiers were used with nested cross-validation at 10-fold numbers.

The test set (10% of the database) was separated in the outer cycle, independent of the model learning. On the remaining data, the best 100 features were selected based on ANOVA F-value [25]. The same features were selected in the test set accordingly.

The validation set (10% of the 90% data) was separated from the training dataset in the internal cycle. This training set was used to train the classifiers, and the optimization was carried out with the validation set. The parameters below were probed available for *sklearn* [26] models:

- *kernel* (linear, rbf), *C* value (0.001, 0.01, 0.1, 1, 10, 100), and *gamma* (10, 1, 0.1, 0.01, 0.001, 0.0001) for **SVM**,
- the *max*_depth (10, 30, 50, 70, 90, None), max*_features* (auto, sqrt), and *min_samples_leaf* (1, 2, 4) for **RF**,
- the *number of neighbors* (2, 3, 5, 10, 20) for **k-NN**.

The test sets were finally estimated with the models trained with the best parameters (on 90% of the data). The sensitivity, specificity, macro f1 score, and area under the receiver operating characteristic (ROC) curve (AUC) values were derived from the estimates.

The predictions of the two different modalities (drawing and movement) were aggregated according to Eq. 1, where $y_s \in \{0,1\}$ is the prediction from spiral drawing and $y_m \in \{0,1\}$ is the prediction from the movement data. $Q \in \{0,1\}$ is the weight factor between the two predictions, $y_{final} \in \{0,1\}$ is the final prediction.

If the value of Q is 0, then the final prediction is equal to the prediction from the movement. In the case of $Q = 1$, the final prediction is equal to the prediction from the drawing. The final prediction for Q between 0 and 1 is a ratio between the two modalities.

$$y_{final} = y_s Q + y_m(1 - Q) \quad (1)$$

The samples are assigned to classes based on the models' estimates with a decision limit 0.5. If the estimate is above this, the sample is classified as positive (PD) and below it as negative (HC).

Mann-Whitney U non-parametric test [27] was used to compare the performance of modalities and aggregations. The comparison was made with the macro f1 scores. The significance level was chosen as 0.05 [28].

IV. RESULTS

A. Separate Classification of Drawings and Movements

TABLE I shows the results achieved on the spiral drawings and the representation of the movement. The marking of the spiral drawing is *Spiral*, while the representations are named according to the names described in the III. Methodology section. The second column contains the names of the feature extractors, and the third column contains the classification algorithms. The last four columns contain the sensitivity (sens), specificity (spec), macro f1 (f1), and AUC (auc) values, respectively. The best results per representation are marked with **bold** style.

In the case of the *Spiral*, the MobileNet feature extractor resulted in the highest macro f1 score on average (67.9%). ResNet50 and Xception achieved an average of 59.7% and 57.6% macro f1, respectively. Regarding the classification algorithms, RB achieved the best macro f1 score on average (66.7%), while k-NN and SVM achieved macro f1 values of 59.5% and 59.0%, respectively. The best result among all cases was achieved with the MobileNet feature extractor and RF classification model for the spiral drawing (macro f1 score 70.6%).

In the case of *RecurrencePlot*, the feature extractors' results no longer deviate spectacularly from each other. MobileNet, ResNet50, and Xception achieved 67.8%, 63.9%, and 66.6% macro f1 scores, respectively. The result is similar according to the classifiers: 64.4% (k-NN), 65.2% (RF), and 68.8% (SVM). Compared with the data of the Spiral drawing, it can be seen that the values have improved on average, whether examining the feature extractors or the classifiers. The average difference between the two modalities is 4.4% in the macro f1 score. The *RecurrencePlot* approach obtained the best result with the MobileNet feature extractor and SVM classification model (74.9% macro f1 score). Compared to *Spiral's* best result, it provided a 4.3% better result. Regarding the other metrics, the specificity improved by an average of 0.7%, the sensitivity by 8.1%, and the auc value by 0.088 compared to the *Spiral* metrics.

In the *GramianAngularField* cases, the averages according to the feature extractors are 67.7% (MobileNet), 59.7% (ResNet50), and 61.2% (Xception) in macro f1 score. According to the classifiers, the average values are 61.1% (k-NN), 65.9% (RF) and 60.9% (SVM). In this case, the average difference compared to the *Spiral* is 0.9% in the macro f1 score. The best result was achieved with MobileNet feature extraction and RF classifier with a macro f1 score of 72.7%. This is 2.1% better than the *Spiral's* best result. Regarding the other metrics, the sensitivity improved by an average of 7.7% and the auc value by 0.039 compared to the *Spiral* metrics, but the specificity decreased by an average of 5.9%.

TABLE I
RESULTS OBTAINED WITH DIFFERENT MODALITIES, FEATURE EXTRACTORS, AND CLASSIFICATION ALGORITHMS.

	Extractor	Classifier	sens	spec	f1	auc
Spiral	**MobileNet**	k-NN	55.6%	78.7%	66.8%	0.686
		RF	**71.1%**	**70.2%**	**70.6%**	**0.757**
		SVM	64.4%	68.1%	66.3%	0.720
	ResNet50	k-NN	46.7%	78.7%	61.9%	0.574
		RF	60.0%	59.6%	59.8%	0.650
		SVM	53.3%	61.7%	57.5%	0.619
	Xception	k-NN	46.7%	53.2%	49.9%	0.533
		RF	68.9%	70.2%	69.6%	0.699
		SVM	53.3%	53.2%	53.3%	0.576
RecurrencePlot	**MobileNet**	k-NN	51.1%	80.9%	65.4%	0.752
		RF	64.4%	61.7%	63.0%	0.750
		SVM	**71.1%**	**78.7%**	**74.9%**	**0.834**
	ResNet50	k-NN	71.1%	46.8%	58.2%	0.689
		RF	66.7%	70.2%	68.4%	0.749
		SVM	64.4%	66.0%	65.2%	0.710
	Xception	k-NN	66.7%	72.3%	69.5%	0.721
		RF	71.1%	57.4%	64.0%	0.709
		SVM	66.7%	66.0%	66.3%	0.689
GramianAngularField	**MobileNet**	k-NN	75.6%	59.6%	67.3%	0.774
		RF	**80.0%**	**66.0%**	**72.7%**	**0.768**
		SVM	64.4%	61.7%	63.0%	0.720
	ResNet50	k-NN	62.2%	46.8%	54.2%	0.606
		RF	64.4%	66.0%	65.2%	0.685
		SVM	57.8%	57.4%	57.6%	0.635
	Xception	k-NN	60.0%	63.8%	61.9%	0.653
		RF	64.4%	55.3%	59.7%	0.664
		SVM	60.0%	63.8%	61.9%	0.659
MarkovTransitionField	MobileNet	k-NN	71.1%	51.1%	60.6%	0.635
		RF	75.6%	61.7%	68.4%	0.749
		SVM	51.1%	63.8%	57.4%	0.661
	ResNet50	k-NN	68.9%	70.2%	69.6%	0.757
		RF	**77.8%**	**74.5%**	**76.1%**	**0.802**
		SVM	75.6%	72.3%	73.9%	0.822
	Xception	k-NN	71.1%	55.3%	62.9%	0.634
		RF	64.4%	74.5%	69.4%	0.736
		SVM	66.7%	76.6%	71.6%	0.727

In the case of *MarkovTransitionField*, the average macro f1 scores are 62.1% for MobileNet, 73.2% for ResNet50 and 68.0% for Xception. Regarding to the classifiers, k-NN achieved 64.3%, RF 71.3%, and SVM 67.6% macro f1 score. Compared to the *Spiral*, ResNet50 and Xception performed better with 13.5% and 10.4% macro f1 scores, respectively. When comparing classification algorithms, on average, all three performed better (4.8%, 4.6%, and 8.6% better results for k-NN, RF, and SVM) in *MarkovTransitionField* representation than in the *Spiral*. Overall, this approach outperformed the *Spiral* by 6.0% in macro f1 score. The best result was obtained with the ResNet50 feature extractor with the RF classifier (76.1% macro f1 value). This is better with a 5.4% macro f1 score than the *Spiral's* best result. Regarding the other metrics, the specificity improved by an average of 0.7%, the sensitivity by 11.4%, and the auc value by 0.079 compared to the *Spiral* metrics.

Overall, it can be seen that the image representation of the movement provides a better result in average value: 4.4% (*RecurrencePlot*), 0.9% (*GramianAngularField*), 6.0% (*MarkovTransitionField*) in macro f1 score. According to the best results, better results were also achieved with movement representations by 4.3% (*RecurrencePlot*), 2.1% (*GramianAngularField*), and 5.4% (*MarkovTransitionField*) macro f1 score. However, ***no significant difference between the two modalities can be established*** with the Mann-Whitney U statistical test. The *p-values* of the tests are 0.331 (*Spiral vs RecurrencePlot*), 0.791 (*Spiral vs GramianAngularField*), and 0.102 (*Spiral vs MarkovTransitionField*). This probably stems from the observation that the two modalities have no clear trend according to the performance.

B. Joint examination of the modalities

Aggregating the predictions achieved on the modalities were based on Eq. 1. Fig. 2 shows the progression of macro f1 scores by connecting *Spiral* and *RecurrencePlot.* The weight factor or voting factor (Q) is shown on the horizontal axis. If Q is zero, then only the prediction of the movement. If $Q = 1$, then the prediction of the drawing applies as the two extreme points of the axis. The macro f1 scores between 0.5 and 0.9 are shown on the vertical axis.

The average macro f1 score achieved with *RecurrencePlot* is 66.1%, while with *Spiral* it is 61.7%. The average of the maximum points shown in the figure was 71.4% macro f1 score. The average of the maximum deviation (maximum point – modality with lower performance) was 11.7% macro f1 score, while the average of the minimum deviation (maximum point – modality with higher performance) was 3.2%. In two cases, no improvement was observed by combining the two modalities: Xception with k-NN classifier and Xception with SVM classifier. The possible reason is that the difference between the two modalities is high (19.6% and 13.0%). In the other cases, the difference between the modalities was minor (6.2% on average). The most considerable improvement was achieved with the MobileNet and SVM classifier. The macro f1 score on the *RecurrencePlot* is 74.9%. On the *Spiral* it is 66.3%, while the maximum value is 82.5%. By comparing the modalities with the maximum points pairwise using the Mann-Whitney U test, we experienced a ***significant improvement***. The p-value is 0.010 between the maximum and the *Spiral*, and 0.047 between the maximum and the *RecurrencePlot*.

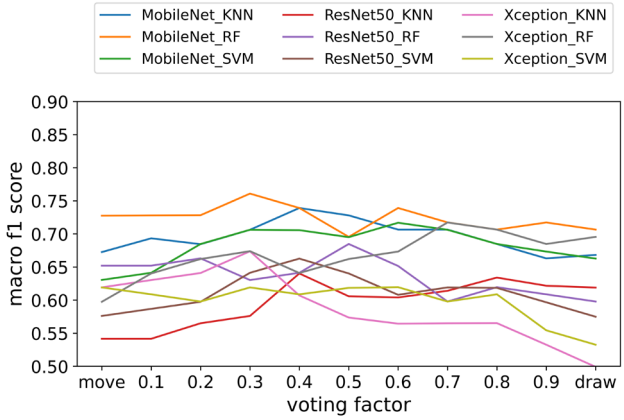

Fig. 3. The aggregated macro f1 scores of *Spiral* and *GramianAngularField* in proportion to the voting factor.

Fig. 3 shows the result of aggregating *Spiral* and *GramianAngularField*. The markings in the figure are the same as those in Fig. 2. In this case, the average macro f1 score of the *GramianAngularField* is 62.6%. This macro f1 score for the *Spiral* is 61.7%. The average of the maximum scores is 69.1%. In all cases, improvement was observed with the prediction aggregation. The maximum improvement is an average of 9.6% macro f1 score, and the minimum improvement is an average of 4.1%. With the present approach, the highest macro f1 score was provided by MobileNet and RF, with 76.1%. The same case achieved 72.7% on *GramianAngularField* and 70.6% on *Spiral*. ResNet50 achieved the most significant improvement (from both modalities) with the SVM classifier. It achieved 57.6% macro f1 score on movement data, 57.5% on drawing data, and 66.3% at the maximum point. We found a ***significant improvement*** by combining the modalities using the Mann-Whitney U test. The p-values are 0.030 between the maximum and the *Spiral* and 0.017 between the maximum and the *GramianAngularField*.

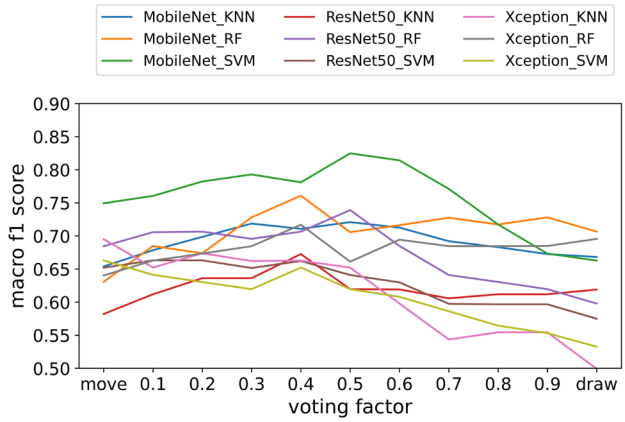

Fig. 2. The aggregated macro f1 scores of *Spiral* and *RecurrencePlot* in proportion to the voting factor.

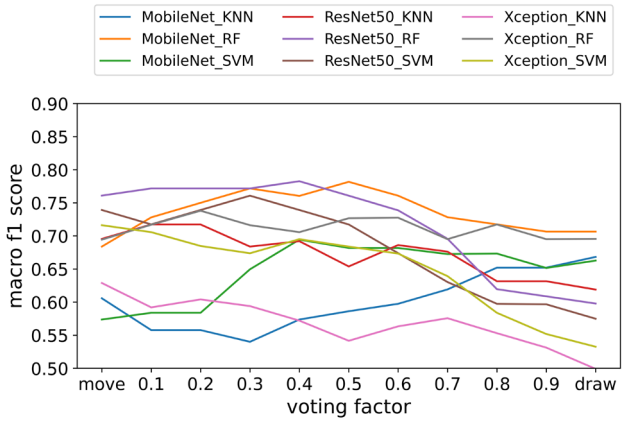

Fig. 4. The aggregated macro f1 scores of *Spiral* and *MarkovTransitionField* in proportion to the voting factor.

Fig. 4 shows the *Spiral* and the *MarkovTransitionField* representation. The markings of the figure are the same as those of the previous two figures. The average macro f1 score of the movement is 67.8%, that of the drawing is 61.7%, while that of the maximum points is 71.5%. The biggest change is 11.7% in macro f1 score on average, while the minimum is 1.8%. In this case too, we experienced cases where aggregation did not improve the results: Xception with k-NN classifier, Xception with SVM classifier and MobileNet with k-NN classifier. The first two showed no improvement using the *RecurrencePlot* either. The results of the two modalities show a similarly large difference compared to each other as with the *RecurrencePlot*. In the case of MobileNet k-NN, the difference between the two modalities is not that large, but it does not show improvement. The highest result was obtained with the ResNet50 feature extraction and the RF classifier in the macro f1 score of 78.3%. It achieved 76.1% on motion and 59.8% macro f1 score on drawing. This was approached by the MobileNet and RB with a macro f1 score of 78.2%.

The same case achieved 68.4% on movement and 70.6% on drawing. With the Mann-Whitney U test, we found a ***significant improvement*** between the maximum and *Spiral*, ***but not*** between the maximums and the *GramianAngularField*. The p-value is 0.009 between maximums and drawing results and 0.102 between maximums and movement results.

TABLE II summarises the best results of the single and joint modalities. The single modalities are marked with the names from TABLE I. The joint results are marked with the name of the representation approaches. The descriptor metrics are analogous to TABLE I. The movement representations appeared superior to the spiral drawing from the single use of the modalities. Utilizing joint predictions, improvements were achieved. The top performance was received with *RecurrencePlot* (sens: 75.6%, spec: 89.4%, macro f1 score: 82.5%).

TABLE II
SUMMARY TABLE OF THE BEST RESULTS OBTAINED FROM THE SINGLE AND JOINT MODALITIES.

	case	sens	spec	bacc	f1	auc
Single	Spiral	71.1%	70.2%	70.7%	70.6%	0.757
	RecurrencePlot	71.1%	78.7%	75.0%	74.9%	0.834
	GramianAngularField	80.0%	66.0%	72.8%	72.7%	0.768
	MarkovTransitionField	77.8%	74.5%	76.1%	76.1%	0.802
Joint	**RecurrencePlot**	**75.6%**	**89.4%**	**82.6%**	**82.5%**	**0.825**
	GramianAngularField	77.8%	74.5%	76.1%	76.1%	0.761
	MarkovTransitionField	80.0%	76.6%	78.3%	78.3%	0.783

V. DISCUSSION

Examining the modalities separately shows that the representations created from movement performed better than the drawing when examining the best results per modality (TABLE II). However, looking at the single results as a whole, we did not find any significant differences using the Mann-Whitney U test. This can be influenced by the methodology for generating the representations and the nature of the feature extraction models. It shows that the different data types generally have the same detection performance with these vision-based examinations on the applied classifiers. This may be a consideration when the physician wants to use a minimal tool in the shortest possible time. Nevertheless, it can be seen from TABLE II that a few percent better results can be achieved by selecting the best-performing models with motion data.

By using the modalities together, we experienced an improvement in all paired cases. This improvement proved significant in the *RecurrencePlot* and the *GramianAngularField*, whether we compared the best values to the drawing or the movement. In the case of the *MarkovTransitionField*, the improvement only reached a significant result compared to the drawing. This suggests that the decision is more confident and accurate when multiple sensors are used for the same task (even with the same processing scheme). However, by observing the voting factor, different optimal points may result. This implies that the modalities may play a different role in the final decision.

The present results provide insight into how the same task can provide better recognition by examining several sensors. There are also studies in the literature where, for example, video cameras and motion sensors help to recognize PD-related episodes. There is a multi-sensory examination of the same task similarly.

The limitation of the research is the heterogeneous database. A database with more elements may be necessary for filtering according to various factors (medication, brain stimulation). Another direction of development is sex equality. Although research [29] shows that the drawing of the Archimedean spiral does not differ significantly between men and women, this requires further support. Studies [30] are underway to investigate the separability of the sexes, where the authors have already shown that there is a significant difference when copying shapes. However, most studies have looked at participants between 18 and 30 years old and have used manual features that have not been linked to Parkinson's disease. However, this raises the need for further investigation.

Finally, it should be mentioned that the present results are based on image data processing with out-of-domain feature extraction algorithms. These play a role in avoiding overlearning by extracting a more general set of features. Presumably, fine-tuning the extractor models on the database specific to the task can increase the performance of the classifiers. However, a database with few elements can also cause overlearning.

VI. SUMMARY AND CONCLUSION

PD is becoming one of the most common neurological diseases of our time. The importance of research related to it is that there is no cure according to current clinical knowledge.

Current clinical knowledge is firmly based on motor symptoms, which are typically limb tremors at rest, bradykinesia, and muscle stiffness (rigidity). Recognizing them in the early stages is also not clear since the appearance of the symptoms and the lateral involvement may differ from person to person.

The support of artificial intelligence (including machine learning algorithms) can be desirable in the diagnostic procedure since it can effectively recognize even minor deviations and can create a more objective evaluation.

Previously, manually extracted features were used to analyze movements and drawings, which proved effective for learning the nature of the disease. On the other hand, deep learning algorithms allow the model to learn the characteristics that support automatic feature extraction. From this point of view, MobileNet, Xception, and ResNet50 were used, and they were previously trained on an extensive database that enabled cross-sectional recognition. We used networks trained in this way to recognize the disease, keeping the original weights.

In the literature research, it can be seen that several modalities are available to recognize PD, such as speech, drawing, or movement. These can achieve significant recognition performance by themselves. Comparing them is difficult because each modality goes through a different processing process, and there are also differences in the databases. In addition, the joint application of several modalities seems to improve PD recognition.

In our work, we examined the spiral drawings in connection with diagnosing the disease through two types of modalities. In the spiral drawing task, the X and Y coordinates of the actual drawing were recorded for each person, as well as the acceleration data along the X, Y, and Z axes with a wrist-mounted sensor.

Image representations were created from the data: 1) production of the actual drawing from X, and Y drawing data, 2) image representations from the resulting vector of movement data (*MarkovTransitionField*, *RecurrencePlot*, *GramianAngularField*). From these input images, we determined features with the pre-trained models, and then we trained classification algorithms on the features using nested cross-validation.

Based on the single modality results, image representations of the ***movement data reached at least a similar performance as the drawing itself*** (no significant difference). This result suggests that using only the spiral drawing alone or with the same processing of only the motion data from the drawing, PD. can be detected with the same efficiency.

Furthermore, the joint approach improved the recognition performance of the PD (significance difference)**, *highlighting the possibility of measuring the same task with various sensors.*** The result shows that even though the task the participant performs is the same, improving detection with different sensors is still possible.

Further investigation is required regarding the database composition.

REFERENCES

[1] R. Balestrino and A. H. V. Schapira, 'Parkinson disease', *Euro J of Neurology*, vol. 27, no. 1, pp. 27–42, Jan. 2020, DOI: 10.1111/ene.14108.

[2] T. Pringsheim, N. Jette, A. Frolkis, and T. D. L. Steeves, 'The prevalence of Parkinson's disease: A systematic review and meta-analysis', *Movement Disorders*, vol. 29, no. 13, pp. 1583–1590, Nov. 2014, DOI: 10.1002/mds.25945.

[3] R. F. Pfeiffer, 'Non-motor symptoms in Parkinson's disease', *Parkinsonism & Related Disorders*, vol. 22, pp. S119–S122, Jan. 2016, DOI: 10.1016/j.parkreldis.2015.09.004.

[4] S. Sveinbjornsdottir, 'The clinical symptoms of Parkinson's disease', *Journal of Neurochemistry*, vol. 139, no. S1, pp. 318– 324, Oct. 2016, DOI: 10.1111/jnc.13691.

[5] M. Gil-Martín, J. M. Montero, and R. San-Segundo, 'Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks', *Electronics*, vol. 8, no. 8, p. 907, Aug. 2019, DOI: 10.3390/electronics8080907.

[6] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, and J. I. Godino-Llorente, 'Advances in Parkinson's Disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects', *Biomedical Signal Processing and Control*, vol. 66, p. 102418, Apr. 2021, DOI: 10.1016/j.bspc.2021.102418.

[7] Z. Li, J. Yang, Y. Wang, M. Cai, X. Liu, and K. Lu, 'Early diagnosis of Parkinson's disease using Continuous Convolution Network: Handwriting recognition based on off-line hand drawing without template', *Journal of Biomedical Informatics*, vol. 130, p. 104 085, Jun. 2022, DOI: 10.1016/j.jbi.2022.104085.

[8] N. Basnin, T. A. Sumi, M. S. Hossain, and K. Andersson, 'Early Detection of Parkinson's Disease from Micrographic Static Hand Drawings', in *Brain Informatics*, vol. 12960, M. Mahmud, M. S. Kaiser, S. Vassanelli, Q. Dai, and N. Zhong, Eds., in Lecture Notes in Computer Science, vol. 12960, Cham: Springer International Publishing, 2021, pp. 433–447. DOI: 10.1007/978-3-030-86993-9_39.

[9] B. Schoneburg, M. Mancini, F. Horak, and J. G. Nutt, 'Framework for understanding balance dysfunction in Parkinson's disease', *Movement Disorders*, vol. 28, no. 11, pp. 1474–1482, Sep. 2013, DOI: 10.1002/mds.25613.

[10] S. Del Din, A. Godfrey, C. Mazzà, S. Lord, and L. Rochester, 'Free-living monitoring of Parkinson's disease: Lessons from the field: Wearable Technology for Parkinson'S Disease', *Mov Disord.*, vol. 31, no. 9, pp. 1293–1313, Sep. 2016, DOI: 10.1002/mds.26718.

[11] J. E. McLennan, K. Nakano, H. R. Tyler, and R. S. Schwab, 'Micrographia in Parkinson's disease', *Journal of the Neurological Sciences*, vol. 15, no. 2, pp. 141–152, Feb. 1972, DOI: 10.1016/0022-510X(72)90002-0.

[12] M. Thomas, A. Lenka, and P. Kumar Pal, 'Handwriting Analysis in Parkinson's Disease: Current Status and Future Directions', *Movement Disord Clin Pract*, vol. 4, no. 6, pp. 806–818, Nov. 2017, DOI: 10.1002/mdc3.12552.

[13] A. W. A. Van Gemmert, H.-L. Teulings, and G. E. Stelmach, 'Parkinsonian Patients Reduce Their Stroke Size with Increased Processing Demands', *Brain and Cognition*, vol. 47, no. 3, pp. 504–512, Dec. 2001, DOI: 10.1006/brcg.2001.1328.

[14] C. Kotsavasiloglou, N. Kostikis, D. Hristu-Varsakelis, and M. Arnaoutoglou, 'Machine learning-based classification of simple drawing movements in Parkinson's disease', *Biomedical Signal Processing and Control*, vol. 31, pp. 174–180, Jan. 2017, DOI: 10.1016/j.bspc.2016.08.003.

[15] P. Sharma, S. Sundaram, M. Sharma, A. Sharma, and D. Gupta, 'Diagnosis of Parkinson's disease using modified grey wolf optimization', *Cognitive Systems Research*, vol. 54, pp. 100–115, May 2019, DOI: 10.1016/j.cogsys.2018.12.002.

[16] S. M. Ali et al., 'Wearable sensors during drawing tasks to measure the severity of essential tremor', *Sci Rep*, vol. 12, no. 1, p. 5242, Mar. 2022, DOI: 10.1038/s41598-022-08922-6.

[17] C. R. Pereira, S. A. T. Weber, C. Hook, G. H. Rosa, and J. P. Papa, 'Deep Learning-Aided Parkinson's Disease Diagnosis from Handwritten Dynamics', in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, Sao Paulo, Brazil: IEEE, Oct. 2016, pp. 340–346. DOI: 10.1109/SIBGRAPI.2016.054.

[18] J. Savalia, S. Desai, R. Geddam, P. Shah, and H. Chhikaniwala, 'Early-Stage Detection Model Using Deep Learning Algorithms for Parkinson's Disease Based on Handwriting Patterns', in *Advancements in Smart Computing and Information Security*, vol. 1759, S. Rajagopal, P. Faruki, and K. Popat, Eds., in Communications in Computer and Information Science, vol. 1759, Cham: Springer Nature Switzerland, 2022, pp. 323–332. DOI: 10.1007/978-3-031-23092-9_26.

[19] C. Taleb, L. Likforman-Sulem, C. Mokbel, and M. Khachab, 'Detection of Parkinson's disease from handwriting using deep learning: a comparative study', *Evol. Intel.*, vol. 16, no. 6, pp. 1813–1824, Dec. 2023, DOI: 10.1007/s12065-020-00470-0.

[20] G. D. Cascarano et al., 'Biometric handwriting analysis to support Parkinson's Disease assessment and grading', *BMC Med Inform Decis Mak*, vol. 19, no. S9, p. 252, Dec. 2019, DOI: 10.1186/s12911-019-0989-3.

[21] J. Faouzi, H. Janati, K. K. LEE, T. Carryer, R. Yurchak, and AvisP, 'johannfaouzi/pyts: Release of version 0.10.0'. Zenodo, Dec. 09, 2019. DOI: 10.5281/ZENODO.3568218.

[22] F. Chollet, 'Xception: Deep Learning with Depthwise Separable Convolutions', 2016, DOI: 10.48550/ARXIV.1610.02357.

[23] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep Residual Learning for Image Recognition', 2015, DOI: 10.48550/ARXIV.1512.03385.

[24] A. G. Howard *et al.*, 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications', 2017, DOI: 10.48550/ARXIV.1704.04861.

[25] H.-Y. Kim, 'Analysis of variance (ANOVA) comparing means of more than two groups', *Restor Dent Endod*, vol. 39, no. 1, p. 74, 2014, DOI: 10.5395/rde.2014.39.1.74.

[26] F. Pedregosa et al., 'Scikit-learn: Machine Learning in Python', 2012, DOI: 10.48550/ARXIV.1201.0490.

[27] P. E. McKnight and J. Najab, 'Mann-Whitney U Test', in *The Corsini Encyclopedia of Psychology,* 1st ed., I. B. Weiner and W. E. Craighead, Eds., Wiley, 2010, pp. 1–1. DOI: 10.1002/9780470479216.corpsy0524.

[28] B. M. Cesana, 'What p-value must be used as the Statistical Significance Threshold? P<0.005, P<0.01, P<0.05 or no value at all?', *BJSTR*, vol. 6, no. 3, Jul. 2018, DOI: 10.26717/BJSTR.2018.06.001359.

[29] M. San Luciano et al., 'Digitized Spiral Drawing: A Possible Biomarker for Early Parkinson's Disease', *PLoS ONE*, vol. 11, no. 10, p. e0162799, Oct. 2016, DOI: 10.1371/journal.pone.0162799.

[30] G. Cordasco et al., 'Gender Identification through Handwriting: an Online Approach,' *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Mariehamn, Finland, 2020, pp. 000197-000202, DOI: 10.1109/CogInfoCom50765.2020.9237863.

Attila Zoltán Jenei was born in Debrecen, Hungary in 1995. He graduated from the Budapest University of Technology and Economics as a Biomedical Engineer (Master's Degree, 2020). Since January 2020, he has been a department engineer and Ph.D. student at the Laboratory of Speech Acoustics, Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Information Technology. His research focuses on diagnostic support for Parkinson's disease with non-invasive medical data. He participated in the Student Research Societies of Budapest University of Technology and Economics and was awarded in 2017 and 2019. From 2021, he is the Vice President, and from 2023, the President of the Department of Engineering Sciences in the National Association of Doctoral Students.

Dávid Sztahó is a research fellow at the Budapest University of Technology and Economics. He completed his MSc studies in 2008 in Infromatics engineering. Since 2018 he is the head of the Laboratory of Speech Acoustics. He completed his PhD studies at the Doctoral School of Computer Science of the Budapest University of Technology and Economics, in the field of emotion recognition by speech signal. He received his PhD degree in 2014. His research interests include: speech technology, speech acoustics, speaker recognition and verification for forensic purposes, biomarker analysis by artificial intelligence, computer analysis of EEG signals.

István Valálik MD., Ph.D., MSc. neurosurgeon and head physician of the Department of Neurosurgery at St. John's Hospital, Budapest, honorary associate professor at the University of Debrecen. In 2011 he defended PhD thesis "CT-guided stereotactic thermolesion and deep brain stimulation in the treatment of Parkinson's disease", in 2015 MSc in Health Services Management. His scientific interest focused on movement disorders, MR-tractography-based surgical planning, psychiatric surgery, acoustic and motion analysis. He developed a planning software for stereotactic brain surgery and portable neuro-navigation system. Since 2010 he is acting in the Executive Committee of the European Society for Stereotactic and Functional Neurosurgery (www.essfn.org). In 2013 he was awarded by the Hungarian Academy of Sciences for the book "Stereotactic and Functional Neurosurgery". In 2019 he participated in mission of successful surgical separation of Bangladeshi craniopagus twins.