Kamila BIAŁEK[1], Anna POTULSKA-CHROMIK[2], Jacek JAKUBOWSKI[1], Monika NOJSZEWSKA[2], Anna KOSTERA-PRUSZCZYK[2]

Military University of Technology, Department of Electronics (1), Medical University of Warsaw, Department of Neurology (2)

doi:10.15199/48.2025.02.03

The application of sentence sequences in the diagnosis of Parkinson's disease from handwriting samples using convolutional neural networks

***Abstract**.. This paper presents the application of convolutional neural networks with the aim to recognize Parkinson's disease from handwriting samples. A method to improve binary classification results in disease diagnosis is presented using a series of images representing individual sentences derived from their sequence recording.*

Streszczenie.** W artykule zaprezentowano zastosowanie konwolucyjnych sieci neuronowych do rozpoznawania choroby Parkinsona na podstawie próbek pisma. Przedstawiono metodę poprawy wyników klasyfikacji binarnej w diagnostyce choroby poprzez użycie serii obrazów prezentujących pojedyncze zdania pochodzące z zapisu ich sekwencji. (**Zastosowanie sekwencji zdań w rozpoznawaniu choroby Parkinsona na podstawie próbek pisma z wykorzystaniem konwolucyjnych sieci neuronowych)

Keywords: convolutional neural networks, neural networks, Parkinson's disease, diagnostics, handwriting
Słowa kluczowe: konwolucyjne sieci neuronowe, sieci neuronowe, choroba Parkinsona, diagnostyka, pismo odręczne

Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by impaired motor function. The main symptoms of Parkinson's disease include resting tremor, limb stiffness, slow movement, and postural instability. However, these symptoms occur to a sufficient degree to clearly diagnose the disease only at a late stage of development [1]. Diagnosing Parkinson's disease based on early symptoms can be difficult, as evidenced by the long delay (on average 10 years) between the first noticeable symptom and the moment of diagnosis [2]. Therefore, numerous studies are being conducted using engineering methods in the field of early diagnosis of Parkinson's disease based on various modalities. A human skill that has diagnostic potential is handwriting. Writing is a simple and natural activity. Additionally, its registration does not require special environmental conditions, a sense of timing, or many exhaustive repetitions. The motivation for research in this area is the fact that Parkinson's disease is a neurological disorder that causes problems with the efficiency of the musculoskeletal system. As a result, the appearance of any of the motor symptoms may affect graphomotor skills, i.e. the ability to write by hand. Studies show that people with Parkinson's disease write slower than healthy controls [3]. Their writing is characterized by greater baseline inaccuracy, is more chaotic, and irregular [4]. People with Parkinson's disease do not control the pressure force when writing to the same extent as healthy people [5]. Therefore, changes in handwriting and problems with writing may be the first noticeable signs of Parkinson's disease.

Related works and proposed method

The first works on handwriting analysis in terms of diagnosing Parkinson's disease concerned the use of statistical analysis tools. Differences in handwriting and the way it was created in people with and without Parkinson's disease were indicated [3], [4], [6]. In these studies, a small set of parameters was calculated based on the handwriting samples, which were the basis for further analysis. The intensive development of machine learning techniques resulted in the automatic classification of people with and without Parkinson's disease in handwriting studies using modern classification algorithms [7,8,9]. The appearance on the market of electronic recording devices, such as graphic tablets, enabled parameterization not only of handwriting as a trace visible on paper but also of the process of its creation, taking into account the kinematics and dynamics of this process. There are many publications on determining an increasing number of handwriting parameters that are supposed to indicate the occurrence of Parkinson's disease. The most frequently used include: length and width of subsequent strokes, writing time, writing speed, acceleration, and average pressure. Finding a suitable set of features for a satisfactory classification of patients requires determining a very large number of handwriting parameters and using an appropriate discriminative algorithm. Modern deep neural network solutions make it possible to take a different approach, based not on feature vectors but directly on entire images, in search of static visual information. The latest studies conducted for diagnosing Parkinson's disease based on handwriting samples use convolutional neural networks, which are one of the types of deep networks that automatically generate (without human intervention) a set of features characteristic of the analyzed image patterns. However, research in the current literature conducted in this area is based on the previously created PaHaW database, which contains handwriting samples recorded during the performance of eight different tasks, including spiral drawing, single letters, and a whole sentence [10, 11, 12, 13]. However, the modern literature lacks the use of convolutional neural networks for handwriting analysis for Parkinson's disease on new handwriting samples presenting words or sentences, without taking into account additional tasks such as spiral drawing, which is not typical handwriting. The use of a tool such as convolutional neural networks in such problems requires having a sufficiently large set of training data, which in the case of obtaining data from patients with PD is not a trivial task. Due to the small number of input data, the authors of the works used transfer learning methods, using pre-trained network structures such as AlexNet [10,12], VGG16 [11,13] and SqueezeNet [13]. To improve recognition results, authors use augmentation techniques to artificially increase the number of data [12,13]. This is done most often by modifying existing examples. In previous works on this topic, due to the availability of a specific database, the focus was solely on the analysis of single words or sentences. The authors of this paper present a different approach consisting of registering more than one sample from one examined person. This allows one to increase the number of authentic data without the need to engage new patients. Such a planned method of data registration results primarily from the fact that changes in handwriting resulting from the presence of the disease in people at an early stage of its occurrence may become visible only as a result of a long-term writing process. However, the use of traditional augmentation methods or the use of a larger number of sentences, but registered at certain intervals, does not take into account changes in handwriting resulting from continuous writing. The use of several images presenting subsequent sentences written by a given person one after another, within the same registration, in the process of training convolutional networks may result in achieving higher disease recognition results. In the following work, sentence images from the recording of their sequence were used to train the structure of the CNN network and compared with the results obtained using only a single sentence.

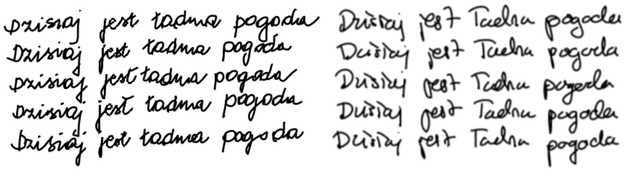
Material

The handwriting image data for the analysis were obtained as part of clinical studies conducted by a medical team at the Warsaw Medical University Clinic and Department of Neurology. A total of 48 people participated in the study. The study group included 24 healthy HC (11 women and 13 men) and 24 with diagnosed Parkinson's disease PD (16 women and 8 men). The group of PD patients with diagnosed Parkinson's disease included people in various stages of the disease. The patients were pharmacologically prepared for the study in such a way that their condition corresponded to the early stage of the disease. All study participants were right-handed. The basic characteristics of both groups are presented in Table 1.

Table 1. Basic characteristics of the research group (PD) and the healthy control group (HC)

volunteers	male	female	total	age
PD	8	16	24	28-84
HC	13	11	24	26-84

The Intous Pro Paper Edition PTH-860 graphics tablet from WACOM was used to record the writing. The tablet allows recording the coordinates of the pen location on the tablet surface, as well as the pressure force and the angles of the pen tilt while writing. During the study, patients were asked to write the sentence five times: "The weather is nice today" (in the national language of the patients, it was "Dzisiaj jest ładna pogoda"). To record changes resulting from the presence of the disease and writing in a spontaneous manner, all sentences were written without a break, one under the other, within the same recording. Also, the size and layout of the writing on the sheet of paper were arbitrary and depended only on the preferences of the person writing.



a) b)

Fig.1. Examples of handwriting samples obtained during the study: a) healthy subjects, b) subjects with Parkinson's disease

Based on the data recorded by the tablet, images presenting single sentences were generated. The drawn images, due to the network requirements regarding the input data dimension, were scaled to be 227x227x3. Example images used as input data for the convolutional network are shown in Figure 2.

In this way, 240 images were obtained, 5 from each of the study participants.

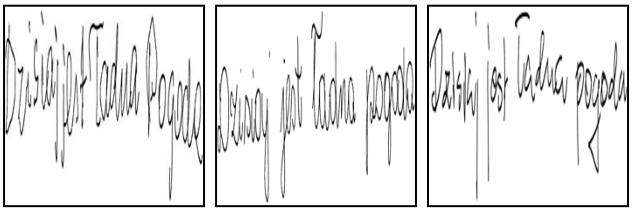


Fig.2. Sample images of single sentences used as network input

Convolutional neural networks (CNN)

Convolutional neural networks (CNNs) are a type of deep network commonly used in image processing today. In fact, these networks are a combination of an unsupervised process of automatic generation of diagnostic image features using many hidden convolutional layers with a final classifier, usually in the form of a softmax solution [14]. Therefore, two parts can be distinguished in the CNN structure: a series of subsequent convolutional layers that provide a set of diagnostic features at the end, whose neurons are connected locally to a small area of input data for a given layer, and a final part of the network with full neuron connection, constituting the actual classifier. The initial convolutional layers are designed to detect certain primitive image elements, the so-called low-level features, e.g. edges or characteristic pixel clusters. Deeper layers, performing similar operations on the output images of the preceding layer, recognize increasingly detailed image features relevant to the task being solved. As a result of this processing, the last convolutional layer contains features of the integrated process that can be input attributes for the actual classifier [15]. The construction of one's own sufficiently developed CNN network structure and its application to solving classification tasks requires the use of a sufficiently large database of patterns. In the absence of such a database, it is possible to use an alternative approach, the so-called transfer learning, which consists of using a network trained to solve a completely different task as a base structure and retraining this structure using a new training database. To do this, the final layers of the network are removed, replacing them with one's own untrained structure, adapted to the current task newly solved, and then the retraining process is carried out using a new database [14]. Fine-tuning can only concern newly added layers and layers of the base model, while freezing the remaining ones. There are many networks trained on a large database, such as AlexNet, GoogleNet, VGG-16, Inception3. Their structures differ in the number of layers and degree of structure development. In this work, the AlexNet network with 5 convolutional layers and 3 fully connected layers was used, which was originally created to distinguish objects belonging to 1000 different classes. In order to adapt the network to the task of recognizing Parkinson's disease, its last three layers were removed, i.e. the fully connected, softmax and classification layers, and replaced with new ones, which, together with the remaining layers, were trained in the network training process. The new fully connected layer contains two neurons, so it distinguishes two categories: "PD" and "Healthy".

Methods Used to Quantify the Recognition Process

In order to reliably evaluate the classification process, the commonly used k-fold cross-validation method was used. This method consists in randomly dividing the entire data set into k subsets of approximately equal size, and then using a single set as validation data and the remaining subsets as training data. Then, the process is repeated k times, changing the validation and training sets each time. The classification results obtained using the validation subsets are then averaged. In the studies described in this article, an

8-fold cross-validation was used. With data from a total of 48 people, in order to use the test and training sets independently of the person, a single subset in the cross-validation contained handwriting images from 6 people, respectively, from 3 people from each class. This means that in each validation, data from 42 people will be used for training and data from 6 people for validation. The recognition results were expressed using the following measures: Accuracy, Sensitivity, Specificity, Precision, defined as:

(1) $$ACC = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100\%.$$

(2) $$Se = \frac{TP}{TP+FN} \cdot 100\%.$$

(3) $$Sp = \frac{TN}{TN+FP} \cdot 100\%.$$

(4) $$Precision = \frac{TP}{TP+FP} \cdot 100\%.$$

where TP (True Positive) represents the number of people correctly classified as patients with PD, FP (False Positive) represents the number of healthy people wrongly classified as ill, TN (True Negative) represents the number of correctly classified healthy people, and FN (False Negative) represents the number of patients with PD wrongly classified as healthy.

Results

At first, only the first sentences written by each of the subjects were used to train the network. The aim was to check the results of the accuracy of the disease recognition on handwriting samples from our own database and compare them with the results of similar studies described in the literature, conducted using data from the PaHaW database. One of the writing tasks available in this data set is a single sentence in Czech: *"Tramwaj dnes už nepojede."* (*The tram will not go today*). The results of Parkinson's disease recognition based on this sentence can be compared to the results obtained in this work. The process of writing the first sentence from the entire sequence of five sentences is adequate to write only one sentence, as was the case with the PaHaW database. Table 2 presents the results of Parkinson's disease recognition based on handwriting samples in the form of average accuracy from 8 validations.

Table 2. Parkinson's disease diagnosis results based on the first sentence

Accuracy	60,42 %
Sensitivity	57,71 %
Specificity	62,50 %
Precision	58,33 %

Using images of only one of the sentences written by each person at fine-tuning, the result of disease recognition accuracy was 60.42%. This result is not satisfactory. However, the results obtained by the authors of other publications are also not favorable. In the work [10], the recognition accuracy obtained by the authors in a similar experiment, depending on the representations used, ranged between 48 and 51%. The authors of [11] obtained a slightly better classification result (67.08%) using a different method of presenting data from the tablet on the image. The authors of [12] obtained a result of 54.95% for the analogue study despite the use of classic augmentation.

In the next experiment, images of all sentences written by each of the subjects were used in the training process. Using a sequence of sentences instead of a single sentence primarily leads to multiplying the number of training data, which is a very desirable phenomenon in the case of learning systems. Additionally, the duration of the writing process can have a positive effect on the classification result. This results from changes in writing related to the presence of the disease, which can appear or become more pronounced during long-term graphomotor effort. Introducing training data to the network in the form of images presenting a full sequence of single sentences allows these changes to be visible and supports the network training process.

The final classification decision for a given individual was made by majority vote from a decision reached for each of the five images of that individual. Figure 4 presents a scheme to generate a classification decision.

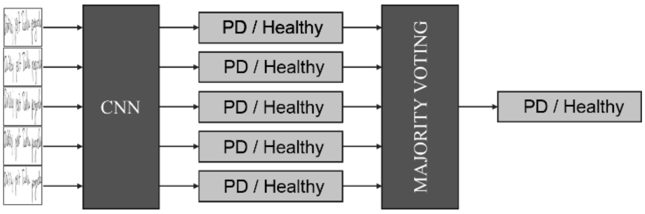


Fig. 4. General scheme of making classification decision

Table 3 presents the results of the diagnosis of Parkinson's disease based on handwriting samples in the form of average accuracy of eight validations.

Table 3. Parkinson's disease diagnosis results based on the sentences sequence

Accuracy	81,25 %
Sensitivity	86,46 %
Specificity	79,17 %
Precision	83,33 %

When analysing the above results, it can be seen that using a sequence of sentences from a single recording in the process of training and testing the network gives better results in the form of higher classification quality measures than using a single sentence. In the case of using only single sentences from each of the subjects, the accuracy of recognizing Parkinson's disease was obtained to be 60.42%. Increasing the amount of data by adding to the set of images four additional ones which extend the number of sentences written by each person to five and using an appropriate decision strategy improved the recognition accuracy to 81.25%.

The question remains whether the increase in disease recognition accuracy by more than 20 percentage points is caused only by the increase in the number of data or also by changes in handwriting resulting from writing more than one sentence. To resolve this, an additional experiment was conducted in which the number of training data was multiplied by using classical augmentation. In the case of handwriting, the number of operations that can be performed on the image to perform augmentation is limited. The images created as a result of such operations must be suitably representative. Using mirror reflection or rotation of the handwriting to a large extent will create data that incorrectly illustrate the handwriting. Also, scaling or shifting in the case of handwriting that presents a full sentence will change the pattern. Therefore, the only rational approach seems to be to rotate the sentence to a certain small extent. The aim of augmentation in this case was to create the same number of training data as in the case of using a sequence of sentences. This meant obtaining four new images from the first sentence. The output images were created by rotating the original image at an angle of: -10°; -5°; 5°; 10°. The inclination of the writing in this range is adequate to that obtained by a human. In this way, the same number of images was obtained as in the case of using sentences derived from the recording of their sequence, thanks to which it will be possible to compare the effectiveness of both methods.

Table 4. Parkinson's disease diagnosis results based on the first sentence with augmentation

Accuracy	64,58 %
Sensitivity	67,08 %
Specificity	70,83 %
Precision	58,33 %

The use of augmentation improved the accuracy of disease recognition compared to its absence. However, not to the same extent as in the case of using a sequence of sentences. Figure 5 shows the error matrices obtained for the three experiments discussed.

Fig. 5. Confusion matrices obtained using training data in the form of: a) a single sentence, b) a single sentence with augmentation, c) a sequence of sentences

Conclusions

The article presents a method for improving binary classification results in Parkinson's disease diagnostics, which involves using more than one image from a given person in the process of tuning a convolutional network. Images were used that show all the sentences written by each of the subjects, one after the other within the same registration. Using a sequence of sentences instead of a single sentence leads to an increase in the number of training data and allows for taking into account changes in writing related to the disease in the training process. Introducing training data into the network in the form of images presenting the full sequence of individual sentences allowed to improve the classification result to a greater extent than the use of augmentation methods.

***Authors**: mgr inż. Kamila Białek, Wojskowa Akademia Techniczna, Instytut Systemów Elektronicznych, ul. Gen. Sylwestra Kaliskiego 2, 00-908 Warszawa, E-mail: kamila.jadczak@wat.edu.pl; dr hab. inż. Jacek Jakubowski, Wojskowa Akademia Techniczna, Instytut Systemów Elektronicznych, ul. Gen. Sylwestra Kaliskiego 2, 00-908 Warszawa, E-mail: jacek.jakubowski@wat.edu.pl; dr hab. n. med. Anna Potulska-Chromik, Warszawski Uniwersytet Medyczny, Klinika Neurologii, ul. Banacha 1a, 02-097 Warszawa, E-mail: apotulska@wum.edu.pl; dr hab. n. med. Monika Nojszewska, Warszawski Uniwersytet Medyczny, Klinika Neurologii, ul. Banacha 1a, 02-097 Warszawa, E-mail: monika.nojszewska@wum.edu.pl; prof. dr hab. n. med. Anna Kostera-Pruszczyk, Warszawski Uniwersytet Medyczny, Klinika Neurologii, ul. Banacha 1a, 02-097 Warszawa, E-mail: anna.kostera-pruszczyk@wum.edu.pl*

REFERENCES

[1] Bloem, B. R., Okun, M. S., & Klein, C. (2021). Parkinson's disease. The Lancet, 397(10291), 2284-2303
[2] Gaenslen, A., Swid, I., Liepelt-Scarfone, I., Godau, J., & Berg, D. (2011). The patients' perception of prodromal symptoms before the initial diagnosis of Parkinson's disease. Movement Disorders, 26(4), 653-658
[3] Teulings, H. L., & Stelmach, G. E. (1991). Force amplitude and force duration in parkinsonian handwriting. In Tutorials in motor neuroscience (pp. 149-160). Dordrecht: Springer Netherlands
[4] Broderick, M. P., Van Gemmert, A. W., Shill, H. A., & Stelmach, G. E. (2009). Hypometria and bradykinesia during drawing movements in individuals with Parkinson's disease. Experimental brain research, 197, 223-233
[5] Broeder, S., Nackaerts, E., Nieuwboer, A., Smits-Engelsman, B. C., Swinnen, S. P., & Heremans, E. (2014). The effects of dual tasking on handwriting in patients with Parkinson's disease. Neuroscience, 263, 193-202
[6] Teulings, H. L., & Stelmach, G. E. (1991). Control of stroke size, peak acceleration, and stroke duration in Parkinsonian handwriting. Human Movement Science, 10(2-3), 315-334
[7] Drotár, P., Mekyska, J., Rektorová, I., Masarová, L., Smékal, Z., & Faundez-Zanuy, M. (2014). Analysis of in-air movement in handwriting: A novel marker for Parkinson's disease. Computer methods and programs in biomedicine, 117(3), 405-411
[8] Impedovo, D., Pirlo, G., & Vessio, G. (2018). Dynamic handwriting analysis for supporting earlier Parkinson's disease diagnosis. Information, 9(10), 247
[9] Rios-Urrego, C. D., Vásquez-Correa, J. C., Vargas-Bonilla, J. F., Nöth, E., Lopera, F., & Orozco-Arroyave, J. R. (2019). Analysis and evaluation of handwriting in patients with Parkinson's disease using kinematic, geometrical, and non-linear features. Computer methods and programs in biomedicine, 173, 43-52
[10] Moetesum, M., Siddiqi, I., Vincent, N., & Cloppet, F. (2019). Assessing visual attributes of handwriting for prediction of neurological disorders—A case study on Parkinson's disease. Pattern Recognition Letters, 121, 19-27
[11] Diaz, M., Ferrer, M. A., Impedovo, D., Pirlo, G., & Vessio, G. (2019). Dynamically enhanced static handwriting representation for Parkinson's disease detection. Pattern Recognition Letters, 128, 204-210
[12] Naseer, A., Rani, M., Naz, S., Razzak, M. I., Imran, M., & Xu, G. (2020). Refining Parkinson's neurological disorder identification through deep transfer learning. Neural Computing and Applications, 32, 839-854
[13] Gazda, M., Hireš, M., & Drotár, P. (2021). Multiple-fine-tuned convolutional neural networks for Parkinson's disease diagnosis from offline handwriting. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52(1), 78-89
[14] Osowski, S. (2020). Sieci neuronowe do przetwarzania informacji (wyd. 4). Oficyna Wydawnicza Politechniki Warszawskiej
[15] Goodfellow, I., Bengio, Y., & Courville, A. (2018). Deep Learning Systemy uczące się. Wydawnictwo Naukowe PWN SA