Ensemble of convolutional neural networks for Parkinson's disease diagnosis from offline handwriting

Matej Gazda [a] , Máté Hireš [a] , Peter Drotár [a]

[a] *Intelligent Information Systems Lab (Department of Computers and Informatics, Technical University of Košice)*
Letná 9,
04200 Košice, SLOVAKIA
matej.gazda@tuke.sk, mate.hires@tuke.sk, peter.drotar@tuke.sk

Abstract. This paper proposes the ensemble of deep convolutional neural networks for diagnosing Parkinson's disease from offline handwriting. The advantage of the offline approach lies in the fact that handwriting acquisition can be performed without any specialized equipment by using only a smartphone camera. The convolutional neural networks ensemble relies on pre-trained networks where the diversity is achieved through the multiple-fine-tuning of individual networks. The experimental results on two handwriting datasets showed that the proposed approach currently provides the highest classification accuracy compared to other strategies for diagnosing Parkinson's disease based on offline handwriting

1. Introduction

There are two strong trends for intelligent engineering systems for next-generation medicine: diagnostic decision support systems and non-invasive monitoring. Non-invasive monitoring and examination are frequently more cost-effective than invasive procedures and require less medical professional effort. Diagnostic decision support systems help to make more objective, knowledge-based decisions. Current technology provides a wide plethora of data that can serve as an input to decision support systems. These can vary from different medical images (such as MRI, retina images, etc.) through electroencephalogram electrocardiogram to speech or handwriting.

Handwriting is a skill that is acquired by learning and practicing. It is the result of several sequential biological processes where the visual information is processed, analyzed, and passed. The main role lies in the motor and cognitive functions of the brain that are responsible for execution. The ability to maintain the constant form of handwriting can be disturbed or lost due to many factors such as aging, comfort during writing, or the writing velocity. Besides these, the injury, disease, or disability can very significantly affect handwriting. As such, handwriting may reflect some disturbances in cognitive or motor functions. This is why handwriting was established as one of the early indicators of Parkinson's disease (PD) [15].

Historically, result of handwriting is represented as a trace on the paper. However, the latest advances in technology allow handwriting to be acquired as a pen tip movement in three dimensions (x, y, and z-captured as pressure). The ability to acquire and process handwriting ignited considerable research attention on handwriting. Initial works of Teulings, Stelmach, and Gemmert [7], [16] showed the PD results in statistically significant deterioration of handwriting captured by changes in size and speed of handwriting. The following works introduced many sophisticated features that were able to capture different hidden aspects of handwriting. The most recent include, for example, cepstral and spectral handwriting features [11]. We do not provide an exhaustive review of the various proposed features, but the interested reader can find more details in [9], [17] and references therein.

In recent years, methodological and technological advances revealed the potential of convolutional neural networks (CNN) in image processing, and CNN penetrated many areas of medical image processing. Therefore, it was only a natural extension to apply CNN for processing the handwriting data. The input of the CNN is an image, so in this case, one has to omit all other parameters that are available in online processing, such as handwriting coordinates, timestamp, pen tilt, and pressure. Even though initial works reported classification accuracy far beyond the capabilities of the methods utilizing online handwriting methods [10], [12] recent results indicate that there is enough of discriminative information also in offline handwriting [1], [6].

2. Proposed approach

We decided to apply the majority voting ensemble method of five convolutional neural networks trained independently.

2.1 Convolutional neural networks

Convolutional Neural Networks achieved significant breakthroughs when used with data in grid format, such as time-series or 2D/3D images. One of the most recognized network is VGG [14] that won ImageNet

2012 challenge [2]. Since both datasets PaHaW and NewHandPD are relatively small, we chose to improve the generability of the model with transfer learning.
Transfer learning (TL) is a method to transfer the knowledge obtained on one domain to another domain. With respect to the CNN, it means to pretrain the model on one task and then reuse the model's weights on the second task. Reused weights can be either frozen, thus not trainable, or not frozen, therefore affected during the fine-tuning process. The general assumption to make transfer learning work is that the domains should be relatively semantically similar. However, a similarity metric between the two domains has not been defined yet.
We divide TL into two categories. TL (a) without mediator dataset, and (b) with mediator dataset.
For models based on traditional TL without mediator dataset we train the network by end-to-end approach on large scale source dataset S and then fine-tune network on the target task – diagnostics of Parkinson's disease on PaHaW/NewHandPD datasets. We denote such networks as CNN_S.
We apply an additional intermediate step for models based on TL with mediator dataset. First, the network is trained on a large source dataset S, then fine-tuned on a dataset A, closer to the final dataset, and finally fine-tuned to the target task. We denote such network as $CNN_{S,A}$.
The assumption is that the mediator can close the semantic gap between the source and target datasets. Mediator dataset is helpful if the source dataset is big but not semantically close to the target task, and the mediator itself is not big enough to be trained from scratch.

2.2 Ensemble of multiple-fine-tuned CNN

The main aim of an ensemble classifier is to build a more robust classifier by combining several base classifiers. Naturally, base classifiers are required to provide some form of diversity to be beneficial for the ensemble. Diversity can be ensured in different ways.
The simplest option to guarantee diversity is to employ different base classifiers, such as Support Vector Machines and CNNs. The second option is to use a classifier with the same structure but trained on different data.
In this paper, we utilize five CNNs as base classifiers. The diversity is derived from different training processes, particularly using different datasets as mediators or not using mediator at all. The whole process is depicted in Figure 1. The training process of all 5 CNNs is presented in the second column.

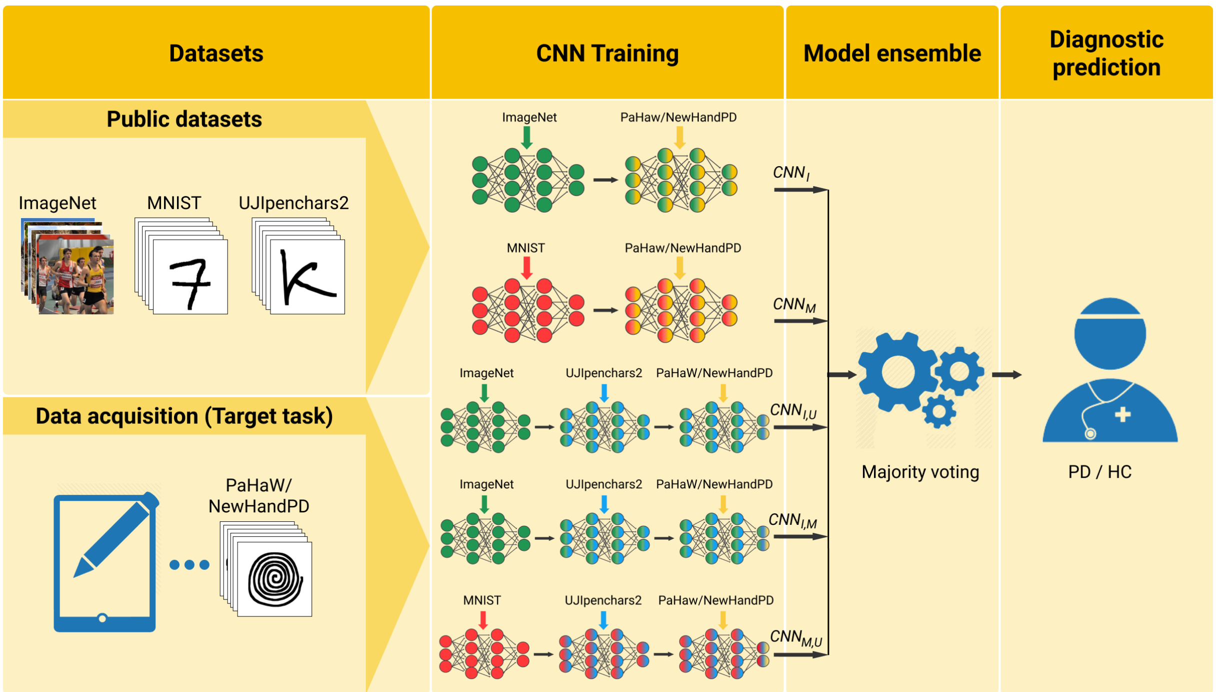

Fig. 1. Concept of the proposed decision support system incorporating MFT CNNs and ensemble voting.

3. Results

The proposed ensemble approach was evaluated on two publicly available datasets: PaHaW [5] and NewHandPD [13]. The PaHaW dataset is a compilation of eight different handwriting tasks, including Archimedean spiral, letters, syllables words, and sentence. We avoided using the sentence task since it is

Table 1
Prediction accuracy of different networks on all evaluated handwriting tasks from NewHandPD and PaHaW datasets.

handwriting task	CNN_I	$CNN_{I,U}$	$CNN_{I,M}$	CNN_M	$CNN_{M,U}$	CNN_{CE}
spiral (HandPD)	88.9 ± 5.9	92 ± 4	89.6 ± 8	81.3 ± 8.4	82.52 ± 8.1	$\mathbf{96.3 \pm 4.59}$
meander (HandPD)	89 ± 10	92.3 ± 6.5	92.7 ± 7.1	89 ± 8.5	89 ± 7.5	$\mathbf{94.38 \pm 8.48}$
spiral	80 ± 10	81.6 ± 8.6	83 ± 8.6	79.5 ± 6	85.3 ± 4.7	$\mathbf{88.54 \pm 3.1}$
l	64.5 ± 6.3	67.6 ± 6.4	66.9 ± 6.2	65 ± 3.2	65 ± 4.6	$\mathbf{71.25 \pm 10.16}$
le	73.8 ± 7.9	71.3 ± 9.5	71.3 ± 8.3	65.1 ± 7.1	66.3 ± 3.4	$\mathbf{78.8 \pm 11.8}$
les	70.7 ± 4	69.9 ± 6.4	70.8 ± 4.1	68.4 ± 7.1	68.6 ± 2.3	$\mathbf{72.5 \pm 11}$
lektorka	72.2 ± 6.2	74.7 ± 4.2	73.4 ± 7.8	68.4 ± 6.2	72.2 ± 8.3	$\mathbf{81 \pm 4.88}$
porovnat	68.1 ± 8.3	67.8 ± 10	68.7 ± 10.6	64.7 ± 6.16	68.5 ± 8	$\mathbf{77.26 \pm 3.94}$
nepopadnout	75.8 ± 4.2	78.4 ± 6	78.5 ± 3.7	72.1 ± 6.2	77.4 ± 5.3	$\mathbf{91.88 \pm 5.02}$

Table 2
Comparison of prediction accuracy of the proposed method and other state-of-the art approaches from literature.

handwriting task	Diaz [4]	Diaz [3]	Moetesum [10]	Pereira [12]	Gazda [6]	This work
spiral (HandPD)	94.44	-	-	76.26	92.7 ± 5.8	96.3 ± 4.59
meander (HandPD)	91.11	-	-	80.75	94.7 ± 7	94.38 ± 8.48
spiral	93.75	75	76 ± 8	-	85.8 ± 7	88.54 ± 3.1
l	96.25	64.16	62 ± 8	-	68 ± 4	71.25 ± 10.16
le	88.75	58.33	57 ± 9	-	74.7 ± 6.9	78.8 ± 11.8
les	90	71.67	60 ± 8	-	72.7 ± 4.7	72.5 ± 11
lektorka	93.75	75.41	60 ± 7	-	76.1 ± 2.8	81 ± 4.88
porovnat	91.25	63.75	51 ± 9	-	76 ± 6	77.26 ± 3.94
nepopadnout	92.5	70	68 ± 7	-	78.5 ± 9.4	91.88 ± 5.02

different from a single word and has more complex structure. The NewHandPD dataset contains two tasks: meander drawing and Archimedean spiral. For the case of PaHaw dataset the handwriting is captured in form of x and y coordinates that were used to render images. If task contains multiple repetitions, every repetition is considered as single image and was used for training and testing as single sample.

As the CNN backbone, we utilize VGG architecture that provides competitive performance and proved itself in our previous experiments. We use stochastic gradient descend (SGD) for pre-training and training on mediator dataset and Adadelta for training on the target task. The learning rate was set up to value 0.01, and we used 300 epochs. All images were resized to 224×224 pixels to match ImageNet size.

Altogether, we trained five networks. Two networks, CNN_I, and CNN_M, were trained on large datasets ImageNet and MNIST. Then, three networks were trained through multiple-fine-tuning procedure introduced in [6]. Networks $CNN_{I,M}$ and $CNN_{I,U}$ were first trained on ImageNet and then further trained on MNIST, and UJIpenchars2 [8], respectively. $CNN_{M,U}$ was pretrained on MNIST and then fine-tuned using the smaller UJIpenchars2 dataset. Finally, all CNNs were fine-tuned on target datasets (PaHaW or NewHandPD). We used only MNIST for training network from scratch since it is much bigger than UJIpenchars2.

The numerical results are presented in Table 1. Stratified five-fold cross-validation was used while ensuring that handwriting samples from one subject were used only in the training dataset or testing dataset, and not in both. The results of individual networks are based on our previous experiments in [6]. In this work we extend the previous results by building the ensemble from all five networks, denoted as CNN_{CE}. This shows a notable boost in the performance of the ensemble classifier.

We compare the performance of the CNN_{CE} ensemble with other state-of-the art results. The comparison is depicted in Table 2. We included in the table the very recent work of [4] that achieves probably the highest overall results in prediction of PD from handwriting. However, it should be noted that this method takes advantage also of kinematic features, such as velocity and pressure, not only imagery data. Even then, we can see that the proposed CNN ensemble outperformed the Diaz's [4] approach on NewHandPD dataset and yielded very competitive results for word *nepopadnout*. The three referenced works [10], [12], [6] provide fair comparison since in this case only imagery data are used in the input. As can be seen, the ensemble proposed in this study outperforms the three competitive approaches on both considered datasets.

4. Conclusions

In this work, we presented the ensemble of CNNs for diagnosing PD from handwriting. To avoid prohibitive computational cost when training the CNN, we utilized the multiple-fine-tuning approach that uses mediator dataset to close the gap between source tasks, such as the classification of natural images and the target task. The mediator dataset also allows to create diversity and to build the ensemble of the CNNs. The proposed approach provides competitive results that even outperform the methods based on the online handwriting on some tasks. The CNNs use only imagery data on the input and no other modalities. This opens new possibilities to bring computerized handwriting analysis closer to real use. The image can be acquired by a high-resolution smartphone camera and then evaluated by CNNs. Naturally, the image acquired by camera can be of different resolution, with possible shadows and some noise, so further experiments need to be performed. Additionally, it is clear that this approach can not provide a complex view on handwriting as the online processing since it does not consider handwriting dynamics and kinematics but can capture significantly more data and screen a larger part of the population.

Acknowledgement

This work was supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under contract VEGA 1/0327/20 and by the Slovak Research and Development Agency under contract No. APVV-16-0211.

References

[1] Mohamad Alissa, Michael A Lones, Jeremy Cosgrove, Jane E Alty, Stuart Jamieson, Stephen L Smith, and Marta Vallejo. Parkinson's disease diagnosis using convolutional neural networks and figure-copying tasks. *Neural Computing and Applications*, pages 1–21, 2021.

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[3] Moises Diaz, Miguel Angel Ferrer, Donato Impedovo, Giuseppe Pirlo, and Gennaro Vessio. Dynamically enhanced static handwriting representation for parkinson's disease detection. *Pattern Recognition Letters*, 128:204–210, 2019.

[4] Moises Diaz, Momina Moetesum, Imran Siddiqi, and Gennaro Vessio. Sequence-based dynamic handwriting analysis for parkinson's disease detection with one-dimensional convolutions and bigrus. *Expert Systems with Applications*, 168:114405, 2021.

[5] Peter Drotar, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdenek Smékal, and Marcos Faundez-Zanuy. Analysis of in-air movement in handwriting: A novel marker for parkinson's disease. *Computer methods and programs in biomedicine*, 117(3):405–411, 2014.

[6] Matej Gazda, Máté Hireš, and Peter Drotár. 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, 2022.

[7] A.w.a.Van Gemmert, H.-L Teulings, J.L Contreras-Vidal, and G.e Stelmach. Parkinsons disease and the control of size and speed in handwriting. *Neuropsychologia*, 37(6):685–694, 1999.

[8] David Llorens, Federico Prat, Andrés Marzal, Juan Miguel Vilar, María José Castro, Juan Carlos Amengual, Sergio Barrachina, Antonio Castellanos, Salvador España, Jon Ander Gómez, et al. The ujipenchars database: a pen-based database of isolated handwritten characters. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, 2008.

[9] Momina Moetesum, Moises Diaz, Uzma Masroor, Imran Siddiqi, and Gennaro Vessio. A survey of visual and procedural handwriting analysis for neuropsychological assessment. *Neural Computing and Applications*, 2022.

[10] Momina Moetesum, Imran Siddiqi, Nicole Vincent, and Florence Cloppet. Assessing visual attributes of handwriting for prediction of neurological disorders—a case study on parkinson's disease. *Pattern Recognition Letters*, 121:19–27, 2019.

[11] Juan A Nolazco-Flores, Marcos Faundez-Zanuy, VM De La Cueva, and Jiri Mekyska. Exploiting spectral and cepstral handwriting features on diagnosing parkinson's disease. *IEEE Access*, 9:141599–141610, 2021.

[12] Clayton R Pereira, Danilo R Pereira, Gustavo H Rosa, Victor HC Albuquerque, Silke AT Weber, Christian Hook, and João P Papa. Handwritten dynamics assessment through convolutional neural networks: An application to parkinson's disease identification. *Artificial intelligence in medicine*, 87:67–77, 2018.

[13] Clayton R Pereira, Silke AT Weber, Christian Hook, Gustavo H Rosa, and João P Papa. Deep learning-aided parkinson's disease diagnosis from handwritten dynamics. In *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 340–346. Ieee, 2016.

[14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[15] H-L Teulings, J L Contreras-Vidal, G E Stelmach, and C H Adler. Adaptation of handwriting size under distorted visual feedback in patients with parkinson's disease and elderly and young controls. *Journal of Neurology, Neurosurgery & Psychiatry*, 72(3):315–324, 2002.

[16] Hans-Leo Teulings and George E Stelmach. Control of stroke size, peak acceleration, and stroke duration in parkinsonian handwriting. *Human Movement Science*, 10(2-3):315–334, 1991.

[17] Mathew Thomas, Abhishek Lenka, and Pramod Kumar Pal. Handwriting analysis in parkinson's disease: current status and future directions. *Movement disorders clinical practice*, 4(6):806–818, 2017.