CNN Based Analysis of the Luria's Alternating Series Test for Parkinson's Disease Diagnostics

Anonymous authors
Paper under double-blind review

Abstract

Deep-learning based image classification is applied in this studies to the Luria's alternating series tests to support diagnostics of the Parkinson's disease. Luria's alternating series tests belong to the family of fine-motor drawing tests and been used in neurology and psychiatry for nearly a century. Introduction of the digital tables and later tablet PCs has allowed deviating from the classical paper and pen setting, and observe kinematic and pressure parameters describing the test. While such setting has led to a highly accurate machine learning models, the visual component of the tests is left unused. Namely, the shapes of the drawn lines are not used to classify the drawings, which eventually has caused the shift in the assessment paradigm from visual-based to the numeric parameters based. The approach proposed in this paper allows combining two assessment paradigms by augmenting initial drawings by the kinematic and pressure parameters. The paper demonstrates that the resulting network has the accuracy similar to those of human practitioner.

1 Introduction

The present paper proposes an approach to support diagnostics of the Parkinson's disease (PD) using convolution neural networks (CNN) to classify the drawing representing the results of digital Luria's alternating series tests (dLAST). Parkinson's disease is the degenerative disorders which most characteristic symptoms rigidity, tremor and non-purposeful motions may severely affect the quality of everyday life of the patient. (Kalia & Lang, 2015), (Louis & Machado, 2015). While there is no known cure from the PD, early diagnoses and proper therapy may relieve the patients from the majority of the symptoms and in turn, improve the quality of the everyday life.

Drawing tests and their digital versions (Vessio, 2019) become more popular in the clinical studies targeted to support early diagnosis of the PD. In the area of kinematic (Marquardt & Mai, 1994), (Drotar et al., 2013) and pressure parameters based analysis (Drotár et al., 2016) spiral drawing test (Danna et al., 2019) is one of the most popular. Some times simpler tests like one described in Kotsavasiloglou et al. (2017) are used. The battery of Luria's alternating series tests (LAST) was proposed by Hodges (2007) and Luria (1995) and later digitised by Nõmm et al. (2018). Machine learning-based approach to study older LAST tests is described inStepien et al. (2019). Analysis of more complex tests, like a clock drawing test (Harbi et al., 2017) or Poppelreuter's Test (Nõmm et al., 2016a) require one to involve neural networks (NN) based techniques either to perform complete analysis or to analyse their parts.

The present research differs from the existing results by the procedure used to incorporate kinematic and pressure parameters into the original drawing. First repeating patterns and thir elements are extracted from the drawing. For each element or pattern, kinematic (velocity, acceleration, jerk, etc.) and pressure parameters are computed. This data is used to colour each segment and change the thickness of its line. Finally, the typical workflow of training and validation of deep neural network models is applied.

The paper is organised as follows. Section 2 explains the symptoms of PD in terms of kinematic and pressure parameters of the fine motor motions, common to the drawing procedure. The same section presents Luria's alternating series tests and their digital version. Formal problem statement is provide by the section 3. Experimental setting is explained in detail in section 4. Transformation technique used to incorporate kinematic and pressure parameters into the original drawing together with the applied data augmentation technique and deep neural network employed are described in section 5. Main results are presented in section 6. Limitations of the proposed approach, together with the interpretation of the achieved results, are discussed in section 7. The final section lists conclusions and possible directions of the future studies.

2 Background

Once one decides to perform a particular action, their brain generates the sequence of impulses to be sent to the spinal cord. Luria referred these sequences as *motion melodies* (Luria, 1995). One may think about the motion melodies as the programs to be executed. Then motion melody is sent to the spinal cord to execute the motions. These two steps are usually referred to in the literature as motion *motion planning* and *motion execution* functions or phases. Progressing PD may affect any of these phases or both of them. The symptoms of the PD such as tremor, rigidity and bradykinesia (Kalia & Lang, 2015) are caused by either distorted planning function or problems on the level of motions implementation. If motion planning function is affected, motion melody would not be optimal to reach the target and would require to be corrected during the motion. Disorders on the level of implementation would disrupt the implementation of the motion melody. From the viewpoint of the motion description, these symptoms of PD are reflected by the features describing velocities, accelerations, and pressure applied the stylus tip to the screen of tablet PC (Drotár et al., 2016), (Nõmm et al., 2018).

2.1 Luria's Alternating Series Tests

LAST tests were proposed in Hodges (2007) and Luria (1995) later their digital version dLAST in Nõmm et al. (2016b). LAST and dLAST require the tested subject to complete, copy and trace the drawing of a repeating pattern. The pattern is designed such that one would have to switch between the motion melodies. Inability to switch between the melodies is referred to as perseveration. Detecting perseveration was the original purpose of the LAST tests. In their digital version battery allows diagnosing PD on the basis of kinematic parameters describing the motion of the stylus tip. Originally the battery consisted of three tests $\Pi\Lambda$, Π and $\sin$ wave and three exercises: *continue*, *follow* and *trace* applied to each test. Such a large battery is difficult and time consuming for some elderly subjects to complete. To optimise the testing procedure, within the frameworks of the present contribution employed only $\Pi\Lambda$ and Π tests are considered. Whereas, only $\Pi\Lambda$ test is used in this paper to explain proposed technique. In Figure 1 the thin blue line represents the reference patterns shown to the tested subject one by one. In the same Figure, the thick yellow line represents the drawings produced by the subject during the testing. Besides the simplicity of the tests, there is one more advantage of this battery. Namely, in some cases, it allows determining if PD has affected motion planning function. If the patient has no difficulty to complete trace tests but fails on the tests requiring to copy or continue the pattern, it is a clear indicator that motion planning function is affected and motion execution function not. The difference between the *copy* and *continue* tests is in their complexity. Also, $\Pi\Lambda$ differs from the Π pattern by its complexity. Sometimes in the literature theses tests are referred to as *Alternating Sequences Tests*, and slightly different patterns may be studied Fountoulakis et al. (2008).

3 Problem Statement

The working hypothesis of the present research is that the machine learning (ML) classifier able to use the shape of the drawn lines together with kinematic and pressure parameters would be able to provide high-level predictions to support diagnostics of the PD. This hypothesis leads to the following problem statement. The main goal of this research is to incorporate the kinematic and pressure parameters describing the motions of the stylus tip to the image of the lines drawn to the test. Then train the classifier to distinguish between the PD patients and healthy control (HC) subjects. This primary goal leads the following sub-problems to be tackled.

- Among available kinematic and pressure parameters chose the subset to be incorporated as part of the image.

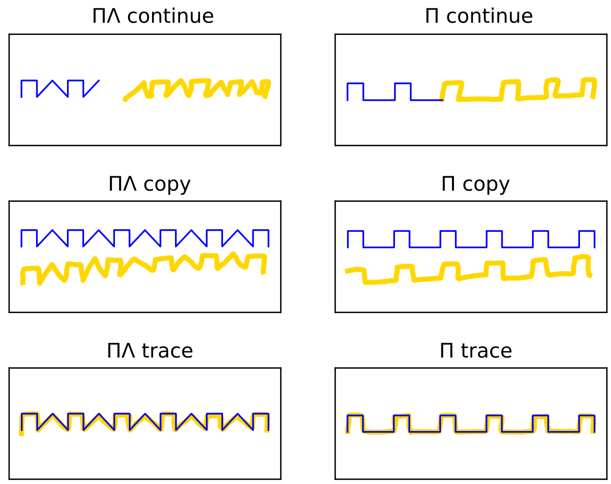

Figure 1: Reference patterns and patient drawings produced during the testing.

- Incorporate chosen parameters without altering the shape of the drawn lines.
- Chose proper classifier architecture, train and validate it.

4 Experimental setting

4.1 Tested subjects

To answer the problem statement and solve sub-problems identified in the previous section, labelled data-set is required. Following the strict personal data protection laws and with the permission of the ethics committee two groups representing 17 PD patients and similar in age and gender distribution group of 33 HC were chosen among those who volunteered to participate in the trials. Mean age of both groups is 69 years old.

4.2 Data acquisition

Tablet computer with stylus and special software developed by work-group was used to conduct the tests. The testing software demonstrates the pattern to be completed, copied or traced, and assignment is demonstrated on the screen and duplicated verbally by the practitioner conducting the test. Using stylus pen tested subject continues, copies and traces the patterns. Tablet PC records the position of the stylus pen with respect to its screen surface together with the pressure applied to the screen two hundred times per second. This information is saved in the form of $N \times 4$ matrix, where N is the total number of observation points per test. Four columns of the matrix are the time stamp, x and y coordinates and pressure.

5 Proposed workflow and methods

Among all the supervised learning techniques known today, convolution neural networks (CNN) are the most suitable choice for image data classification []. This choice immediately poses the requirement of to have a large dataset, which in turn requires one to use data augmentation procedure. Proposed workflow is depicted in Figure 2.

5.1 Enhancement and augmentation

The first step: based on the coordinates and time stamps velocity, acceleration and jerk are computed for each observation point. Together with the pressure, this gives four kinematic parameters to choose from. On the second step, the Shi-Tomasi corner detection algorithm (Shi et al., 1994) is

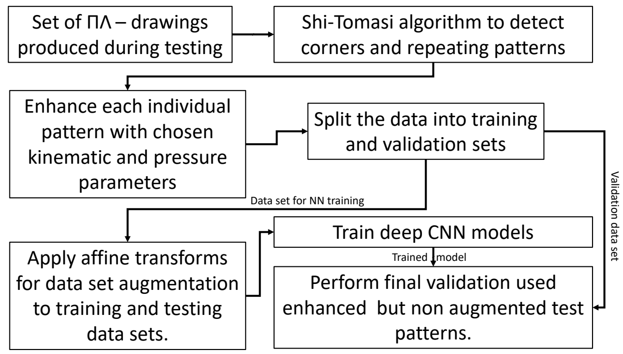

Figure 2: Research workflow.

applied to detect corners of the patterns drawn by the patient. In Figure 3 thin blue line drawn by the patient and yellow points are the corners detected by Shi-Tomasi algorithm.

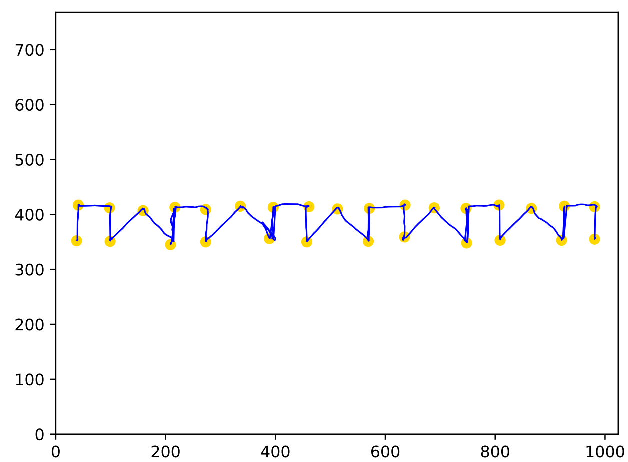

Figure 3: Corners detect by Shi-Tomasi algorithm.

Knowledge of the corner coordinates allows extracting straight segments of the drawing. Then combine them into the repeating patterns. Figure $\Pi\Lambda$ depicts one such pattern where line thickness corresponds to the pressure applied by the stylus tip to the tablet screen and colour is generated by the *jet* colour map on the basis of acceleration values. The thin black line represents one repeating pattern drawn by the tested subject.

On the third step, the width of the line is changed to reflect the pressure. Colour of the line is changed then to reflect acceleration. Then the data is split into training and validation data sets. On the fourth step, the data augmentation procedure (Shorten & Khoshgoftaar, 2019) is applied to the training set only. Since drawings are not real-life images, there is no need to apply the noise procedure. Also, the colouring of the image was left unchanged. Remaining augmentation transformations belong to the set of affine transforms; stretching and squeezing along the axis and counterclockwise and clockwise rotations. Augmentation parameters then consist of stretching and squeezing parameters and rotation angles. Stretching and squeezing parameters are taken from the interval $(0.85, 1.15)$ whereas rotation angle from $(-5°, 5°)$. For each segment, six values are chosen from each transformation. The number of recognised segments vary between five and four; these lead more around 33000 images to be used for the training and testing. As the last step, each image was resized to 224×224 pixels.

Figure 4: Pattern enhanced by the acceleration and pressure

Table 1: Goodness of the different deep CNN architectures

	LeNet-5	AlexNet
Accuracy	0.8427	0.9917
Recall	0.9753	0.9976
Precision	0.7709	0.9858
F1 - score	0.8611	0.9917

5.2 Workflow of CNN Training Testing and Validation

Augmentation procedure also was used to balance the data-set between (PD) and (HC) Augmented data-set is split for training and validation in the proportion of $70/30$. Among the most popular deep CNN architectures LeNet-5 (Lecun et al., 1998), AlexNet (Krizhevsky et al., 2012) and Vgg16 (Simonyan & Zisserman, 2015) were chosen to be evaluated upon their description and suitability for the particular time of the images. These architectures were chosen based on their popularity and description, which is at least in theory, fits the type of images representing drawings of the dLAST.

6 Main results

LeNet-5 and AlexNet usually converge after four epochs whereas Vgg16 not only took much longer time to converge but also demonstrated poor accuracy of just 0.62. Based on its performance for the particular type of task studies in the present paper Vgg16 was excluded from further consideration. More complex AlexNet has demonstrated a better performance. Figure 5 depicts evaluation of the accuracy and loss for the LeNet-5 architecture. Confusion matrix for the LeNet-5 model is pre-

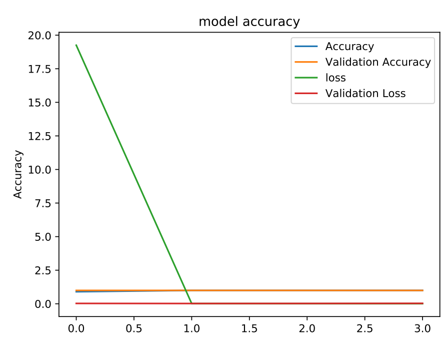

Figure 5: Training of LeNet-5.

sented by Figure 6 Figure 7 depicts evaluation of the accuracy and loss for the AlexNet architecture. Confusion matrix for the AlexNet model is presented by Figure 8 Since each pattern was analysed

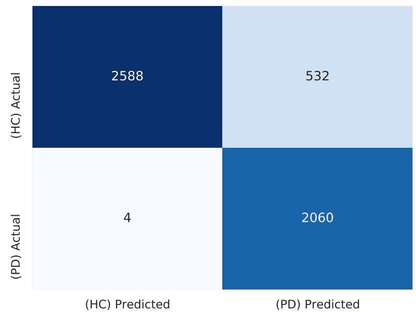

Figure 6: Confusion matrix for the LeNet-5 model.

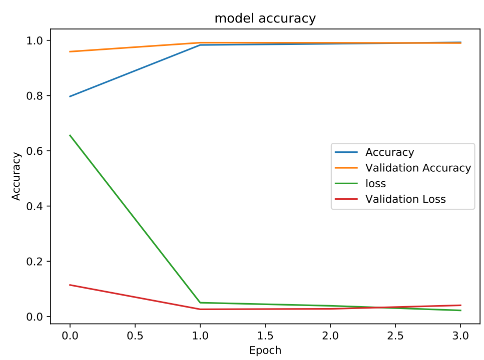

Figure 7: Training of AlexNet architecture

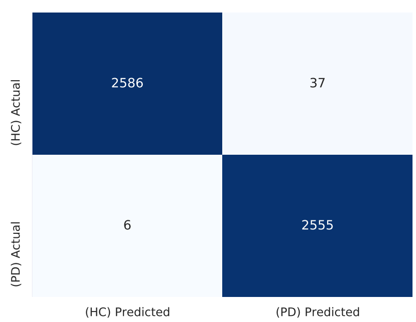

Figure 8: Confusion matrix for the AlexNet model.

separately and each test consists of five full patterns (some times subjects draw a lesser or greater number of patterns), it is essential to see how the position of the pattern affects the performance of the classifiers. Figures 9 and 10 represent numbers of false positives and false negatives (for a small separate selection previously unseen by the network) for LeNet-5 and AlexNet architectures respectively.

7 Discussion

The proposed technique is based on analysing each pattern of the drawing separately, which requires one to summarise classification results for each pattern. Observing classification results for each pattern, one can see that these may be done employing computing the mode of the classes. On the one hand accuracy prediction accuracy for each pattern is in line with Nomm et al. (2019), which demonstrates that patterns in different positions have different discriminating power. On the other

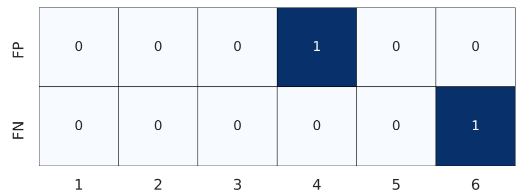

Figure 9: LeNet-5, interval-wise prediction errors.

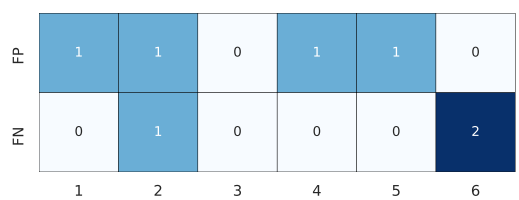

Figure 10: AlexNet, interval-wise prediction errors.

hand, positions of most informative patterns are not the same as in Nomm et al. (2019), which may be due to the difference between machine learning and deep CNN classifiers.

Performance of only three deep CNN structures was evaluated in this paper, whereas, the architectures were used without any tuning or adjustments. The specific nature of the analysed images may require to adjust or tune one of the existing structures to suit particularities of the dLAST images better.

8 Conclusions

The present paper has proposed a novel way to enhance drawing tests used to diagnose Parkinson's disease. Main results of the paper demonstrate the combined with the image augmentation technique proposed approach allows efficient use of the deep convolution neural networks to support diagnostics of Parkinson's disease. Main results have identified the necessity to pay attention to choosing and tuning architecture of the CNN, which will constitute the subject of future studies.

References

Jérémy Danna, Jean-Luc Velay, Alexandre Eusebio, Lauriane Véron-Delor, Tatiana Witjas, Jean-Philippe Azulay, and Serge Pinto. Digitalized spiral drawing in parkinson's disease: A tool for evaluating beyond the written trace. *Human Movement Science*, 65:80 – 88, 2019. ISSN 0167-9457. doi: https://doi.org/10.1016/j.humov.2018.08.003. Special issue: Articles on graphonomics.

P. Drotar, J. Mekyska, Z. Smekal, I. Rektorova, L. Masarova, and M. Faundez-Zanuy. Prediction potential of different handwriting tasks for diagnosis of parkinson's. In *E-Health and Bioengineering Conference (EHB), 2013*, pp. 1–4, Nov 2013. doi: 10.1109/EHB.2013.6707378.

Peter Drotár, Jiří Mekyska, Irena Rektorová, Lucia Masarová, Zdeněk Smékal, and Marcos Faundez-Zanuy. Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease. *Artificial Intelligence in Medicine*, 67:39 – 46, 2016. ISSN 0933-3657. doi: https://doi.org/10.1016/j.artmed.2016.01.004.

Konstantinos N Fountoulakis, Panagiotis T Panagiotidis, Melina Siamouli, Stamatia Magiria, Stavroula Sokolaki, Sotiris Kantartzis, Klairi Rova, Natalia Papastergiou, George Shoretstanitis, Timucin Oral, et al. Development of a standardized scoring method for the graphic sequence test suitable for use in psychiatric populations. *Cognitive and behavioral neurology*, 21(1):18–27, 2008.

Zainab Harbi, Yulia Hicks, and Rossitza Setchi. Clock drawing test interpretation system. *Procedia Computer Science*, 112:1641 – 1650, 2017. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2017.08.259. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference, KES-20176-8 September 2017, Marseille, France.

John R Hodges. *Cognitive Assessment for Clinicians Second Edition*. Oxford Medicine, 2007.

Lorraine V Kalia and Anthony E Lang. Parkinson's disease. *The Lancet*, 386(9996):896 – 912, 2015. ISSN 0140-6736.

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*, 31:174 – 180, 2017. ISSN 1746-8094. doi: https://doi.org/10.1016/j.bspc.2016.08.003.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Elan D Louis and Duarte G Machado. Tremor-related quality of life: a comparison of essential tremor vs. parkinson's disease patients. *Parkinsonism & related disorders*, 21(7):729–735, 2015.

Aleksandr Romanovich Luria. *Higher Cortical Functions in Man*. Springer, 1995. doi: 10.1007/978-1-4684-7741-2.

C. Marquardt and N. Mai. A computational procedure for movement analysis in handwriting. *Journal of Neuroscience Methods*, 52(1):39 – 45, 1994. ISSN 0165-0270. doi: http://dx.doi.org/10.1016/0165-0270(94)90053-1.

S. Nõmm, K. Bardõš, I. Mašarov, J. Kozhenkina, A. Toomela, and T. Toomsoo. Recognition and analysis of the contours drawn during the poppelreuter's test. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 170–175, Dec 2016a. doi: 10.1109/ICMLA.2016.0036.

S. Nõmm, A. Toomela, J. Kozhenkina, and T. Toomsoo. Quantitative analysis in the digital luria's alternating series tests. In *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1–6, Nov 2016b. doi: 10.1109/ICARCV.2016.7838746.

S. Nõmm, K. Bardõš, A. Toomela, K. Medijainen, and P. Taba. Detailed analysis of the luria's alternating seriestests for parkinson's disease diagnostics. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1347–1352, Dec 2018. doi: 10.1109/ICMLA.2018.00219.

S. Nomm, T. Kossas, A. Toomela, K. Medijainen, and P. Taba. Determining necessary length of the alternating series test for parkinson's disease modelling. In *2019 International Conference on Cyberworlds (CW)*, pp. 261–266, 2019.

Jianbo Shi et al. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pp. 593–600. IEEE, 1994.

Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1409.1556.

P. Stepien, J. Kawa, D. Wieczorek, M. Dabrowska, J. Slawek, and Sitek E.J. Computer aided feature extraction in the paper version of luria's alternating series test in progressive supranuclear palsy. In E. Pietka, P. Badura, J. Kawa, and W. Wieclawek (eds.), *Information Technology in Biomedicine*, pp. 561–570. Springer, 2019. doi: https://doi.org/10.1007/978-3-319-91211-0_49.

Gennaro Vessio. Dynamic handwriting analysis for neurodegenerative disease assessment: A literary review. *Applied Sciences*, 9(21):4666, 2019.