2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Novel Computer Vision Approach to Kinematic Analysis of Handwriting with Implications for Assessing Neurodegenerative Diseases

Ron Nachum[1], Kyle Jackson[2,3], Zoran Duric[2], and Lynn Gerber[4,5]

Abstract— **Fine motor movement is a demonstrated biomarker for many health conditions that are especially difficult to diagnose early and require sensitivity to change in order to monitor over time. This is particularly relevant for neurodegenerative diseases (NDs), including Parkinson's Disease (PD) and Alzheimer's Disease (AD), which are associated with early changes in handwriting and fine motor skills. Kinematic analysis of handwriting is an emerging method for assessing fine motor movement ability, with data typically collected by digitizing tablets; however, these are often expensive, unfamiliar to patients, and are limited in the scope of collectible data. In this paper, we present a vision-based system for the capture and analysis of handwriting kinematics using a commodity camera and RGB video. We achieve writing position estimation within 0.5 mm and speed and acceleration errors of less than 1.1%. We further demonstrate that this data collection process can be part of an ND screening system with a developed ensemble classifier achieving 74% classification accuracy of Parkinson's Disease patients with vision-based data. Overall, we demonstrate that this approach is an accurate, accessible, and informative alternative to digitizing tablets and with further validation has potential uses in early disease screening and long-term monitoring.**

Clinical relevance— **This work establishes a more accessible alternative to digitizing tablets for extracting handwriting kinematic data through processing of RGB video data captured by commodity cameras, such as those in smartphones, with computer vision and machine learning. The collected data has potential for use in analysis to objectively and quantitatively differentiate between healthy individuals and patients with NDs, including AD and PD, as well as other diseases with biomarkers displayed in fine motor movement. The developed system has potential applications including providing widespread screening systems for NDs in low-income areas and resource-poor health systems, as well as an accessible form of disease long-term monitoring through telemedicine.**

I. INTRODUCTION

The current diagnostic process for neurodegenerative diseases (NDs), such as Alzheimer's Disease (AD) and Parkinson's Disease (PD), is complex and taxing on patients. The diagnostic process involves multiple specialists relying on their judgment and leveraging a variety of approaches such as mental status exams [1], cognitive assessment [2], and brain imaging [3] to build a case and rule out alternative causes for symptoms. This process is often delayed two to three years after symptom onset and takes several months to reach a conclusion [4]. Because of these barriers to diagnosis, up to 50% of patients with NDs are not diagnosed during their lifetime [5]. Even for patients who receive a diagnosis, an accurate conclusion is not guaranteed; studies have shown that the clinical diagnostic process for NDs is typically only 75-80% accurate [6].

Fine motor movement has been demonstrated as a biomarker, or measurable indicator of disease presence, for NDs, including AD and PD [7]. Quantification and kinematic analysis of fine motor movements has applications for providing early screening assessments to improve and optimize the diagnostic process, as well as monitoring change over time that for long-term monitoring and assessing treatment response [7]. Moreover, kinematic analysis of fine motor movements is applicable to assessing a variety health conditions with biomarkers displayed in fine motor movement, including strokes [8] and early developmental disorders [9], as well as depression and anxiety [10].

Handwriting tasks are commonly used for assessing fine motor movement ability, with specific tasks including tracing of Archimedean spirals and cursive 'l's and 'e's, as well as writing of words and short sentences [7]. During these movements, the pen's position is tracked, which can be used to compute kinematic values of speed, acceleration, and jerk [7]. These kinematic features can be further analyzed to produce measures of movement fluidity and fine motor skill which can be used compare groups of people with different health conditions and as supporting information for disease state classification.

Currently, data for studies in this field are usually collected by specialized digitizing tablets [7]. These digitizing tablets are expensive and can often be inaccessible in poor-resource health systems or telemedicine settings due to their cost. Furthermore, since the use of electronic pens can be unfamiliar to patients, a time-consuming training phase must be completed to acquaint patients with their use. Digitizing tablets collect strictly pen position and pressure, and are unable to capture other available data (e.g., hand pose) that could improve assessment accuracy. By contrast, a computer vision system to quantify these movements offers a fast, easy-to-use, and more widely accessible screening solution due to the pervasiveness of cameras in smartphones and laptops. Furthermore, vision-based systems would be able to collect more data than just pen position, acquiring information about pen grip, arm pose, and compensatory

[1]Thomas Jefferson High School for Science and Technology (TJHSST), Alexandria, VA 22132, USA.

[2]Department of Computer Science, George Mason University, Fairfax, VA 22030, USA

[3]The MITRE Corporation, McLean, VA 22102.

[4]College of Health and Human Services, George Mason University, Fairfax, VA 22030

[5]Inova Health System, Falls Church, VA 22042

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

movements with potential to improve assessment accuracy. This data could be used to augment tablet-based systems which collect accurate pressure data, or potentially to replace them as a more accessible solution.

In this work, we propose a computer vision system to capture handwriting kinematic information with commodity cameras. We tested this system's accuracy through direct comparison to data produced by digitizing tablets during common handwriting tasks. Since commodity cameras capture frames at a lower frequency (typically 30 or 60 Hz) compared to sampling rate of digitizing tablets (typically 100 Hz), we investigated the viability of lower-frequency kinematic data for ND screening assessments using machine learning. To achieve this, we down-sampled the PaHaW dataset of handwriting movements captured by a digitizing tablet and trained classifiers on the resultant information to assess their accuracy.

II. MATERIALS AND METHODS

A. Materials

The primary experimental objectives were assessing accuracy of extracted kinematic data from videos and classification accuracy of resultant assessments that can be made based on this data.

To best determine accuracy and statistically assess the developed computer vision-based system for kinematic data extraction, handwriting tasks were simultaneously captured in a video format by a smartphone camera and quantified by a Wacom Intous Medium digitizing tablet. These synchronized data streams enabled the comparison of handwriting kinematics captured by the computer vision system and digitizing tablet. This system is shown in Figure 1, consisting of a digitizing tablet overlaid with a writing template and connected to a laptop as well as a smartphone on a small tripod. 214 handwriting movements were captured from a single neurotypical test subject to demonstrate feasibility of extracting kinematic information from videos. Measured tasks included Archimedean spiral drawing (124 videos), tracing of l's and e's (60 videos), and tracing of words (30 videos) on the PaHaW study writing template [11].

The PaHaW dataset consists of digitizing tablet data of 8 different handwriting tasks from 38 healthy controls (HCs) and 37 PD patients (total 75 individuals) [11]. The collected position data were utilized at the originally sampled at 100 Hz, typical of digitizing tablets, and also at down-sampled frequencies of 30 and 60 Hz, which are typical of commodity cameras. The resultant kinematic data were then filtered with a Gaussian filter with a sigma value of 5.

B. Computer Vision Quantification of Fine Motor Movement

The computer vision data collection system consists of a few different structures to extract handwriting information from videos, primarily making use of a recurrent system for determining pen position. The entire computer vision system is outlined in Figure 2.

The central objective of the computer vision system for quantifying fine motor movements, in addition to producing vision-specific features, is to extract kinematic information with accuracy comparable to that collected by digitizing tablets. This requires pen tip x and y coordinates tagged with timestamps.

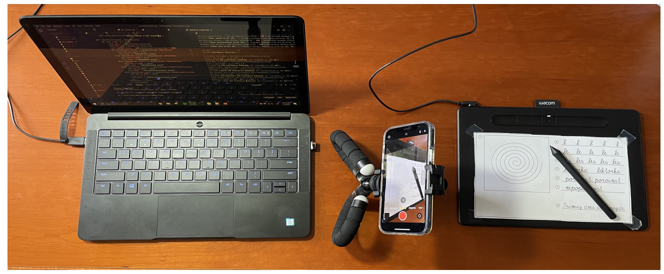

Fig. 1. Experimental setup to collect synchronized data from smartphone videos and digitizing tablet quantification, enabling statistical comparisons to assess accuracy of the vision-based system

1) Preprocessing: In the preprocessing stage, the video frames are prepared for data extraction using thresholding, contour detection, and key point selection, followed by a perspective transform and capture of a pen template image.

To determine the location of the paper template, the OpenCV adaptive thresholding function was used to detect lighter regions of the image [12]. OpenCV contour detection with default parameters was then applied to these thresholded frames, and the largest contour detected was chosen as that of the paper template [12]. With this contour, the OpenCV polygonal approximation method with an epsilon value of 1% of contour arc length was used to identify the 4 corners of the paper [12].

From the camera vantage point, this polygon would appear trapezoidal or irregular when in reality it is a rectangle. To correct for differences in camera perspective, OpenCV can be used to calculate a perspective transform matrix, which can then be used to transform the image into a top-down view of the rectangular paper [13]. Lastly, a template image of the pen is captured to be used for later feature matching in the coordinate extraction [12].

2) Coordinate Extraction: The coordinate extraction phase consists of tracking of the pen-tip using perspective-transformed images of the paper template, using a recurrent approach to produce region of interests for pen tip location.

Feature matching is used to determine a region of interest for the pen in each frame based on the original capture template image. The region of interest is then sharpened using OpenCV's detail enhance method, and blurred using a median filter with a size of 11 [12]. OpenCV's threshold is then applied to increase contrast between the pen tip and the background, followed by contour detection to outline the pen tip geometry in the image and enable precise detection of the tip [12].

With these extracted coordinate data and the known, consistent capture rate of cameras, kinematic features such as speed, acceleration, and jerk can be calculated. As the next frame is processed, the previous position of the pen and calculated kinematic information can be used to decrease the search area for the pen tip with feature matching,

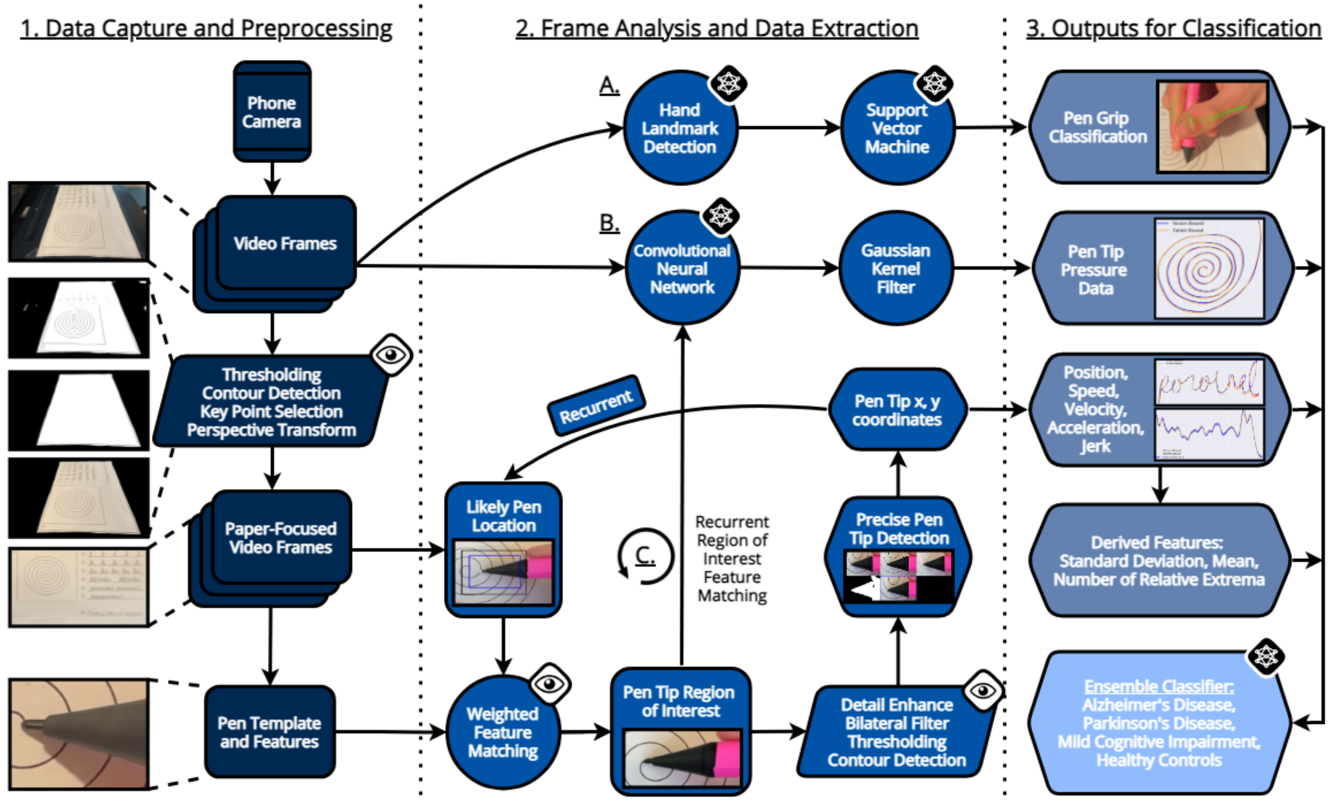

Fig. 2. Computer vision system for data extraction from videos, consisting of three sections: (1) preprocessing, (2) data extraction, and (3) outputs for classification. This paper places particular emphasis on the preprocessing in section 1 and coordinate extraction with recurrent region of interest feature matching algorithm in section 2C. Sections 2A and 2B have been investigated in a preliminary manner and will be investigated for future work.

implementing a *recurrent region of interest feature matching* algorithm. This modification makes this tracking algorithm less computationally expensive and also more accurate, as it has a smaller search area and is less prone to single-frame errors caused by vision jitter and varying lighting conditions.

3) Comparison to Digitizing Tablet: To assess accuracy of kinematic data produced by the vision-based system, the timestamps associated the with computer vision data were matched in a pairwise fashion to digitizing tablet data with the closest timestamp. The aligned time series data were then used to calculate errors and determine accuracy of the vision-based system.

Mean absolute error (MAE) for position was calculated using the following formula across the entire length n of each time series, where (x_i, y_i) represent digitizing tablet coordinate data, and (x'_i, y'_i) represent vision-based data:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \tag{1}$$

Kinematic features of speed, acceleration, and jerk were calculated using symmetrical differences using the following formulas:

$$s_i = \frac{\sqrt{(x_{i+1} - x_{i-1})^2 + (y_{i+1} - y_{i-1})^2}}{t_{i+1} - t_{i-1}} \tag{2}$$

$$a_i = \frac{s_{i+1} - s_{i-1}}{t_{i+1} - t_{i-1}} \tag{3}$$

$$j_i = \frac{a_{i+1} - a_{i-1}}{t_{i+1} - t_{i-1}} \tag{4}$$

C. Assessment of Vision-Based Data for Classification

The PaHaW dataset was used to demonstrate the potential of vision-based data in discriminative ND classification. The collected coordinate information in the dataset was down-sampled from the 100 Hz collected by digitizing tablets to 30 Hz and 60 Hz, typical frame rates produced by cameras. The adjusted data were then used to calculate kinematic features, including speed, acceleration, and jerk. A total of 176 derived features were produced, including mean, minimum, maximum, standard deviation, and number of extrema for profiles of each kinematic feature during a handwriting task. These features were then tested for statistical significance using t-tests to produce the final feature set, consisting of the features with p-values less than 0.10 for each data capture rate.

An ensemble classifier, consisting of a neural network [14], support vector machine [15], and random forest [15] was trained on these data using 10-fold cross validation [14] to prevent overfitting. Each machine learning structure casts a prediction vote for the patient, and the outcome with the most votes (PD or HC) is chosen.

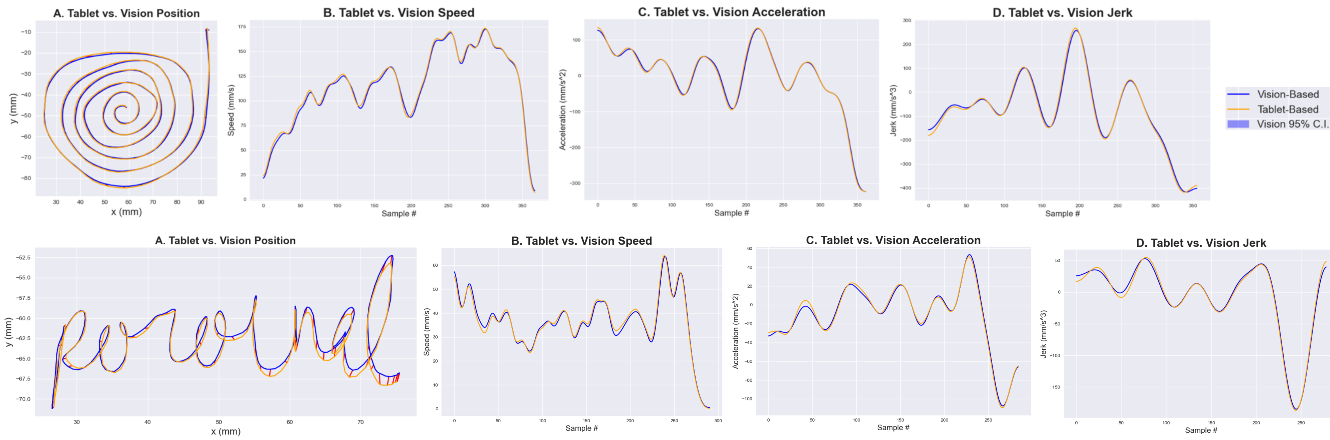

Fig. 3. Relative comparison of our computer vision-based data collection system to digitizing tablet control. Demonstrated by overlaid graphs of extracted kinematic features in an Archimedean spiral and word writing task: (A) Pen tip position (B) Pen tip speed (C) Pen tip acceleration (D) Pen tip jerk. Graphs B-D display narrow 95% confidence intervals. Note the slight position drift in the word writing task, which is due to the pen's increasing distance from the camera.

III. RESULTS

A. Computer Vision Fine Motor Kinematic Data Extraction

Quantitative comparisons of the vision-based system for quantifying fine motor kinematic data from videos to the digitizing tablet are summarized in Tables II and I. Most important to note are the position MAEs, which are less than 0.5 mm for both spirals (n=124) and writing (n=90). Furthermore, the speed and acceleration MAEs were under 1.1% for spiral tasks (n=124), and under 2% for handwriting tasks (n=90). Figure 3 shows a graphical comparison of these kinematic features for representative Archimedean spiral and handwriting tasks, demonstrating the nearly identical kinematic information captured by our computer vision approach compared to the digitizing tablet.

TABLE I

ACCURACY OF VISION-BASED KINEMATIC DATA - SPIRAL

Data	MAE	% MAE	95% CI
Position	0.48 mm	N/A	±0.088 mm
x-coordinate	0.29 mm	N/A	±0.088 mm
y-coordinate	0.31 mm	N/A	±0.041 mm
Speed	1.54 mm/s	1.05%	±0.169 mm/s
Acceleration	0.93 mm/s^2	1.08%	±0.177 mm/s^2
Jerk	4.16 mm/s^3	2.76%	±0.837 mm/s^3

TABLE II

ACCURACY OF VISION-BASED KINEMATIC DATA - WRITING

Data	MAE	% MAE	95% CI
Position	0.40 mm	N/A	±0.055 mm
x-coordinate	0.24 mm	N/A	±0.056 mm
y-coordinate	0.27 mm	N/A	±0.041 mm
Speed	2.39 mm/s	1.95%	±0.286 mm/s
Acceleration	1.88 mm/s^2	1.78%	±0.399 mm/s^2
Jerk	3.71 mm/s^3	1.87%	±0.924 mm/s^3

B. Machine Learning ND Classification with Vision Data

The ensemble learning classification system accuracy was assessed using data down-sampled to three rates of capture: the tablet-collected 100 Hz, and down-sampled values of 60 and 30 Hz to simulate vision-based data. The findings are shown in Table III.

An accuracy of 74% (n=75) was achieved with the 60 Hz capable of capture by many modern, accessible vision-based systems, which is nearly identical to the 75% (n=75) achievable with 100 Hz offered by digitizing tablet data and very similar sensitivity and specificity values. Furthermore, even at a capture rate of 30 Hz, which is attainable with nearly all commodity cameras, an accuracy of 71% (n=75) was achieved in distinguishing PD patients from HCs, with slightly lower sensitivity at specificity values compared to the higher frequencies.

TABLE III

ND ASSESSMENT PERFORMANCE BY DATA CAPTURE RATES

Frequency (Hz)	Accuracy	Sensitivity	Specificity
30	71%	75%	65%
60	74%	79%	70%
100	75%	80%	72%

IV. DISCUSSION

The results of this study demonstrate the viability of our framework using commodity cameras, in particular those in smartphones, to accurately quantify kinematic information of fine motor movements with computer vision algorithms. The significance of this is further compounded by the accuracy achieved in classifying PD patients and HCs using data at frequencies that can be captured by commodity cameras, with accuracy and potential increased emphasis on sensitivity enabling this tool to be used for widespread early screening of NDs.

The vision-based aspect of this system, in combination with modern widespread access to cameras with capability of capturing these data in mobile phones and other devices, make it a prime candidate to enable wider access to ND screening, especially in lower-income populations and resource-poor health systems. Furthermore, the system's at-home accessibility enhances long-term monitoring of disease state, including treatment effects, clinical deterioration, and disease progression, via telemedicine. This ease of use also allows for larger-scale data collection of handwriting movements of patients with NDs as well as HCs to develop and improve our understanding of differences between these groups and increase the accuracy of assessments.

In this paper, we have focused primarily on this data extraction and analysis system's uses for ND screening. However, the framework for vision-based kinematic analysis of fine motor movement is versatile and can be utilized to screen for any health conditions in which biomarkers are displayed in handwriting movements, including strokes, early developmental disorders (e.g., dysgraphia), and arthritis. An accessible and easy-to-use tool for assessing these movements is a necessary step to better understand these biomarkers' significance in differentiating diseased patients, while the resultant expedited optimized screening process has potential to improve treatment outcomes for these conditions.

Digitizing tablets are capable of collecting both pen position and pressure data. Currently, vision based systems are unable to collect high accuracy pressure data, which has been shown to increase classification accuracy of NDs by 5-10% when combined with kinematic features [11]. However, digitizing tablets are limited in their scope of data collection with computer vision providing more types of data collection. Computer vision systems are capable of quantifying hand pose and body movements and also classifying pen grip types, which have potential to improve assessment accuracy and require further research to support their use. For example, hand pose and pen grip type can be quantified using Google's MediaPipe library for hand landmark detection [16].

V. CONCLUSION

In this study, we developed an accessible, vision-based system for analyzing fine motor movements in handwriting tasks to provide ND screening assessments. Our results show that accurate quantification of fine motor movement kinematic features is possible with low-cost commodity cameras. We further demonstrate that kinematic data sampled at frequencies of commodity cameras is viable for distinguishing between ND patients and HCs on the PaHaW data set, with high sensitivity and specificity achieved in ND assessments. This system has potential to increase ND screening access in lower-income populations and resource-poor health systems, provide a long-term disease monitoring solution through telemedicine, and offer a quantifiable screening tool to support clinical diagnosis of NDs.

Future work primarily centers around data collection to further test the accuracy of the vision-based system for quantifying kinematic information. Data collection would also allow for further testing of the significance of vision-specific features such as pen grip and body pose during writing, and exploring the estimation of pen pressure from video data. Additionally, new data collection opportunities would enable observing more complex handwriting tasks such as the clock drawing test (CDT) which cannot be done on a digitizing tablet and determining if higher sensitivity can be achieved to improve this system's utility as a screening tool for NDs.

REFERENCES

[1] M. Grossman and D. J. Irwin, "The mental status examination in patients with suspected dementia," *CONTINUUM: Lifelong Learning in Neurology*, vol. 22, no. 2 Dementia, p. 385, 2016.

[2] R. S. Wilson, A. W. Capuano, L. Yu, J. Yang, N. Kim, S. E. Leurgans, M. Lamar, J. A. Schneider, D. A. Bennett, and P. A. Boyle, "Neurodegenerative disease and cognitive retest learning," *Neurobiology of aging*, vol. 66, pp. 122–130, 2018.

[3] S. Shimizu, D. Hirose, H. Hatanaka, N. Takenoshita, Y. Kaneko, Y. Ogawa, H. Sakurai, and H. Hanyu, "Role of neuroimaging as a biomarker for neurodegenerative diseases," *Frontiers in neurology*, vol. 9, p. 265, 2018.

[4] M. Balasa, E. Gelpi, A. Antonell, M. Rey, R. Sanchez-Valle, J. Molinuevo, and A. Llado, "Clinical features and apoe genotype of pathologically proven early-onset alzheimer disease," *Neurology*, vol. 76, no. 20, pp. 1720–1725, 2011.

[5] L. Boise, R. Camicioli, D. L. Morgan, J. H. Rose, and L. Congleton, "Diagnosing dementia: perspectives of primary care physicians," *The Gerontologist*, vol. 39, no. 4, pp. 457–464, 1999.

[6] T. G. Beach, S. E. Monsell, L. E. Phillips, and W. Kukull, "Accuracy of the clinical diagnosis of alzheimer disease at national institute on aging alzheimer disease centers, 2005–2010," *Journal of neuropathology and experimental neurology*, vol. 71, no. 4, pp. 266–273, 2012.

[7] D. Impedovo and G. Pirlo, "Dynamic handwriting analysis for the assessment of neurodegenerative diseases: a pattern recognition perspective," *IEEE reviews in biomedical engineering*, vol. 12, pp. 209–220, 2018.

[8] B. Simpson, A. McCluskey, N. Lannin, and R. Cordier, "Feasibility of a home-based program to improve handwriting after stroke: a pilot study," *Disability and rehabilitation*, vol. 38, no. 7, pp. 673–682, 2016.

[9] S. Rosenblum and M. Livneh-Zirinski, "Handwriting process and product characteristics of children diagnosed with developmental coordination disorder," *Human movement science*, vol. 27, no. 2, pp. 200–214, 2008.

[10] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clémençon, and G. Cordasco, "Emothaw: A novel database for emotional state recognition from handwriting and drawing," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 273–284, 2017.

[11] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson's disease," *Artificial intelligence in Medicine*, vol. 67, pp. 39–46, 2016.

[12] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[13] "Geometric transformations of images," accessed: 2021-02-17. [Online]. Available: docs.opencv.org/4.5.0/da/d6e/tutorial_py_geometric_transformations .html

[14] F. Chollet. (2015) Keras. Accessed: 2021-02-17. [Online]. Available: https://github.com/fchollet/keras

[15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[16] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. G. Yong, J. Lee, W. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for building perception pipelines," *CoRR*, vol. abs/1906.08172, 2019. [Online]. Available: http://arxiv.org/abs/1906.08172