Analyzing Neurodegenerative Diseases with Web Chatbot Typing Behavior

Bachelor's Thesis (15 ECTS)
Informatics

Supervisor: Prof. Dr.-Ing. J org Ott
Advisor: Leon Nissen, M.Sc.
Advisor: Yifan Yang, M.Sc.
Author: Nicolas Othmar Theodarus

August 8, 2024

Abstract

Neurodegenerative diseases are chronic conditions that destroy and damage part of nervous system of the sufferer over time, especially the brain. This diseases pose a significant challenge for general public health, since the damages are permanent and incurable. This condition happens mainly on elderly people, given that aging is the greatest risk factor. Moreover, early detection of these diseases are inefficient, impractical and only have minuscule success percentage. There is a need for detection methods that are cost-effective, user-friendly and accurate.

This thesis proposes a solution that involve developing a mobile optimized web application to gather typing data from the elderly. A clean and robust architecture structure is utilised to guarantee reliability, scalability and maintainability. It is also need to be ensured that the program is able to effectively process and save the collected data, so that the data can be used for research purposes. Once the typing behaviors of the correspondents are collected, the data will be analysed with the goal to find mathematical properties. This mathematical properties can then be used as a base for later researches, for example, to find what kind of deviation from the mathematical norm can be used as early sign of these neurodegenerative diseases.

The expected result of the project includes the finding of a certain statistical distribution that can be derived from the typing behavior of a healthy person. With this finding, it can then be concluded that typing behavior is one of the possible metrics to detect early neurodegenerative diseases. The application can also then be developed further with features addition. An example of such additional feature would be an analysis section, where typing behaviour data of a person can instantly be analysed with a click of a button.

1 Problem Statement

It is estimated by The World Health Organization that 50 million people worldwide are affected by neurodegenerative disorders, which are primarily characterized by motor neuron dysfunction and loss [1]. Neurodegenerative diseases, such as Alzheimer's disease (AD) and Parkinson's disease (PD) are primarily characterized by progressive degradation of structure or loss of function of neurons [2]. For both diseases, the main known risk factor is age [1]. AD, although being the most common neurodegenerative disease, has no effective treatment yet [3]. Billions of dollars have been spent to find a treatment of Alzheimer's disease with no success [4]. An alternative to curing AD would be to find a way to diagnose the disease effectively. The current methods for diagnosing AD, however, are currently still subpar with a 20% misdiagnosis rate [5].

PD, the second most common neurodegenerative disease [6], also has a similar story. It has no known cure and only has treatments to mitigate the symptoms. Diagnosis for PD is usually done by analyzing motoric signs and symptom through neurological examination. After the symptoms are diagnosed, some medications will be used as initial treatment. As the disease worsens, effectiveness of these medications also lessens.

For neurodegenerative diseases such as AD and PD, it is therefore critical that the symptoms are recognized early. Luckily, an early diagnosis accompanied with early treatment can help greatly with controlling the symptoms. As aforementioned, however, diagnosis of early symptoms of these diseases are still unpractical and have a significant misdiagnosis rate.

2 Proposed Solution

This thesis aims to address the problem of impractical and ineffective diagnosis for early symptoms of neurodegenerative diseases. A possible solution is to develop a mobile optimized web application chatting application to capture and analyze typing behavior of its users. It is a practical solution, since nowadays more and more people are using chatting applications on their smartphone. The researcher aims to build a chatting application for the purpose of gathering data of typing behavior from correspondents. After the data is gathered, it will then be analysed.

This project consists of two parts. The first part is building the chatting application. This application is utilizing Llama3, a large language model (LLM), as a chatting partner for the user. As it is important to have a lot of samples to be able to deduce a convincing conclusion, the LLM will be prompted to try to get as much response from the user, since the more the user types, the more data can be gathered. It is also important to consider that this thesis is focusing solely on the elderly group, since age is the main risk factor for the diseases, as mentioned in the previous section. Therefore, the app needs to be adjusted to cater to the needs of the elderly, such as by using bigger fonts and avoiding excessive contents. As for the technical part of the application, a standard also needs to be adhered to, such as clean architecture and secure coding practices. This is done to ensure that the application can be maintained or developed further when such occasions ensue.

The second part of the project is to process and analyze the accumulated data. It is to be analyzed whether a certain statistical distribution or pattern can be derived from the data. This conclusion can then be used as a base for further research. Possible future research could find out whether certain anomalies from the pattern or statistical distribution found in this thesis can be used as an early detection for Alzheimer's or Parkinson's disease.

3 Related Work

Similar research has been done in the work of Kapsecker et al[7]. In this research, data of typing behaviors are also accumulated, such as typing speed and variation in character usage. Kapsecker's

research, however, used a modified version of the iOS default keyboard to able to gather these data. This modified version of the iOS keyboard brings forwards a limitation in this research, specifically that it shows deviation from the standard keyboard which cause more frequent use of backspace due to typos and different typing behaviors in general. It is because the default keyboard of iOS devices is highly optimized. Differences in structure and layout from this default keyboard, however minor, could cause noticeable changes in the behavior of the users and thus the gathered data.

This research has three main findings. The first finding show that the uniform statistical property can be found in the subjects' typing patterns, i.e. their typing speed and their associated overall distribution. The second finding is also regarding the typing speed. The results of the research shows that there is a strong consistency in typing speed between healthy subjects regardless of potential impact factors, such as daytime. This implies that the method of recording typing behavior, in this case with a custom keyboard, is suitable for measuring baseline deviations for both short and long term. The third finding shows that there is a high correlation of approximately 0.8 between frequency and average transition time. It implies that subjects show different transition time during typing characters that are rarely used and more often used. The system used in this research seems sensitive enough to notice these differences.

These findings suggest that the likelihood of detecting cognitive and psycho-motor impairments through recording and analyzing typing behavior is increased. In the effort of extending this research, the author is hoping to achieve similar results while decreasing the limitations, specifically the limitation caused by using custom keyboard.

Van Waes et al. suggested in 2017 that typing tasks might provide a more accessible alternative for both patients and clinicians. Additionally, the research explores the use of keystroke dynamics as digital biomarkers, which could enhance the diagnostic accuracy for detecting fine motor decline associated with neuropsychiatric disorders [9]. The research highlights how these tasks could serve as a valuable tool in assessing typing and motor skills, which may decline in patients with Alzheimer's disease.

Mastoras et al. suggested in their research in 2019 that typing patterns can be indicative of psychomotor impairment associated with depressive tendencies [10]. This research contributes to the development of unobtrusive, high-frequency monitoring tools for depressive tendencies, providing a potential method for early detection and intervention in everyday settings. The findings highlight the potential of using everyday interactions with mobile devices as a source of data for mental health monitoring.

A newer research in 2023 by Tripathi et al. showed the recognition of neurodegenerative diseases, such as Parkinson's Disease (PD), using typing patterns is an emerging field that leverages keystroke dynamics [11]. This approach involves analyzing the time it takes for individuals to press and release keyboard keys during typing, known as hold time, as well as the time between keystrokes, referred to as flight time. These metrics can be used to detect signs of PD in an ecologically valid setup, such as at the subject's home.

These researches highlighted the possibility of detecting fine motor decline and psychomotor impairment through typing pattern on a keyboard. They showed that through analysing keystroke dynamics, flight time and hold time, psychomotor impairment can be detected. To the researcher knowledge, there has not been a study about using typing behavior on a mobile optimized application to detect signs of Parkinson's and Alzheimer's Diseases.

Time Schedule

Timeframe: August 15 to December 15, 2024

- August
  - 3^{rd} - 4^{th} week: deploy application on UKBonn's server
- September

  - 1st - 2nd week: add still missing and additional feature
  - 3rd - 4th week: last check on the application in preparation for searching for correspondents
- October
  - 1st - 2nd week: search for correspondents and conduct study
  - 3rd - 4th week: research on analysis methods for the data
- November
  - 1st - 2nd week: analyse the data
  - 3rd - 4th week: write thesis
- December
  - 1st - 2nd week: finish writing thesis

References

[1] Northwestern University Feinberg School of Medicine, *Uncovering the Culprits of Neurodegenerative Disorders*, April 11, 2024. Available at: `https://news.feinberg.northwestern.edu/2024/04/11/uncovering-the-culprits-of-neurodegenerative-disorders/`. [Accessed: August 5, 2024].

[2] Lamptey RN, Chaulagain B, Trivedi R, Gothwal A, Layek B, Singh J. (February 2022). "A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics". *Int J Mol Sci.* 23 (3): 1851. doi:10.3390/ijms23031851. PMC 8837071. PMID 35163773.

[3] Erkkinen, Michael G.; Kim, Mee-Ohk; Geschwind, Michael D. (April 2018). "Clinical Neurology and Epidemiology of the Major Neurodegenerative Diseases". *Cold Spring Harbor Perspectives in Biology.* 10 (4): 20. doi:10.1101/cshperspect.a033118. PMC 5880171. PMID 28716886.

[4] J. L. Cummings, D. P. Goldman, N. R. Simmons-Stern, and E. Ponton, "The costs of developing treatments for Alzheimer's disease: A retrospective exploration," *Alzheimer's & Dementia*, vol. 18, no. 3, pp. 469–477, Mar. 2022. doi: 10.1002/alz.12450.

[5] Archer MC, Hall PH, Morgan JC. (2017). "[P2–430]: Accuracy of Clinical Diagnosis of Alzheimer's Disease in Alzheimer's Disease Centers (ADCS)". *Alzheimer's & Dementia.* 13 (7S_Part_16): 800–1. doi:10.1016/j.jalz.2017.06.1086. S2CID 54359937.

[6] Elbaz A, Carcaillon L, Kab S, Moisan F. (January 2016). "Epidemiology of Parkinson's disease". *Revue Neurologique.* 172 (1): 14–26. doi:10.1016/j.neurol.2015.09.012. PMID 26718594.

[7] Maximilian K., Simon O., Jonas, Stephan M. (July 2022). "Analysis of Mobile Typing Characteristics in the Light of Cognition". *Proceedings of the International Conference on Digital Health (ICDH)*, Barcelona, Spain. doi:10.1109/ICDH55609.2022.00022.

[8] H. Jimison, M. Pavel, and J. McKanna, "Unobtrusive computer monitoring of sensory-motor function" in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2006, pp. 5431-5434.

[9] L. Van Waes, M. Leijten, P. Mariën, and S. Engelborghs, "Typing competencies in Alzheimer's disease: An exploration of copy tasks," *Computers in Human Behavior*, vol. 73, pp. 311-319, 2017.

[10] R.-E. Mastoras, D. Iakovakis, S. Hadjidimitriou, V. Charisis, S. Kassie, T. Alsaadi, *et al.*, "Touchscreen typing pattern analysis for remote detection of the depressive tendency," *Scientific Reports*, vol. 9, no. 1, pp. 1-12, 2019.

[11] S. Tripathi, T. Arroyo-Gallego, and L. Giancardo, "Keystroke-dynamics for Parkinson's disease signs detection in an at-home uncontrolled population: A new benchmark and method," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 1, pp. 182–192, Jan. 2023. doi: 10.1109/TBME.2022.3187309.