NeuroTrace: A Novel, Machine-Powered System Software to Detect Neurodegeneration through Handwriting Kinematics Analysis

AP ID: 8Z30X542

AP Research

30 April 2024

Word Count: 5420

All figures created by student unless otherwise specified and accredited.

Abstract

NeuroTrace is a machine-learning model and software package that detects the early onset of neurodegenerative diseases (NDs), such as Parkinson's and Alzheimer's, through handwriting kinematics. Handwriting deterioration is often an early sign of the onset of NDs, affecting the motor and nervous systems. Trained on the publicly-available DARWIN dataset, which contains handwriting kinematics data from 174 participants, NeuroTrace analyzes 10 features from several tracing tasks, including pressure and speed, among others. The initial classifier was trained and tested with 75% and 25% of the dataset, respectively, to evaluate effectiveness. Then, the NeuroTrace software was designed and coded, a lightweight multi-page web program that tracks coordinates, pressure, and tilt metrics, outputting processed numerical data. Willing participants (n=18) signed informed consent forms, completed a demographic survey about vocational/educational data and attempted 6 different tracing tasks on a WACOM Intuos Tablet, which were then calculated into metrics for machine learning (speed, jerk, etc).

The results of this study were promising, with the NeuroTrace model demonstrating accuracy and precision up to 88% and 80%, respectively, in pilot testing involving consenting seniors (n=18), indicating its potential as a reliable tool for ND detection. The NeuroTrace software is continuously updated to patch bugs and increase tracking accuracy. Future research will focus on integrating multimodal data and expanding the dataset to create a more robust model., NeuroTrace represents a significant advancement in early NDs detection, offering a quick and non-invasive screening tool and improving the quality of life for millions worldwide.

Introduction

Neurodegenerative diseases (NDs), such as Parkinson's, Alzheimer's, and multiple sclerosis, affect more than 62 million people worldwide, who are mostly seniors (Hu et al., 2023). These neurodegenerative diseases occur when nerve cells in the brain, central nervous system, and peripheral nervous system deteriorate and die over time, leading to a progressive decline in cognitive function, motor skills, and overall neurological health (National Institute of Environmental Health Sciences, 2018). As the human body ages, cell and nerve health also naturally weaken, which is why neurodegenerative diseases are most common in the older population group. Additionally, as the average life expectancy continues to lengthen due to improvements in health care and medical technology, cases of neurodegenerative disease will only continue to rise. While there is currently no known permanent cure for neurodegenerative diseases, early detection is key to slowing down the development of "severe clinical symptoms" that lead to the rapid deterioration of nerve cells and cognitive functions later in the disease cycle. More specifically, in recent years, pre-disease assessments, such as in-person screenings where NDs can be diagnosed, have emerged as a central research focus for neurodegenerative disorders, where these are able to use physical and holistic data to determine the chance of the presence of neurodegenerative diseases (Shusharina et al., 2023).

NDs significantly decrease the quality of life of diagnosed patients, including symptoms like memory loss, motor degeneration, and communication problems. One crucial motor impairment is hand motion and control. Normal functioning human hands require precise finger and thumb dexterity, grip and wrist strength, and object-to-hand manipulation through three-dimensional space for everyday life, work, and leisure activities (Gopal et al., 2022).

However, those affected by NDs (or chronic disorders of the central nervous system) have impaired hand functions that hinder daily activities – even simple tasks such as teeth brushing, typing on a computer, or writing a letter. Therefore, if the loss of hand motor function is correlated with the onset and development of neurological diseases, then the exploration and establishment of artificial intelligence-powered early-detection assessments with respect to hand function can prove to be a useful and accurate diagnostic in discerning the multifaceted progression of neurological diseases.

Literature Review

Typically, damage to the central nervous system (CNS) precedes the eventual deterioration of kinesthetic and visuospatial motor skills in the hand. For example, most patients with Parkinson's ultimately develop hand tremors and have difficulty making precise finger movements (Gopal et al., 2022). In this study regarding remote-monitored Parkinson's detection, several researchers evaluated the effectiveness of external devices and telerehabilitation (online rehabilitation), such as smartphones, tablets, and online physical therapy, to track Parkinson's disease progressions in subjects through the quantification of hand tremor, finger tapping, and dexterity data (Gopal et al., 2022). While the researchers found high feasibility and reliability for remote progression-logging smartphone apps, they only utilized online data without collecting it firsthand, and more research is needed to verify the results and to ensure the smartphone apps are working correctly. Nevertheless, Gopal et al found that it is likely that remote assessments can facilitate ND tracking in the future, which connects to the common theme of early detection, diagnosis, and intervention throughout this paper.

One visible symptom of NDs is the (often undesired) alterations in handwriting, which demonstrates a loss of motor control. For instance, Judie Walton, a medical PhD who focuses on Alzheimer's research, conducted a study to examine the changes in handwriting due to aging versus Parkinson's disease, measuring characteristics such as letter size, pen pressure, tremors, rhythm, and variabilities in letter formations (Walton, 1997). The results displayed the typically similar handwriting of age-matched control subjects over long periods of time, juxtaposed by the rapid deterioration of handwriting in Parkinson's patients after just a few months to years. The Parkinson's patients' handwriting was characterized by increased letter variability, slower writing, higher error rate, and the characteristic hand tremor (Walton, 1997). This is particularly evident in and relevant to seniors with Parkinson's, Alzheimer's, and other neurodegenerative diseases, as it supports the notion that handwriting and the physical process of writing may drastically change, but mainly only for those suffering from NDs. Yet, outdated for modern technological times as it employs obsolete methodologies, such as taking pictures of handwriting samples with a traditional digital camera, Walton's study reveals a need for the utilization of modern technologies to detect and analyze handwriting in potential NDs patients.

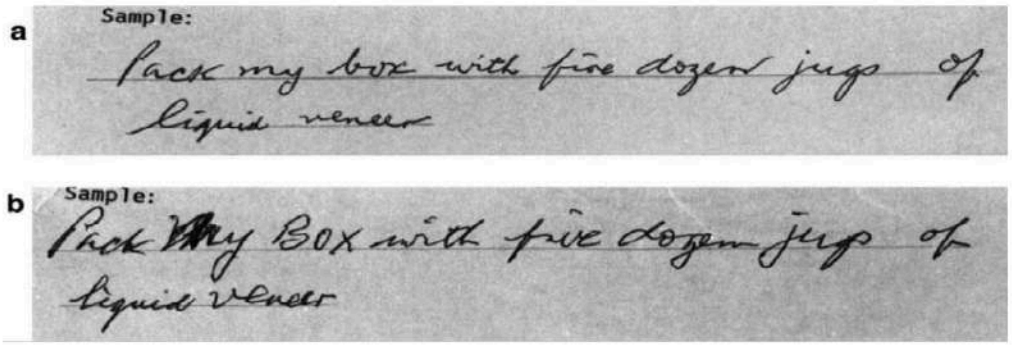


Figure 1. The deterioration of a Parkinson's subject's handwriting:

a) before 5 years and b) after 5 years. Credits: Walton et. al, 1997

Amidst technological advancements, handwriting detection and analysis have become easier. Smart pens, specialized software, and artificial intelligence now play significant roles in modern-day research regarding not just handwriting and neurodegenerative diseases, but the scientific community overall. For example, several researchers from different experiments utilized a specific electronic tablet brand, WACOM, a producer of smart pens, tablets and software with computer data syncing and processing capabilities, which streamlined data collection during the writing process. The combination of hardware and software can quantify not just static handwriting, but also raw data, such as the x and y coordinates of the pen position, time stamps, pen orientation such as azimuth and altitude, and pen pressure, and manipulate this data further to find derivatives (speed, acceleration, and jerk) with respect to time, direction, and curvature of the handwriting process (WACOM, 2023).

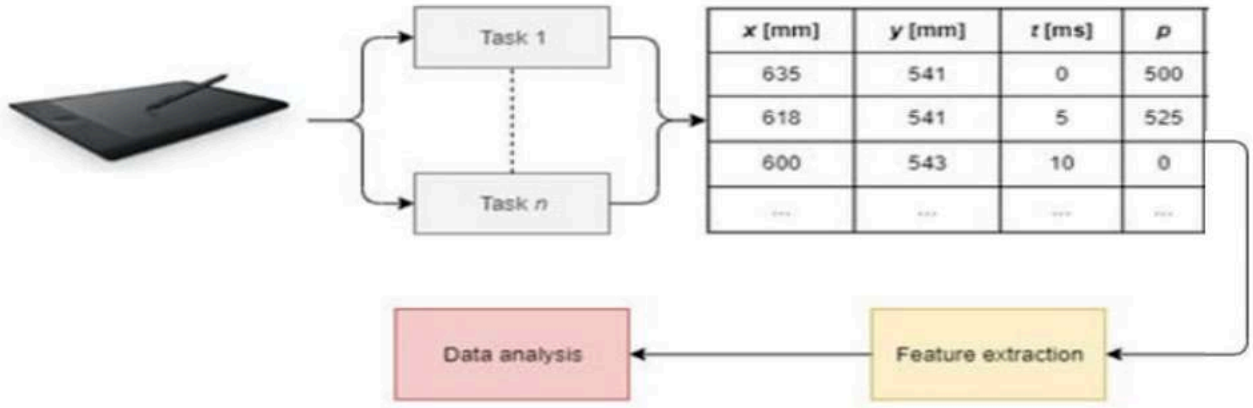


Figure 2. Typical workflow of software handwriting analysis (t = time, p = pressure). Credits: Joseph Garre-Olmo et al., 2023.

Two different research projects utilized the Wacom tablet in a similar manner to analyze the handwriting process. In one experiment conducted in Spain, participants (either control or suffering from Alzheimer's) were instructed to perform several tasks using the smart pen and tablet, such as draw a spiral, copy a sentence, and construct the Clock Drawing Test, where the

subjects had to sketch an analog clock that represented a certain time of day (Josep Garre-Olmo et al., 2023). In comparison, the other experiment, carried out in France, involved a group of control subjects and groups of seniors suffering from early to moderate Parkinson's or Alzheimer's, and mild cognitive impairment. The tasks, which were also carried out on a Wacom tablet and smart pen, encompassed gradually writing in French and Arabic as well as tracing spirals over lines (Ibtissame Aouraghe et al., 2017). It is important to note that there may be discrepancies between the intended and the interpreted meanings of this particular paper as the original file was published nearly completely in French – for the best analytical purposes, this file was translated into English first. Nonetheless, both research projects utilized the software features on the Wacom software which quantified the parameterized coordinates $<(x(t),y(t),p(t)>$ where t is in seconds, into kinematics, including horizontal/vertical speed, acceleration, and jerk, spatial features, including direction, curvature, length of letter spaces, and height, and dynamics, including pressure and pressure variation over time (Ibtissame Aouraghe et al., 2017). A comparison of results and procedures revealed gaps in both experiments. To illustrate, Ibtissame Aouraghe et al. did not have a results section, nor did they explain their findings in the context of their subjects. One of the biggest flaws with the experiments was a bias introduced due to the lack of variety of ages of the participants. The researchers did not list the ages of their subjects, which is necessary for representative semi-random sampling. Although they did include a key, the data visualization section was unclear as the researchers failed to separate the control subjects' data from the Alzheimer's or Parkinson's subjects, leading to overgeneralization (Ibtissame Aouraghe et al., 2017). On the other hand, Josep Garre-Olmo et al. did capably include a results section, discuss their findings in the context of their subjects, and list age ranges

for their subjects. However, a drawback of this study was the small sample size, which could create less generalizable results (Josep Garre-Olmo et al., 2023). There is a need for an effective program that easily collects handwriting kinematics data and records it in an accessible database through software and data collection.

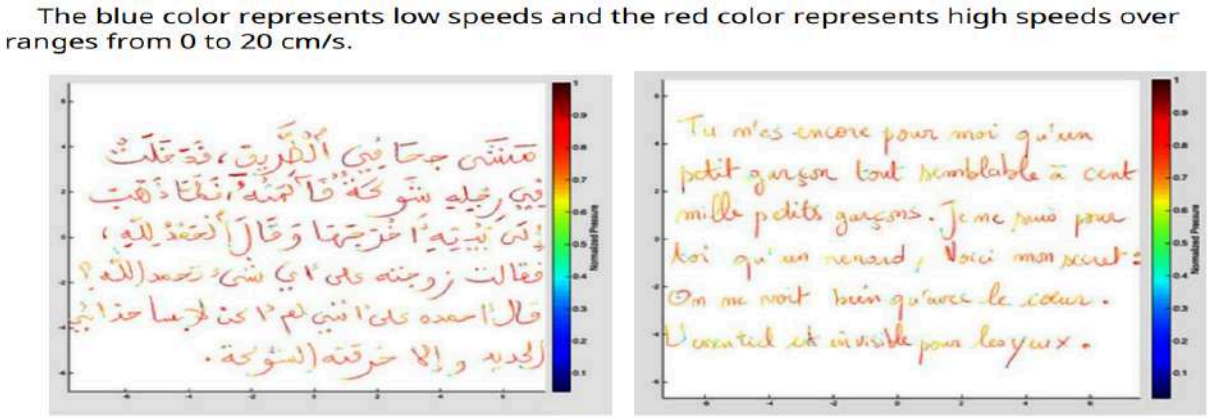


Figure 3. Ibtissame et al.'s visualization of writing pressure. Credits: Ibtissame Aouraghe et al., 2017.

Furthermore, a literary review on dynamic handwriting analysis by computer science student Gennaro Vessio revealed that handwriting analysis studies conducted up to 2019 focused mostly on Alzheimer's and Parkinson's detection. For Parkinson's detection, studies looked at using handwriting analysis to gain insights into motor control changes, monitor disease progression and treatment effects, and develop automated diagnostic tools. For Alzheimer's detection, the focus was more on understanding motor and cognitive changes and early diagnosis (Vessio, 2019). This reveals trends between previously conducted studies on handwriting analysis and NDs further, increasing the sample size of the research.

Research Gap & Questions

During brainstorming, it was hypothesized that using an AI-powered deep-learning classifier algorithm would streamline the screening process of individuals at risk of neurodegenerative diseases such as Alzheimer's, Parkinson's, and multiple sclerosis through quantified handwriting analysis. Furthermore, no current methods for NDs detection are effective; most involve manual assessments by certified experts, which are slow, ineffective, and universally inapplicable for accurate NDs detection and diagnosis. Therefore, this led to the essential research questions: how do neurodegenerative diseases affect kinematic and motor functions of individuals, and how can that data be quantified? Additionally, how can an AI algorithm, such as a random forest classifier, utilize this data to accurately and efficiently predict individuals at risk of NDs?

To address these gaps in the research field of neurodegeneration, the NeuroTrace application was designed and programmed. By determining the relationship between handwriting and neurodegenerative diseases, NeuroTrace has the potential to heighten awareness and promulgate the importance of early detection and treatment for those who may be destined to suffer from neurodegenerative diseases, improving the quality of life for those who need and deserve it most in the near future.

Methods

Overview

Throughout the past few months, NeuroTrace has constantly been in development. The goal of this project was to train a machine-learning algorithm backend and a software package frontend with implications as a quick, accurate, and efficient NDs testing/checkup tool, especially for the senior population, and then test this prototype on real-life seniors (with and without NDs) to verify its accuracy.

Considering the lack of prior personal experience, especially in the development of an entire machine-learning algorithm and application, an immense amount of learning has occurred over the past few months of strenuous work. Not only does this develop a personal portfolio of experiences and skills, but, more importantly, also harnesses personal passions and hobbies to contribute towards a world where advancements in technology are harnessed for the quality of life for individuals, particularly the elderly.

Ethical Considerations

Since this project involved engaging with human participants, it was imperative that IRB (Institutional Review Board) approval was obtained to ensure all experimental procedures were ethical. Every potential subject was notified that participation was optional and their information would be kept completely confidential; those who volunteered to participate in the NeuroTrace pilot study willingly signed informed consent forms. Moreover, certain participants gave photo consent for their body parts (i.e. hands, arms) to be photographed for the pilot study, and may be featured in this paper. There was no compensation of any kind involved.

Outline

Throughout NeuroTrace development, to keep organized through this multifaceted project, a rough chronological roadmap was followed, which is as listed below:

1. Preprocess the DARWIN dataset, visualize the data, and remove unnecessary features
2. Utilize Python to develop a machine-learning model with decent ND detection (via handwriting kinematics data), which could be trained with a public dataset, and engage in model hyperparameter fine-tuning
3. After obtaining a WACOM Intuos tablet, design and code a portable, lightweight software program, NeuroTrace, that could:
    a. Receive pen input from the tablet at a constant polling rate and display the input accurately on a computer screen
    b. Reliably record the raw pen positional data (x, y, pressure, pen ups/pen downs) at a constant polling rate (~120hz), write the data constantly to a CSV (comma-separated values) formatted file, and download that file locally when prompted.
    c. Process and normalize the raw CSV file into analyzable metrics (speed, acceleration, jerk, pressure mean, etc) using numerical approximations.
    d. Plug the processed CSV file back into the trained model in step 1 to obtain the algorithm's predicted results.
4. Verify NeuroTrace model accuracy by conducting pilot testing at senior facilities nearby

1. Data Processing and Visualization

The NeuroTrace model is based on the publicly available DARWIN (Diagnosis AlzheimeR WIth haNdwriting) dataset (n=174), which is open-source on both Kaggle and the University of California - Irvine's Machine Learning Repository. The largest, most comprehensive database for handwriting kinematics in relation to NDs, the DARWIN dataset involved 174 demographics-matched (gender, age, level of education, vocational history) participants, comprising 89 Alzheimer's and 85 healthy control subjects. They were recruited using standard clinical tests for neurodegeneration, such as the Mini-Mental State Examination, Frontal Assessment Battery, and Montreal Cognitive Assessment (Cilia et. al). The participants were then subjected to 25 handwriting tasks, where their handwriting kinematics were tracked, collected, and processed into 18 metrics (*features*, in technical terms).

Table 1
List of tasks performed. Task categories are: memory and dictation (M), Graphic (G), and Copy (C).

#	Description	Category
1	Signature drawing	M
2	Join two points with a horizontal line, continuously for four times	G
3	Join two points with a vertical line, continuously for four times	G
4	Retrace a circle (6 cm of diameter) continuously for four times	G
5	Retrace a circle (3 cm of diameter) continuously for four times	G
6	Copy the letters 'l', 'm' and 'p'	C
7	Copy the letters on the adjacent rows	C
8	Write cursively a sequence of four lowercase letter 'l', in a single smooth movement	C
9	Write cursively a sequence of four lowercase cursive bigram 'le', in a single smooth movement	C
10	Copy the word "foglio"	C
11	Copy the word "foglio" above a line	C
12	Copy the word "mamma"	C
13	Copy the word "mamma" above a line	C
14	Memorize the words "telefono", "cane", and "negozio" and rewrite them	M
15	Copy in reverse the word "bottiglia"	C
16	Copy in reverse the word "casa"	C
17	Copy six words (regular, non regular, non words) in the appropriate boxes	C
18	Write the name of the object shown in a picture (a chair)	M
19	Copy the fields of a postal order	C
20	Write a simple sentence under dictation	M
21	Retrace a complex form	G
22	Copy a telephone number	C
23	Write a telephone number under dictation	M
24	Draw a clock, with all hours and put hands at 11:05 (Clock Drawing Test)	G
25	Copy a paragraph	C

Figure 4. DARWIN's set of 25 tasks for data collection. Credits: Cilia et. al, 2020.

The published DARWIN dataset is in the format of a CSV (comma-separated values) file, with 175 rows and 451 columns (18 features * 25 tasks + *class* column). The *class* column concludes the dataset and is vital to the dataset by defining whether each participant is a control or an actual Alzheimer's patient (P).

To gain a more thorough understanding of the dataset once it was obtained, a series of data visualizations were performed, where the sets of numbers are represented graphically. For example, scatterplots serve as an excellent indicator of overall trends within a certain task, where the x-axis represents the 174 participants, and the y-axis represents a specific measured feature.

```
['air_time', 'disp_index', 'gmrt_in_air', 'gmrt_on_paper', 'max_x_extension',
 'max_y_extension', 'mean_acc_in_air', 'mean_acc_on_paper', 'mean_gmrt',
 'mean_jerk_in_air', 'mean_jerk_on_paper', 'mean_speed_in_air', 'mean_speed_on_paper',
 'num_of_pendown', 'paper_time', 'pressure_mean', 'pressure_var', 'total_time']
```

Figure 5. Each of the 18 metrics (or features) the DARWIN dataset collected (for each of the 25 tasks), where disp_index is the displacement index and GMRT is the generalized mean relative tremor (both calculated with special formulas).

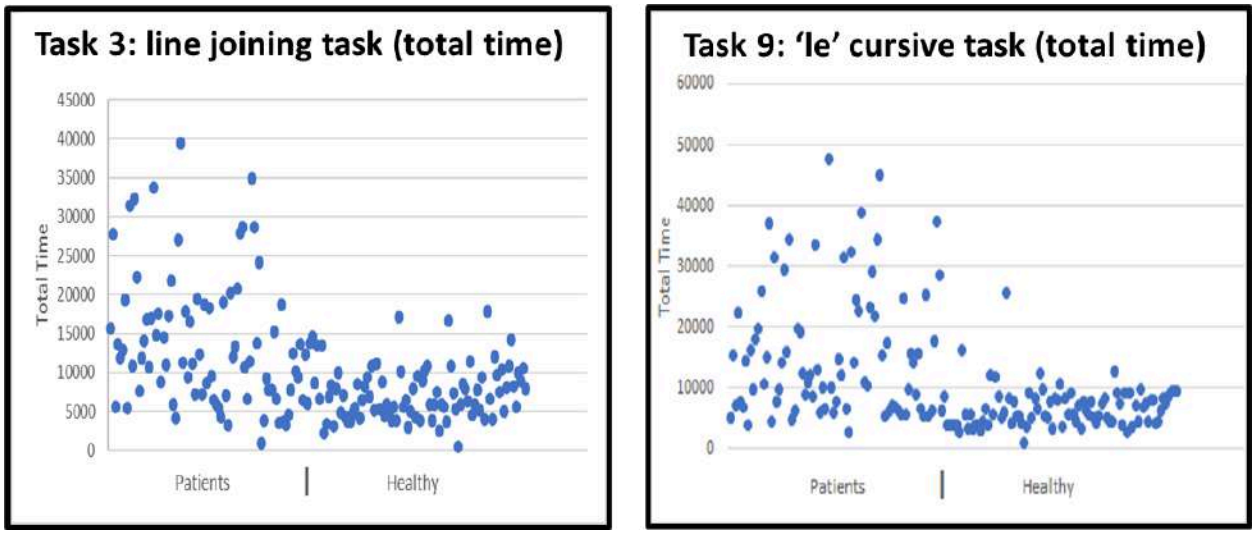


Figure 6. Scatterplot visualizations of subjects' total time taken to perform tasks 3 and 9.

In an effort to simplify the DARWIN dataset to reduce unnecessary data noise, another series of visualizations were performed, involving analyzing and comparing the individual tasks and features themselves to their neighboring tasks and features ("*task and metric importance*"), respectively; this facilitated in "weeding out" unnecessary tasks and features, leaving behind an easier model-training process. By doing this, the 25 tasks and 18 metrics were reduced to 6 tasks and 11 metrics, maintaining appropriate and acceptable model accuracy in the process while dramatically decreasing the workload for data collection and calculations.

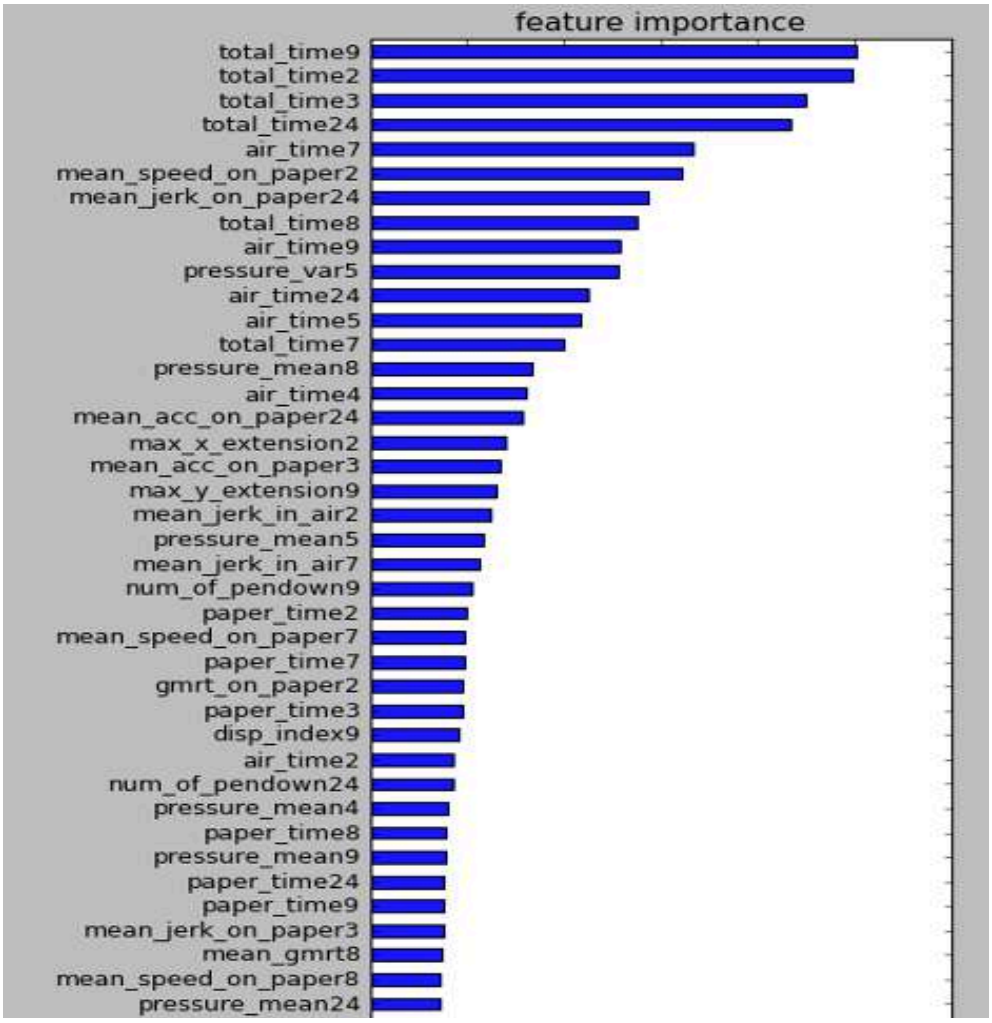


Figure 7. DARWIN task and metric importance. Notice the prevalence of certain tasks (9, 2, 3, 24, 7) and metrics (total_time, air_time, mean_speed) in the bar chart.

2. Model Training and Optimization

Having set some limitations for the dataset, model training could begin. The actual *training* process of the machine learning model was rather trivial, requiring little brainpower, as the process itself was easily replicated from how-to websites accessible online. The model was trained and hosted on Google Colab's Python 3; this was the best solution for portability and cost. The *Scikit-learn* library of machine-learning algorithms – more specifically, the Random Forest Classifier (an ensemble-learning classifier which bases its predictions off the majority voting by a series of feature-built "decision trees") – was utilized, due to several reasons:

a. There are extensive resources on the web for Random Forest model training/tuning.
b. While there exists a multitude of other classifiers that could prove practical for this project, the Random Forest is less prone to overfitting, which occurs when the model is trained *so well* on a specific dataset that it fails to generalize to new datasets.
c. Generally, Random Forest balances accuracy with data specificity (Kho).

With that said, the programming aspect could warrant a logbook of its own. Moreover, given the personal minimal coding experience, a sizable portion of the code throughout the NeuroTrace developmental stage was formulated with occasional assistance from online searches, forums, and artificial intelligence, like ChatGPT. This assistance streamline the workflow massively for debugging.

```
# Prepping the Random Forest
X_train,X_test,Y_train, Y_test = train_test_split(X,Y,test_size=0.75)
model = RandomForestClassifier(n_estimators = 70, max_depth = 40,
                               bootstrap = True, max_features = 12)
model.fit(X_train, Y_train)
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Using X_train_scaled and X_test_scaled for training and testing
model.fit(X_train_scaled, Y_train)
Y_pred = model.predict(X_test_scaled)
```

Figure 8. Code snippet for Random Forest Classifier training/testing (75% & 25% split)

Several important metrics exist for the Random Forest model, which define how "successful" the algorithm is at completing its prediction task:

- *Accuracy (%)* represents the overall correctness of the classifier's predictions.
- *Precision*, the positive predictive value, answers: "Of all the instances predicted as positive, how many are actually positive?"
- *Recall*, the sensitivity value, answers: "Of all the actual positive instances, how many did the model correctly predict?"
    - In this project, recall is more important as it is more costly and dangerous to falsely label a potential patient as "healthy" than the other way around.
- *F1 Score*, the weighted mean of precision and recall, provides a balanced metric for rates of both false positives and negatives.
- *Area Under the Precision-Recall Curve (AUC-PR)* provides insights into the classifier's performance across different levels of precision and recall.
- *Cross-Validation score* evaluates the generalizability of a predictive model, indicating how the model will perform on unseen data.

With this information, the hyperparameter tuning process can proceed, which will involve considerable data analysis. Every machine-learning algorithm has adjustable parameters, which, if tinkered with correctly, can boost performance significantly.

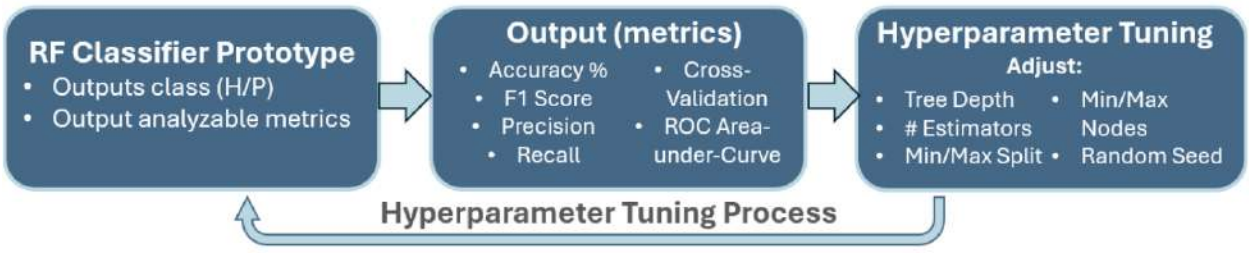


Figure 9. Model tuning flowchart

For example, one adjustable hyperparameter is the *number of features*, which accounts for the number of features taken into consideration during each decision-tree splitting process and contributes to the randomness or overfitting of the model. By utilizing *for* loops of increasing increments of *max_features*, the optimal range of this parameter was determined:

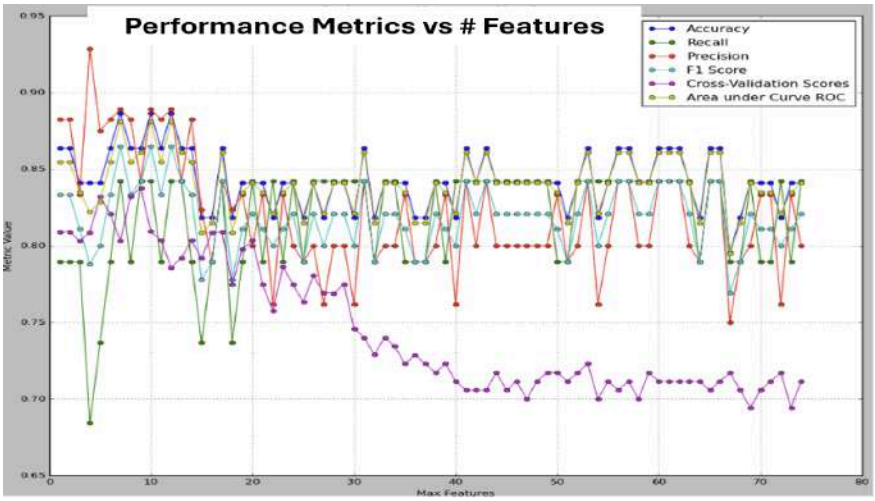


Figure 10. Performance metrics vs # of features (0-80)

From Figure 10, it was determined that 5-fold Cross-validation (generalizability) scores drop significantly after about a value of 10 for *max_features*, and that accuracy also decreases slightly around a *max_features* value of 15. Therefore, a *max_features* value of 9 was selected.

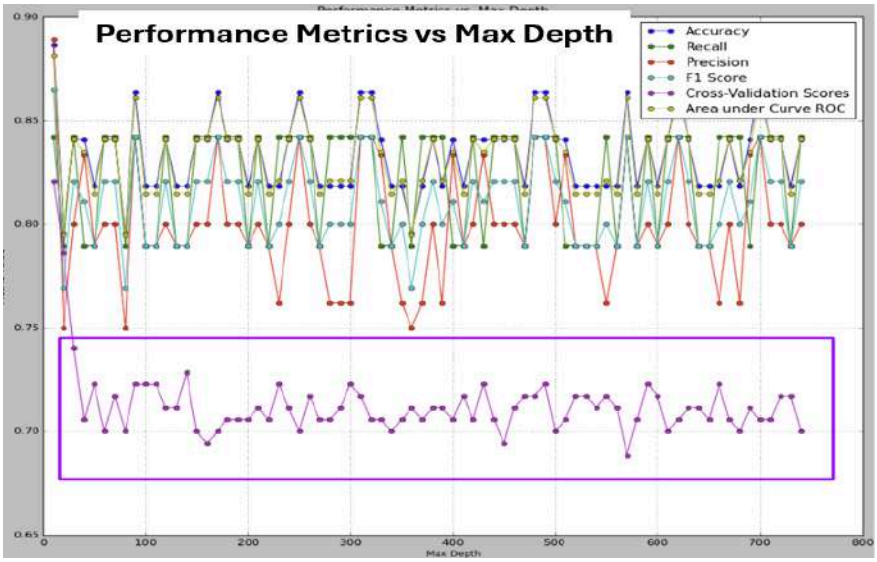


Figure 11. Performance metrics vs max depth (0-750)

Similarly, the optimal range for another parameter, *depth,* which controls how complex each individual decision-tree becomes, was experimented with. From Figure 11, it is seen that after a *max_depth* value of about 50, the 5-fold cross-validation score continually decreases and flattens out, indicating that a *max_depth* value that is less than 50 would create a model that generalizes better towards new datasets. It is also important to note that increasing the max depth gradually increases the computation time of the model since there are more nodes and decisions to consider, with a complexity of *O(log[n]).* Therefore, a *max_depth* value of 40 was selected.

While the hyperparameter tuning process can be elaborated on extensively, its exhaustive documentation exceeds the scope of this paper. However, before the next section, some preliminary data analysis regarding the trained and tuned model is necessary – it is imperative to have a "baseline" model to compare with when the real-life data is collected and tested.

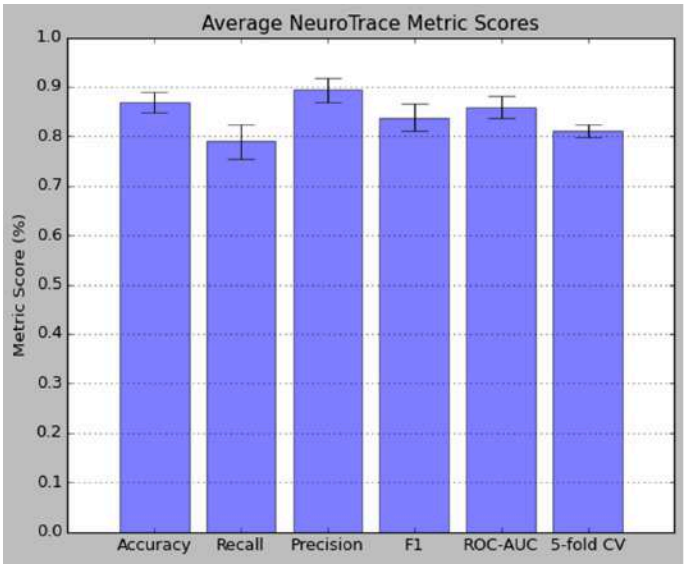


Figure 12.1. Average performance metrics for tuned NeuroTrace model

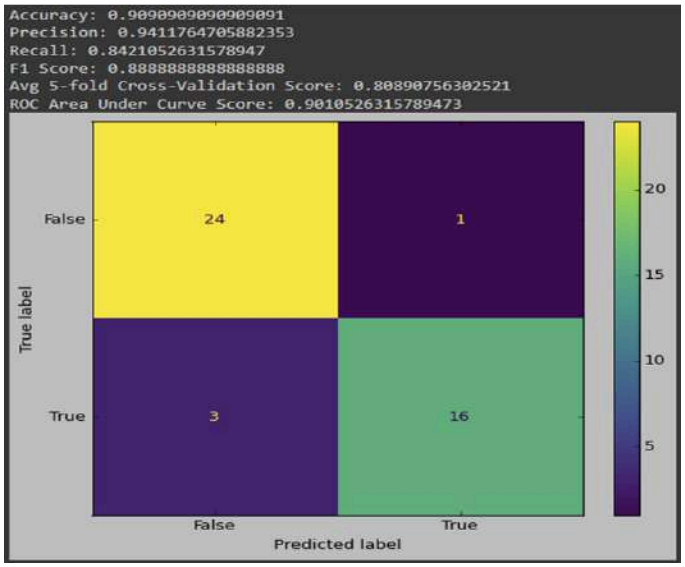


Figure 12.2. Confusion matrix for tuned model when testing with a 75/25 split

Thus, Figure 12.1 indicates that the model has high accuracy (90%) and precision (94%), with fairly good recall (84%). Intuitively, the model is very reliable at identifying/flagging cases,

with some false positives. The 5-fold Cross-validation score (80%) and ROC Area Under Curve score (90%) indicate that NeuroTrace generalizes well across unknown datasets and real-life data. Moreover, Figure 12.2 proves that NeuroTrace results are statistically significant, as the error bars do not overlap with 0.5, yielding significantly better results than chance (~50%).

3. NeuroTrace Frontend Development

The third major milestone was developing the NeuroTrace frontend application. Written completely in JavaScript and HTML, this prototype is portable (requires no local installation) and fast. Since NeuroTrace is currently designed to work with the WACOM Intuos Tablet, a portion of NeuroTrace was borrowed from the WACOM Developer GitHub page (WACOM cited in GitHub), a portal for open-source code showcasing device-integration capabilities of their tablets (more specifically, the *wacom-device-kit-web* webpage).

Figure 13. Testing device: WACOM Intuos Tablet Medium

The initial borrowed API already had tracing capabilities, where movements on the WACOM tablet would be mirrored instantaneously onto the computer screen attached, allowing for pen-tracking. However, the actual drawings were buggy and did not poll at a fast enough cycle, leading to straight lines and inaccurate reported data. Several weeks were spent ensuring that both the drawing projections and the handwriting kinematics data reporting were accurate.

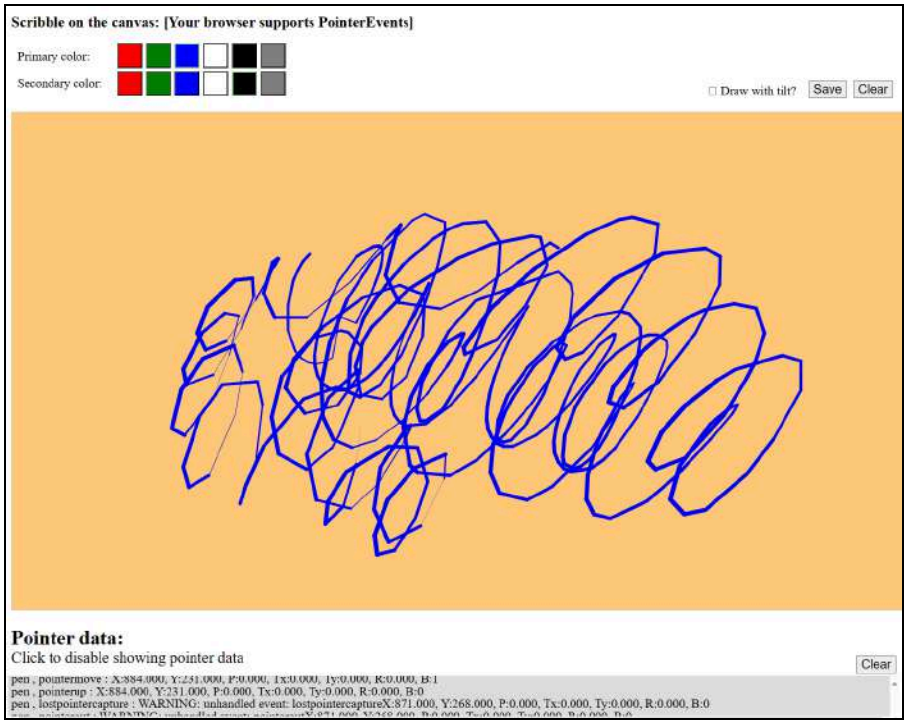


Figure 14. Initial wacom-device-kit-web API, often buggy and reporting inaccurate data.

In addition to fixing the buggy tracking, extra capabilities to the software were added, allowing the pen tracking of the handwriting kinematics to be exported via a CSV file at the press of a key. This involved:

a. Tracking the pen movements (x, y, pressure, tilt, pen lift ups and lift downs) at 120hz
b. Appending this continuous stream of data into an array that updates in real-time
c. Post-process said array with a header and converting the .txt format into .csv format

Then, several tracing guideline pictures (replicating the 6 NeuroTrace tasks) were created to replace the orange background in the figure above. The tasks were replicated to the closest dimensions possible (e.g. explicitly stated in DARWIN database that circle-tracing tasks were 3cm and 6cm in diameter, etc.) so the total time to complete these tasks would be normalized and similar to the original DARWIN data collection deviance. These pictures were also printed out and taped on top of the WACOM tablet, so participants could have a guideline of what to draw.

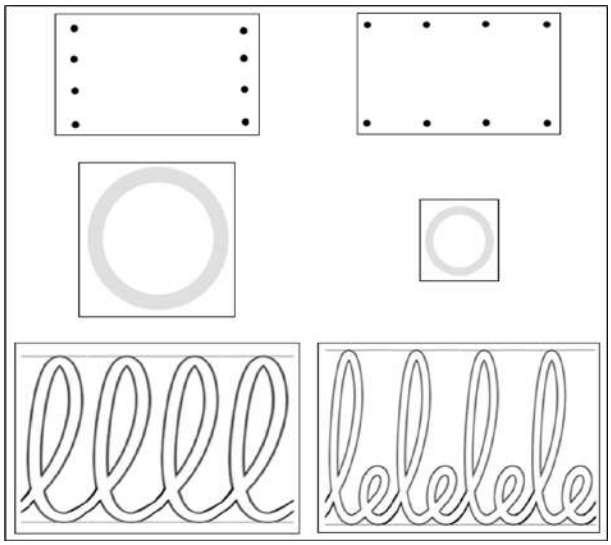


Figure 15.1. The 6 NeuroTrace tasks

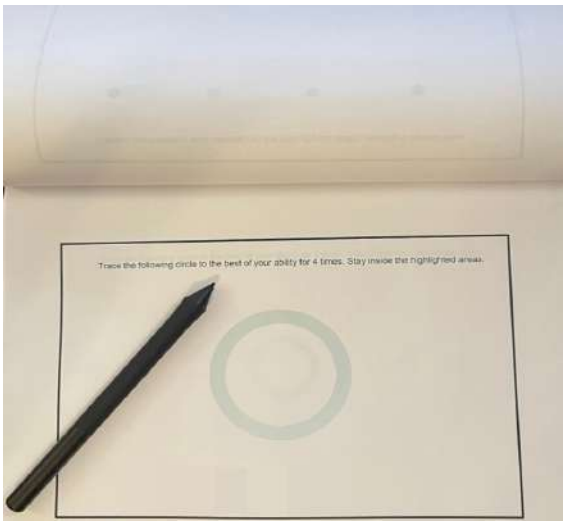


Figure 15.2. Experimental tablet setup

The WACOM Intuos tablet has a built-in software that allows screen-mapping to specific pixels on the computer screen. Therefore, the WACOM tablet mapped it exactly to the active regions of the NeuroTrace application on the computer, which is ~1050 x 1620 pixels.

So, ultimately, the NeuroTrace frontend process outline follows a process:

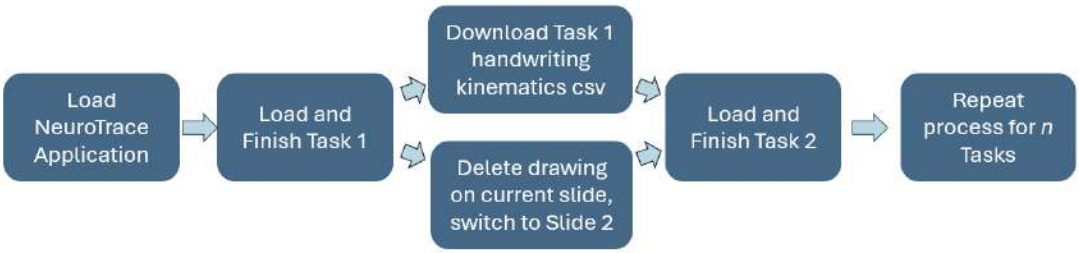


Figure 16. NeuroTrace application flowchart.

Some screenshots of the finished application are displayed:

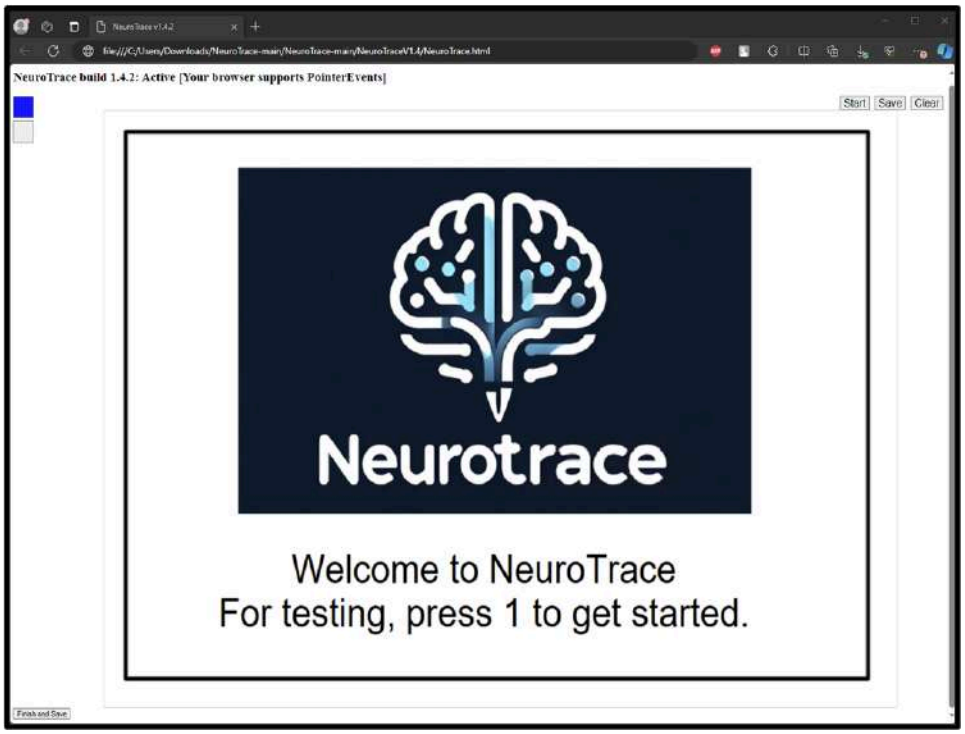


Figure 17. NeuroTrace home page.

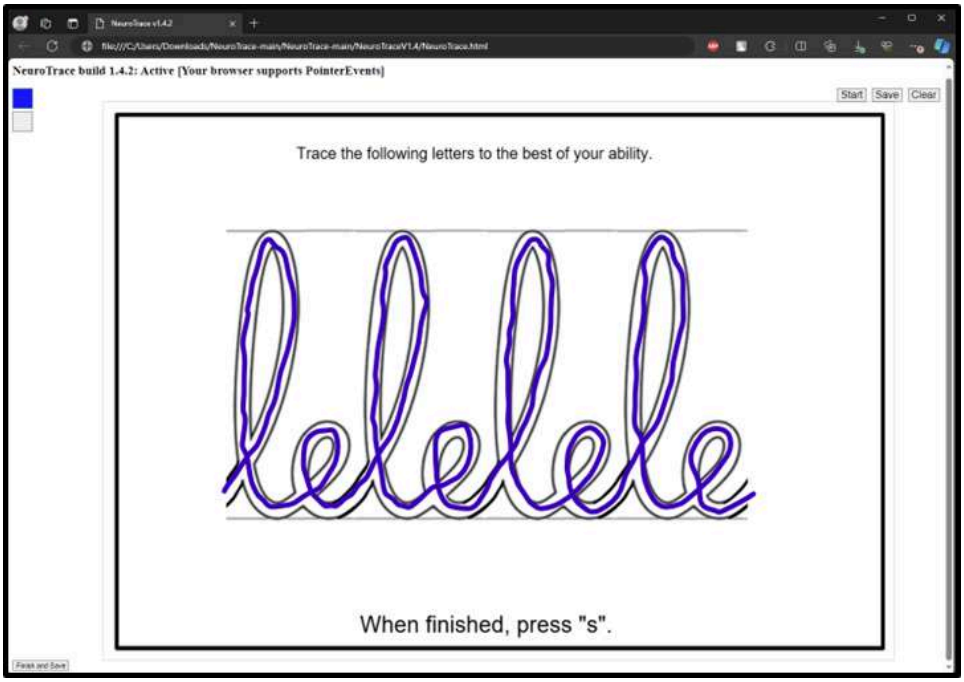


Figure 18. NeuroTrace 'le' tracing task

After the 6 raw CSV files are exported per user, they are uploaded back into Python to finish mathematical post-processing via *NumPy* and *Pandas*, since the raw CSVs are unusable. Conceptually, since these numbers are in the form of parametric equations, the following approximations can be used to find necessary metrics:

$$\text{avg. speed} = \frac{1}{n}\sum_{i=0}^{n}\sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2}$$

$$\text{avg. acceleration} = \frac{1}{n}\sum_{i=0}^{n}\frac{\sqrt{(x_{i+1} - 2x_i + x_{i-1})^2 + (y_{i+1} - 2y_i + y_{i-1})^2}}{120}$$

$$\text{avg. jerk} = \frac{1}{n}\sum_{i=0}^{n}\frac{\sqrt{(x_{i+2} - 3x_{i+1} + 3x_i - x_{i-1})^2 + (y_{i+2} - 3y_{i+1} + 3y_i - y_{i-1})^2}}{120^2}$$

Figure 19. Central difference approximation formulas for parametric equations (120hz)

The CSVs are then converted into 11 usable metrics/features:

[air_time, mean_acc_in_air, mean_acc_on_paper, mean_jerk_in_air, mean_jerk_on_paper, mean_speed_in_air, mean_speed_on_paper, num_of_pendown, paper_time, pressure_mean, total_time]

Finally, the array of 6 * 11 usable metrics (for the 6 tasks) of each participant (n=18, in the pilot-testing case) are appended together and exported directly to the trained Random Forest model for testing.

4. Real-Life NeuroTrace Pilot Testing

Verifying the accuracy of NeuroTrace necessitated volunteers that would be willing to participate in a trial-run of NeuroTrace. After drafting up informed consent forms and an overall experimental plan, nearby senior facilities, such as memory cares, independent living, and public

senior centers, were contacted. After weeks of back-and-forth contact with several facilities, 18 participants were secured by the end of the project. The experimental procedure is outlined below, and took less than 10 minutes on average per participant:

1. The willing participant was notified that they (or a Power of Attorney) were required to sign an informed consent form (which explained the purpose of study and data collection/protection procedures), that participation was optional, and all sensitive information would be kept confidential.
2. After signing, the participant (or a Power of Attorney) filled out a quick, 5-min demographic questionnaire, which included content such as:
    - Self-diagnosing whether they
        a. Currently suffer from a neurodegenerative disease

           OR

        b. Think they suffer from symptoms of a neurodegenerative disease
    - Recall their former vocational/educational history, including:
        i. Number of years of high school/college (or other secondary education)
        ii. A description of their longest past career/field & avg. hours worked/week
        iii. Age of retirement
3. Utilized the WACOM tablet and digital pen to complete the 6 NeuroTrace tracing tasks.

Overall, the participants were receptive towards the study, with many enthusiastically finishing the handwriting tasks and quip me with humorous remarks such as:

- “So… Did I pass the handwriting test?”
- “Don’t let *him* trace! His handwriting is so bad it’ll diagnose him with every disease!”

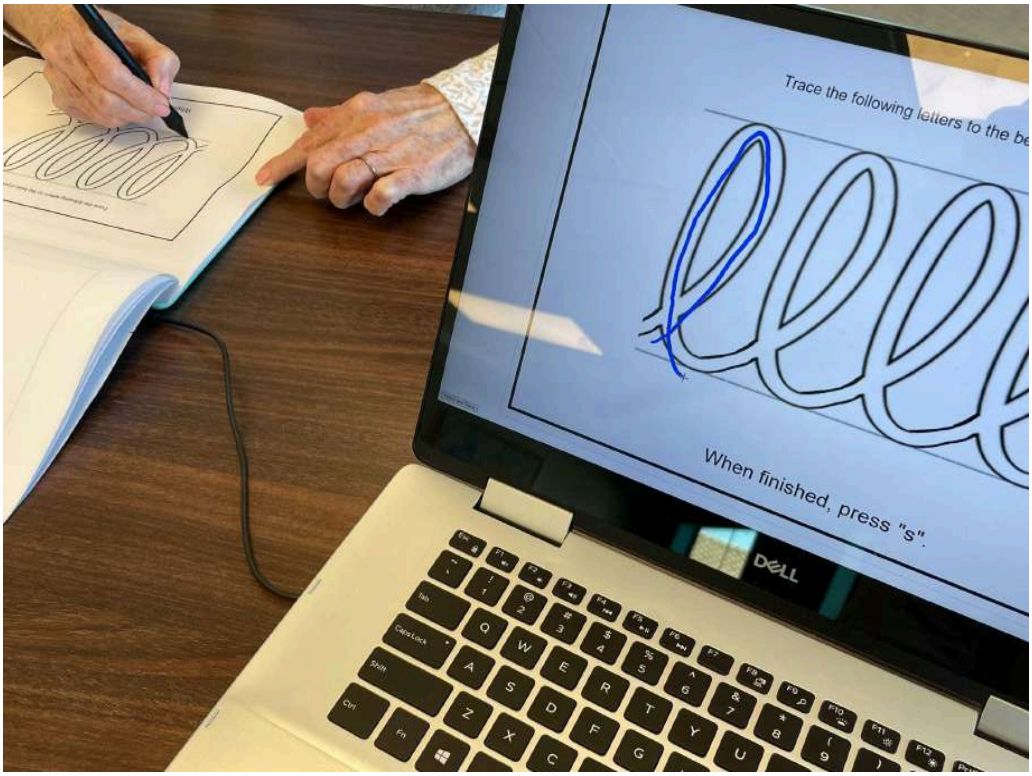


Figure 20. Participant completing the 'cursive l' task. Depicted is the typical experimental setup.

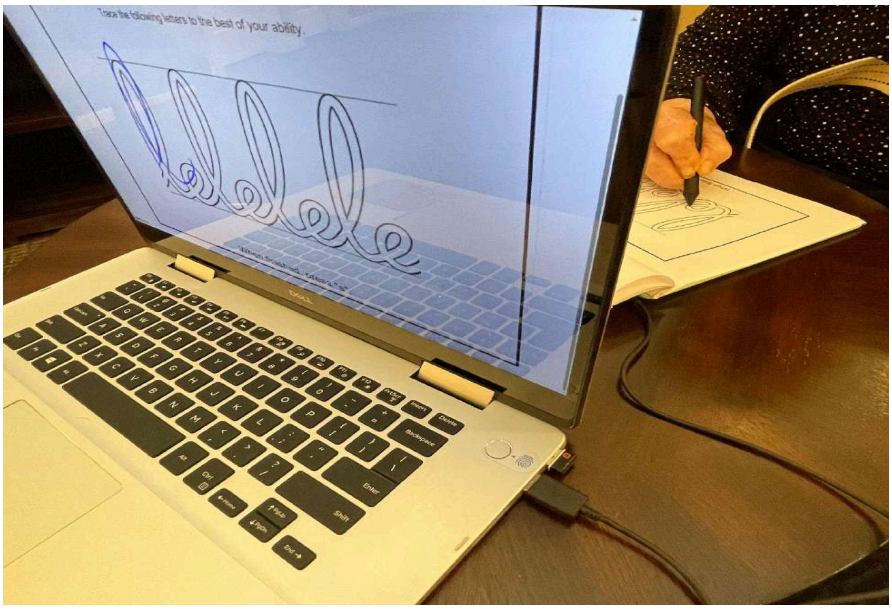


Figure 21. Another participant (Alzheimer's patient). Although she (age 97) couldn't speak or communicate well, she followed directions perfectly!

After analysis of the demographic questionnaires by the 18 participants, some insights appeared:

- There were 6 participants self-reporting a neurodegenerative disease, primarily:
    - Alzheimer's, Parkinson's, Multiple Sclerosis, and Essential Tremor
- The rest of the participants (12) self-reported as healthy.
- The average age of the participants was 80.

Subsequently, the 18 participants' raw handwriting kinematics CSV data was entered into the NeuroTrace Python preprocessing program, exported as one CSV file (labeled with *Class* as well to indicate healthy control or patient):

ID	air time2	mean acc in air2	mean acc on paper2	mean jerk in air2	mean jerk on paper2	mean speed in air2	mean speed on paper2	num of pendown2	paper time2	pressure mean2	total time2	Class (0 = H, 1 = P)
1	139	20.0645	0.3163	0.023	0.0464	5.157	8.4577	5	2656	2600.309	2794.375	0
2	139	0.1894	0.2187	0.031	0.0361	1.9908	6.8318	4	3141	1569.507	3279.35	0
3	1080	0.345	0.1967	0.04	0.0312	4.5166	3.8983	7	4688	1205.668	5767.729	1
4	2090	0.3594	0.203	0.0383	0.0239	7.1328	3.2209	5	7696	1695.401	9786.087	1
5	583	0.8118	0.5028	0.1173	0.0787	20.2619	17.2233	3	1195	1652.087	1778.239	0
6	803	0.5746	0.5324	0.1153	0.0741	7.8917	8.7556	5	2927	1722.367	3729.683	0
7	1518	0.5074	0.2165	0.0632	0.0291	12.1568	5.4828	5	4284	1341.825	5802.37	0
8	1634	0.3875	0.3551	0.0349	0.0336	3.8881	4.9947	5	4474	1485.496	6108.366	1

Figure 22. A snippet of processed CSV file, ready for NeuroTrace analysis testing

Data Analysis

By testing the processed data with the trained NeuroTrace Random Forest model, it is evident that the model was a success, with an 88% accuracy rate at identifying the correct *class* of the participants (i.e. whether healthy or patient), with 5-fold cross-validation. While the real-life collected dataset is slightly skewed, with more healthy controls (12) than patients of NDs (6), the NeuroTrace model was able to adjust to this and produce a decently accurate result nonetheless.

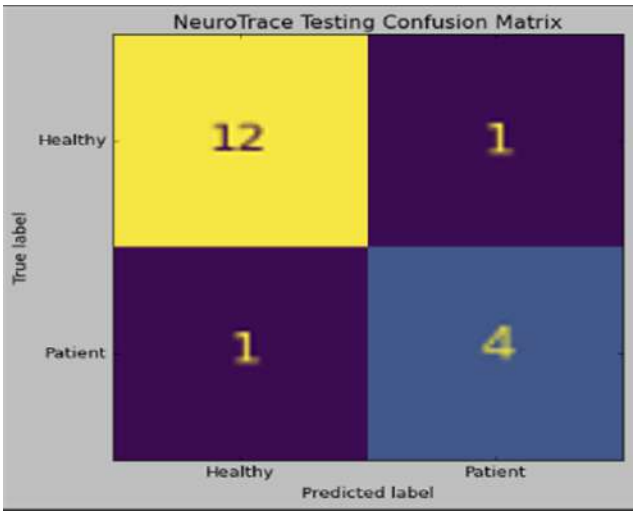


Figure 23. Confusion matrix for NeuroTrace validation data.

From the confusion matrix above, it is evident that NeuroTrace identifies most patients (positives) as patients, and most healthy controls (negatives) as healthy.

	Accuracy	Precision	Recall	F1 Score
Training	90%	94%	84%	88%
Senior Testing	88%	80%	80%	80%

Figure 24. NeuroTrace model training vs. model testing accuracy metrics.

Additionally, from Figure 24 above, the accuracy only decreased slightly, from 90% to 88%, while the precision, recall, and F1 (weighted average of precision and recall scores) dropped from 94%, 84%, and 88% to 80% for all three. This is normal, as training datasets are expected to be more accurate than testing datasets, as testing datasets' performances will vary based on the degree of normalization of the newly unseen data.

Thus, from solely the participant data (n=18), with accuracy metrics above 80%, NeuroTrace is reliable at detecting neurodegeneration, proving that a distinct difference between healthy control subjects' and NDs subjects' handwriting kinematics exist.

Limitations

One challenge for NeuroTrace is detecting NDs when hand or motor control is not impacted. For example, some patients with dementia may have motor control comparable to the average healthy senior of the same age. In these cases, NeuroTrace would be unable to detect or flag the individual for potential NDs. Therefore, for the most accurate classifier, more data and samples are still needed. 18 samples absolutely cannot represent an entire region, country, or worldwide population of seniors with and without NDs.

Subsequently, while one asset of NeuroTrace is its capability to detect NDs with 6 simple tracing tasks, some may argue whether only these 6 tasks provide enough complexity and depth to accurately detect neurodegeneration. This was evident during some patient testing, where some subjects with NDs traced (qualitatively speaking) steadily and as well as healthy controls.

Likewise, training databases and collected samples that currently train and serve as reference NeuroTrace are limited to certain populations. For example, the utilized DARWIN dataset in this classifier contained only Italian seniors, and may not generalize well to all populations across the world. Finally, the single-dimensionality of numerical data is limiting; there is only so much that can be analyzed and interpreted from just numerical data – this will lead into the future research section.

Future Research

In the future, there will be a need to utilize other datasets in conjunction with DARWIN to train and amalgamate a more robust model. This is where multimodal software integration,

such as qualitative image analysis or other sensory data (speech, hearing, etc), can provide deeper insights and more accurate results. For example, while one subject with Parkinson's had steady tremors but had no difficulty completing the tasks, another subject with Alzheimer's could understand the tasks but would get confused by the tracing tasks (and ultimately be unable to complete certain tasks). With more data collection and model training, it could be possible for NeuroTrace to even differentiate between the different types of neurodegeneration in the future.

Furthermore, analysis of seniors' individual lifestyles through user personalization (profiling) in NeuroTrace has the ability to deepen model complexity and heighten accuracy. Personalization methods could involve asking for personal information such as demographics (age, ethnicity), genetic history (to check for genetic NDs), vocational/educational history (as some careers have statistical correlations with NDs), education history, and overall lifestyle (diet, hobbies, health). By utilizing more online databases such as the *GEO (Gene Expression Omnibus),* the users' details could be matched to preexisting profiles in the database to warn the users/healthcare providers as a preemptive measure.

Additionally, future development could focus on extending device support to all digital platforms. Since devices like WACOM digitizing tablets are less common, broader device support – such as iPads, Galaxy Tab, or Microsoft Surface – would streamline setup, implementation, and testing processes, while still maintaining the tracking accuracy and sensitivity of a digitizing tablet (following some workarounds and likely additional coding).

Conclusion & Implications

Ultimately, based on real-life participant data, NeuroTrace will ensure early neurodegeneration detection with an accuracy greater than 80%. The continuously-improving artificial intelligence model will contribute to early detection efforts, enabling timely interventions, improving patient outcomes, and enhancing quality-of-life. Additionally, NeuroTrace will greatly contribute to handwriting and kinematics relevance in the research field of neurodegeneration – emphasizing the relevance of handwriting kinematics as a potential biomarker for neurodegenerative diseases.

A novel system software that can streamline medical procedures for hundreds of millions worldwide, NeuroTrace has several implications and applications, as a quick, accurate, and non-invasive screening tool for seniors worldwide. For instance, with the recent implementation of in-store digital (iPad-based) self-screening/checkup systems at a nearby Walgreens, users can perform wellness checkups and screenings more efficiently than before. In this case, NeuroTrace can be perfectly applied as an addition to these check-up systems where it is targeted towards the senior or soon-to-be-senior population. Additionally, it can serve as a cheap, efficient application for low-resource groups, with support for all devices, browsers, and operating systems. Moreover, NeuroTrace will effectively act as a remote diagnostic in areas without nearby healthcare or neurology professionals and serve as a virtual medical monitor in a world that is growing more online and distanced than ever. **Ultimately, NeuroTrace creates a future where the burden of neurodegenerative diseases is alleviated, increasing quality-of-life for those who need and deserve it most.**

Works Cited

Cilia, Nicole D.10, et al. "Diagnosing Alzheimer's Disease from On-Line Handwriting: A Novel Dataset and Performance Benchmarking." *Engineering Applications of Artificial Intelligence*, vol. 111, no. 104822, May 2022, p. 104822, https://doi.org/10.1016/j.engappai.2022.104822.

GitHub.com. "Wacom-Developer/Wacom-Device-Kit-Web." *GitHub*, 13 Mar. 2021, github.com/Wacom-Developer/wacom-device-kit-web.

Gopal, Arpita, et al. "Remote Assessments of Hand Function in Neurological Disorders: Systematic Review." *JMIR Rehabilitation and Assistive Technologies*, vol. 9, no. 1, 9 Mar. 2022, p. e33157, https://doi.org/10.2196/33157.

Hu, Yihan, et al. "Infections among Individuals with Multiple Sclerosis, Alzheimer's Disease and Parkinson's Disease." *Brain Communications*, vol. 5, no. 2, 16 Mar. 2023, www.ncbi.nlm.nih.gov/pmc/articles/PMC10053639/#:~:text=An%20estimated%2050%20million%20people, https://doi.org/10.1093/braincomms/fcad065.

Ibtissame Aouraghe, et al. *Automatic Analysis of On‑Line Handwriting for the Early Detection of Neurodegenerative Pathologies*. 19 June 2017.

Josep Garre-Olmo, et al. "Kinematic and Pressure Features of Handwriting and Drawing: Preliminary Results between Patients with Mild Cognitive Impairment, Alzheimer Disease and Healthy Controls." *Current Alzheimer Research*, vol. 14, no. 9, 2023, pp. 960–968, doi.org/10.2174%2F1567205014666170309120708.

Kho, Julia. "Why Random Forest Is My Favorite Machine Learning Model." *Medium*, Towards Data Science, 19 Oct. 2018,

towardsdatascience.com/why-random-forest-is-my-favorite-machine-learning-model-b97651fa3706.

Mashio, Yuhei, and Hideo Kawaguchi. “Detecting Early Symptoms of Mental Health Deterioration Using Handwriting Duration Parameters.” *Neuropsychopharmacology Reports*, vol. 40, no. 3, 4 July 2020, pp. 246–253, https://doi.org/10.1002/npr2.12123.

National Institute of Environmental Health Sciences. “Neurodegenerative Diseases.” *National Institute of Environmental Health Sciences*, 2018, www.niehs.nih.gov/research/supported/health/neurodegenerative/index.cfm.

Shusharina, Natalia, et al. “Modern Methods of Diagnostics and Treatment of Neurodegenerative Diseases and Depression.” *Diagnostics*, vol. 13, no. 3, 3 Feb. 2023, p. 573, www.ncbi.nlm.nih.gov/pmc/articles/PMC9914271/, https://doi.org/10.3390/diagnostics13030573.

TeamAI. “Parkinson Disease Spiral Drawings - Database.” *Www.kaggle.com*, 2016, www.kaggle.com/datasets/team-ai/parkinson-disease-spiral-drawings.

Vessio, Gennaro. “Dynamic Handwriting Analysis for Neurodegenerative Disease Assessment: A Literary Review.” *Applied Sciences*, vol. 9, no. 21, 1 Nov. 2019, p. 4666, https://doi.org/10.3390/app9214666..

WACOM. “OpenGraph Data.” *Wacom*, 2023, www.wacom.com/en-us/about-wacom.

Walton, Judie. “Handwriting Changes due to Aging and Parkinson’s Syndrome.” *Forensic Science International*, vol. 88, no. 3, Aug. 1997, pp. 197–214, https://doi.org/10.1016/s0379-0738(97)00105-9.

Appendices

Note: Certain portions of appendices blacked out to protect privacy

Appendix A.

Informed Consent Forms for Participants with Self-Power of Attorney

Human Informed Consent Form

Instructions to the Student Researcher(s): The informed consent/assent/permission form should be developed in consultation with the Adult Sponsor, Designated Supervisor or Qualified Scientist.
This form is used to provide information to the research participant (or parent/guardian) and to document written informed consent, minor assent, and/or parental permission.

- When written documentation is required, the researcher keeps the original, signed form.
- Students may use this sample form or may copy ALL elements of it into a new document.

If the form is serving to document parental permission, a copy of any survey or questionnaire must be attached.

Student Researcher(s): [redacted]

Title of Project: NeuroTrace: Utilizing a Machine-Powered Approach to Detect the Presence of Neurodegenerative Diseases through Vocational/Education History and Handwriting Kinematics Data

I am asking for your voluntary participation in my science fair project. Please read the following information about the project. If you would like to participate, please sign in the appropriate area below.

Purpose of the project:
To determine a correlation between subjects' (especially seniors') handwriting, educational & vocational histories (how long/industry worked in) and whether they suffer from or exhibit symptoms of neurodegenerative diseases like Alzheimer's or Parkinson's.
The survey and kinematics data of the participant will be used to train and test the coded Python AI classifier.

If you participate, you will be asked to:
1. They (or a POA) needs to answer a questionnaire to i) self-diagnose whether they a) currently suffer from a neurodegenerative disease OR b) think they suffer from symptoms of a neurodegenerative disease AND ii) to recall their former vocational/educational history, including number of years of high school/college (or other secondary education), listing their longest past career, age of retirement, etc.
2. Utilize an electronic tablet and pen to trace-draw several pictures/shapes and trace-write several brief sentences. The kinematics (speed, pressure, etc) of the handwriting will be collected for further processing.

Time required for participation:
5-15 minutes

Potential Risks of Study:
Moderate risk - survey and tracing may trigger unwarranted and unexpected behaviors, such as stress/anxiety, in some patients
Supervision during the entirety of seniors' participation may be needed.

Benefits:
The surveyed vocational data, along with the raw handwriting kinematics data, will be processed and normalized into features such as average tremor, speed, acceleration, etc. This processed data (csv file) will be directly analyzed through a Python-based machine-learning Random Forest Classifier to train and test the model (already trained using the online DARWIN dataset) to classify new samples as "healthy" or "patient"
Ultimately, this contributes to future machine-powered technology where neurodegenerative diseases can be efficiently detected earlier, increasing quality of life (especially for the growing senior population).

How confidentiality will be maintained:
Questions are completely optional. Participants will not be identified by name. All data is anonymous, and the actual handwriting itself is not analyzed - only the kinematics (motion) behind it. Sensitive educational, vocational, and medical data will remain password-encrypted, and all data will be prompted deleted after experiment.

If you have any questions about this study, feel free to conta[redacted], NeuroTrace8@gmail.com

Adult Sponsor/QS/DS: [redacted] Phone/email: [redacted]

Voluntary Participation:
Participation in this study is completely voluntary. If you decide not to participate there will not be negative consequences. Please be aware that if you decide to participate, you may stop participating at any time and you may decide not to answer any specific question.

By signing this form I am attesting that I have read and understand the information above and I freely give my consent/ assent to participate or permission for my child to participate.

Adult Informed Consent or Minor Assent Date Reviewed & Signed: _____ (mm/dd/yy)

Research Participant Printed Name: _____ Signature: _____

Parental/Guardian Permission (if applicable) Date Reviewed & Signed: _____ (mm/dd/yy)

Parent/Guardian Printed Name: _____ Signature: _____

Page 40 International Rules: Guidelines for Science and Engineering Fairs 2023-2024, **societyforscience.org/ISEF**

Appendix B.

Informed Consent Forms for Participants WITHOUT Self-Power of Attorney

Human Informed Consent Form

Instructions to the Student Researcher(s): rmed consent/assent/permission form should be developed in consultation with the Adult Sponsor, Designated Supervisor or Qualified Scientist.
This form is used to provide information to the research participant (or parent/guardian) and to document written informed consent, minor assent, and/or parental permission.

- When written documentation is required, the researcher keeps the original, signed form.
- Students may use this sample form or may copy ALL elements of it into a new document.

If the form is serving to document parental permission, a copy of any survey or questionnaire must be attached.

Student Researcher(s): [redacted]

Title of Project: NeuroTrace: Utilizing a Machine-Powered Approach to Detect the Presence of Neurodegenerative Diseases through Vocational/Education History and Handwriting Kinematics Data

I am asking for your voluntary participation in my science fair project. Please read the following information about the project. If you would like to participate, please sign in the appropriate area below.

Purpose of the project:
To determine a correlation between subjects' (especially seniors') handwriting, educational & vocational histories (how long/industry worked in) and whether they suffer from or exhibit symptoms of neurodegenerative diseases like Alzheimer's or Parkinson's.
The survey and kinematics data of the participant (whom has already been diagnosed with a neurodegenerative disease) will be used as a baseline/"positive" outcome to train the coded Python AI classifier.

If you participate, you will be asked to:
1. They (or a POA) needs to answer a questionnaire to i) self-diagnose whether they a) currently suffer from a neurodegenerative disease OR b) think they suffer from symptoms of a neurodegenerative disease AND ii) to recall their former vocational/educational history, including number of years of high school/college (or other secondary education), listing their longest past career, age of retirement, etc.
2. Utilize an electronic tablet and pen to trace-draw several pictures/shapes and trace-write several brief sentences. The kinematics (speed, pressure, etc) of the handwriting will be collected for further processing.

Time required for participation:
15-25 minutes

Potential Risks of Study:
Moderate risk - survey and tracing may trigger unwarranted and unexpected behaviors, such as stress/anxiety, in some patients
Supervision during the entirety of seniors' participation may be needed.

Benefits:
The surveyed vocational data, along with the raw handwriting kinematics data, will be processed and normalized into features such as average tremor, speed, acceleration, etc. This processed data (csv file) will be directly analyzed through a Python-based machine-learning Random Forest Classifier to train and test the model (already trained using the online DARWIN dataset) to classify new samples as "healthy" or "patient". Ultimately, this contributes to future machine-powered technology where neurodegenerative diseases can be efficiently detected earlier, increasing quality of life (especially for the growing senior population).

How confidentiality will be maintained:
Questions are completely optional. Participants will not be identified by name. All data is anonymous, and the actual handwriting itself is not analyzed - only the kinematics (motion) behind it. Sensitive educational, vocational, and medical data will remain password-encrypted, and all data will be promptly deleted after experiment.

If you have any questions about this study, feel free to conta[redacted] NeuroTrace8@gmail.com

Adult Sponsor/QS/DS: [redacted] Phone/email: [redacted]

Voluntary Participation:
Participation in this study is completely voluntary. If you decide not to participate there will not be negative consequences. Please be aware that if you decide to participate, you may stop participating at any time and you may decide not to answer any specific question.

By signing this form I am attesting that I have read and understand the information above and I freely give my consent/ assent to participate or permission for my child to participate.

Adult Informed Consent or Minor Assent Date Reviewed & Signed: _____ (mm/dd/yy)

Research Participant Printed Name: _____ Signature: _____

Parental/Guardian Permission (if applicable) Date Reviewed & Signed: _____ (mm/dd/yy)

Parent/Guardian Printed Name: _____ Signature: _____

Appendix C.

NeuroTrace Study Recruitment Flyer

For more information, contact Alex at NeuroTrace8@gmail.com

NeuroTrace

A Novel Machine-Learning Study For Neurodegeneration Detection

NeuroTrace is an innovative study aimed at **early detection** of neurodegenerative diseases (NDs) like **Parkinson's, Alzheimer's,** and **Multiple Sclerosis** through **handwriting kinematics analysis.**

Why Handwriting?

Hand movements, affected early in NDs, can indicate underlying issues. NeuroTrace seeks to identify these diseases earlier using machine learning and artificial intelligence (AI) by gathering training/testing data.

How It Works:

1. Participants fill out an **quick, anonymous survey** about their background for demographic purposes
2. Participants perform **handwriting tracing tasks** on a tablet
3. Participants' handwriting kinematics data is analyzed, like **velocity, acceleration, and pressure**
4. Participants contribute towards training of an AI (computer machine learning model) **that will be able to distinguish between healthy individuals and those at risk with high precision and accuracy, increasing patient quality-of-life**

All collected data will be anonymous. Informed consent forms provided upon request & before study participation.

Appendix D.

Demographic Survey Completed by all Participants

For questions or concerns, contact us at NeuroTrace8@gmail.com.

Research Questionnaire: Utilizing a Machine-Powered Approach to Detect the Presence or Early Onset of Neurodegenerative Diseases through Health, Vocational & Educational History and Handwriting Data

Dear Participant,

Thank you for your willingness to contribute to this research study! Your participation is invaluable in advancing society's knowledge of neurodegenerative diseases.

Please be assured that all information collected is **strictly confidential, password-protected,** and **will be used for research purposes only**. Data will be **permanently deleted** after the completion of this research project.
You have the right to opt out at any time.

- ☐ I consent to participate in the study and consent the usage of my data for research purposes.
- ☐ I do not consent to participate in the study.

By signing below, you acknowledge that you voluntarily consent to participate in the research study.

Participant's Name: _____

Participant's Signature: _____
Date: _____

Section 1: Personal Information

Please provide the following information for demographic purposes.

1.1. Age: _____ years old

1.2. Gender:

- ☐ Male
- ☐ Female
- ☐ Other (please specify: _____)

1.3. Ethnicity:

- ☐ White
- ☐ Asian / Pacific Islander
- ☐ Indigenous / Native
- ☐ Hispanic
- ☐ African
- ☐ Other: _____

Section 2: Self-Diagnosis and Medical History

(All responses are optional. You have the right to opt out at any time.)

2.1. Do you know if any of your parents, siblings, or children have been diagnosed with a neurodegenerative disorder (ND), such as Alzheimer's, Parkinson's, Huntington's, dementia, or Amyotrophic Lateral Sclerosis (or any other NDs) before?

- ☐ Yes
- ☐ No

2.2. Have YOU been diagnosed with a neurodegenerative disorder (ND), such as Alzheimer's, Parkinson's, Huntington's, dementia, or Amyotrophic Lateral Sclerosis (or any other NDs)?

- ☐ Yes
- ☐ No

2.3. If you answered 'No' to 2.2 above, do you believe you exhibit any of the following symptoms below?

☐ Yes ☐ No

- **Memory Loss:**
- **Cognitive Impairment:**
  - Difficulty concentrating or focusing on a task
- **Motor Function Issues:**
  - Difficulty writing properly
  - Tremors or shaking in hands or other body parts
  - Stiffness in muscles
- **Changes in Speech:**
  - Slurred speech
  - Difficulty articulating thoughts
  - Trouble finding the right words
- **Personality and Mood Changes:**
- **Sleep Disturbances:**
- **Visual Disturbances:**
  - Difficulty reading or recognizing familiar objects
  - Blurred or double vision
  - Changes in perception of colors
- **Loss of Smell or Taste:**
- **Difficulty with Everyday Tasks**

(go on to the next page.)

Section 3: Vocational and Educational History

3.1. Did you graduate high school? If not, what was the highest level of education you completed?

☐ Yes, I graduated high school.

☐ No, my highest level of education is: _____

3.2. If you answered yes to 3.1, how many years did you spend in a post-high school secondary education institution, such as college, trade/vocational school, apprenticeship, and/or the army?

☐ _____ years

☐ Not applicable (did not pursue higher education)

3.3. Please indicate your longest-held job position. Include the position title, career industry, and average hours worked per week, If you were self-employed, provide a brief explanation.

Longest-held position title	Industry worked (education, finance, health, tech, retail, etc)	Hours worked per week
		hours/ week

3.4. At what age did you retire from your last formal position?

_____ years

Section 4: Additional Information and Conclusion

4.1. Do you have any additional information you would like to share? If so, please comment in the box below.

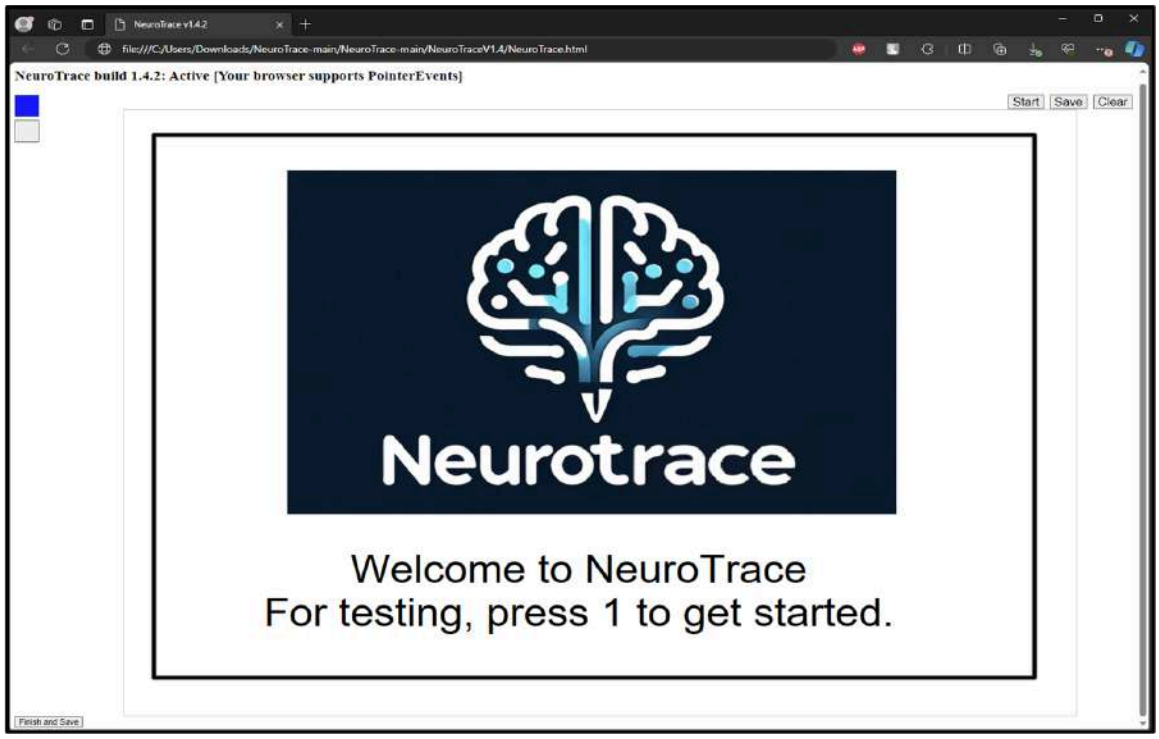
4.2 Do you have any concerns or feedback regarding the questionnaire? If so, please comment in the box below.

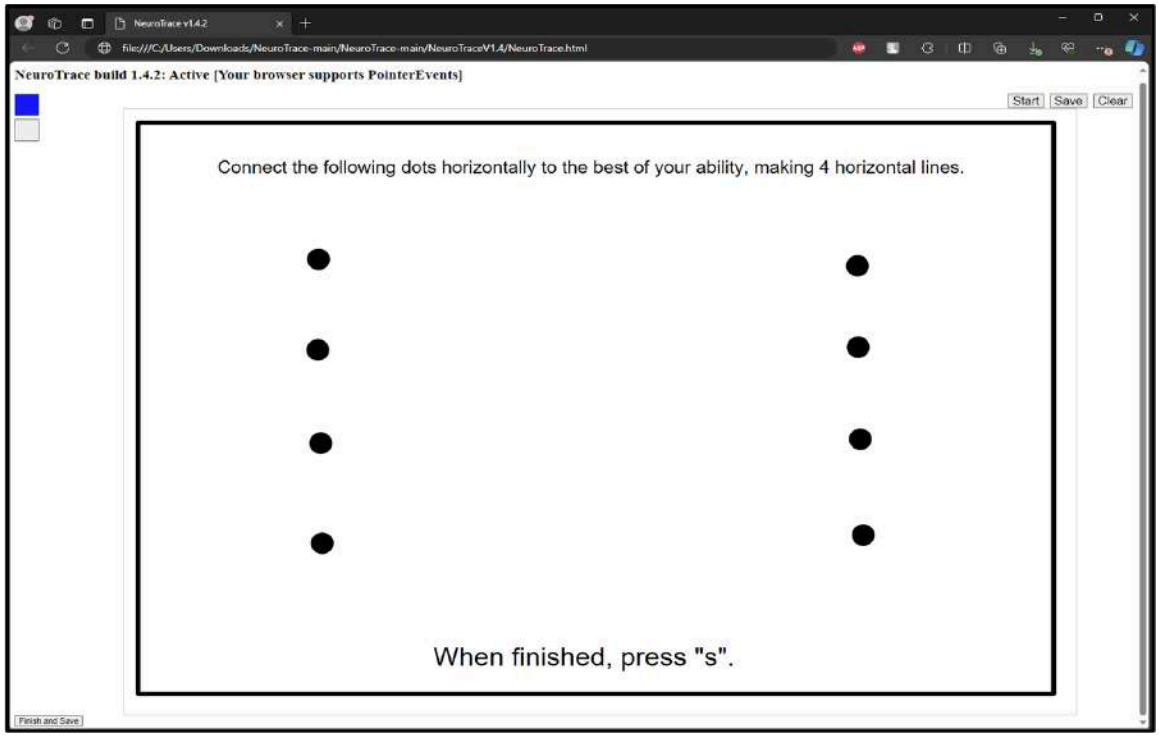
Thank you for your time and participation in this research study. Your input is vital to advancing our understanding of neurodegenerative diseases, saving lives in the future.

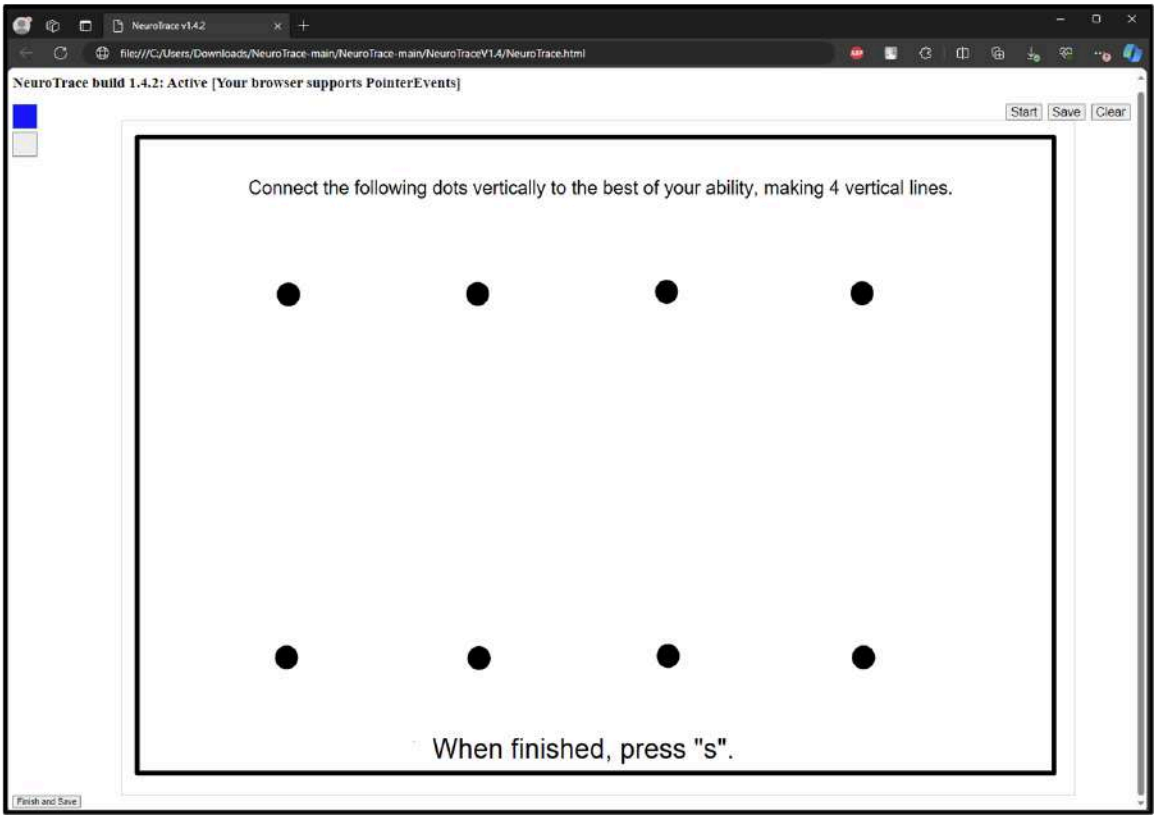
If you have any questions or concerns, please contact the researcher at NeuroTrace8@gmail.com.

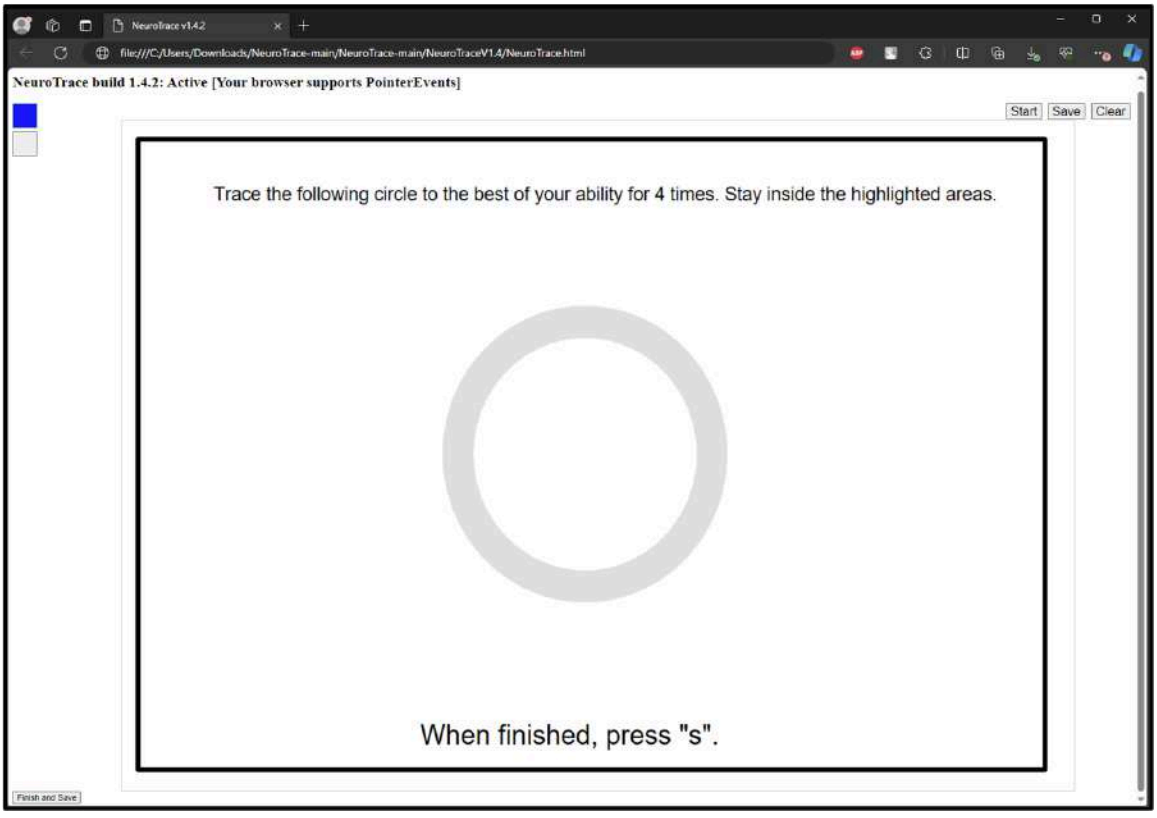
Appendix E.

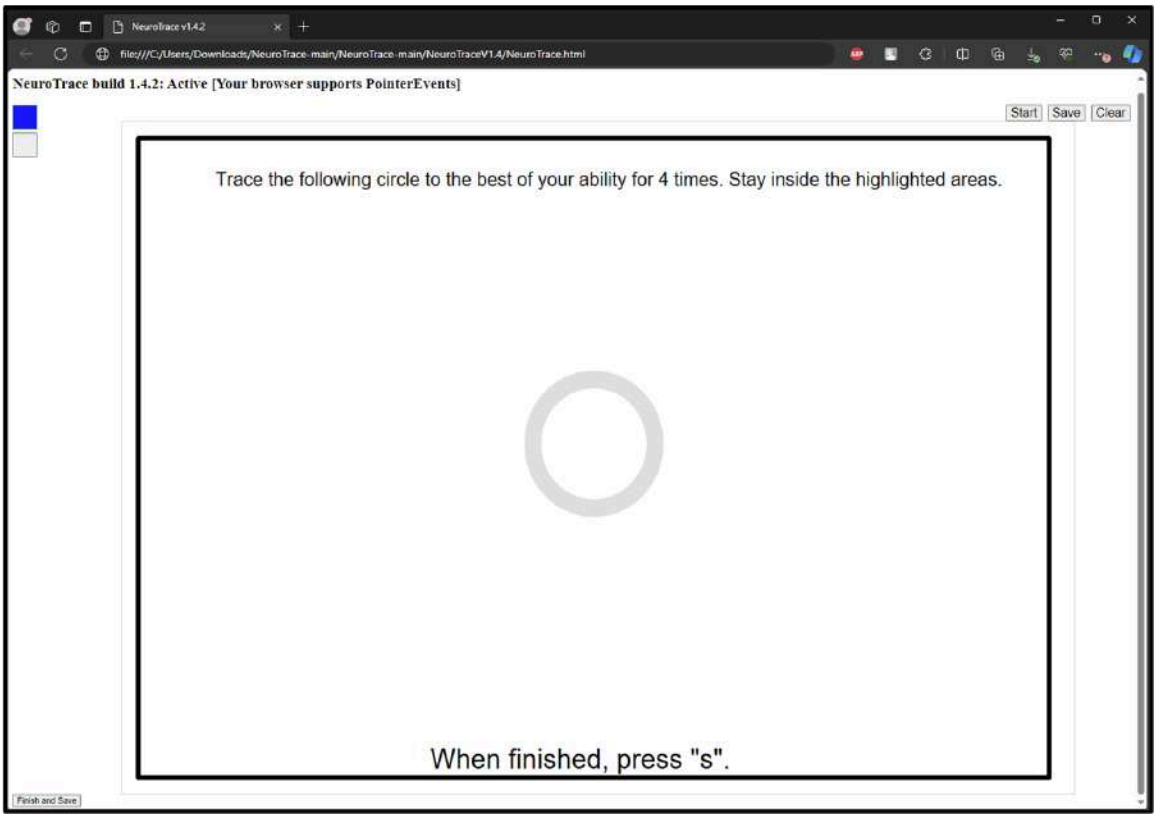
NeuroTrace Application Interface

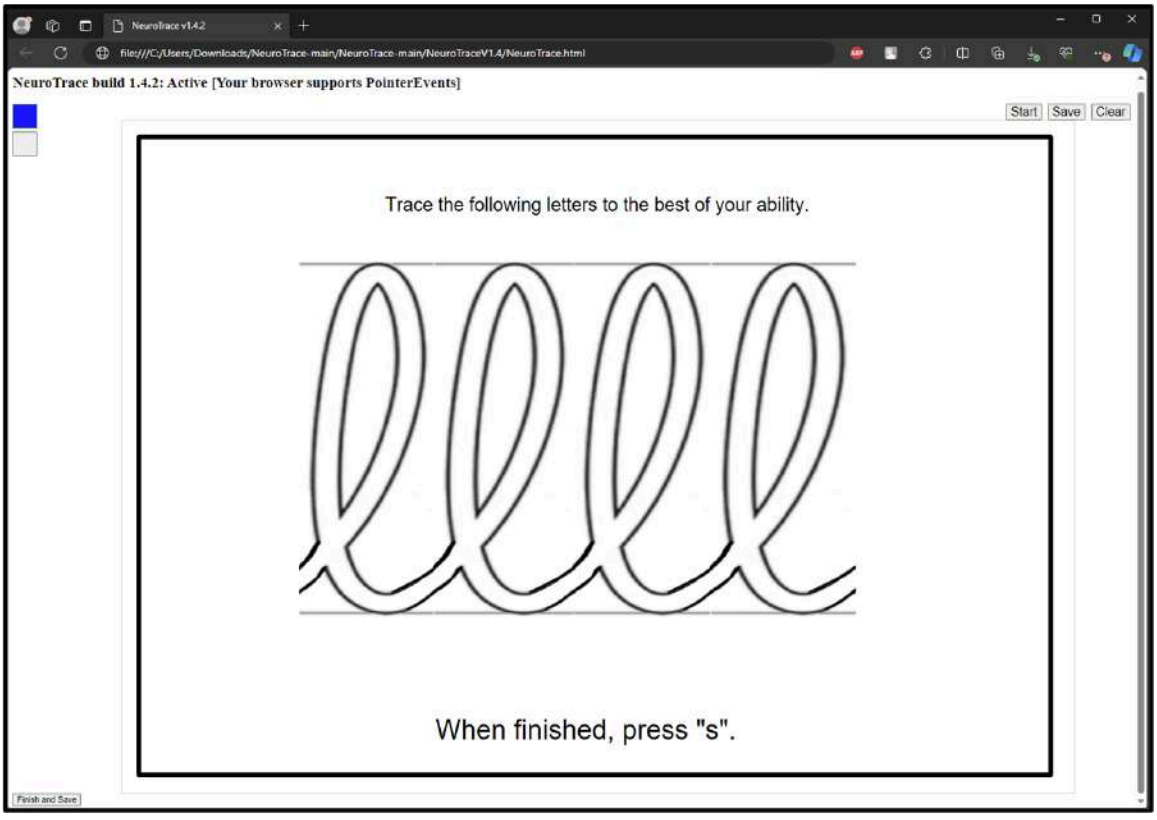


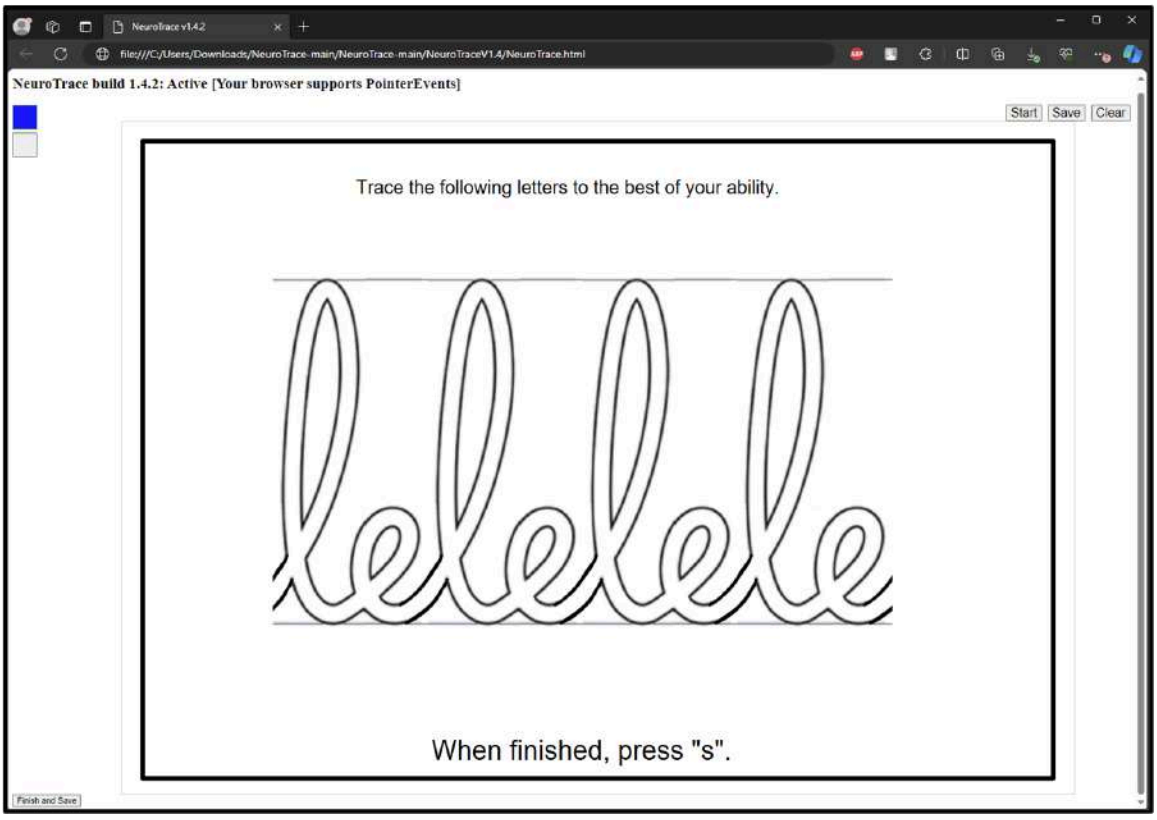


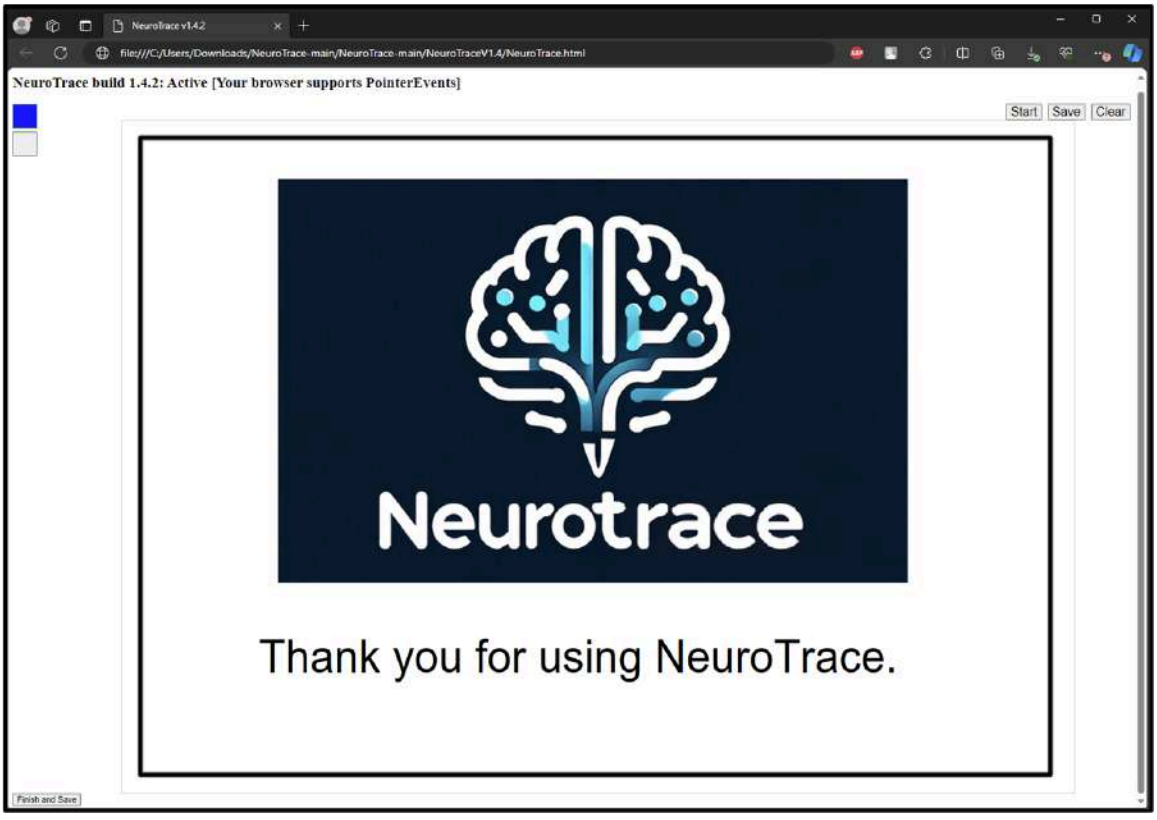

NeuroTrace v1.4.2
file:///C:/Users/Downloads/NeuroTrace-main/NeuroTrace-main/NeuroTraceV1.4/NeuroTrace.html
NeuroTrace build 1.4.2: Active [Your browser supports PointerEvents]
Start
Save
Clear
Connect the following dots vertically to the best of your ability, making 4 vertical lines.
When finished, press "s".
Finish and Save


NeuroTrace v1.4.2
file:///C:/Users/Downloads/NeuroTrace-main/NeuroTrace-main/NeuroTraceV1.4/NeuroTrace.html
NeuroTrace build 1.4.2: Active [Your browser supports PointerEvents]
Start
Save
Clear
Trace the following circle to the best of your ability for 4 times. Stay inside the highlighted areas.
When finished, press "s".
Finish and Save


NeuroTrace v1.4.2
NeuroTrace build 1.4.2: Active [Your browser supports PointerEvents]
Start
Save
Clear
Trace the following circle to the best of your ability for 4 times. Stay inside the highlighted areas.
When finished, press "s".
Finish and Save


NeuroTrace v1.4.2
NeuroTrace build 1.4.2: Active [Your browser supports PointerEvents]
Start
Save
Clear
Trace the following letters to the best of your ability.
When finished, press "s".
Finish and Save


NeuroTrace v1.4.2
file:///C:/Users/Downloads/NeuroTrace-main/NeuroTrace-main/NeuroTraceV1.4/NeuroTrace.html
NeuroTrace build 1.4.2: Active [Your browser supports PointerEvents]
Start
Save
Clear
Trace the following letters to the best of your ability.
When finished, press "s".
Finish and Save


NeuroTrace v1.4.2
file:///C:/Users/Downloads/NeuroTrace-main/NeuroTrace-main/NeuroTraceV1.4/NeuroTrace.html
NeuroTrace build 1.4.2: Active [Your browser supports PointerEvents]
Start
Save
Clear
Neurotrace
Thank you for using NeuroTrace.
Finish and Save

Appendix F.

Email Correspondence to Cadence Chandler Senior Living's Activities Director

Gmail

School Project Request

5 messages

Abigail [redacted] Fri, Mar 1, 2024 at 12:26 PM

To: [redacted]

Cc: [redacted]

Hi [redacted],

I am the activities director at Cadence Chandler. You and another high school student have both expressed interest in using our residents as participants for your projects. I have spoken with a group of residents, and they would like to hear more about the projects and decide to participate once they've asked their questions. Are you free on Monday 3/4 at 1PM to explain your project for them?

Thanks!

Abby

Resident Lifestyle Director

Cadence Chandler

100 W Queen Creek Rd

Chandler AZ 85248

Office | 480.219.9657

Fax | 480.306.6689

Email | [redacted]

website https://cadencechandler.com/

cadence LIVING® Chandler

Life In The Key Of Beautiful …

BEST INDEPENDENT LIVING U.S.News & WORLD REPORT 2023-2024

BEST ASSISTED LIVING U.S.News & WORLD REPORT 2023-2024

[redacted] Fri, Mar 1, 2024 at 12:43 PM

To: Abigail [redacted]

Hi Abigail,

Thanks for reaching out! I can make it on Monday at 1PM (or anytime after 12PM). Let me know if I need to bring anything.

However, you mentioned that the seniors will decide. I just wanted to double check since my project focuses on seniors already in memory care/suffering from a neurodegenerative disease such as Alzheimer's or Parkinsons. However, if the seniors are in memory care, will they not need a Power of Attorney?

I've attached consent forms, the survey that participants will fill out, and my research overview.

Thank you so much!
-[redacted]

Research Overview:

I am currently working on my AzSEF project, "NeuroTrace: Utilizing a Machine-Powered Approach to Detect the Presence of Neurodegenerative Diseases through Vocational/Education History and Handwriting Kinematics Data". **NeuroTrace recently won Best of Fair at HISEF (Hamilton Invitational Science and Engineering Fair) at Chandler Innovation Fair 2024.**

Purpose of the project:

This project will involve data collection at nearby senior homes, memory cares, and hospitals to determine a correlation between subjects' (especially seniors') handwriting kinematics (speed, pressure, etc), educational & vocational histories (how long/industry worked in) and whether they suffer from or exhibit symptoms of neurodegenerative diseases, like Alzheimer's and Parkinson's, through machine learning data analysis. By utilizing machine learning, coded in Python with the help of scikit-tools, AI can efficiently analyze enormous amounts of numerical and even image data to accurately predict the presence or early onset of neurodegenerative diseases in seniors.

The survey and kinematics data of the participant (who has already been diagnosed with a neurodegenerative disease) will be used as a baseline/"positive" outcome to train the coded Python AI classifier.

If the seniors participate, they will be asked to:

1. They (or a POA) needs to answer a questionnaire to i) self-diagnose whether they a) currently suffer from a neurodegenerative disease OR b) think they suffer from symptoms of a neurodegenerative disease AND ii) to recall their former vocational/educational history, including number of years of high school/college (or other secondary education), listing their longest past career, age of retirement, etc.

2. Utilize an electronic tablet and pen to trace-draw several pictures/shapes and trace-write several brief sentences. The kinematics (speed, pressure, etc) of the handwriting will be collected for further processing.

Time required for participation:

15-25 minutes

Potential Risks of Study:

Moderate risk - survey and tracing may trigger unwarranted and unexpected behaviors, such as stress/anxiety, in some patients Supervision during the entirety of seniors' participation may be needed.

Benefits:

The surveyed vocational data, along with the raw handwriting kinematics data, will be processed and normalized into features such as average tremor, speed, acceleration, etc. This processed data (csv file) will be directly analyzed through a Python-based machine-learning Random Forest Classifier to train and test the model (already trained using the online DARWIN dataset) to classify new samples as "healthy" or "patient" Ultimately, this contributes to future machine-powered technology where neurodegenerative diseases can be efficiently detected earlier, increasing quality of life (especially for the growing senior population). How confidentiality will be maintained:

Questions are completely optional. Participants will not be identified by name. All data is anonymous, and the actual handwriting itself is not analyzed - only the kinematics (motion) behind it. Sensitive educational, vocational, and medical data will remain password-encrypted, and all data will be prompted deleted after experiment.

[Quoted text hidden]

3 attachments

Vocational, Health, Educational Research Survey (3).pdf
79K

NeuroTrace Overview.pdf
1812K

Human Consent Form - Patient Baseline.pdf
97K

Abigail [redacted] Fri, Mar 1, 2024 at 12:49 PM

To: [redacted]

Hi [redacted],

It's a bit more tricky getting the memory care residents to participate since, as you know, the POA will need to sign off. Some residents on our AL/IL side have Parkinson's. I think this is the best way to present your project to a large group to start. If you don't get enough participants, we will move forward with trying to reach out to POA's of our memory care residents.

[Quoted text hidden]

[redacted] Fri, Mar 1, 2024 at 12:59 PM

To: Abigail [redacted]

Abigail,

I'd be happy to present my project to the large group!

I understand your proposal, but I'd prefer to test more participants that have already been diagnosed with a neurodegenerative disease since I am trying to use these participants as a data baseline (a positive test case) to train my machine learning model (since I can go out to normal senior centers or even in public and test on healthy senior participants). This is why I felt like

reaching out directly to AL/Memory Care was more effective since it is easier to find pre-diagnosed patients (without having any staff potentially violate HIPAA or other medical confidentiality regulations)

Nevertheless, I'm still excited to come in and present to the big group and we can proceed from there, but I thought it would be best to be clear and let you know my intent as to what type of participants I am primarily looking for (Those already diagnosed with NDs) and how their data will be used (as a "Positive" test case baseline).

Let me know if you have further questions and thank you for your help/understanding,
-[redacted]

[Quoted text hidden]

[redacted] Mon, Mar 4, 2024 at 1:19 PM
Draft To: Abigail [redacted]

[Quoted text hidden]

Vocational, Health, Educational Research Survey.pdf
79K

Appendix G.

Email Correspondence to Parkland Memory Care's Activities Director, who Emailed Families

Gmail

Neurodegenerative Disease Research
8 messages

PK Life Enrichment Coordinator <mcactivities@parklandmemorycare.com> Wed, Feb 14, 2024 at 9:36 AM
To: PK Executive Director <ed@parklandmemorycare.com>, PK Life Enrichment Coordinator <mcactivities@parklandmemorycare.com>
Bcc: [redacted]

Hello Parkland Families,

We have a senior high school student who is doing a research project on Detecting the Presence or Early Onset of Neurodegenerative Diseases through Health, Vocational & Educational History and Handwriting Data.

The project would start with you, the POA, signing the consent form and answering the Research Questionnaire (both attached to this email). After both of those have been completed, the researcher would come to Parkland to complete the next part, which is utilizing an electronic tablet and pen to trace-draw several pictures/shapes and letters.

Please read his proposal and see the attachment to see if this is something that you would like your loved one to participate in.

If you are interested in participating, please reply to this email immediately so that I can get an accurate count of how many residents will be participating. Then, complete the two attached forms and return them to Parkland by the end of February. If you need to pick up a printed copy of either of these forms, they will be at the front desk at Parkland.

Thank you for your help!
Rachel [redacted]

Subject: Neurodegenerative Disease Research

Hello,

I'm [redacted] I am researching neurodegenerative diseases. Overall, I have completed my research plan but still need to finalize the details.

This project will involve data collection at nearby senior homes, memory cares, and hospitals to determine a correlation between subjects' (especially seniors') handwriting, educational & vocational histories (how long/industry worked in) and whether they suffer from or exhibit symptoms of neurodegenerative diseases, like Alzheimer's and Parkinson's through machine learning data analysis. By utilizing machine learning, coded in Python with the help of scikit-tools, AI can efficiently analyze enormous amounts of numerical and even image data to accurately predict the presence or early onset of neurodegenerative diseases in seniors. The actual data collection will involve two parts, and will take about 15-25 minutes of a participant's time:

1. Answering a questionnaire to
   i) self-diagnose whether they a) currently suffer from a neurodegenerative disease OR b) think they suffer from symptoms of a neurodegenerative disease AND
   ii) to recall their former employment/vocational/educational history, including number of years of college/work, avg. hours worked per week, listing their main career, and age of retirement.
2. Utilizing an electronic tablet and pen to trace-draw several pictures/shapes and trace-write several brief sentences.

While there will be sensitive medical, vocational, and educational data being collected, confidentiality will be prioritized. Questions are completely optional. Participants will not be identified by name. All

data is anonymous. Sensitive educational, vocational, and medical data will remain password-encrypted, and all data will be deleted after the experiment.
Ultimately, this project will contribute to future machine-powered technology where neurodegenerative diseases can be efficiently detected earlier, increasing the quality of life, especially for the growing senior population suffering from neurodegenerative diseases.

Rachel
PK Life Enrichment Director

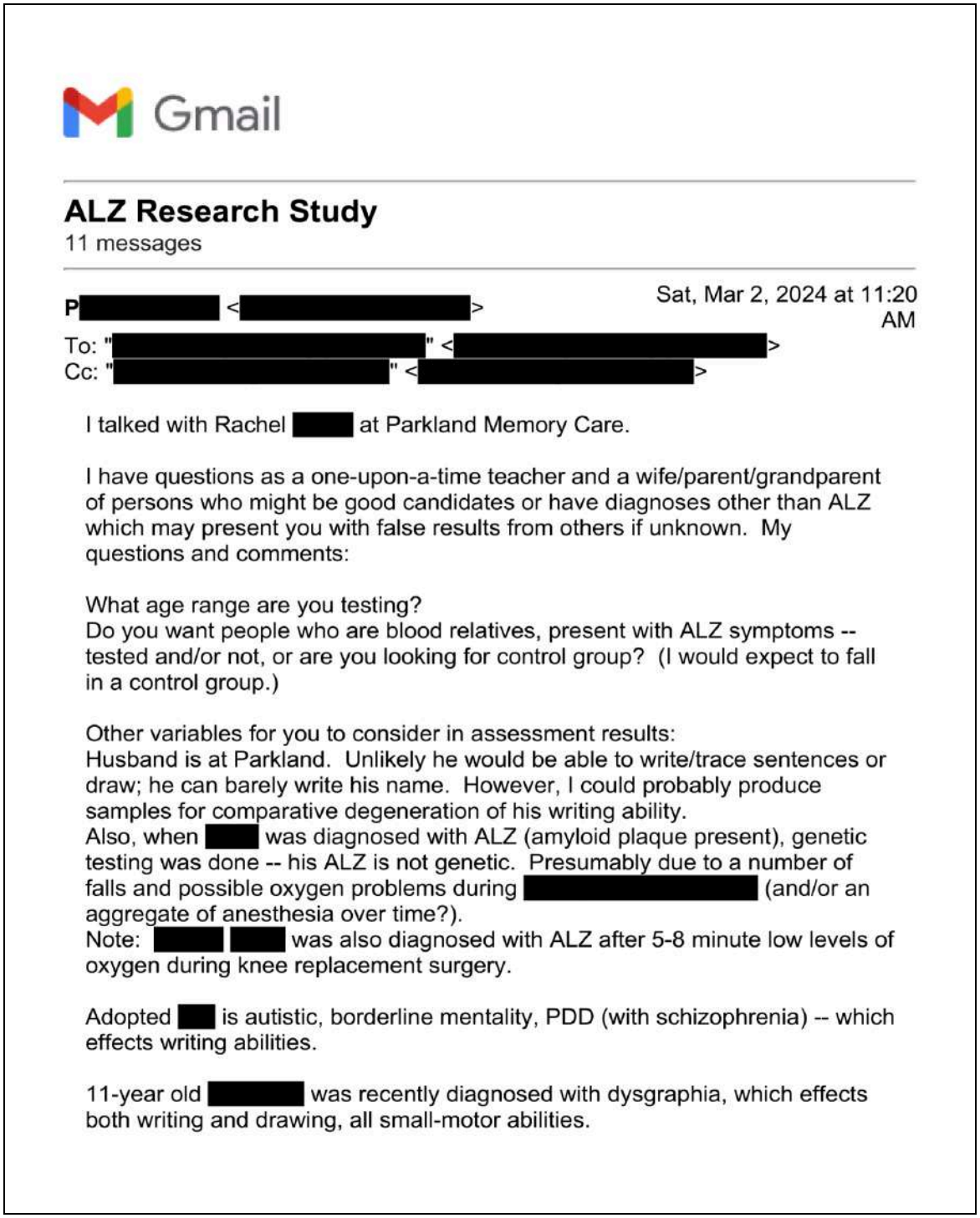
e: mcactivities@parklandmemorycare.com
o: 480-857-4984 | f: 480-571-1380
a: 3500 S. Arizona Ave, Chandler, AZ 85248
w: Parklandmemorycare.com

2 attachments

Human Consent Form.pdf
95K

Appendix H.

Email Correspondence to "P", wife and POA of "D", one of the Residents at Parkland Memory Care. She was very interested in my project, and very willing to get "D" to help out



Gmail

ALZ Research Study
11 messages

P[redacted] <[redacted]> Sat, Mar 2, 2024 at 11:20 AM
To: "[redacted]" <[redacted]>
Cc: "[redacted]" <[redacted]>

I talked with Rachel [redacted] at Parkland Memory Care.

I have questions as a one-upon-a-time teacher and a wife/parent/grandparent of persons who might be good candidates or have diagnoses other than ALZ which may present you with false results from others if unknown. My questions and comments:

What age range are you testing?
Do you want people who are blood relatives, present with ALZ symptoms -- tested and/or not, or are you looking for control group? (I would expect to fall in a control group.)

Other variables for you to consider in assessment results:
Husband is at Parkland. Unlikely he would be able to write/trace sentences or draw; he can barely write his name. However, I could probably produce samples for comparative degeneration of his writing ability.
Also, when [redacted] was diagnosed with ALZ (amyloid plaque present), genetic testing was done -- his ALZ is not genetic. Presumably due to a number of falls and possible oxygen problems during [redacted] (and/or an aggregate of anesthesia over time?).
Note: [redacted] [redacted] was also diagnosed with ALZ after 5-8 minute low levels of oxygen during knee replacement surgery.

Adopted [redacted] is autistic, borderline mentality, PDD (with schizophrenia) -- which effects writing abilities.

11-year old [redacted] was recently diagnosed with dysgraphia, which effects both writing and drawing, all small-motor abilities.

If you would like to talk or test, let me know.

[redacted]

[redacted] <[redacted]> Sat, Mar 2, 2024 at 2:59 PM

To: [redacted] <[redacted]>

Hi [redacted],

I hope you're doing well! I appreciate your insightful questions and comments. Allow me answer your questions:

What age range are you testing?

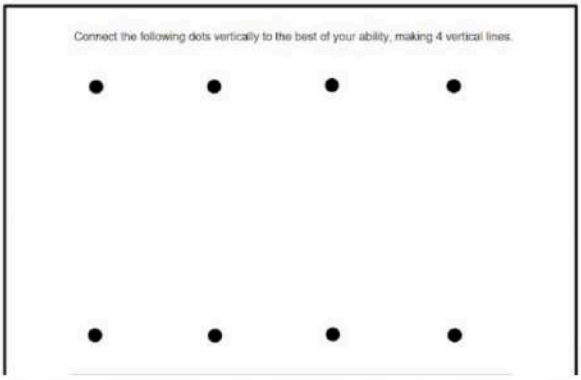
- As of now, I am primarily testing on seniors for my science project. I understand that it is possible for people of all ages to become diagnosed with NDs, but it is more prevalent among the older population, so I am focusing on seniors for now.
- When I contacted Parkland (and other nearby memory cares), my goal was to find participants who have already tested positive for Alzheimer's/Parkinson's or other NDs, and to use the handwriting kinematics data that these participants provide as a "positive" baseline to train, test, and verify my machine-learning model.
- Likewise, I would expect healthy individuals (like yourself as you said) to be "negative" baseline controls to also train, test, and verify my machine-learning model. (Of course, it would be awesome if you could be a participant too, under the control/negative category!)

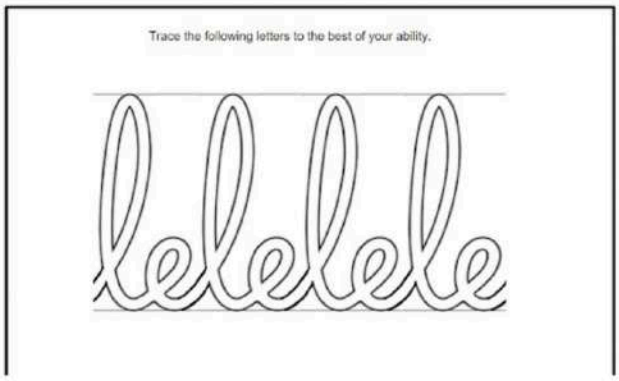
Unlikely he would be able to write/trace sentences or draw; he can barely write his name. However, I could probably produce samples for comparative degeneration of his writing ability.

- My project involves participants writing/tracing on a digital electronic tablet (since I can capture the kinematics data for further analysis), so I

would need him to be the one drawing.

- However, I have revised my procedures to make the drawing tasks easier and straightforward. Since you know his physical/mental state, please let me know if these tasks are achievable for him (I can provide assistance during the drawing process). I've provided some examples below of what the software/user interface looks like:





Finally, I understand that my experiments won't be the "end-all-be-all" of future neurodegenerative disease research and diagnosis since it is too general and

there are too many factors at play for an 100% accurate diagnosis/result. However, there's always room for growth, and I'd like to think that I am improving the quality of life for the future of our generations. Therefore, I encourage [redacted] (and you too!) to participate as he could be part of the development of new and pivotal discoveries.

I appreciate the support for NeuroTrace! I'd love to talk to you more about my project if you're willing. Please reply with any comments or questions.

Best,
-[redacted]
[Quoted text hidden]

[redacted] <[redacted]> Tue, Mar 5, 2024 at 11:17 AM
To: [redacted] <[redacted]>

Hi [redacted],

Just floating this email to the top of your inbox. Please read the email attached and let me know if this is something you would give Dave consent to participate in.

Thanks!
Alex
[Quoted text hidden]

[redacted] <[redacted]> Thu, Mar 7, 2024 at 8:57 AM
To: [redacted] <[redacted]>

[redacted], I found your Saturday e-mail in spam just this am.
Will talk to my husband and try a little of the samples on him today to see how he responds.
Will get back to you later today about whether I think he can effectively participate or not.

You're the researcher I would be, if in another lifetime!

[Quoted text hidden]

To: [redacted] <[redacted]>

[redacted] was apprehensive; not sure a good idea to try and test him.

I am willing and would like to help, if you need another control. Am with [redacted] Monday through Friday until 2:00 pm; could stay a little later, if I know when. Or, I will be at the Dementia Decline seminar in the [redacted] from 1-2 pm on March 19th; could stay there a little later.

[Quoted text hidden]

[redacted] <[redacted]> Fri, Mar 8, 2024 at 11:22 AM
To: [redacted] <[redacted]>

[redacted],

Thanks for testing on [redacted]. I was hoping [redacted] could participate, but I will continue to look for potential participants. I'm out of town from 3/11 to 3/20, so unfortunately I will not be at the Dementia seminar. It'd be great if I could still meet with you and [redacted] sometime (let me know what times work for you).

By the way, do you know of any potential other facilities (like hospitals or memory cares) where I could have more positive responses and participants? I'm running on crunch time for finishing my project before AZsef in April, so any help at all would be awesome.

Thanks for your help!
[redacted]
[Quoted text hidden]

[redacted] <[redacted]> Sat, Mar 9, 2024 at 8:17 AM
To: [redacted] <[redacted]>

Three possibles come to mind that might be willing to be involved in study/research:

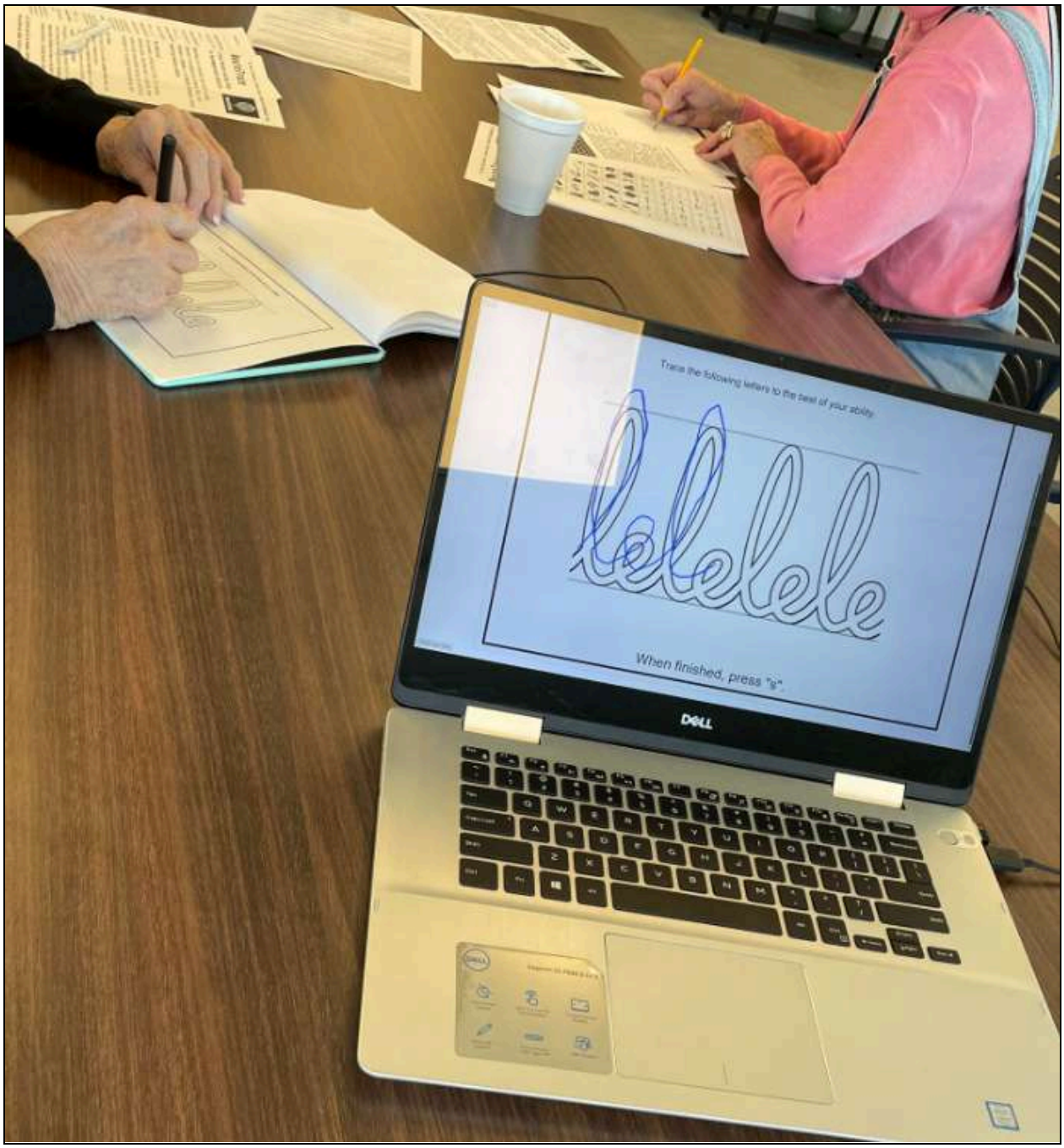
- Neurology, e.g. Gilbert Neurology participated in Brain Health Registry studies that [redacted] was in at the beginning of his diagnoses.
- VA, perhaps Conrad (Southwest is on Val Vista) or another location. Dave went to Val Vista for routine health care.

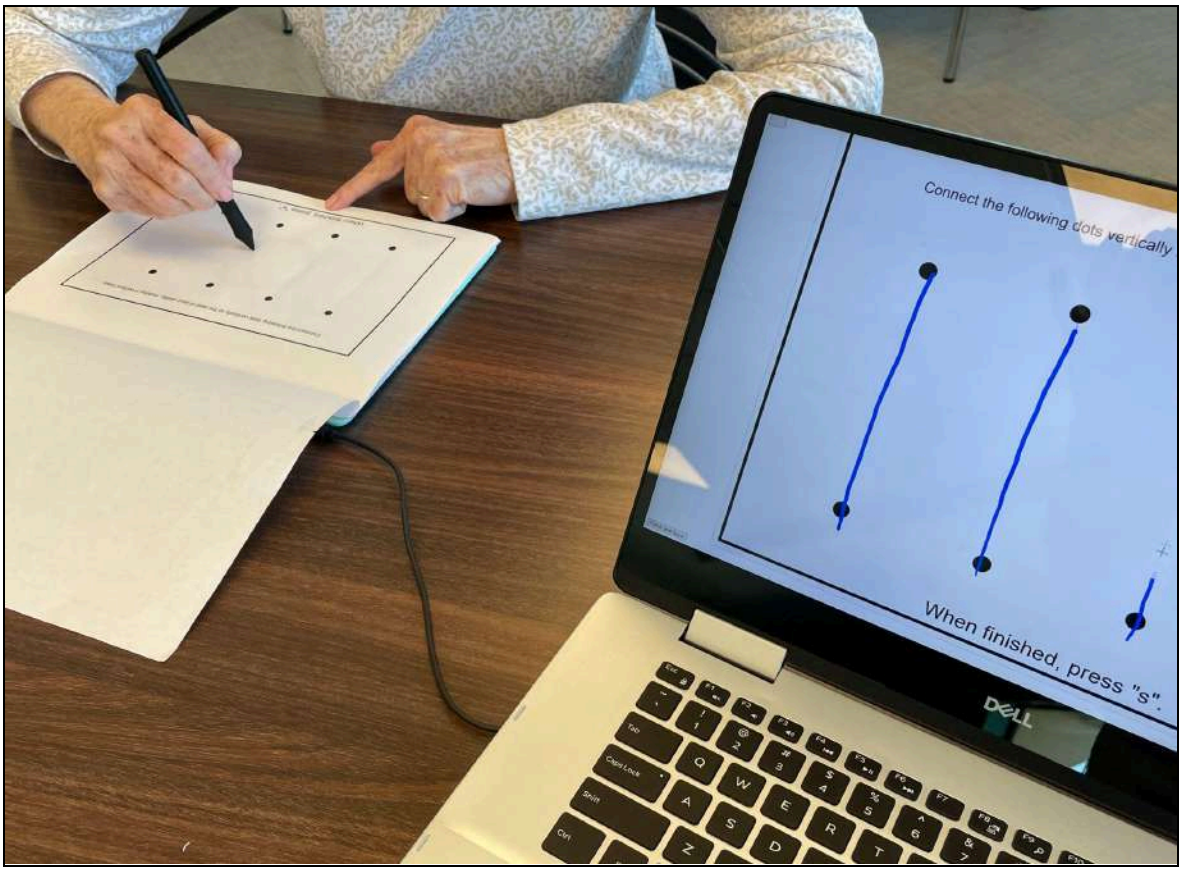
Appendix I.

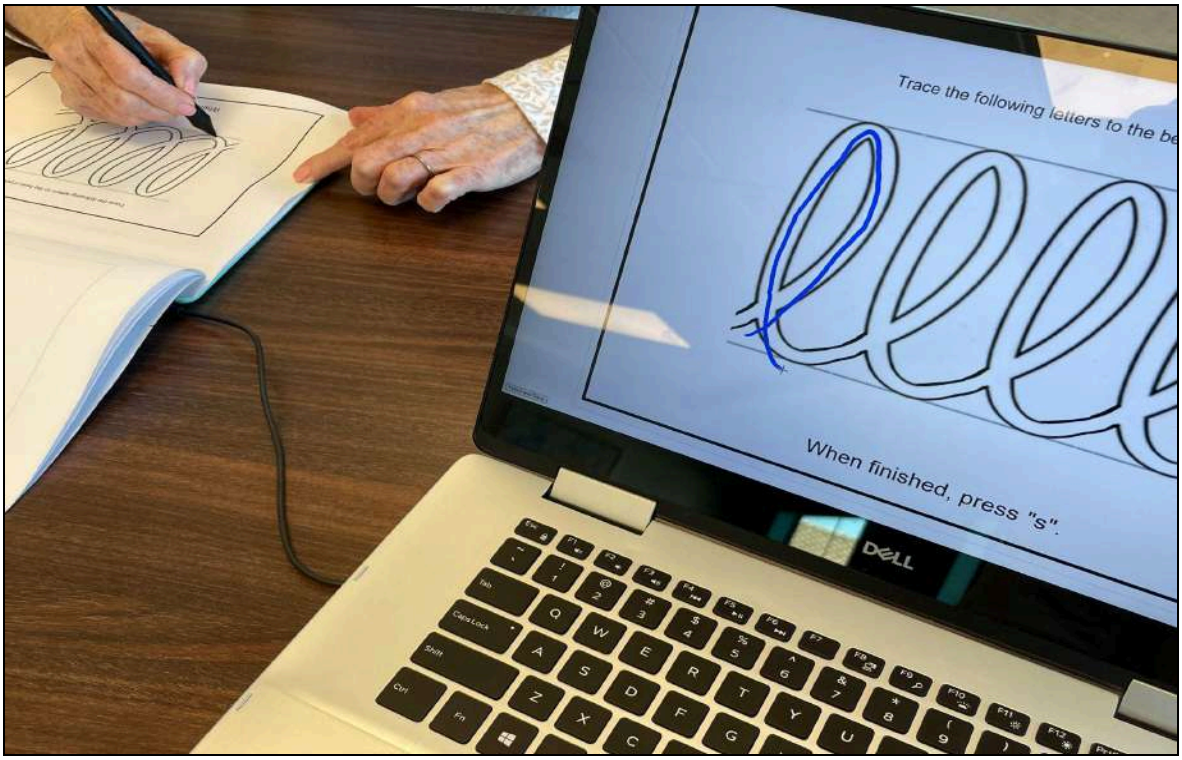
Participant Testing Process - Pictures

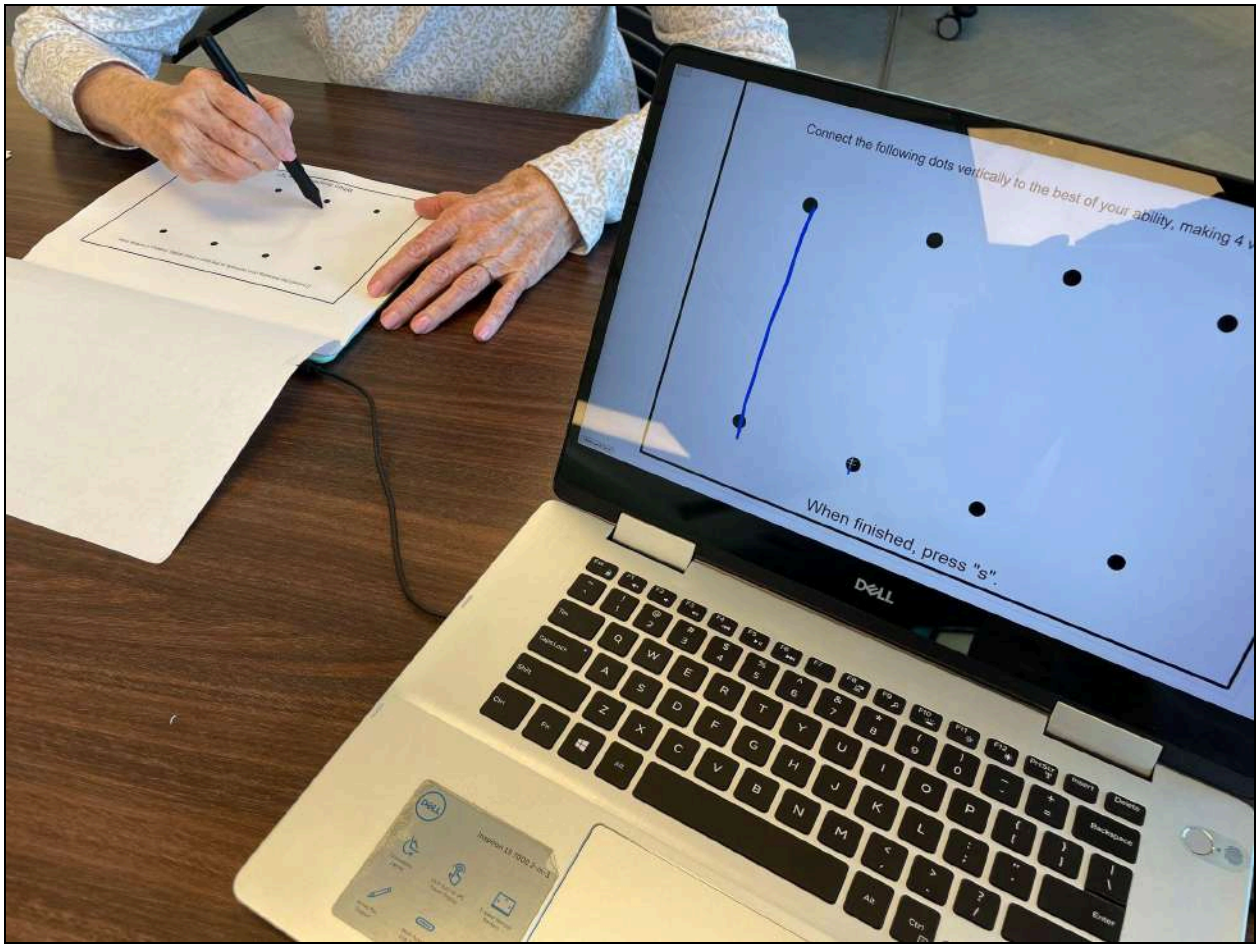
**For video, click on link:*

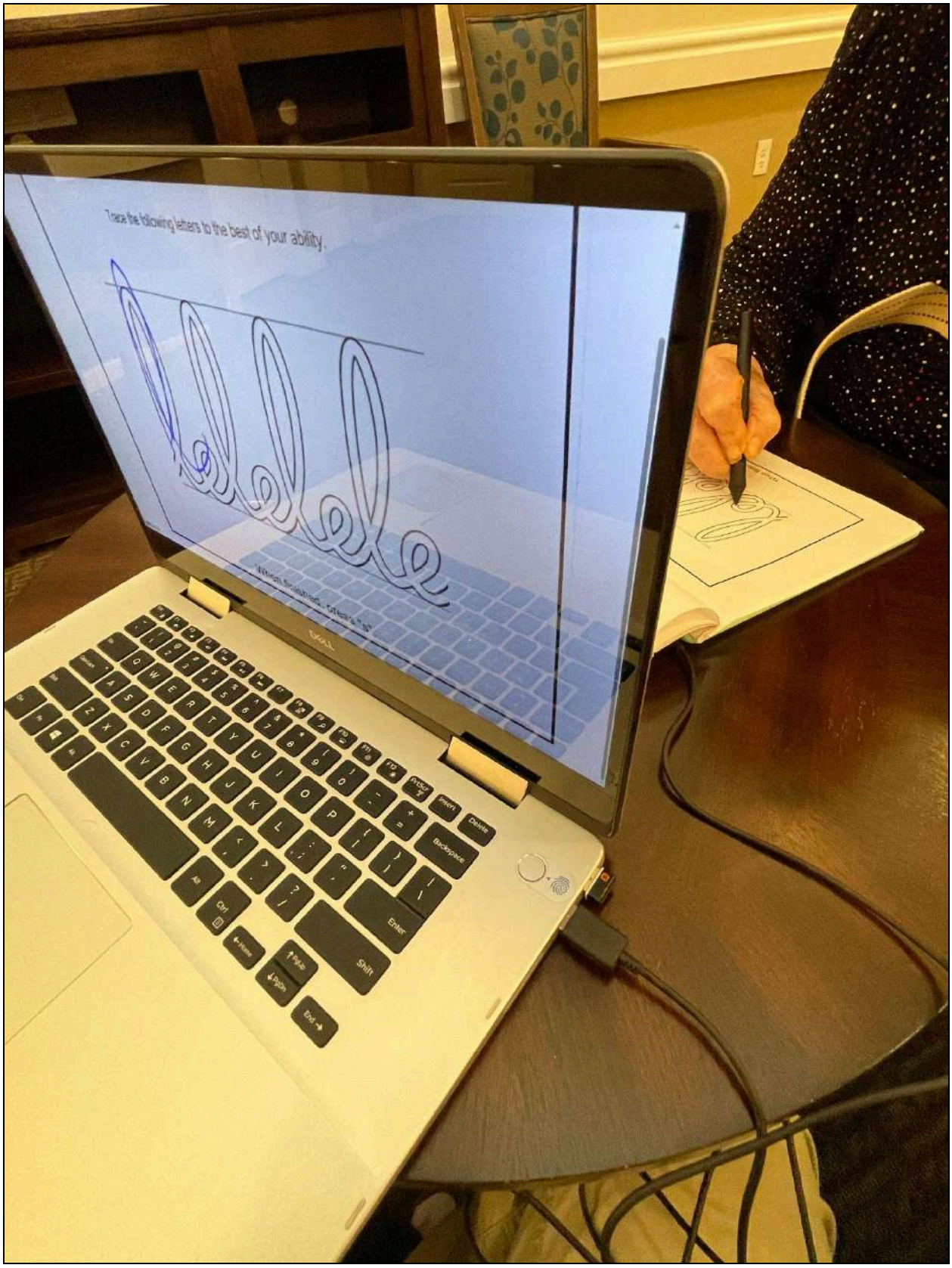
https://drive.google.com/file/d/13miprDh7Zmr13lyIDRuevlw5oBGinXh8/view?usp=sharing




Connect the following dots vertically
When finished, press "s".


Trace the following letters to the be
When finished, press "s".


Connect the following dots vertically to the best of your ability, making 4
When finished, press "s".
DELL


Trace the following letters to the best of your ability.

Appendix J.

NeuroTrace Project Slides (submitted for AzSEF*, the Arizona Science and Engineering Fair)*

Neurotrace

NeuroTrace

A Novel, Machine-Powered Program to Detect Early Onset of Neurodegenerative Diseases through Vocational/Education Data Handwriting Kinematics Analysis

SR-TMED-003

All images self-created unless specified

Introduction

Fig.1. news-medical.net [1]

Neurodegenerative diseases (NDs), such as Parkinson's, Alzheimer's, dementia, and multiple sclerosis, affect **more than 62 million people**, mostly seniors, **worldwide** [1]. These diseases occur due to **progressive degeneration and breakdown** of **nerve cells** in the brain and the nervous system, and can be caused by exposure to certain hazards or stresses in a workplace.

These diseases share a common pattern of **permanently damaging the nervous system**, which can lead to:

- **Cognitive decline**
- **Motor degeneration**
- **functional impairment**

One specific motor degeneration is **hand motion and control**, which impede daily activities, like **writing**. Hand movements often become **segmented and jagged**.

Therefore, it may be possible to detect early onset of neurodegenerative diseases through a combination of past career and educational history as well as **handwriting analysis.**

There is a need for an effective program that accurately easily detects neurodegenerative diseases, especially Alzheimer's and Parkinson's, through handwriting kinematics tracking and analysis for at-risk individuals, saving lives and suffering.

Previous Studies

- Mergl et. al [2] utilized a **digitizing graphic tablet** to analyze handwriting and drawing kinematics, but the researchers focused on **correlating depression and mental disorders with handwriting,** not neurodegenerative diseases.
- Impedovo et. al [3] used an **earlier (PaHaW) dataset**, which contains 70 samples, as the baseline for their handwriting kinematic analysis. However, the researchers **only sought to use classification to support an earlier diagnosis**, which may have skewed the data as they **only utilized 29 patient samples**.
- Isenkul et. al [4] also experimented using the WACOM digitizing tablet, testing for Parkinson's and **contributing a dataset on Kaggle**. However, their database and experiment **only used the Archimedes Spiral test (and no other tasks)**, which can be limiting for a comprehensive understanding of neurodegenerative diseases.
- Nachum et. al [5] designed a similar project involving computer vision and its applications to diagnose Alzheimer's and Parkinson's. However, they utilized camera tracking instead and were unable to obtain participant samples.

In the exploration of handwriting kinematics and its potential applications in health diagnostics, several notable studies have paved the way. However, the existing research landscape exhibits certain gaps and limitations.

By utilizing **a more complex dataset**, incorporating **a wide array of tasks,** and developing **a detection program** utilizing the latest machine learning technology, **a more thorough correlation between handwriting kinematics and neurodegenerative diseases can be discovered.**

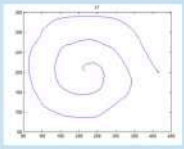


Figure 2.1: control subject's spiral drawing

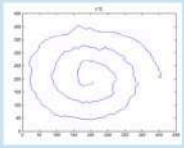


Figure 2.2: Parkinson's subject's spiral drawing

Source: Isenkul et. al [4]

Framework

Figure 3.1: Random Forest Classifier visualization Source: JavaTPoint [6]

- **Handwriting Kinematics**
  - Properties of handwriting that can be recorded and analyzed by software, including position (x,y), mean velocity, acceleration, and jerk, average tremor, number of pendowns, and total time.
- **Random Forest Classifier**
  - Ensemble-learning classifier that operates by building decision trees of features during training.
  - During testing, each tree independently predicts the class of input data to provide a final result.
- **Metrics to assess classifier accuracy:**
  - **Accuracy** represents the overall correctness of the classifier's predictions.
  - **Precision (positive predictive value)** = True Positives / (True Positives + False Positives)
    - Answers: "Of all the instances predicted as positive, how many are actually positive?"
  - **Recall (sensitivity)** = True Positives / (True Positives + False Negatives)
    - Answers: "Of all the actual positive instances, how many did the model correctly predict?"
    - In this project, recall is more important as it is more costly and dangerous to falsely label a potential patient as "healthy" than to falsely label a healthy subject as "patient"
  - **F1 Score** = 2 * (Precision * Recall) / (Precision + Recall)
    - The mean of precision & recall, providing a balanced metric for rates of both false positives and negatives.
  - **Area Under the Precision-Recall Curve (AUC-PR)** provides insights into the classifier's performance across different levels of precision and recall.
  - **Cross-Validation score** is a metric used to evaluate the generalizability of a predictive model towards new datasets, giving a better indication of how the model will perform on unseen data.

All images self-created unless specified

Raw Data Preprocessing

NeuroTrace is based on the **DARWIN dataset** [7], which is **the largest, most extensive handwriting kinematics database for NDs research.** In DARWIN, **174 Italian demographics-matched subjects** (89 patients (P), 85 control (H)) participated in **25 different handwriting tasks**.

- NeuroTrace seeks to utilize reproducible tests, and some DARWIN tasks (ones without language barriers) are well suited:
  - **Tasks 2, 3, 4, 5, 6, 8, 9 selected** (Figure 4.1)
- Derived from raw handwriting data, each task produces 18 features:
  - airtime, displacement index, general tremor in air/paper, maximum x & y extension, mean velocity/acceleration/jerk, number of pendowns, pressure mean/variation, and time taken
  - *pressure_varN, total_timeN (where N is the # of task, from 1-25)*
  - However, not all features are equally useful.

#	Description
1	Signature drawing
2	Join two points with a horizontal line, continuously for four times
3	Join two points with a vertical line, continuously for four times
4	Retrace a circle (6 cm of diameter) continuously for four times
5	Retrace a circle (3 cm of diameter) continuously for four times
6	Copy the letters 'l', 'm' and 'p'
7	Copy the letters on the adjacent rows
8	Write cursively a sequence of four lowercase letter 'l', in a single smooth movement
9	Write cursively a sequence of four lowercase cursive bigram 'le', in a single smooth movement
10	Copy the word "foglio"
11	Copy the word "foglio" above a line
12	Copy the word "mamma"
13	Copy the word "mamma" above a line
14	Memorize the words "telefono", "cane", and "negozio" and rewrite them
15	Copy in reverse the word "bottiglia"
16	Copy in reverse the word "casa"
17	Copy six words (regular, non regular, non words) in the appropriate boxes
18	Write the name of the object shown in a picture (a chair)
19	Copy the fields of a postal order
20	Write a simple sentence under dictation
21	Retrace a complex form
22	Copy a telephone number
23	Write a telephone number under dictation
24	Draw a clock, with all hours and put hands at 11:05 (Clock Drawing Test)
25	Copy a paragraph

Figure 4.1: a list of the 25 DARWIN handwriting tasks collected into the database.
Source: Cilia et. al [7]

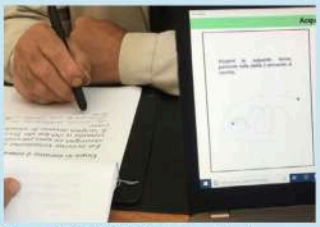
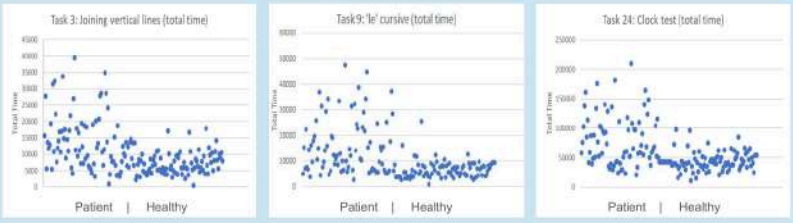


Figure 4.2: DARWIN data collection process
Source: Cilia et. al [7]

Raw Data Processing

- This project seeks also to **simplify DARWIN**'s dataset, **removing unnecessary tasks** and noisy features of little importance
- Utilizing Python Scikit, **feature importance** can be generalized (Figure 5.1). The features that were more significant were:
  - Air & total time, pressure mean/variance, avg speed, acceleration, jerk
- Visually, it is easy to see **significant deviation within the patient data** (vs. control), especially with the 'total_time' feature (Figures 5.2-5.4)



Figures 5.2, 5.3, 5.4: scatterplots of subjects' total time taken to perform DARWIN tasks 3, 9, and 24 respectively.

All images self-created unless specified

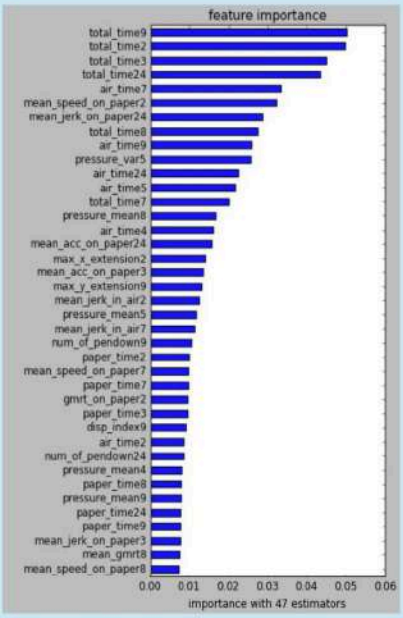
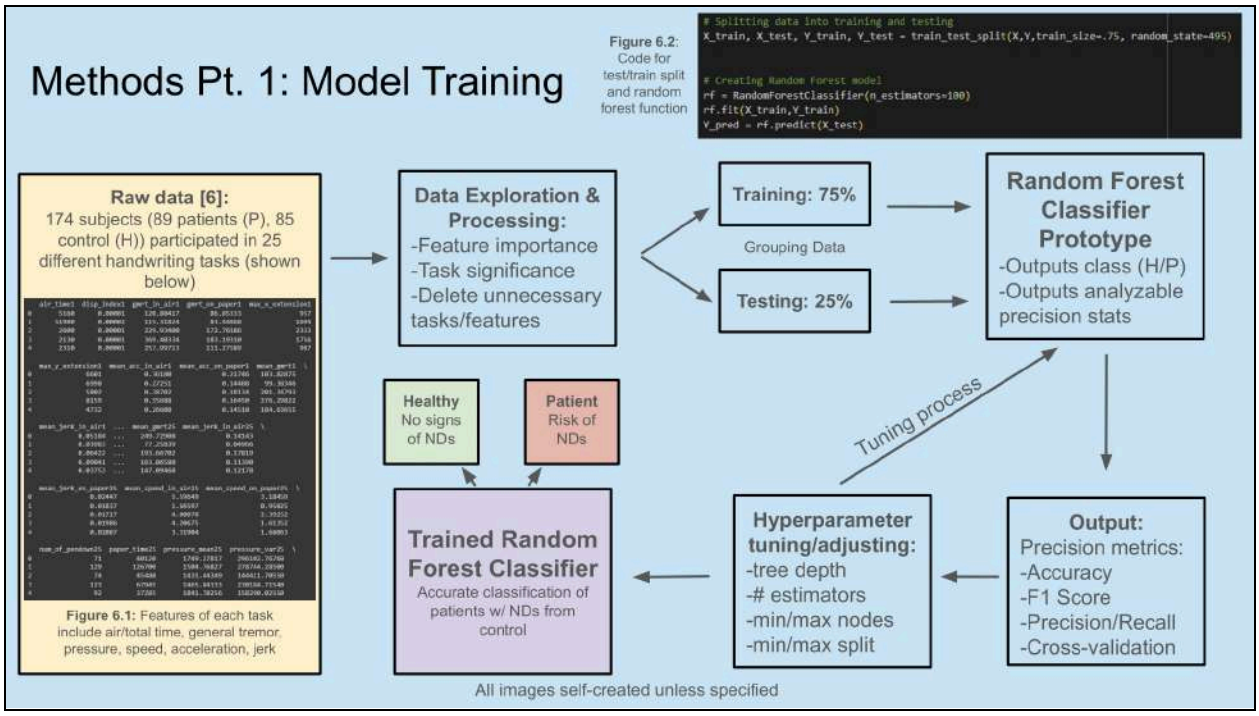
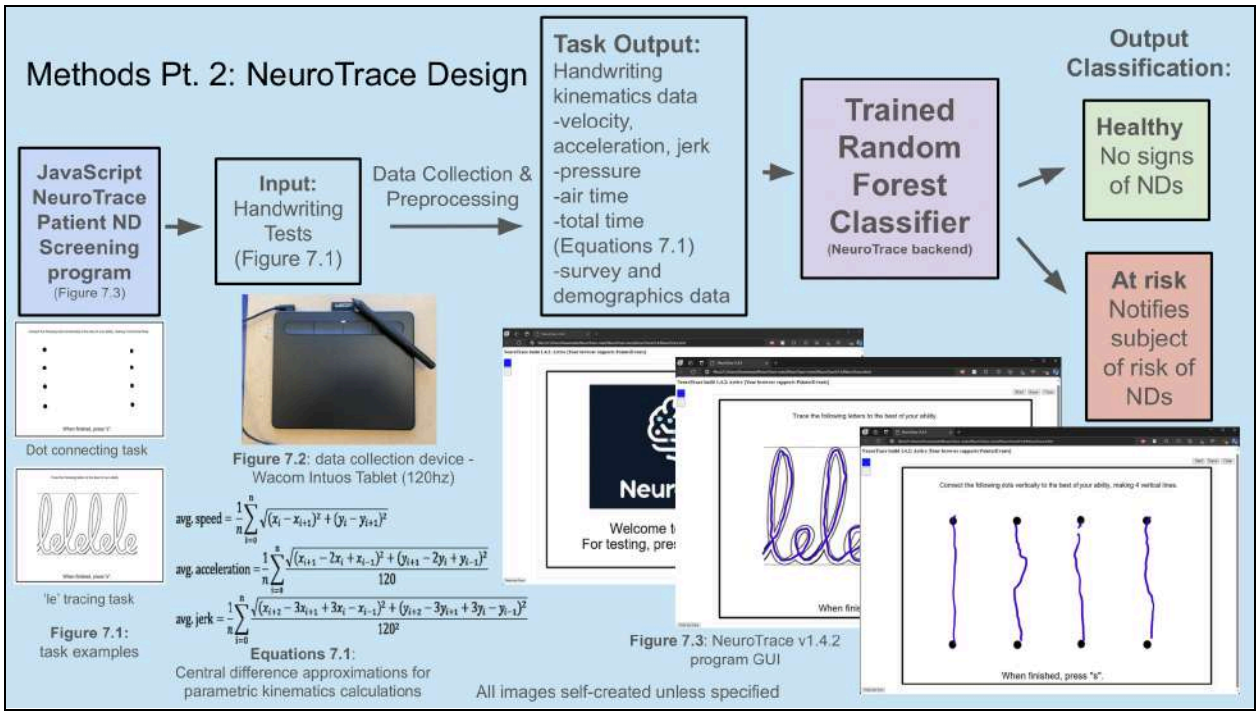


Figure 5.1: a snippet of the DARWIN feature importance graph generated with *scikit-importance* and *pandas*. Some of the graph is omitted for visibility


Methods Pt. 1: Model Training
Figure 6.2: Code for test/train split and random forest function
Splitting data into training and testing
X_train, X_test, Y_train, Y_test = train_test_split(X,Y,train_size=.75, random_state=495)
# Creating Random Forest model
rf = RandomForestClassifier(n_estimators=100)
rf.fit(X_train,Y_train)
Y_pred = rf.predict(X_test)
Raw data [6]:
174 subjects (89 patients (P), 85 control (H)) participated in 25 different handwriting tasks (shown below)
Figure 6.1: Features of each task include air/total time, general tremor, pressure, speed, acceleration, jerk
Data Exploration & Processing:
-Feature importance
-Task significance
-Delete unnecessary tasks/features
Training: 75%
Grouping Data
Testing: 25%
Random Forest Classifier Prototype
-Outputs class (H/P)
-Outputs analyzable precision stats
Healthy
No signs of NDs
Patient
Risk of NDs
Tuning process
Trained Random Forest Classifier
Accurate classification of patients w/ NDs from control
Hyperparameter tuning/adjusting:
-tree depth
-# estimators
-min/max nodes
-min/max split
Output:
Precision metrics:
-Accuracy
-F1 Score
-Precision/Recall
-Cross-validation
All images self-created unless specified


Methods Pt. 2: NeuroTrace Design
JavaScript NeuroTrace Patient ND Screening program
(Figure 7.3)
Input:
Handwriting Tests
(Figure 7.1)
Data Collection & Preprocessing
Task Output:
Handwriting kinematics data
-velocity, acceleration, jerk
-pressure
-air time
-total time
(Equations 7.1)
-survey and demographics data
Trained Random Forest Classifier
(NeuroTrace backend)
Output Classification:
Healthy
No signs of NDs
At risk
Notifies subject of risk of NDs
Dot connecting task
'le' tracing task
Figure 7.1: task examples
Figure 7.2: data collection device - Wacom Intuos Tablet (120hz)
$\text{avg. speed} = \frac{1}{n}\sum_{i=0}^{n}\sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2}$
$\text{avg. acceleration} = \frac{1}{n}\sum_{i=2}^{n}\frac{\sqrt{(x_{i+1} - 2x_i + x_{i-1})^2 + (y_{i+1} - 2y_i + y_{i-1})^2}}{120}$
$\text{avg. jerk} = \frac{1}{n}\sum_{i=0}^{n}\frac{\sqrt{(x_{i+2} - 3x_{i+1} + 3x_i - x_{i-1})^2 + (y_{i+2} - 3y_{i+1} + 3y_i - y_{i-1})^2}}{120^2}$
Equations 7.1: Central difference approximations for parametric kinematics calculations
Figure 7.3: NeuroTrace v1.4.2 program GUI
When finished, press "s".
All images self-created unless specified

Findings: Hyperparameter tuning

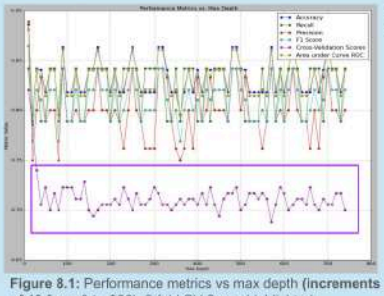


Figure 8.1: Performance metrics vs max depth **(increments of 10 from 0 to 800).** 5-fold CV Score highlighted.

Figure 8.2: Performance metrics vs max depth **(increments of 1 from 0 to 75).** 5-fold CV Score highlighted.

Several hyperparameters of the Random Forest were tuned to train the machine, including:

- Max depth: max depth of each decision tree in the forest, which can increase accuracy but lead to overfitting (where it cannot generalize to new data).
  - A max depth > 100 led to CV score decline, representing overfitting **(Fig. 8.1)**
  - **A max depth of 40 was selected.**
- # estimators: number of trees in the forest, which will improve accuracy but utilize more computation power
  - A max depth > 75 saw diminishing returns in most metrics, also taking longer to compute **(Fig. 8.4). 70 estimators chosen.**
- Max features: the max # of features considered for splitting nodes, and creates randomness for model
  - A max features > 10 saw CV score decline also, representing overfitting (Fig 8.5). **9 max features chosen.**

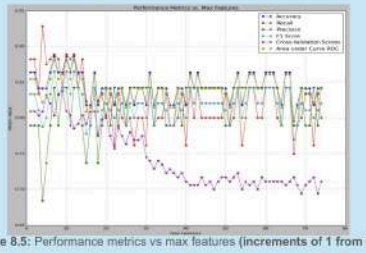


Figure 8.5: Performance metrics vs max features **(increments of 1 from 0 to 75).**

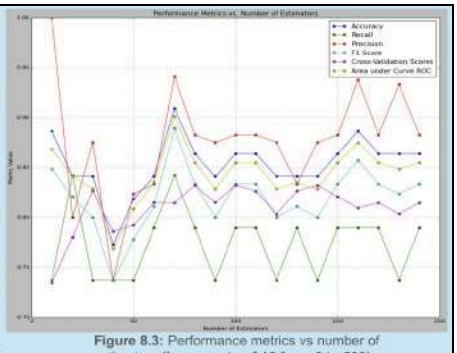


Figure 8.3: Performance metrics vs number of estimators **(increments of 10 from 0 to 200).**

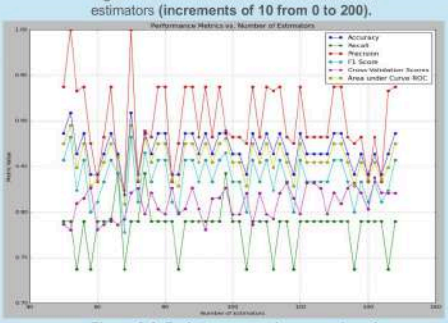


Figure 8.4: Performance metrics vs number of estimators **(increments of 3 from 50 to 150).**

Findings: Data Analysis

Metrics to determine NeuroTrace precision:

- The output (Fig. 9.1) indicates that the **model has high accuracy (90%)** and **precision (94%),** with a **fairly good recall (84%).**
- The F1 score, which balances precision and recall, is also high (88%), indicating that the **model is reliable in identifying true positives and true negatives** for the condition it is predicting.
  - Intuitively, the model is very good at identifying the cases it is designed to detect (precision), and a large proportion of actual cases (recall).
- The 5-fold Cross-validation score (80%) and ROC Area Under Curve score (90%) indicate that **NeuroTrace generalizes well across unknown datasets, such as the data generated by real participants** (Figures 9.3 & 9.4).

Statistical Significance:

- The high values of these metrics, especially the F1 score and ROC-AUC, indicate that the results **are statistically significant** as they yield significantly better results than random chance (~50%).
- In Figure 9.1, the **error bars are small** and do not overlap with 0.5, also **indicating statistical significance.**

All images self-created unless specified

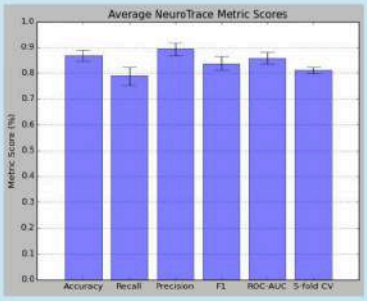


Figure 9.1: Average performance metrics for trained NeuroTrace model

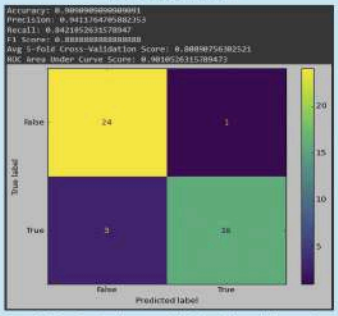


Figure 9.2: Confusion matrix for trained NeuroTrace model

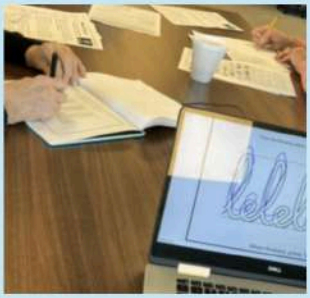


Figure 9.3: Participant completing the 'le' task (task 9) utilizing the WACOM tablet

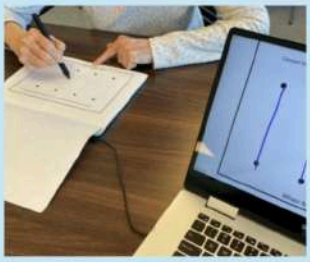


Figure 9.4: Participant completing the 'connecting vertical dots' task (task 3) utilizing the WACOM tablet

Conclusion & Applications

Based on the data analysis, with more data training, **NeuroTrace will:**

- **Ensure early neurodegenerative disease detection with an accuracy > 80%:**
  - The model will contribute to early detection efforts, enabling timely interventions, improving patient outcomes, and enhancing quality-of-life for longer.
- **Contribute to the relevance of handwriting kinematics in the field of NDs:**
  - The model will emphasize the relevance of handwriting kinematics as a potential biomarker for neurodegenerative diseases, as it can leverage subtle motor patterns for accurate predictions.

NeuroTrace has several implications and applications, as a:

- Quick, accurate, and non-invasive **screening and checkup tool**
- A **cheaper, more efficient application** for low-resource/income groups
- **Remote** NDs **diagnosis & monitoring**
- Effective **longitudinal study tool,** with quick & simple tasks
- **Contribute to further research** about handwriting kinematics and NDs

All images self-created unless specified

Limitations

- For the most accurate classifier, **more data and samples are still needed**.
  - Databases and samples are limited to certain populations. For example, the utilized DARWIN datasets in this classifier contains only Italian seniors, and **may not generalize well to all populations** across the world
- At this time, the NeuroTrace software is still in development. There have been roadblocks, especially with C++ coding.
- The **dimensionality of numerical data is limiting**; there is only so much that can be analyzed and interpreted from just numerical data.

Future Research

- **Multimodal software integration**, such as **image analysis** or **real-time monitoring** processes, can provide deeper insights and more accurate results.
- **Utilize other datasets** in conjunction with DARWIN to create a more robust model
- Analysis of seniors' **individual lifestyles** to reveal more correlations, such as:
  - **Career** - may have an impact in onset of NDs, in **blue vs. white vs. pink collar jobs**
  - **Education** - may also have an impact in onset of NDs due to **life/health choices**.

References

1. News-Medical. "Alzheimer's Disease | Definition, Causes, Diagnosis & Treatment." News-Medical.net, 25 Oct. 2019, www.news-medical.net/health/Alzheimers-Disease-Definition-Causes-Diagnosis-Treatment.aspx.
2. Mergl, R., et al. "Kinematical Analysis of Handwriting Movements in Depressed Patients." Acta Psychiatrica Scandinavica, vol. 109, no. 5, May 2004, pp. 383–391, https://doi.org/10.1046/j.1600-0447.2003.00262.x.
3. Impedovo, Donato, et al. "Dynamic Handwriting Analysis for Supporting Earlier Parkinson's Disease Diagnosis." Information, vol. 9, no. 10, 3 Oct. 2018, p. 247, https://doi.org/10.3390/info9100247.
4. Isenkul, Muhammed, et al. "Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinson's Disease." The 2nd International Conference on E-Health and Telemedicine , vol. pp: 171-175, May 2014, https://doi.org/10.13140/RG.2.1.1898.6005.
5. Nachum, Ron, et al. "A Novel Computer Vision Approach to Kinematic Analysis of Handwriting with Implications for Assessing Neurodegenerative Diseases." IEEE Xplore, 1 Nov. 2021, ieeexplore.ieee.org/document/9630492.
6. JavaTpoint. "Machine Learning Random Forest Algorithm - Javatpoint." Www.javatpoint.com, www.javatpoint.com/machine-learning-random-forest-algorithm.
7. Cilia, Nicole, et al. "DARWIN Dataset." UCI Machine Learning Repository, 5 Nov. 2018, https://archive.ics.uci.edu/dataset/732/darwin.

Appendix K.

Final in-person Project Board (36x48 in.), presented at AzSEF