TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Information Technology
Department of Computer Engineering

Thomas-Bairam Toodo 163403IAPM

ASSESSMENT OF PARAMETERS FROM THE HANDWRITTEN SENTENCE TEST USED TO DIAGNOSE PARKINSON'S DISEASE

Master Thesis

Supervisors: Sven Nõmm PhD
Kadri Medijainen MSc
Aaro Toomela PhD

Tallinn 2018

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond
Arvutitehnika instituut

Thomas-Bairam Toodo 163403IAPM

PARKINSONI TÕVE DIAGNOOSIMISEL LÄBITUD KÄEKIRJALIKU LAUSE TESTI PARAMEETRITE ANALÜÜS

Magistritöö

Juhendajad: Sven Nõmm PhD
Kadri Medijainen MSc
Aaro Toomela PhD

Tallinn 2018

Author's declaration of originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication. All works and major viewpoints of the other authors, data from other sources of literature and elsewhere used for writing this paper have been referenced.

Olen koostanud antud töö iseseisvalt. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud. Käsolevat tööd ei ole varem esitatud kaitsmisele kusagil mujal.

Author: Thomas-Bairam Toodo

May 14, 2018

Abstract

The main focus of the present thesis is the analysis and evaluation of the handwritten sentence test used in diagnosing the Parkinson's disease. Main goal is to determine the set of features that can be used to distinguish healthy controls from the patients with Parkinson's disease.

In scope of the thesis, word recognition system was created to determine the validity of the written sentence. Motion mass, kinematic and geometrical parameters of the sentence were extracted and analysed.

Classification was conducted on the basis of analysed features. In the process of classification, three different classifiers were trained and tested with the goal of producing a classifier which could give feedback about the test to the assessing physician and estimate the health condition of the patient.

The thesis is in English and contains 37 pages of text, 7 chapters, 9 figures, 3 tables.

Annotatsioon

Käesolevas magistritöös analüüsitakse parameetreid käekirjalisest lause testist, mida kasutatakse Parkinsoni tõve diagnoosimisel. Põhiline eesmärk on välja selgitada parameetrid, mida saab kasutada, et teha vahet Parkinsoni tõvega ja tervetel patsientidel.

Magistritöö raames loodi sõnatuvastus süsteem, mille abil määratakse kindlaks kirjutatud lause õigsus. Analüüsiti erinevaid Motion Mass, kinemaatilisi ja geomeetrilisi lause parameetreid ning selgitati välja, millised parameetrid on sobivad patsiendi terviseseisundi tuvastamiseks.

Magistritöö tulemusena selgitati välja asjakohaseimad parameetrid ning loodi klassifikaatorid, mis võimaldavad ennustada patsiendi terviseseisundit.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 37 leheküljel, 7 peatükki, 9 joonist, 3 tabelit.

Contents

List of Figures

List of Tables

1. Introduction

The purpose of present thesis is to analyse the handwritten sentence test used in diagnosing the Parkinson's disease and determine the set of features which can distinguish healthy people from the people with Parkinson's. The end goal is to construct a classifier which gives feedback and assistance to overseeing physician about the presence of the disease.

Parkinson's disease is a neurodegenerative disorder of the central nervous system named after the English surgeon James Parkinson. In most cases, notable symptoms include dementia and deterioration of motor functions which can manifest as shaking, rigidity across the body, slowness of movement and difficulty with walking. The cause of the disease as well as a definitive cure is not currently known to today's medical community.

The most effective way to contest the disease is an early diagnosis and respective treatment to alleviate the symptoms. Early detection is a matter of utmost importance as it has the biggest potential to mitigate the course of the disease.

However, there is no certain test to diagnose the Parkinson's disease in patients. Instead, the prevalent method is to conduct series of smaller tests to determine the existence of the disease. Tests of the testing suite are not definitively set and can vary between different situations. Some of the tests often included in the set are Luria's alternating series tests, the drawing of clock, Poppelreuter's test and handwritten sentence test. The latter one being the focus of the present thesis.

In handwritten sentence test, the patient is asked to write a specific sentence. Sentence asked of the patient varies but is most importantly in the patient's native language. Physician or doctor is overseeing the writing process and looks for different signs which could indicate the state of patient's condition. Some of the more common signs include decreased hand swings, rigidity and micrographia which is a disorder where the handwriting is abnormally small or cramped together [2].

Another sign that could indicate the existence of the disease is related to dementia. As was mentioned in [3], many of the Parkinson's disease patients have sentence comprehension deficit which can influence the speed and correctness when going through the handwritten sentence test.

Earlier, these tests were performed with pen and paper. Today, we have the opportunity to conduct these tests on a digital tablet, using a stylus as a pen. Using a tablet has many advantages as we can track the pen more accurately. Thanks to that, different parameters can be extracted, like the velocity and acceleration of the pen throughout the writing process. Additional parameters dependant on the specific hardware can also be available, such as the accurate angles at which the pen is held and pressure that is applied by the tip of the pen.

In present thesis, the digitalized sentence data from the test is analysed by performing a word recognition to determine the correctness of the sentence, extracting of kinematic, motion mass and geometrical features from the sentence, performing statistical analysis of the extracted features with the end result being the creation of a classifier that uses the most relevant features. The function of the classifier is to help the assessing physician to make a more accurate diagnosis.

1.1. Motivation

One of the reasons for the interest in this area is the low general accuracy of diagnoses. It is reported that one in every fifth Parkinson's disease diagnosis is a misdiagnosis [4]. Even a slight improvement in the accuracy of the diagnoses could make a huge difference to this field as Parkinson's disease is one of the most common neurodegenerative diseases affecting over 6 million people worldwide.

Present thesis also complements previous and ongoing research of tests used in diagnosing the Parkinson's disease in Tallinn University of Technology. Further information about the specific studies is presented in the chapter "Linked studies". Results from this work will hopefully take the research one step closer to understanding the differences between the healthy people and people with Parkinson's disease.

1.2. Related work

The notion of motion mass is introduced in [5]. Motion mass initially described the smoothness of the human limb movement but later found use in detecting incorrect performance of the therapeutic exercises. The initial notion of motion mass could then be adjusted to fit the specific research problems. This was done in [6], where the three initial motion mass variables: combined Eucleidian distance, trajectory mass and acceleration mass were adjusted to handwriting rather than limb movement. As a result, following variables were calculated: trajectory length, acceleration mass, velocity mass and length of the action in time.

Other notable works include "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease" by Peter Drotar [7]. A study with 75 subjects evaluated the kinematic and pressure features produced by the handwritings from numerous handwriting tasks, such as the drawing of Archimedean spiral and writing of a sentence.

Results varied from task to task, achieving the lowest classification accuracy in the handwritten sentence test. Present thesis will hope to surpass these results with the help of incorporation of other features - most specifically motion mass features.

1.2.1. Handwriting recognition

According to [8], research in handwriting recognition seemed to have peaked in the late 1990s, but is currently experiencing a renewal for the following reasons: advancements in machine learning techniques, increasing number of smart devices (with a stylus) and a desire to not to use keyboards.

There are two mainstream approaches to the recognition of handwritten characters. Earlier results dated back to the end of the XX century mainly use topologic approach, where geometric features of the characters play a key role. For example, this approach has been used by Tou and Gonzalez in [9]. Nowadays, deep learning techniques like convolutional neural networks are prevailing.

Many of the commonly used features in handwriting recognition are normalized coordinates, curvature, aspect ratio, curliness, linearity, inflection points, stroke crossings, velocity, ascenders and descenders, directional features, moments, number of strokes, rendered bitmaps, and orientation maps [8].

Handwriting recognition accuracy and error rates vary heavily depending on the method, that is used to calculate accuracy, text language and testing group. In general, handwriting recognition yields much lower accuracy results than typed text recognition. In [10], one in every six letters were wrongly recognized. This study can be paralleled to present thesis because of its target group. Children and Parkinson's disease patients do represent minority of the society with relatively different handwriting than the rest of the people.

1.3. Linked studies

Present thesis belongs to a larger group of similar studies conducted in the Tallinn University of Technology which are focusing on the analysis of the tests taken during the Parkinson's disease testing. More closely linked studies are shown on the following figure.

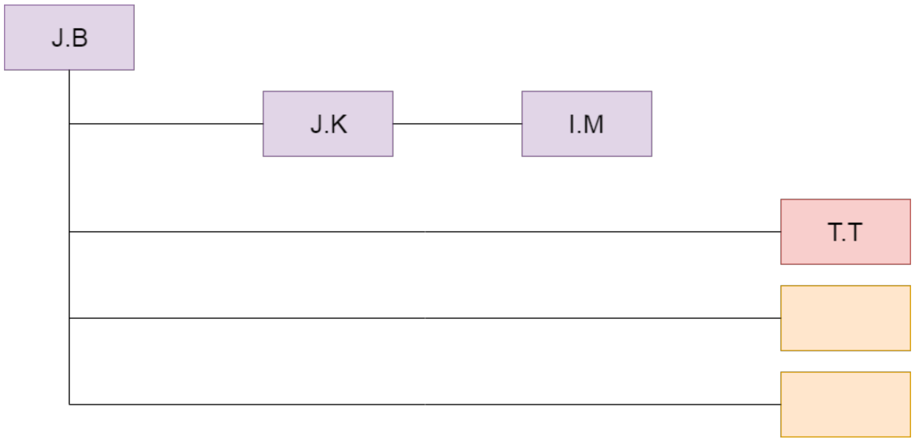

Figure 1. Linked studies

Current thesis is represented by the red rectangle. Orange rectangles are representing studies which are in the works at the moment. Rectangles are positioned chronologically from left to right.

"Alternative Approach to Model Changes of Human Motor Functions" by Jevgenii Borushko [11] is marked with the initials J.B. This Master's thesis focused on numerically describing the changes of human motor functions.

Initials J.K correspond to Master's thesis named "Quantitative Analysis of the Kinematic Features for the Luria's Alternating Series Test" by the author of Julia Kozhenkina [6]. Thesis focused on evaluating and determining the set of parameters from the Luria's alternating drawing tests that would indicate differences between Parkinson's disease patients and healthy controls.

I.M corresponds to Ilja Mašarov's Master's thesis named "Digital clock drawing test implementation and analysis" [12]. In this thesis, the goal was to digitalize the digital clock drawing test with the side goal of gathering and analysing data from the implemented application.

Present thesis will carry the purpose of evaluating one of the tests used in diagnosing the Parkinson's disease, similar to Julia Kozhenkina's evaluation of Luria's alternating series test and Ilja Mašarov’s evaluation of digital clock drawing test.

2. Problem Statement

Handwritten sentence test is one of the most popular fine motor tests which is widely used in neurology to assess the existence and severity of the neurodegenerative disorders like Parkinson's or Alzheimer's diseases. The main goal of the present thesis is to extract and evaluate information generated from this test.

Handwritten sentence test consists of the patient writing a given sentence in their native language. Writing process is examined by the overseeing physician who then makes an assessment about the health condition of the patient.

This kind of testing was previously done on paper with a pen. Today, we have the opportunity use digital tablets with a stylus as a pen. This allows us to gather information previously unavailable for the physician such as the exact velocities and acceleration of the pen, pressure that is applied on the tip of the pen and many more.

Digital testing platform including tablet computers and styluses is already in use in the Tartu University Hospital. This platform is also the source for the raw data used in the present thesis.

Tallinn University of Technology students and researchers have already evaluated some of the tests used in the diagnosis of the disease, for example Poppelreuter's test [13], Luria's alternating series tests [6] and drawing clock test [12]. Results from these works are already in clinical testing. Analysis of the handwritten sentence test and resulting classifier produced is meant to complement and improve the digitalized testing platform.

Analysis consists of multiple parts. First part is the word recognition which determines the validity of the written sentence. Then, other parameters such as the kinematic features and geometrical parameters of the sentence are extracted and calculated.

Motion mass parameters introduced by [5] are also incorporated in present thesis. Specific motion mass parameters describing the smoothness of the writing process are calculated.

All of the features and parameters are then statistically analysed in regard to determining the best set of features to distinguish between patients with Parkinson's disease and healthy controls.

The end results is a classifier which could give feedback about the test to the assessing physician and help the physician with the estimation of the patients health condition.

Integration with the existing infrastructure is not in the scope of present thesis. Further objectives include performing regression analysis in addition to classification.

3. Methods

3.1. Hardware

Patients use Apple iPad tablet in combination with the Apple Pen to execute the test suite. Mentioned hardware with the needed iPad application is already in place in the Tartu University Hospital. Data is stored in the Amazon Web Services server. From there, the data is manually collected to be analysed in the present thesis.

Integration between the software created as result of the work done in present thesis and existing infrastructure was not in the scope of the current thesis but is part of the further objectives.

3.2. Subjects

Total of 26 subjects performed the handwritten sentence test. Out of 26 subjects, 14 were from the Parkinson's disease group and 12 were from the healthy control group. Patients were anonymous and in the age group of 55 and older.

3.3. Data Parsing

Data is acquired in JavaScript Object Notation format (JSON format). Each sentence test is represented by a JSON object that includes the writing hand, identification code, session code, starting time and another JSON object for the writing data.

Writing data is represented in a time series format as the data points are chronologically ordered. Each data point include abscissa and ordinate values of the pen tip, altitude and azimuth of the pen (2), pressure that is applied to the tip of the pen and a time stamp.

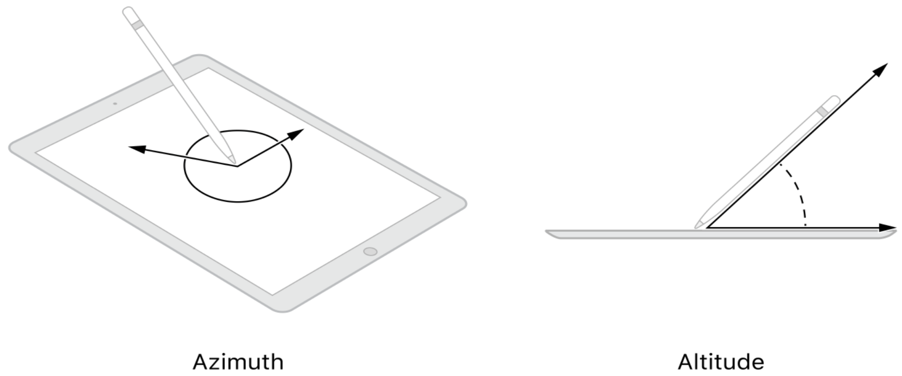

Figure 2. Azimuth and altitude of the pen [1]

3.4. Software

Programming language Python was used to process and analyse the data. Python features a stellar selection of data processing and analysing libraries. This and the author's familiarity with previously mentioned programming language were the reasons that Python was used. Coding was done in Spyder and default Python integrated development environment.

From libraries, Json and Numpy were used to store and process initial data. SciPy and sklearn libraries were used for mathematical functions and classifiers. MatPlotLib and GraphViz libraries were used for various visualization tasks. Software library Tensorflow was used for image recognition.

3.5. Mathematical methods

Statistical hypothesis testing in conjunction with independent two sample t-test for calculating the T-statistics are used to determine relevant features of the handwriting.

Curve fitting is used to describe the change in some of the parameters throughout the sentence. Data points are interpolated on to a line equation where the slope of the line indicates the increase or the decrease of the specific parameter in the writing process.

Results are validated using cross-validation as the amount of existing data is insufficient to compile training and test datasets normally.

4. Methodology

In this chapter, methodology is divided into two parts. First part is the methodology for the handwriting recognition process which describes the original and final methodologies chosen. Second part describes the overall methodology for extracting the features.

4.1. Methodology for Handwriting Recognition

4.1.1. Original methodology

Originally, methodology for recognizing handwriting would follow the general process used in other recognition systems as the following steps would be implemented [14]:

1. Pre-processing - converting initial information to suitable data.
2. Character segmentation - detecting the transition areas where characters are starting and ending. Ultimately discovering unrecognised characters.
3. Character recognition - performing character recognition, converting unrecognised characters to specific letters.
4. Overall sentence recognition - Assessing the correctness and value of the words and in turn, the whole sentence.

However, original approach proved to be unsuitable for multiple reasons. One of the reason being that most of the sentences lacked in quality and in some cases, for example, suffered from micrography (3). This made character segmentation unreliable and yielded unacceptable results as used segmentation algorithms could not distinguish characters in a normal handwriting and at the same time, in a small, cramped together handwriting.

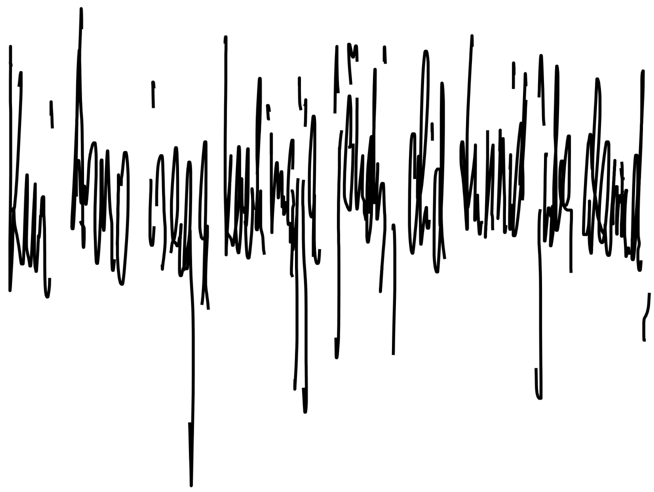

Figure 3. Sample sentence with micrographia

Algorithms that were used for character segmentation included modified DBSCAN and classic grid with the latter algorithm delivering slightly better, yet still inadequate results. However, these algorithms were used to extract some of the geometrical features of the sentence such as the areas of the characters and vertical alignments of the sentence, albeit being not too reliable.

Other problem presented during the segmentation was related to the vertical alignment of the sentence, more specifically, the change in vertical alignment (4). Some of the sentences were abnormally rising in their text alignment which made it difficult to segment the characters and also to distinguish different lines in the text.

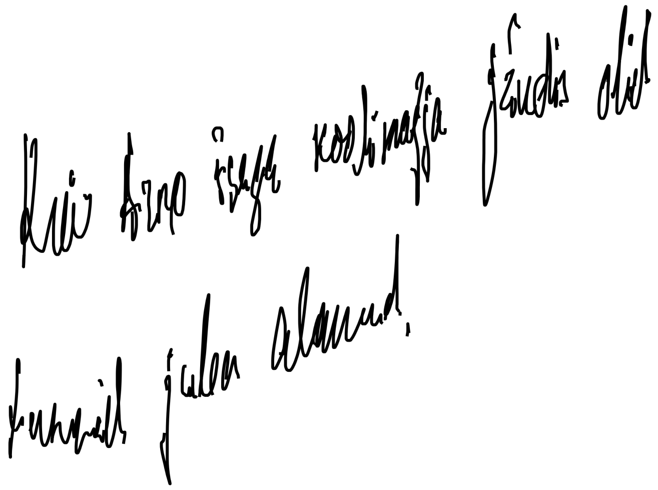

Figure 4. Sample sentence with unusual text alignment

Additionally, a number of existing recognition systems were tested, such as the Tesseract software. Tesseract (sometimes called TesseractOCR) is one of the earliest optical character recognition engines originally developed at HP between the years of 1984 and 1994 [15]. Since then, Tesseract has gone being to an open source software with obtained sponsorship from Google since 2006. However, Tesseract performed poorly in current task because it is directed to typed text rather than the handwritten text recognition.

Other existing solution tested was MyScript software. MyScript concentrates on real time character recognition by taking account the coordinates of the writing. MyScript performed the best recognition but unfortunately introduced an API call limit which was unsuitable for the present thesis.

Other problems with off-the-shelf solutions included previously mentioned lack of support for handwriting and also a lack of support for the Estonian language as almost all the today's solutions use dictionary of words to validate the result when performing the recognition process. When missing the Estonian words, the recognition results yield much lower accuracy and therefore makes them difficult to use.

4.1.2. Final methodology

Eventually, the final approach was reached - building our own recognition system featuring machine learning incorporated image recognition. The fact that the sentence is same in each and every test allows us to train a model where only the words from this specific sentence are recognized. This eliminates the need to perform character segmentation and recognition. Instead a word segmentation is performed, which is a considerably easier task producing more reliable results.

System is based on recognizing images with text on them rather than using coordinate values of the strokes. This approach was chosen in hope of more accurate results as the sentences contained a lot of unwanted or out of place micro-strokes. Since so-called micro-strokes are generally intersecting or completely on the main contour of the character, the impact of those strokes is reduced and better results are produced when recognizing text from the image rather than coordinates of the strokes. Also, using images makes text scaling issues considerably smoother as images can be scaled much more easily than strokes.

Our word-based model had nine different classes, each corresponding to specific word. These words were "Kui", "Arno", "isaga", "koolimajja", "jõudis", "olid", "tunnid", "juba", "alanud". System takes one input - image with single word, and outputs confidence ratings for classes which are most likely to feature the word from the input image. Data to train the model was gathered manually from the original dataset.

Tensorflow Python library and API were used in conjunction with Inception v-3 image recognition model. Tensorflow is an open source software from Google which is often used for image recognition applications [16]. Inception v-3 is an computer vision model which utilizes convolutional neural networks and features fast training process with accurate results [17].

4.2. Methodology for feature extraction

General knowledge about the features that were important to extract were given by the previous works from the field, specifically [7], but also from an eye test.

Firstly, basic features such as the number of strokes, number of data points, total time, average azimuth of the pen, average altitude of the pen, average pressure and others are extracted. These features are easy to gather as in they are directly available from the dataset or need a slight effort including math to gather.

After the extraction of basic parameters, specific handwriting data is collected. This is acquired with the help of handwriting recognition. Then, geometrical and motion mass parameters are gathered.

For some features, its values are curve fitted on a straight line, meaning that line equation which describes the data points the best is constructed. Acquired line equation (most importantly the slope or gradient of the line) describes the changes within that feature throughout the sentence.

4.2.1. Geometrical features

In some of the sentences, unusual geometrical features were presented. Therefore it was needed to extract relevant geometrical properties from the sentence, such as the rise and fall of the vertical alignment of the sentence, average areas of the characters and angles between top and bottom of the text. Previously mentioned curve fitting was used for describing the fall and the rise of the sentence.

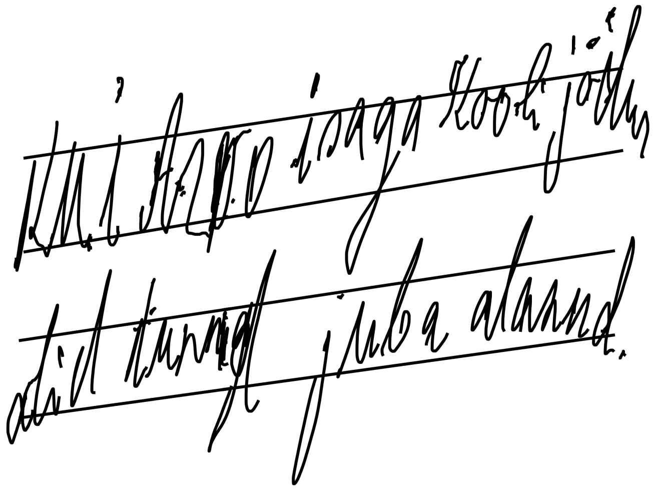

Figure 5. Lines describing the geometrical features of sentence

4.2.2. Motion mass features

The notion of motion mass was introduced by [5] and [11]. Originally, it described the smoothness of the human limbs or joints, but was modified in later works to describe the smoothness of the handwriting.

In present thesis, the following motion mass parameters were collected and calculated:

- Velocity mass
- Acceleration mass
- Jerk mass
- Pressure mass

Motion mass parameters are calculated by summing the parameters at each measuring point. For example, calculation for the velocity mass V_T is done like this:

$$V_T = \sum_{i=1}^{n} |v_i| \tag{1}$$

where v_i is the velocity in the ith observation point and n is the total number of observation points. Similarly, acceleration mass A_T and jerk mass J_T were calculated:

$$A_T = \sum_{i=1}^{n} |a_i| \tag{2}$$

$$J_T = \sum_{i=1}^{n} |j_i| \tag{3}$$

where a_i and j_i are respectively representing acceleration and jerk in the ith observation point.

Additionally, total time T ratio to motion mass parameters was calculated:

$$V_T/T \quad A_T/T \quad J_T/T \tag{4}$$

and similarly total time spent with the pen tip on the screen T_w ratio to motion mass parameters:

$$V_T/T_w \quad A_T/T_w \quad J_T/T_w \tag{5}$$

5. Analysis

5.1. Statistical Analysis

After the extraction of the features, the data is divided into two groups. One group is for the healthy controls and other one for the patients with Parkinson's disease. Statistical analysis is performed on the basis of those two groups. For each specific feature p-value (also known as probability value or asymptotic significance) and fisher score are calculated. These parameters describe the variation in between both of the patient groups.

Independent two sample t-test is used for calculating the p-value. In the process of calculating the p-value, t-statistic is calculated which also describes the variation between the two groups.

P-value is then used to perform statistical hypothesis testing. Null-hypothesis H_0 and alternative hypothesis H_a are stated:

$$H_0 : \mu_C = \mu_{PD} \tag{6}$$

$$H_a : \mu_C \neq \mu_{PD} \tag{7}$$

where μ_C represents the data from the healthy control groups and μ_{PD} represents the data from the patients with Parkinson's disease.

Significance level for p-value is 1 percent. This means that the difference between the datasets is deemed statistically significant and alternative hypothesis H_a is accepted when the p-value is lower than 0.01.

Selected features are then manually tested and used to train the classifiers. Three classifiers were tested, these included decision tree classifier, K-nearest neighbours (KNN) classifier and linear support vector machine (SVM) classifier. Choice for these classifiers was made because of the support and compatibility with smaller datasets.

Small dataset also directed us to using cross-validation method. In the validation phase, the training data for the classifiers consisted of 25 sentences and the test data of 1 sentence. Training was done 26 times (the number of sentences in the original dataset) for the purpose of each sentence being in the test dataset once.

5.2. Results

Best results according to statistical analysis were related to time and velocity or to derivatives of velocity. Other features where alternative hypothesis was accepted included average length of one stroke, pressure parameters and average height of characters. These results are displayed in the table below.

Parameter name	t-statistic	p-value	fisher score
Number of data points	-4.0889	0.0004	0.6966
Velocity mass	2.8044	0.0098	0.3277
Velocity mass divided by total time	4.6087	0.0001	0.885
Velocity mass divided by on-screen time	5.4473	0.0001	1.2364
Acceleration mass	2.9106	0.0077	0.353
Acceleration mass divided by total time	4.7249	0.0001	0.9302
Acceleration mass divided by on-screen time	5.5162	0.0001	1.2679
Jerk mass	-3.1423	0.0044	0.4114
Jerk mass divided by total time	-4.8942	0.0001	0.9981
Jerk mass divided by on-screen time	-5.6643	0.0001	1.3369
Average time to complete one stroke	-3.1179	0.0047	0.4051
Vertical velocity mass	3.0087	0.0061	0.3772
Average vertical acceleration	-3.2557	0.0034	0.4416
Average height of characters	3.0766	0.0052	0.3944
Pressure mass divided by on-screen time	3.7473	0.001	0.5851
Pressure mass divided by total time	3.7392	0.001	0.5826
Average length of stroke	3.2955	0.003	0.4525

Table 1. Selected features when p-value significance level is 1 percent

Number of data points is related to velocity since iPad tablets are gathering data points periodically after a certain time interval. Therefore, a faster writer will record fewer data points than a slower one.

Additional features when the p-value significance level is 5 percent are shown in the table below.

Parameter name	t-statistic	p-value	fisher score
Average area of characters	2.1804	0.0393	0.1981
Horizontal velocity mass	2.1994	0.0377	0.2016
Average horizontal acceleration	-2.3903	0.025	0.2381
Average altitude of the pen	2.0681	0.0496	0.1782
Total time	-2.6374	0.0144	0.2898

Table 2. Selected features when p-value significance level is 5 percent

5.3. Classification

Decision tree (DT) classifier, support vector machine (SVM) classifier and K-nearest neighbours (KNN) classifier were tested with various features for the purpose of producing the highest accuracy possible.

Accuracy of the classifiers was between 92 and 62 percent. Highest results were achieved with the decision tree classifier and lowest results with SVC classifier. KNN classifier was close to decision tree classifier, achieving results of 84 percent.

Classifier	Highest accuracy	Percentage
DT	24/26	92
SVM	16/26	62
KNN	22/26	84

Table 3. Classification results

Most accurate results were obtained, when selected features for the training included the feature "Velocity mass divided by pen on-screen time" by itself or in conjunction with "Average height of the characters" or "Average altitude of the pen". "Velocity mass divided by pen on-screen time" by itself produced the highest results with all of the three classifiers. Adding "Average altitude of the pen" of "Average height of the characters" reduced the accuracy of DT classifier to 88 percent but stayed the same with other classifiers.

Generally, acceptable results were produced when following the formula which uses two features in training the classifiers, while one feature being related to velocity and other one not - this include geometrical and other various features such as the previously mentioned altitude of the pen. Increasing the number of features used in the training seemed to have had a negative effect on the accuracy.

The reason for the poor performance of SVC classifier remained unknown to us as the best guess would be the size of the dataset which is considerably smaller than suggested size for this classifier.

Confusion matrices of the classifiers trained with the same set of features ("Velocity mass divided by pen on-screen time" and "Average altitude of the pen") are shown below.

		Prediction Outcome	
		PD	**H**
Actual Value	**PD′**	13 True Positives	1 False Negative
	H′	2 False Positives	10 True Negatives

Figure 6. Confusion matrix of DT classifier

		Prediction Outcome	
		PD	**H**
Actual Value	**PD′**	13 True Positives	1 False Negative
	H′	3 False Positives	9 True Negatives

Figure 7. Confusion matrix of KNN classifier

		Prediction Outcome	
		PD	**H**
Actual Value	**PD′**	4 True Positives	10 False Negative
	H′	0 False Positives	12 True Negatives

Figure 8. Confusion matrix of SVC classifier

False classifications of the SVM classifier appeared entirely made up of false positives.

In contrast, KNN classifier produced more false positives while the overall amount of misclassification were considerably lower.

Additionally, visual representation of the decision tree produced by the DT classifier looked as following:

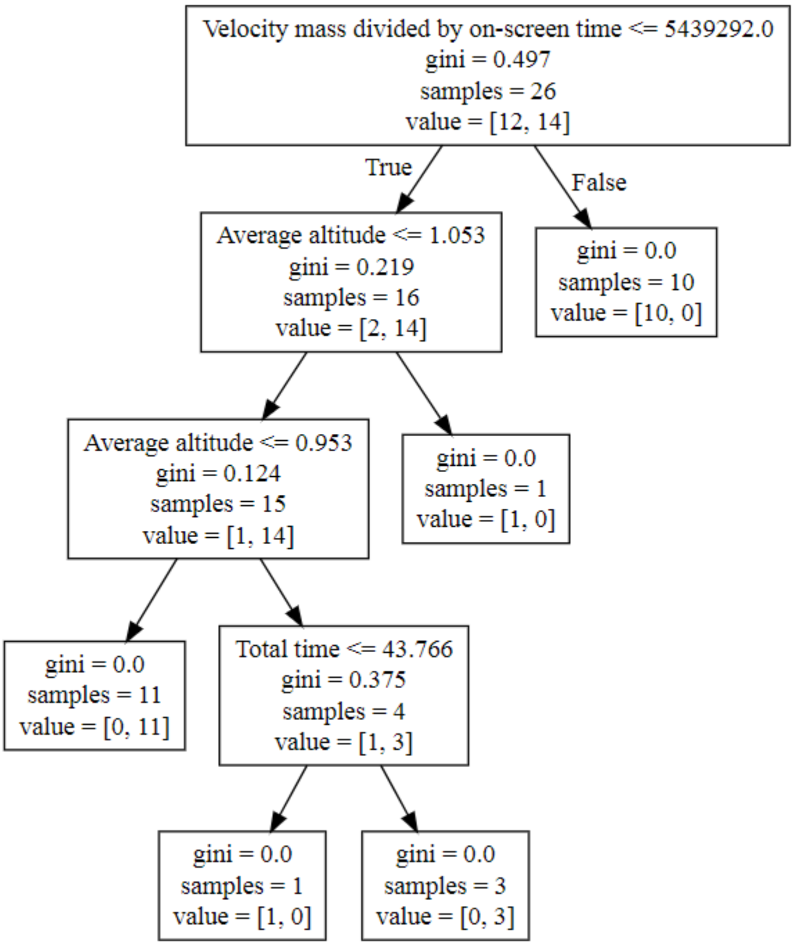

Figure 9. Decision Tree

While the results from the classification process are unusual, there is a clear indication that distinguishing PD patients and healthy controls by the velocities, motion mass parameters and geometrical features is possible.

6. Discussion

In conclusion, the results of this study indicate strongly that there is a difference in the writing speeds between the healthy patients and patients with Parkinson's disease. This is illustrated by the very low p-values of the speed and velocity related features. Even with selected significance level of only 1 percent, alternative hypotheses were accepted with thirteen velocity and speed related features.

As medical evidence suggests, there is an indication that geometrical features of the characters such as height and area are also different in sentences written by the Parkinson's disease patients. However, there is no such indication about the widths of the characters. This is interesting as micrographia which is generally expressed by the patients, include taller, smaller but also narrower characters. In present thesis, this result can be caused by the insufficient amount of training data or mistakes during the feature extraction process.

All of the motion mass parameters gathered in this study showed a considerable difference between the two groups. This can be indicative of smoother writing process in the case of healthy people and falls in line with previously done studies which incorporated motion mass parameters.

The biggest room for improvement is in the handwriting recognition part of this thesis. In future, the program should use character recognition instead of word recognition to reliably determine the correctness of the written sentence. Also, that could allow us to gather character specific features which can lead us to discovering differences in the writing of single characters.

Current word recognition system shortcomings are mostly due to lack of data in the training dataset. Words from 26 sentences are clearly not enough to train a reliable recognition system as most of the advanced solutions used today have been trained on thousands of data points. It is expected that increasing the training sentences, the overall accuracy of the system will go up considerably.

At the moment, the word recognition system could have a better integration with the rest of the software. However, the modular structure of the software is not necessarily a negative characteristic as this allows us to easily change the recognition system which as previously mentioned, is in the future objectives.

Hopefully, classification results of around 90 percent encourage following studies as it shows that differentiating patients on the basis of handwritten sentence test is possible and could be even effective. However, it has to be mentioned that these results are by no means exhaustive and numerous improvement affecting the results could and should be made.

Further objectives include increasing the size of the dataset for more reliable results, regression analysis in addition to classification and integration between the software created in present thesis and the existing infrastructure.

7. Conclusion

The main goal of this thesis was to analyse and evaluate handwritten sentences from the handwritten sentence test used in diagnosing the Parkinson's disease.

In the process, handwriting recognition was needed to perform. Multiple approaches were tested but finally settled on building our own handwriting recognition system which utilized image recognition and machine learning.

Kinematic and geometrical features of the sentence were extracted and then analysed with respect to their suitability to our goal.

Motion mass parameters were calculated. Motion mass parameters included velocity mass, acceleration mass, jerk mass and pressure mass. Corresponding ratios to time and writing time were calculated.

All of the extracted features were statistically analysed. P-values and fisher scores were calculated. Statistical hypotheses were created and the most suitable features were selected.

Three classifiers were tested with different set of parameters. These three classifiers included decision tree, K-nearest neighbours and support vector machine classifier. The best performing classifier was decision tree classifier, producing results in the accuracy level of over 90 percent.

Overall, results achieved in this thesis would strongly indicate that patients with Parkinson's disease write slower. Other indications gathered are that they are also writing smaller and their hand movements are not as fluid as the hand movements of healthy people.

In future, handwriting recognition system needs to improve. This can be achieved with adding data to training dataset. This research can be used as a foundation to explore handwriting differences between the healthy people and Parkinson's disease patients. Next objective would be to perform regression analysis on the basis of this research.

References

[1] "Handling input from apple pencil." http://developer.apple.com/documentation/uikit/touches_presses_and_gestures/handling_input_from_apple_pencil.

[2] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[3] M. Grossman, "Sentence processing in parkinson's disease," *Brain and Cognition*, vol. 40, no. 2, pp. 387–413, 1999.

[4] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, and G. Logroscino, "Accuracy of clinical diagnosis of parkinson disease a systematic review and meta-analysis," *Neurology*, vol. 86, no. 6, pp. 566–576, 2016.

[5] S. Nõmm and A. Toomela, "An alternative approach to measure quantity and smoothness of the human limb motions," *Estonian Journal of Engineering*, vol. 19, no. 4, p. 298, 2013.

[6] S. Nomm, A. Toomela, J. Kozhenkina, and T. Toomsoo, "Quantitative analysis in the digital luria's alternating series tests," in *Control, Automation, Robotics and Vision (ICARCV), 2016 14th International Conference on*, pp. 1–6, IEEE, 2016.

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

[8] D. Keysers, T. Deselaers, H. A. Rowley, L.-L. Wang, and V. Carbune, "Multi-language online handwriting recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1180–1194, 2017.

[9] J. T. Tou and R. C. Gonzalez, "Recognition of handwritten characters by topological feature extraction and multilevel categorization," *IEEE Transactions on Computers*, vol. 100, no. 7, pp. 776–785, 1972.

[10] D. Das, "Handwriting characters recognition errors and recognizing errors of children on the tablet pc," *International Journal of Science, Environment and Technology*, vol. 1, no. 1, pp. 36–40, 2012.

[11] S. Nomm, A. Toomela, and J. Borushko, "Alternative approach to model changes of human motor functions," in *Modelling Symposium (EMS), 2013 European*, pp. 169–174, IEEE, 2013.

[12] I. Mašarov, "Digital clock drawing test implementation and analysis," Master's thesis, Tallinn University of Technology, 2017.

[13] S. Nõmm, K. Bardõš, I. Mašarov, J. Kozhenkina, A. Toomela, and T. Toomsoo, "Recognition and analysis of the contours drawn during the poppelreuter's test," in *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on*, pp. 170–175, IEEE, 2016.

[14] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: a comprehensive survey," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 1, pp. 63–84, 2000.

[15] R. Smith, "An overview of the tesseract ocr engine," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, vol. 2, pp. 629–633, IEEE, 2007.

[16] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning.," in *OSDI*, vol. 16, pp. 265–283, 2016.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.