MACHINE LEARNING AND DIGITAL SKETCH RECOGNITION METHODS TO SUPPORT

NEUROPSYCHOLOGICAL DIAGNOSIS AND IDENTIFICATION OF COGNITIVE

DECLINE

A Dissertation

by

RANIERO AARON LARA GARDUÑO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Tracy Hammond
Committee Members,	Anxiao Jiang
	James Caverlee
	Gerard Coté
Head of Department,	Scott Schaefer

May 2021

Major Subject: Computer Engineering

ABSTRACT

With approximately 15 to 20 percent of adults aged 65 and older living with Mild Cognitive Impairment (MCI), researchers in neuropsychology have placed increasing emphasis in early detection to best preserve quality of life for MCI patients before the turnover to full Alzheimer's or similar types of dementia. Current methods of diagnosis in clinical neuropsychology involve a lengthy process of developing a full behavioral profile, in which tests are administered to examine the existence and extent of cognitive impairment. Efforts to digitize and semi-automate the process have emerged, but solutions tend to be limited in scope and there exists ample room for improvement in developing new and faster diagnosis techniques.

This dissertation presents digital neuropsychological diagnosis tools by adapting existing neuropsychological tests or presenting new prototypes based on existing methodology. Digitizing these examinations allows for the collection of high-granularity data that allows for the creation of machine learning algorithms as well as automated graders designed to be supported with various intuitions about the test-taker's behavior in a manner consistent with existing literature in neuropsychology. We sought to produce knowledge on the mapping of sketch recognition analysis to identifying MCI in patients. On a technical level, we present novel sketch recognition techniques that identify specific degrees of shape correctness for automated grading of complex-figure tests, as well as sketch segmentation techniques specific to the TMT.

To that end, we present three digitized systems that collect fine-grained behavioral usage data through cognitive examinations: 1) a digital TMT that simulates the paper-and-pencil experience on a tablet, 2) an automated grader for digitized ROCF tests, and 3) a new examination that leverages modern touch tablet technology to expand on the inherent diagnosis weaknesses of paper-based examinations as patients manipulate a virtual 3-dimensional object. We present our technical approach, results from experiments, expected goals, and the intellectual contributions of this dissertation research.

Our studies produced over 350 digital sketches from neuropsychological examinations across

our three presented systems. Data segmentation techniques have resulted in over 6000 pieces of segmented sketch data. We believe the findings presented in this dissertation can provide valuable insight into the process of developing semi-automated or fully automated neuropsychological diagnosis tools.

DEDICATION

To my research team: parents, siblings, colleagues, and friends across three continents

just as it takes a village to raise a child, it takes a community to build a dissertation.

ACKNOWLEDGMENTS

CONTRIBUTORS AND FUNDING SOURCES

Contributors

Funding Sources

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Clinical neuropsychology focuses on identifying and managing behavioral abnormalities that might lead to severe conditions such as Alzheimer's disease and dementia, both of which are conditions that continue to climb in prevalence in recent years [3–12]. Neuropsychologists identify and track cognitive decline through observation of a patient's behavior, and this frequently involves exercises and tests, some of which are handwritten. Handwriting in particular offers a window into a person's cognitive state [13, 14]. Handwriting is affected by a wide variety of conditions that affect cognition, ranging from temporary (tiredness and inebriation [15–20]) to permanent (Alzheimer's disease [21–26]).

In recent years neuropsychology and modern digital technologies have intersected. Aside from obvious applications that focus on digitizing records, modern mobile applications and the development of analytical and predictive algorithms are increasingly being leveraged as assistants for diagnostic and therapeutic purposes [27–30].

Of the many neuropsychological tools to diagnose and track cognitive decline, tests involving drawing on paper with a pen or pencil garnered our interest, as there exists a multitude of digital sketch recognition technologies that leverage machine learning to assist in handwriting recognition, labeling, ideation, and technologies for creative expression. Existing research has extended to identify properties of those who are writing with these technologies; research has identified sketching behaviors correlated to development in young children, domain expertise in creative design problems, the trajectory of experience with perspective sketching activities, and the tracking of aptitude in handwriting foreign-language characters and symbols among many others. It stands to reason that if these technologies can be successfully leveraged for cognitive development, they may also be used to identify cognitive decline. The fact that drawing activities remain in use to this day as a diagnostic tool to measure cognitive decline further motivates a research focus in intersecting both domains of research.

This research focuses on two sketching activities that are frequently used in clinical neuropsy-

chology: the Trail-Making Test [31], and the Rey-Osterrieth Complex Figure Test [32]. The former is a connect-the-dots exercise that tests executive function, while the latter is a test involving the memorization and copying of a large complex figure. Trail-Making Tests have been in use for several decades, initially as part of a general intelligence test but quickly adapted for diagnosis purposes once it became clear it was also sensitive to cognitive decline. Rey-Osterrieth Complex Figure Tests are but one of several complex-figure examinations developed and used over several decades as the sensitivity to cognitive decline was also identified. This research dissertation documents the process of adapting these tests to a digital format, and how these applications were built and adapted to ensure the task of drawing is as unaffected as possible in the transition to digital technology. We developed recognition technology for predictive purposes in two fronts. For ROCFT, we focus on the recognition of the same geometric shape a neuropsychologist looks for when grading complex figure tests. For TMT, we focus on the recognition of cognitive decline in order to develop a predictive model. The former is meant to create an assistive technology that aids neuropsychologists by automating an existing process. The latter aims to create a predictive technology to identify cognitive decline itself.

This document also outlines research into a third option: the creation of entirely new tests using current advancements in mobile technology. We present our findings and document the design of a new test, called the 3D-Trail-Making Test, that uses touch tablet technology to adapt the TMT to a more dynamic examination. We outline the reasoning behind design decisions and draw comparisons in performance between participant groups and how these show similar sensitivity to specific conditions.

2. INTELLECTUAL MERIT

Digitizing existing neuropsychological tests and creating new ones necessitates domain knowledge in neuropsychology as well as expertise in digital sketch recognition algorithms and techniques. This intersection of both domains is non-trivial and requires careful considerations for application and interaction design for participants who use the presented solutions. Existing solutions for general-purpose sketch recognition techniques are not adequate for the task given the specificity of a neuropsychologist's needs in automating sketch analysis. Similarly, resulting sketches for both the TMT and ROCFT tasks are intentionally designed to not resemble everyday writing tasks. This requires solutions specifically tailored for these tasks if the recognition techniques are expected to work well for this domain.

This dissertation focuses on the contributions to human-computer interaction, specifically in two broad categories: 1) the creation of a comprehensive set of recognition heuristics to present an automated analyzer that grades existing neuropsychological examinations with largely the same rules as applied by an actual neuropsychologist, and 2) the creation of predictive algorithms that uses high-granularity digital sketch input data to generate predictions of possible signs of cognitive decline. The former leverages heuristic rules for recognition of geometric shapes, while the latter leverages the computation of geometric features of sketches to produce predictive algorithms. These two categories are representative of the two major categories of diagnosis and treatment automaton: a system that analyses tests with already established grading heuristics (albeit originally designed to be graded by humans), and a system that can generate its own heuristics for predictions via machine learning.

A third class of contributions is orthogonally developed in this dissertation: design considerations developed for the adaptation of digital neuropsychological tests, both existing and novel. Early designs in digitizing the TMT, for example, consisted of the manipulation of computer mice to connect dots. This task, especially in the 1990s when the computer mouse was a relatively more novel input modality for the general consumer, introduced a confounder in the requirement

to manipulate a mouse pointer successfully. Our intent is to mimic the test taking experience as closely as possible to the pen-and-paper experience, both on a a tablet application and via "smart pen" use on a printed piece of paper. We also intentionally created a novel, digital-only task and explored how input would be designed. These considerations, while not the focus of this dissertation, may provide key insights into how to develop these digital applications to aid, rather than hinder or replace, existing diagnosis methodology in neuropsychology.

We present the following contributions to the field of human-computer interaction as it pertains to sketching and cognition:

- This dissertation work presents a more granular analysis of sketches produced by the TMT and ROCFT neuropsychological examinations, as well as the novel 3D-TMT task. The segmentation and analysis techniques developed for these tasks is intended to be applied to other similar tasks, such as different abstract shape layouts for complex-figure tests, and other dot layouts for TMT variations.

- For the TMT task, we present predictive algorithms derived via machine learning techniques to explore the classification of participants based on cognitive decline. We perform analysis on general binary classification tasks for "non-MCI" vs. "MCI" participants, as well as predictions of more continuous levels of cognitive decline as reflected in the Montreal Cognitive Assessment task (MoCA).

- For the ROCFT task, we present a comprehensive set of geometric rules for automatically grading complex-figure examinations, striking a balance between leniency for hand-drawn shapes while simultaneously having rigidity through which to determine distortion or absence of shapes.

- For the 3D-TMT task, we present design considerations for a novel neuropsychological examination and a predictive algorithm to determine whether a cognitively healthy participant is a native of the United States vs. a native of Japan. The original TMT task has been shown to be sensitive to background such as nationality, as subtle cultural changes can result

4

in different test completion behaviors. We present how these differences are reflected in a novel neuropsychological examination.

3.   PRIOR WORK

Prior work relevant to the study of cognitive decline reflected in digital sketches from neuropscyhological examinations can be divided into domain-specific information on clinical neuropsychology, as well as existing research in digital sketch recognition, digitizing cognitive examinations, and the current state of research in both.

Clinical neuropsychology has a long and rich history with sketching, as the act of sketching and tasks related to the act of drawing have been shown over the years to be reflective of a patient's cognitive state. The Compendium of Neuropsychological Examinations [2] provides an in-depth look at common examinations and how they are administered, and serves as repository of acquired normative data. Among the many examinations provided, several consist of sketching or drawing tasks, ranging from making patients draw small marks such as in the Double Letter Cancellation tes, to recreating entire sets of abstract geometric shapes such as the Rey-Osterrieth Complex Figure Test.

3.1   Clinical Neuropsychology

3.1.1   Mild Cognitive Impairment

The characteristics of MCI were initially established as part of the Global Deterioration Scale (DGS) [33], defining it as a syndrome where an individual's cognitive decline is greater than expected for their age [34–36]. It is generally somewhat as a precursor to further decline that may progress into dementia, with Alzheimer's in particular being likely. The existence of MCI in itself is not an indication that decline will progress into a more severe disease or condition, and in fact many MCI patients do not decline further. Additionally, unlike these more severe forms of cognitive decline, MCI does not severely impact one's daily quality of life [37] and can thus be challenging to diagnose. Often the signs are subtle and can be easily dismissed as expected decline in executive function for an individual's age. MCI itself is characterized as not having a significant impact on daily activities, and may not manifest in a noticeable way for years.

However, MCI over the longer term can decline into more severe deficits. Characteristics of severe cognitive decline in this condition can vary depending on the severity along with background and genetic conditions particular to the patient. Reisberg *et al.* [33] specifies the emergence of "behavioral disturbances," neurological abnormalities, electrophysiological changes, motor deficits, balance and coordination deficits, and deficits in general active daily living activities. Such activity can be identified from analyzing biomarkers like MRI data with machine learning [38–40], but more subtle behavioral recognition remains an active field of study. With an increase in life expectancy seeing a rise in the prevalence of dementia and Alzheimer's disease [41], research attention has turned to the successful identification of MCI, both in the physical and digital spaces [42, 43], and how existing tools can be improved to assist. Constructional abilities are also significantly affected [44, 45], which have led to the development of tasks aimed at probing said abilities to help identify abnormal decline.

3.1.2 Trail-Making Test

Clinicians have historically relied on paper-and-pencil neuropsychological examinations as one of the primary methods to diagnose MCI, as handwriting and cognitive decline have been shown to be related [46]. These typically involve a series of simple tasks for an individual to complete, and have been shown to be sensitive to the same cognitive functions affected by MCI through several decades of research [47]. We focus on Trail-Making Test, a connect-the-dots task that tests executive function and active memory. Initially conceived as a test to assess general intelligence, the TMT is known to be sensitive to cognitive decline and possible early signs of dementia. Currently the Trail-Making Test is widely used in neuropsychologists' test battery to assess for various signs of cognitive decline, including MCI [2]. The switching between numbers and letters found in the TMT-B relies on frontal lobe function [48–52], and is one of the primary reasons for the sensitivity of the test to MCI.

The test consists of two separate connect-the-dots tasks. The participant is handed a piece of paper with a series of labeled dots printed on it, and are handed a pen or pencil with which to connect the dots. The "A" variant of this test consists of a participant connecting dots in ascending

numerical order (1, 2, 3, and so on), while the "B" variant of the test requires connecting dots alternating between numbers and letters in ascending order (1, A, 2, B, and so on). The participant is typically asked to complete the test as a pair, starting with variant A and immediately followed by variant B. Multiple layouts of these tests exist and are used when a clinician wishes to test the participant more than once, since a different arrangement of labeled dots is necessary to avoid the learning effect. Dot layout has been observed to directly affect time to completion on healthy populations [53]. Participants are asked to not lift their pen or pencil whenever possible. While it is necessary to lift the pen or pencil to correct mistakes should the participant connect the incorrect dot in the sequence, at every other time the participant is asked to leave the pen touching the paper, including when the pen is stationary while they look for the next dot. The participant is not allowed to make erasures of lines already drawn. In some instances of the pen-and-paper version of the TMT, the participant will be asked to cycle between several colored pencils every set number of seconds to aid the proctor in determining how long the participant would take to connect certain dots.

Assessment of Trail-Making Tests is primarily done in two ways: comparing the test score with established normative data, and the qualitatively observed behavior of a participant as the test is being completed. The test score is calculated as the test's time to completion rounded to the nearest whole number. The fact that a score is reported as a single numerical value necessitated the qualitative observation, and over the decades clinicians have devised multiple methods for assessing a participant's performance as they complete the test. Colored pencils were introduced to have participants switch between colors every few seconds, to color-code their progress and visualized how often it takes for a participant to connect the next dot in the test. Clinicians have also recorded the video of the participants to be able to repeatedly analyze their behavior during the test. These behaviors include but are not limited to: sitting posture, ability to follow directions, reactions to mistakes, process to identify the next dot if participants lose their place, and how a participant holds the pen.

These measures taken highlight the notion that the behavior a participant exhibits during the

test is just as important, if not more so, as the single time-to-completion reported score. The subtle nature of MCI, however, has historically meant that clinicians rely on their own expertise and experience for qualitative observations. Most recent advancements to digital sketching technology have made it feasible for these tests to be assessed with much higher granularity than in previous decades.

3.1.3 The Montreal Cognitive Assessment

The Montreal Cognitive Assessment (MoCA) is among the most widely used assessment protocols for gauging an individual's cognitive function. It consists of various short tasks, both written and verbal, aimed at testing various functions of a person's cognition and is frequently administered as a triage to help determine whether a patient requires further diagnosis and possible treatment, and is also frequently used to determine whether patients have symptoms of MCI [54, 55] and more serious diseases like Alzheimer's [56]. Originally developed in 1995 by Ziad Nasreddine [57], it has since been the subject of various validation studies [48, 58], and normative data has been collected and analyzed for patients of various populations [59–62], diseases [63, 64], cognitive states [64, 65], education level [66], and post-trauma conditions [67, 68]. The primary conditions that it has been validated for include MCI, Alzheimer's disease and Parkinson's Disease dementias [54, 58, 69], and has been shown to be more sensitive than other similar examinations such as the Mini-Mental State Examination (MMSE) at detecting early signs of decline related to MCI [70]. Hobson describes that the MoCA can assess cognitive domains including but not limited to "Visuospatial/Executive, Naming, Memory, Attention, Language, Abstraction, Delayed Recall and Orientation (to time and place)" [57].

The MoCA is completed in around 10 minutes and can be completed by anyone who understands the instructions, with the test being translated and validated to over 30 languages [57]. Julayanont *et al.* [48] describes the tasks of the MoCA as follows:

- *Modified Trail Making Test*. A shortened, modified version of TMT-B with 10 dots.

- *Copy of the Cube*. Converting a two-dimensional contour into a 3-dimensional object.

MONTREAL COGNITIVE ASSESSMENT (MOCA)
Version 7.1 Original Version

NAME :
Education :   Date of birth :
Sex :   DATE :
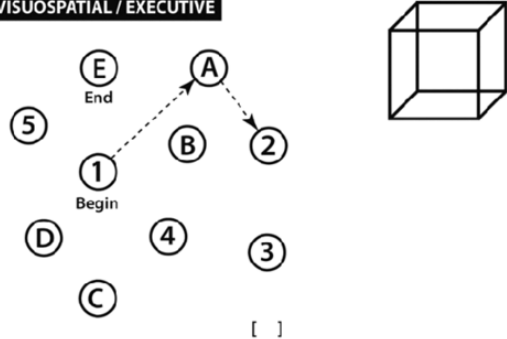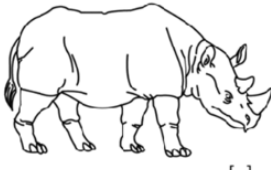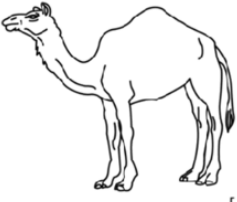
VISUOSPATIAL / EXECUTIVE		POINTS

Copy cube

Draw CLOCK (Ten past eleven)
(3 points)

E End
A
5
1 Begin
B
2
D
4
3
C

[ ]   [ ]   [ ]   [ ]   [ ]   ___/5
                              Contour   Numbers   Hands

NAMING

[ ]   [ ]   [ ]   ___/3

MEMORY Read list of words, subject must repeat them. Do 2 trials, even if 1st trial is successful. Do a recall after 5 minutes.

		FACE	VELVET	CHURCH	DAISY	RED	No points
	1st trial						
	2nd trial						

ATTENTION Read list of digits (1 digit/ sec.).
Subject has to repeat them in the forward order   [ ] 2 1 8 5 4
Subject has to repeat them in the backward order [ ] 7 4 2   ___/2

Read list of letters. The subject must tap with his hand at each letter A. No points if ≥ 2 errors
[ ] F B A C M N A A J K L B A F A K D E A A A J A M O F A A B   ___/1

Serial 7 subtraction starting at 100   [ ] 93   [ ] 86   [ ] 79   [ ] 72   [ ] 65   ___/3
4 or 5 correct subtractions: **3 pts**, 2 or 3 correct: **2 pts**, 1 correct: **1 pt**, 0 correct: **0 pt**

LANGUAGE Repeat : I only know that John is the one to help today. [ ]
The cat always hid under the couch when dogs were in the room. [ ]   ___/2

Fluency / Name maximum number of words in one minute that begin with the letter F   [ ] _____ (N ≥ 11 words)   ___/1

ABSTRACTION Similarity between e.g. banana - orange = fruit   [ ] train – bicycle   [ ] watch - ruler   ___/2

DELAYED RECALL

Has to recall words WITH NO CUE	FACE [ ]	VELVET [ ]	CHURCH [ ]	DAISY [ ]	RED [ ]	Points for UNCUED recall only	___/5
Optional Category cue							
Multiple choice cue							

ORIENTATION [ ] Date   [ ] Month   [ ] Year   [ ] Day   [ ] Place   [ ] City   ___/6

© Z.Nasreddine MD   www.mocatest.org   Normal ≥ 26 / 30   TOTAL   ___/30
Administered by: _____   Add 1 point if ≤ 12 yr edu

Figure 3.1: The Montreal Cognitive Assessment (MoCA). Reprinted from mocatest.org [1]

10

- *Clock Drawing Test.* Drawing a circle representing the lock, and setting the hour and minute hands to half-past eleven.

- *Naming Test.* Subject sees three animal pictures and must name correctly a lion, rhinoceros, and camel.

- *Digit Span.* Proctor reads two sets of digits and subject must repeat them. The first set is read in forward order, and the second set read in backwards order.

- *Letter A Tapping Test.* Proctor reads a list of letters that do not spell any word, and subject must tap their hand with each mention of the letter A.

- *Serial 7 Subtractions.* Subject must specify how many times the number 7 must be subtracted from 100 to achieve 5 numbers that the proctor reads.

- *Sentence Repetition.* Proctor reads two sentences and the subject must repeat them back.

- *Letter F Frequency.* Subject must list aloud as many words as they can remember that begin with the letter F.

- *Object Similarity.* Subject must describe the difference between two pairs of objects that the proctor reads.

- *Delayed Recall.* After several minutes of being read a list of words, the subject must repeat the words back.

- *Time and Place Orientation.* Subject must identify date, month, year, day, place, and city in which they currently are.

The MoCA is frequently used in tandem with other neuropsychological examinations for ensuring that its results are consistent across various other examinations such as the Trail-Making Test. In effect, one might consider performance of the MoCA and the TMT to be correlated, such that exceptionally well or poor performance of one test is likely to lead to poor performance in the

other. Indeed, a brief TMT-B appears in the MoCa as one of the tasks [48], and they both test of the same frontal lobe function. Given that the TMT isolates this particular task to its own test and introduces a TMT-A variant along with a longer set of dots to connect provides a more thorough test of the same functions that the MoCA's TMT-B task tests. For this reason comparisons can be made between performances for both, although little work has been done in machine-learning systems that could predict the MoCA based on the performance of the TMT.

3.1.4 Rey-Osterrieth Complex Figure Test

The Rey-Osterrieth Complex Figure Test (ROCF), developed by Rey [71] in 1941 and refined by Osterrieth [32] in 1944, is a neuropsychological test that evaluates several cognitive functions including visuospatial abilities, memory, attention, planning, working memory and executive functions [2, 72–74]. The ROCF is characterized as a complex cognitive task [75], and is known in the field of neuropsychology as a useful metric for the frontal lobe function [76] and to possibly detect conditions such as Alzherimer's [77, 78] and schizophrenia [79]. A participant is asked to copy the figure into a piece of paper, then copy it again two more times from memory. The shape is specifically designed to be abstract so that participants cannot associate it with any common object or concept. A clinician then grades all three sketches on whether 18 separate sub-shapes (henceforth called "details") exist and, if they do, how neatly they were drawn. A clinician grants up to 2 points for each detail that totals to 36 points, with partial credit given to distorted or misplaced shapes. Points for overall neatness of individual details is subjective and is generally dependant on an expert's intuition, especially for shapes that exist but might be drawn poorly. This results in different ROCF graders potentially producing two different scores. The proliferation of digital sketch recognition techniques and a push to digitize clinical neuropsychological examinations motivated our creation of an automated ROCF that can grade itself on the existing grading scheme. The ROCF test is not only used to detect cognitive decline but is also used to gauge executive function across age populations from adults [80–82] to children [83–85], to study conditions from learning disabilities [86], eating disorders [87], and decline due to age [88] among many others. The ROCF's sensitivity to various participant's cognitive and behavioral states have

thus necessitated developing various methods to grade the test.

From a digital sketch recognition standpoint, the task of grading an ROCF is non-trivial due to the complexity of the figure and test conditions resulting in inherently fuzzy sketch data. Assessing participant performance and how it might inform their behavior has been an ongoing topic of research [89, 90], with proposals for new grading rubrics being compared against each other [91–93]. When grading the ROCF tests, no two completed sketches are drawn in the same order, and very frequently shapes are drawn using portions from other shapes [94]. Bottom-up approaches tend to classify shapes as soon as their constraints are met, but shapes in an ROCF may in fact be only part of a detail or may end up as a portion of an entirely different one. A top-down approach not only more closely resembles a human grading an ROCF, but it also simplifies the recognition process by not needing to re-classify a shape at every step of the hierarchical recognition process.

3.1.5 Uses in Clinical Neuropsychology

Neuropsychologists employ a variety of cognitive examinations through which to observe a patient's behavior. Some of these involve manipulating physical objects, answering a series of oral questions, solving exercises that integrate drawing exercises. These tests typically involve paper and pencil, and the patient is required to complete an examination by sketching, drawing on objects, crossing out or circling shapes, and other forms of interaction using pen on paper.

The TMT was conceived and originally introduced as a testing tool in the field of clinical neuropsychology to assess general intelligence. Shortly thereafter, clinicians found utility in the test's ability to aid in the detection of cognitive abnormalities. Clinicians who administer the Trail-Making test observe the patient's behavior according to a predetermined checklist. Among the items include sitting posture, how well they maintain eye contact, reaction to mistakes, and their efforts in remembering the next item in the sequence among others [2]. Tests produce one quantitative performance metric in the form of the time taken to complete the test, rounded to the nearest second. A clinician compares the score against established normative data depending on which age classification the patient belongs to. The lengthy diagnosis process combined with the increased emphasis for early detection due to the climbing proliferation of Alzheimer's disease and

other types of dementia[1] has maintained the TMT's relevance in modern diagnoses.

3.1.6 Uses in Other Fields

One of the landmark traits of the TMT is in its reliable utility as a general tool to test cognitive function. It has been shown to be sensitive to a wide variety of differing behaviors. It is used in sports to assess the extent of effects from mild head injuries [95]. These tests have shown to be an improvement over a subjective report of mild traumatic brain injury symptoms following injuries in sports [96]. The TMT has been shown to be sensitive to both age and education among healthy participants in various populations such as Japan [97], Brazil [98], and Portgual [99], exemplifying its versatility across cultures and geographic regions. It is used to gauge the effects of drug addiction such as cocaine and alcohol on cognitive functions among participants [100]. It is also sensitive to sleep deprivation such as when studying sleep apnea/hypopnea [101]. The TMT also is used in the military, with extensive studies being done to gauge the effect of PTSD on combat veterans [102]. In short, the TMT is sensitive to changes in behavior based on brain injuries, traumatic events, sleep deprivation, age, education, drug addiction, cultural background, and geographic location. This has allowed the TMT's expansion of utility across a wide range of purposes over the decades of its existence.

3.2 Digital Sketch Recognition

3.2.1 Sketch Recognition Systems

Digital sketch recognition techniques favor bottom-up approaches that employ computational geometry to classify shapes [103–107] or gestures [108–111]. Hierarchical sketch recognition systems such as LADDER [112–127], Sketchread [128], Cheminx [129] and Mechanix [130–139, 139–144] generate composite figures by re-classifying shapes into more complex shapes in every step of the sketching process. In early bottom-up sketching approaches "steps" were typically separated by a UI button that explicitly instructed the system to create a recognition step. More modern systems, however, automatically separate "steps" by single-stroke actions and usually

---

[1]Centers for Disease Control and Prevention: www.cdc.gov/features/alzheimers-disease-deaths/index.html

triggered when the user lifts their pen. This allows the system to continuously check the sketch to see whether the user is drawing a composite sketch made up of shape primitives.

Popular applications for bottom-up recognition of composite shapes using geometric primitives is especially popular in digital recognition of hand-drawn diagrams [145–151], tutoring systems [152–166], artistic drawings requiring technical skill [167–172], distinguishing text from images [173], teaching and recognizing handwritten characters of various languages including Chinese [174–179], Japanese [180,181], Urdu [182], and Arabic characters [183,184], handwritten mathematical formula recognition [185–187], military course-of-action diagrams [188, 189] and drone interaction [190, 191], music [192, 193], document notation recognition [147, 194, 195], and converting offline images into digital stroke data [196, 197]. For hand-drawn diagram recognition, researchers seek to digitize hand-drawn flowchart and system design diagrams, interpreting diagram structure, flow of information, and preservation of variable and state checks through digital sketch recognition techniques [148, 149, 198–202]. Circles, rectangles, diamonds, rhombi, and directional arrows [106] are used in diagrams to denote specific system or algorithm states or commands [150, 151, 203]. Indeed, these projects originally served as the basis of *Auto ReyO*'s recognition due to the emphasis in recognizing primitives as part of a larger composite system of shapes. However, a chief difference between these projects and an ROCF sketch is that the ROCF by design has a large number of overlapping shapes, and specific details can be as granular as a single line within a specific area of other shapes. Diagrams and flowcharts, by contrast, are required to have clear spacing between its components and recognizing missing or distorted shapes is not a focus of these automated systems. While some form of composite figure recognition is necessary for automatically grading the ROCF, a top-down approach as explored in other systems [204] proved ultimately the most viable for *Auto ReyO*.

Corner detection [205–207] also helps characterize digital shapes, with lightweight systems such as ShortStraw [208] and iStraw [209] being among the most efficient. *Auto ReyO* uses the open-source ShortStraw library in its recognition of corners and endpoints to generate the vertices that will be used in the graph creation as explained in 5.1.3. This is used in tandem with line-

intersection algorithms to segment the sketch lines such that individual shapes can be recognized. A frequent use case of this is recognizing details 4 and 6 of the ROCF (see Fig. 5.1). A user typically draws a single long line at once across the ROCF shape, so we are unable to use individual stroke order to recognize details, but rather need the segmentation that a line-intersection algorithm combined with ShortStraw is able to provide.

Hierarchical sketch recognition approaches generally check drawn lines to see if they meet requirements for a composite shape [210, 211]. However, this creates a need to check for every possible shape component at every step of the sketching process. This becomes prohibitive for large, complex line drawings such as the Rey-Osterrieth complex figure, especially if there is no guarantee that every line is present. Although important foundational work has been established, no one system has presented fully automated grading of ROCFs that also identifies absent and distorted details.

3.2.2 Cognition in Digital Sketch Recognition

One of the prevalent methods of digital sketch recognition is through the analysis of sketches as "gestures" comprising of geometric properties of sketches. This includes but is not limited to line length, speed, acceleration, line straightness, and various trigonometric properties of line strokes. Individual features were calculated in early efforts from Rubine *et al.* [103,212], and later expanded by others [105,213–219]. Digitial sketch recognition initially leveraged machine learning to afford developers tools to recognize simple geometric shapes. Shape recognition expanded to alphabets, scaffolded recognition to identify components of complex composite shapes, and entire sketches. Machine learning algorithms have allowed these analyses to be made feasible over a large corpus, resulting in models that are able to distinguish between objects depending on subtle changes in sketching behavior.

An increasingly common application of digital sketch recognition does not identify the shapes drawn, but rather characteristics of those who draw them. In addition to identifying writers based on their drawing style [220, 221], Kim *et al.* identified strong correlations between sketching behavior and early cognitive development in infants [222–226], with Paulson *et al.* also exploring

16

the relationship between digital sketching and children's cognitive development [227]. Davis *et al.* [228] and Muller *et al.* [229] similarly has focused on cognitive decline by analyzing sketches from Clock-Drawing tests [230]. Zham *et al.* identified the presence of Parkison's disease through the way a participant drew spirals with a smart-pen [228]. Digital variants on existing neuropsychological tests are numerous, with various proposed systems designed for test automation, diagnosis assistance, or self-administration [231–234]

3.3 Intersecting Neuropsychology with Human-Computer Interaction

3.3.1 Template Matching Shape Classification Systems

The "Dollar" family of recognition systems [235–239] remains among the most well known single and multi-stroke gesture classification algorithms, and serve as the basis for our own template-matching recognition algorithm presented as part of our system. While most techniques rely on stroke order, geometric properties, and physical characteristics such as speed, acceleration, etc., the "$P+" recognizer calculates similarity via "point cloud" approximation [240]. A point cloud is generated by resampling both a template shape and an input shape on the same resampling parameters, overlaying the input shape on top of the template sketch matching its shape, centering, and orientation as close as possible, and iterating through every point finding the closest match between template points and input points. The distance between the points that are closest together are added cumulatively and are presented as the overall "distance" metric between the template shape and the input shape. The "$P+" recognizer returns the closest template match, identifying what kind of shape the user has drawn. This is especially flexible when the application in question necessitates recognition that is agnostic to stroke order. Our technique for shape recognition as described in Section 5.1.5 is based on the "$P+" recognizer, particularly the technique of calculating a "distance."

Our technique differs, however, in that rather than calculating distance via point-for-point comparison, we generate a fixed-resolution matrix of point density for both the template and the input shapes and calculate distance between cells of both matrices. This allows us to generate a

17

more accurate grader for shape neatness. Indeed, "$P+" only focuses on finding the *closest match* to a template since it is a shape classifier, but its internal distance metric value does not perform well to gauge whether an input shape is poorly drawn next to its provided "ideal" template shape.

3.3.2 Hierarchical Sketch Recognition

Hierarchical sketch recognition approaches generally check drawn lines to see if they meet requirements for a composite shape [210, 211]. Layered hierarchical systems for graph creation have been applied to both bottom-up and top-down systems [204], and involve the decomposition of a drawn sketch to specific broad categories by analyzing sub-graphs [241–243]. This is typically used in the fields of computer vision to help decompose a system to primitive parts and represent them as a tiered graph. We envisioned a similar hierarchical tiered approach to the recognition of an ROCF due to the nature of the drawn details. To draw detail 10 in an ROCF, for example, the user needs to have drawn both details 2 and 3 to be able to connect the line properly (see Figure 5.1). Similarly, detail 14 requires the existence of detail 13 to receive full marks for both correct placement and shape neatness. Rather than represent the entirety of the sub-shape as a single vertex in a graph, however, we envisioned the vertices of a graph being represented by intersecting lines and endpoints, and applied the concepts behind sub-graph composite object recognition to identify the ROCF details themselves. The cited foundational work on graph implementations to supplement computer vision and object recognition informed our own approach to automatically grade ROCFs using a graph itself as the vehicle for tiered object recognition.

3.3.3 Efforts to Automate Neuropsychological Examination Analysis

Since interactions with digital interfaces is often impacted by a decline in a user's cognition from handwriting tasks [244–247] to tasks involving physical movement [248–252] and wayfinding [253–255], an active area of research is developing automated tools that can track decline from a digital system's usage metrics [256–264] while taking into account the non-equivalence of computerized and paper-based tests [265, 266]. Tracking systems also have been developed with the intent to aid in post-diagnosis therapy in an information-rich, unobtrusive manner [267–

273] Efforts to automate other neuropsychological tests has renewed interest in sketch sub-object detection [274–279]. Object recognition ranges across various neuropsychological examinations including clocks [280, 281] and general handwriting tasks [282–286]. However, whereas recognized objects for these tests tend to have heavily distinct characteristics, ROCF details are mostly composed of simple primitives that appear frequently. For example, detail 5 shown in Figure 5.1 is defined not only as any vertical line, but rather a specific vertical line within the sketch. Work presented by Prange *et al.* [287] cites Rey-Osterrieth figures as a motivating factor in the need to identify geometric shapes inside complex abstract figures. Existing attempts to automatically grade ROCFs are semi-automated or do not implement detection of all 18 details [94, 288]. The most recent attempt automates grading using a deep-learning neural network but leaves ample room for improvement of individual segment detection, most notably single-line details [289]. Additionally, our system is able to produce a recognizer from only five training sketches to serve as templates, whereas neural networks require exponentially higher amounts of training data to function properly.

3.3.4   Challenges of Assessing TMT Participants

The TMT's sole quantitative measure is the participant's completion time, rounded to the nearest second [31]. While this facilitates analysis of participant performance, it also discards large amounts of context-sensitive data that may yield insight into the participants' state of mind. Whether the administrator is attempting to identify brain injury, mild cognitive impairment, cultural or education differences, PTSD or inebriation, much of the assessment is left to subjective evaluation. The only quantitative comparison that can be made is to whether their score is within the range of "normal" scores according to established normative data. As mentioned above, the test's primary use in clinical neuropsychology still relies on external factors such as sitting posture and reactions to mistakes to gauge a patient's cognitive state, resulting in a lengthy testing process.

The TMT also does not test a participant's perception of objects in a 3-dimensional space and any cognitive functions related to 3-dimensions such as image recognition of rotated objects, object permanence, and depth perception. Since the TMT's introduction in the 1940s, newer research

19

into human cognition has more recently identified general perception of 3-D space as an important indicator in a participant's cognitive state [290].

3.3.5 Leveraging Technology for Existing Cognitive Tests

Researchers have previously used segmentation and recognition techniques of drawn shapes for the purposes of assessing cognitive testing performance. Souillard-Mandar et. al developed machine learning algorithms to detect behavioral abnormalities on a digital version of the Clock Drawing Test [230]. Moetesum et. al uses distorted shape recognition to help automatically grade the widely used Bender Gestalt Test (BGT), a drawing test similar in style as the TMT [275, 291], but does not share the TMT's same versatility or ubiquity. Nazar et. al advanced the computerized grading process of the BGT by employing convolutional neural networks [292]. Similar techniques have been employed for the Rey-Osterrieth Complex Figure Test, which consists of copying a complex figure consisting of various simple geometric shapes arranged together. Canham et. al performed some early work in identifying only some individual shapes of the larger complex figure [94], while more recent work from the same author further developed recognition of structures from the naturally distorted Rey-Osterrieth figure drawn by a human [288].

These works establish the concept of leveraging modern machine learning techniques to recognize behavior from neuropsychological exams, but have not worked with the TMT previously. We have previously been able to perform these sketch recognition technologies on a digital version of the TMT to correctly classify a participant's age based on their digital pen input [293].

4. RESEARCH QUESTIONS

4.1 R1. Can we create a predictive algorithm for estimating cognitive decline using digital sketch recognition features?

H1. A combination of existing general-purpose sketch recognition features in combination with novel features aimed at capturing participant performance while completing a digital Trail-Making Test can help create a predictive algorithm that can predict cognitive decline in two ways:

- A binary classification between a participant with Mild Cognitive Impairment (MCI), and a participant without it

- A continuous classification that can predict a participant's MoCA scores depending on their performance of the TMT within 5 points.

4.2 R2. How closely can an automated grader for a complex figure examination mirror a human grader for ROCFTs?

H2. We expect an automated grader for the inherently subjective task of ROCFT grading to perform in two aspects:

- A recognition F1-score of greater than 0.8 for the majority (at least 14 of the 18) of ROCFT details when gauging whether the automted grader can **recognize the existence** of individual drawn details.

- An average difference of no greater than 5 points for the entire test's score between expert graders and our automated grader

4.3 R3. How well can we differentiate between participants of different nationalities from a novel cognitive examination?

H3. We expect a novel neuropsychological examination based on the TMT to be sensitive to nationality, much like the original TMT. In conducting a binary classification when analyzing

movement metrics between American and Japanese participants, we expect a significantly higher than random-chance performance (greater than 0.75) F1-score for the classification between nationality of said participants.

5. AUTOMATICALLY GRADING SKETCH-BASED NEUROPSYCHOLOGICAL EXAMINATIONS

The Rey-Osterrieth Complex Figure Test (ROCF) is among the most widely used neuropsychological examinations to analyze visual spatial constructional ability and memory skills, but grading the patient's sketched complex figure is subjective in nature and can be time consuming. With increasing demand for tools to help detect cognitive decline, there is a need to leverage sketch recognition research to assist in detecting fine details within an ROCF's inherently abstract figure. We present a series of recognition algorithms to detect all 18 official ROCF details using a top-down sub-shape recognition approach. This automated grader transforms a sketch into an undirected graph, identifies and isolates detail sub-shapes, and validates sub-shape neatness via a point-density matrix template matcher. Experimental results from hand-drawn ROCFs confirm that our approach can automatically grade ROCF Tests on the same 18-item sketch detail checklist used by neuropsychologists with marginal error margin.

5.1 Auto Rey-O: A System For Automated Grading For Rey-Osterrieth Complex Figure Tests

This chapter contains research work currently in submission.*

5.1.1 Contribution

Significant research has been produced in analyzing the reliability of current rubrics [93, 294–296]. Automating the process started over two decades ago [94], but even recently surveys have cited a lack of contributions towards grading all 18 details at once. Moetsum *et al.* in research published in 2020 [297] specifies that "due to the unconstrained nature, these drawings, localization and segmentation of individual scoring sections become a highly challenging task" and existing work localizes only a "small subset of ROCF scoring sections."

---

*Auto Rey-O: A System For Automated Grading For Rey-Osterrieth Complex Figure Tests. Raniero Lara-Garduno, Benton Guess, Thomas Merner, Zengxiaoran Kang, Nancy Leslie, and Tracy Hammond. Review in Progress
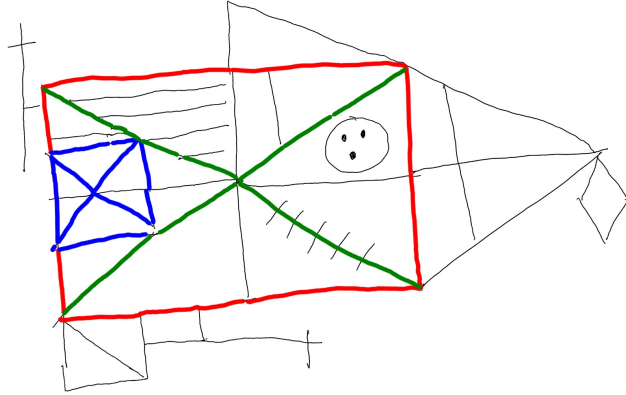
Figure 5.1: Our automated grader highlighting Details 2, 3, and 6 in red, green, and blue respectively.

Whereas previous efforts in automatically grading the ROCF can identify only a subset of the complex figure's details, we present the first fully automated ROCF grader that does not require user input to point to baseline shapes from which to begin recognition. Our contribution widely expands on Field's truss recognition technique [130] by introducing several graph traversal algorithms in order to isolate specific sub-shapes or regions from a given sketch. In addition to triangles, we also recognize squares, parallel lines, crosses, straight horizontal and vertical lines, and diamonds as well as shapes specific to the ROCF such as detail 6 (Cross with Square), detail 14 (Circle and 3 Dots), and 18 (Square with Line). Our system uses a multi-step recognition process that can identify shapes whether by crawling the resulting graph, by using template-matching shape recognition, or a combination of both, resulting in a more accurate and robust sub-shape recognition system for ROCF grading. Many of our recognition algorithms utilize well-known graph traversal [298] and optimization algorithms (such as Dijkstra's Shortest Path [299–301] and Depth-First Search [302]). Our system represents the first fully-automated ROCF grader that recognizes the existence or absence of each of the 18 details and checks individual shapes for distortion.

To test our recognizer's performance, the system graded 141 digitized Rey-Osterrieth tests from participants and we compare how closely our system's grades correlate with those of two expert

graders. The experimental results demonstrates the proposed approach is successful in identifying the existence of sub-shapes within a large abstract shape.

We wish to emphasize the purpose of this automation is to re-create the grading process to similar to a hand-grader to the best of this technology's ability. It is not intended to automatically detect MCI or similar conditions on its own. This is an important distinction due to the nature of the examinations, since normative data points to correlations between exam performance and possible signs of decline, but require a more holistic analysis of a patient's behavior and cognition to draw a formal diagnosis.

Additionally, the top-down recognizer does not intend to counteract existing physiological conditions that might exist in patients. For example, poor eyesight, hand tremors due to trauma or diseases such as Parkinson's can have a significant effect on the patient performance of this test in that the produced shapes might be heavily distorted. In these cases the distortion may not be due to cognitive decline, and in those cases the Rey-Osterrieth Complex Figure's grading would not be able to distinguish it apart from partial credit due to a correct but distorted shape. While it might seem feasible to have an automated system recognize and work around it, our intent with this work is to automate an existing grading scheme rather than significant adapt it. Future work might be able to identify the difference between distorted shapes due to physiological conditions versus those as a result of cognitive decline.

5.1.2 Generalizability of Recognition Technique

An important consideration of novel recognition and automation techniques in sketch recognition lies in articulating the generalizability and defining the constraints under which a presented technique aims to perform well.

Automating the ROCF motivates a brief discussion on generalizability due to the inherently "hard-coded" nature of its automation. Indeed, the complexity of the ROCF shape coupled with the requirements of detecting very specific lines necessitates a certain specificity of location and shape composition requirements. Some details, for example, are a single horizontal or vertical line, but of most importance is the location of the line relative to other details and the starting and
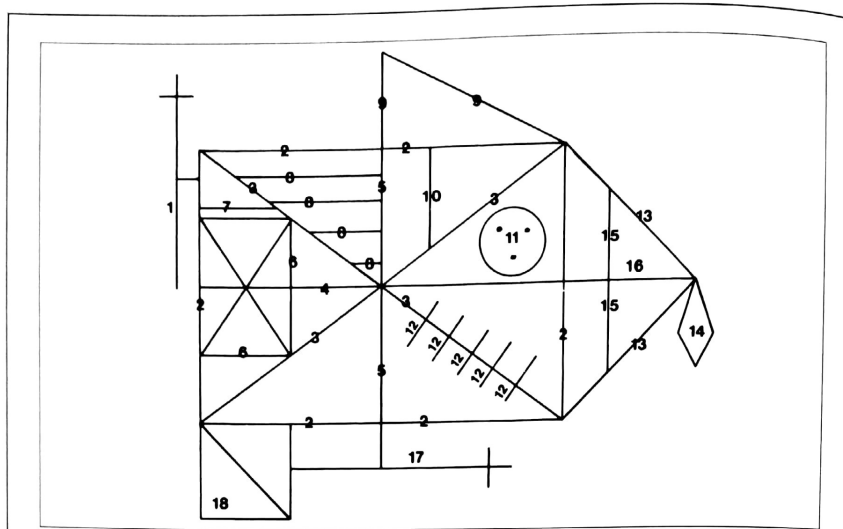
REY-OSTERRIETH COMPLEX FIGURE TEST
FORM A (Rey Figure)

Details:	COPY	DELAY
1. Cross upper left corner, outside of rectangle	_____	_____
2. Large rectangle	_____	_____
3. Diagonal cross	_____	_____
4. Horizontal midline of 2	_____	_____
5. Vertical midline	_____	_____
6. Small rectangle, within 2 to the left	_____	_____
7. Small segment above 6	_____	_____
8. Four parallel lines within 2, upper left	_____	_____
9. Triangle above 2 upper right	_____	_____
10. Small vertical line within 2, below 9	_____	_____
11. Circle with three dots within 2	_____	_____
12. Five parallel lines with 2 crossing 3, lower right	_____	_____
13. Sides of triangle attached to 2 on right	_____	_____
14. Diamond attached to 13	_____	_____
15. Vertical line within triangle 13 parallel to right vertical of 2	_____	_____
16. Horizontal line within 13, continuing 4 to right	_____	_____
17. Cross attached to low center	_____	_____
18. Square attached to 2, lower left	_____	_____
TOTAL SCORE	_____	_____

Scoring:
Consider each of the eighteen units separately, and appraise accuracy of each unit and relative position within the whole of the design. For each unit count as follows:

Correct	{ placed properly	2 points
	placed poorly	1 point
Distorted or incomplete but recognizable	{ placed properly	1 point
	placed poorly	½ point
Absent or not recognizable		0 points
Maximum		36 points

Figure 5.2: Grading rubric for the Rey-Osterrieth Complex Figure Test. Reprinted from *A Compendium of Neuropsychological Tests* [2]
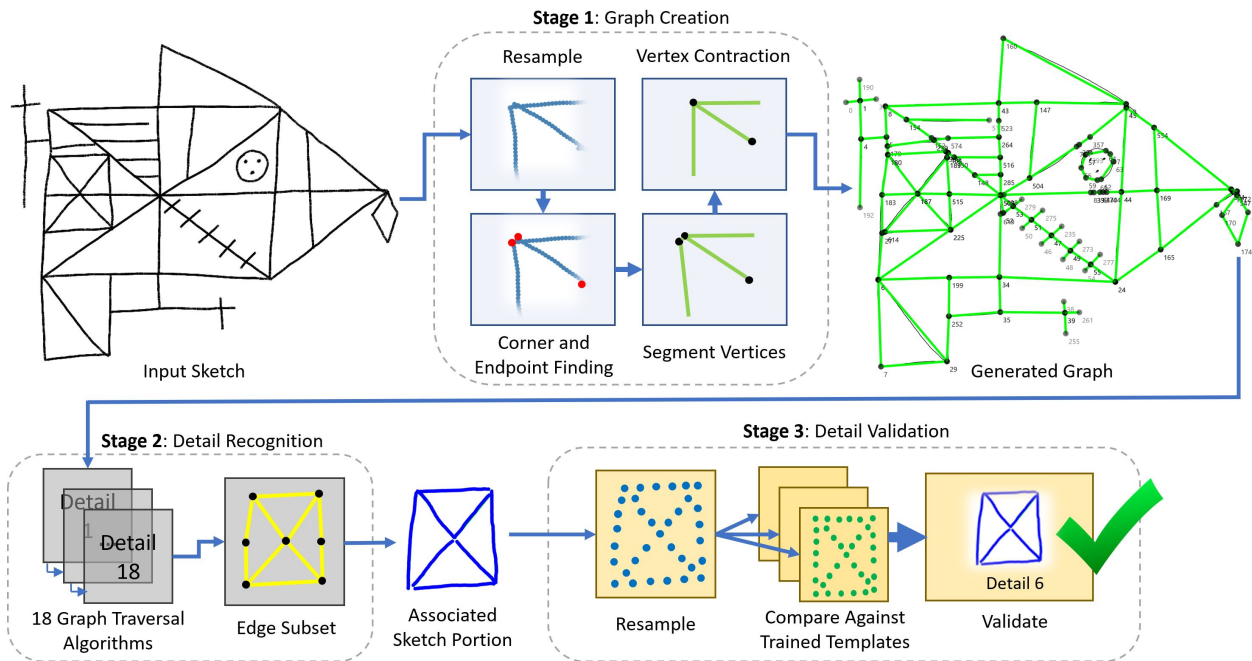
Figure 5.3: Description of ROCF sub-shape recognition system. Stages 2 and 3 shown in the figure are repeated for each of the 18 details of a Rey-Osterrieth complex figure.
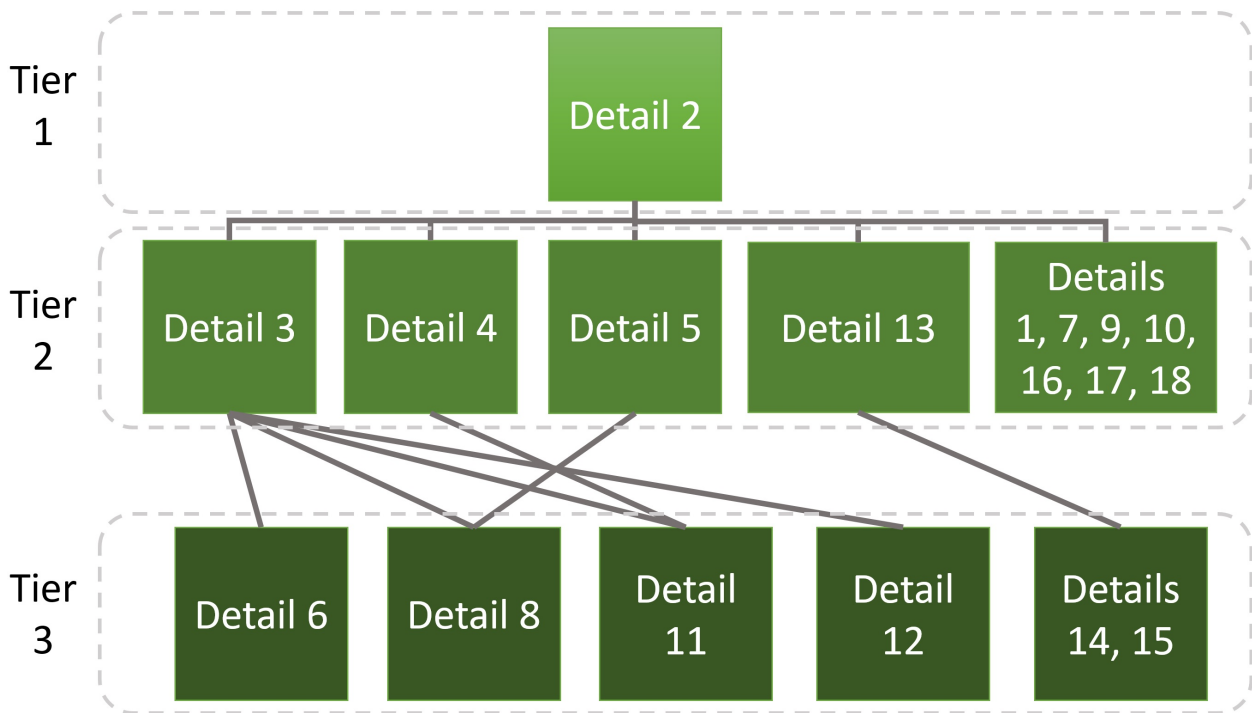
Figure 5.4: Auto ReyO's recognition hierarchy, designed to have as few dependencies as possible.
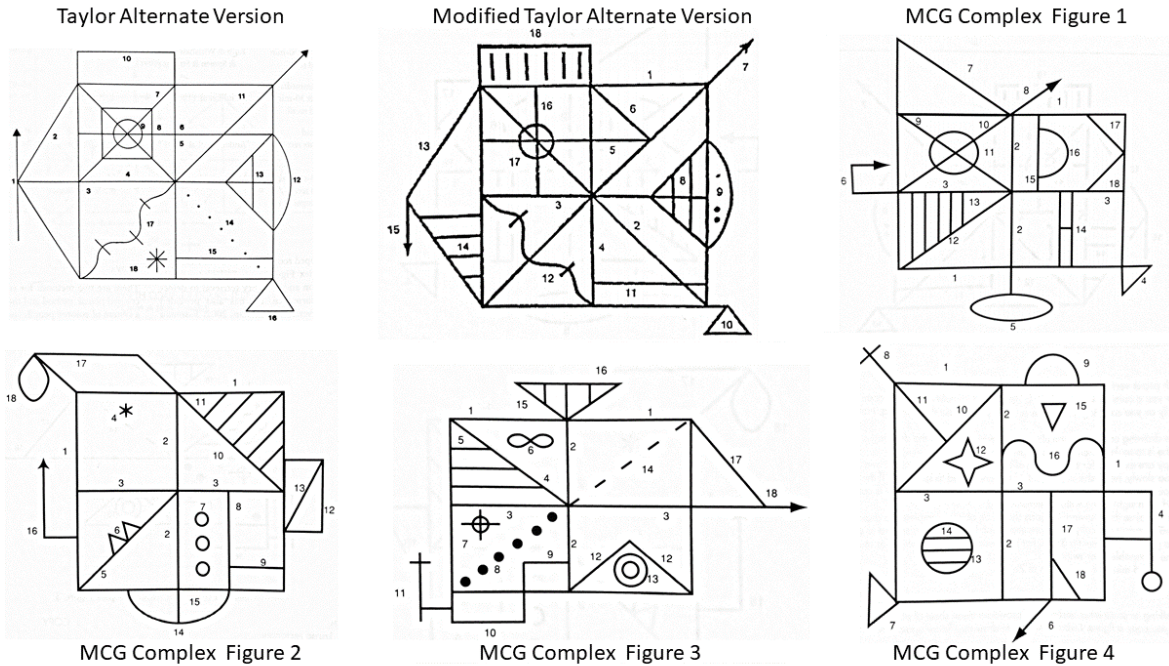
Figure 5.5: The six other complex figure tests. Our 3-stage technique can also be applicable for complex-figure tests, provided the exact heuristics are implemented for the details. Reprinted from *A Compendium of Neuropsychological Tests* [2]

stopping points. It is, in fact, this specificity in requirements that allows our method to recognize all 18 details, opposing previous work that only detects a subset of them.

At the same time, however, generalizability was taken into account when designing the recognizers that will be described in the following section. Generalizability was considered for two primary reasons:

- **Our set of recognizers is intended to work with imperfect hand-drawn ROCFs**. Our algorithm must be general enough to recognize details despite a varying list of imperfections including but not limited to: crooked lines, shapes not entirely closed, various lines intersecting at different points, sharp angles accidentally being curved, the same line being drawn over several times, etc. The algorithm must also be able to, within reason, identify as many shapes as possible even in the absence of other shapes. Unless the shapes are directly dependent on each other for recognition, the absence or heavy distortion of one unrelated

detail should not prevent the recognition of the other. These design considerations were aligned with existing research for digital applications that involve sketching modalities [303].

- **As many recognition techniques as possible should be easily adaptable for other complex figure tests**. As per the Compendium of Neuropsychological Tests [2], seven complex figures are recognized as valid and tested figures for this purpose, and the Rey-Osterrieth Complex Figure test is the most popular. New variants with small changes are uncommon. The remaining six figures are: Taylor Alternate Version, Modified Taylor Complex Figure, and four Medical College of Georgia Complex Figures [304]. All have similar size, complexity, and are a combination of straight lines, triangles, and simple geometric shapes. All contain a "detail 2": a large rectangle that serves as an anchor for the rest of the shapes. Our system was designed to be adaptable to recognize the 18 details of the remaining six complex figure tests by applying variations on the pathfinding algorithms on Table 1. Our three-stage method detailed in Fig. 3 can be adapted for all six remaining complex figure tests, so that extent we consider this approach generalizable for other complex figure tests of this type. Location heuristics need to be tailored for each detail, since the rules themselves are inherently specific and unique to the ROCF. We believe our three-step approach can be usable for any hierarchical sketch recognition problem involving complex figures where multiple sub-shapes must be discretely recognized but may share any number of lines.

To this end we intend to highlight the balance that we aimed to strike between a set of shape recognition heuristics specific enough to separate detail shapes from the more complex ROCF, while generalizable enough to provide effective shape recognition techniques for other complex shapes and hand-drawn imperfections of the ROCF. Section 5.1.4.1 distills the recognition techniques into four distinct categories that should be applicable for the other listed complex shapes.

5.1.3 Stage 1: Graph Creation

The graph creation stage is divided into four distinct steps. First, we prepare the sketch for corner detection by resampling the sketch points of the entirety sketch to a uniform interspace
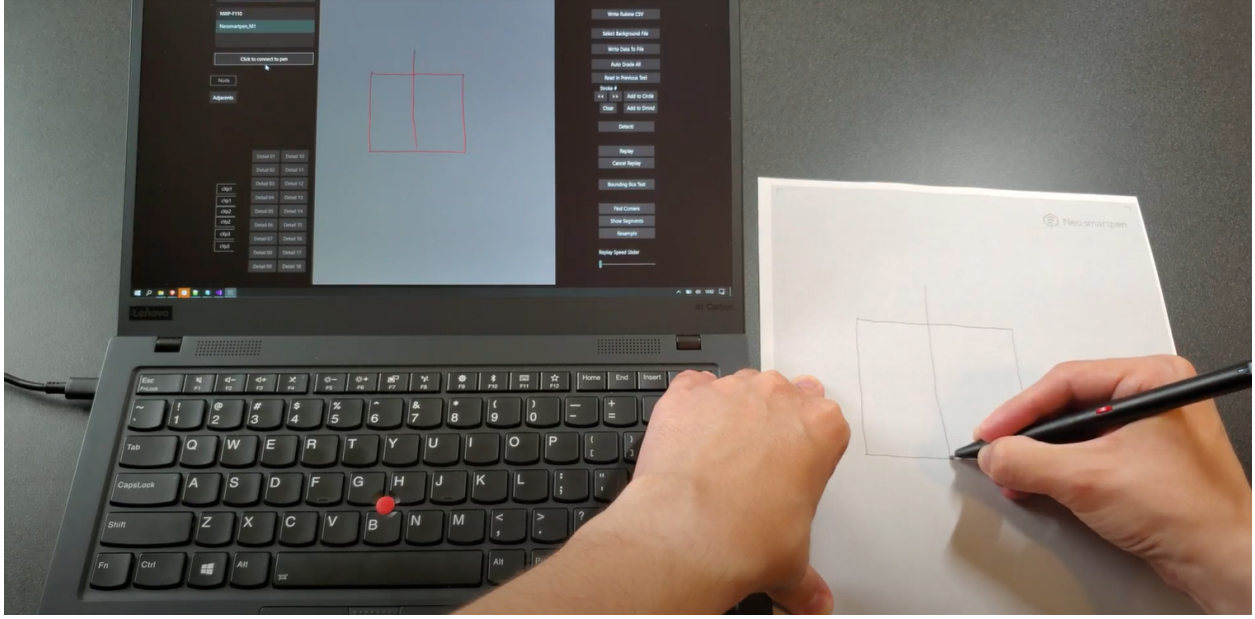
Figure 5.6: *Auto ReyO* in use, digitizing the ROCF test in real-time for saving and processing. The same app can be used to load existing ROCFs and automatically grade the tests with our recognizer.

length S as follows:

$$S = \frac{\sqrt{(x_m - x_n)^2 + (y_m - y_n)^2}}{c} \tag{5.1}$$

where (x_m, y_m) is the lower-right corner of the sketch, (x_n, y_n) is the upper-left corner of the sketch, and c is a constant $c = 40$. The value of c was empirically determined to produce enough granularity of the resulting sketch to preserve detail while being spaced far enough such that the following step of corner finding is able to perform properly.

The second step utilizes the corner-finding algorithm from Wolin *et al.* [208] to identify any "corner" from drawn strokes. To detect line intersections, two straight-line segments are compared with the target segment $y_a = a_2x + a_1$ checked for intersection against comparison segment $y_b = b_2x + b_1$ with equation 5.2.

$$\frac{a_1 + b_1}{a_2 + b_2} \in \left(x_1 - \frac{(0.15l)^2}{1 + a_2^2}, x_n + \frac{(0.15l)^2}{1 + a_2^2} \right) \tag{5.2}$$

where x_1 and x_n represent the x values of the less and greater vertices of the target segment

respectively with l being its segment length. Comparing every line with every other line in this way will return a series of points where the lines themselves intersect. This along with the detected corners from the algorithm of Wolin *et al.* will serve as the vertices of the undirected graph, which is constructed in the following step.

The third step creates undirected graph G, where every vertex v is a line endpoint, corner, or intersection, and every edge e is a drawn line connecting each v. Each v contains a point from the sketch, and each e contains the sampled points that connect the two vertices.

The fourth and final step performs vertex contraction on the created graph. Each vertex is iterated over and checked for near vertices that fall below a predetermined distance threshold. If two vertices are joined, their respective sampled points s_i including points from an edge that might fall between them are combined into a single vertex v containing all the sampled points. The distance measure is determined through complete distance used in hierarchical clustering, taking the maximum of the set in equation 5.3.

$$\left\{ \, ||s_i - s_j||_2 \mid s_i \in v_1, s_j \in v_2 \, \right\} \tag{5.3}$$

This serves both to connect segmented or near vertices and to reduce the overall complexity of the graph by eliminating edges. Finally, the vertices are iterated over a second time, checking if near vertices fall below a distance threshold, where the distance is determined through taking the minimum of the set in equation 5.3 referred to as single distance used in hierarchical clustering. If the distance falls below a predetermined threshold, the points are then are linked through an edge.

5.1.4 Stage 2: Detail Recognition

All 18 ROCF details are recognized by applying a graph traversal algorithm to identify a "shape" within the graph. Each detail has an associated algorithm that is called in the hierarchical order defined by Figure 5.4. Every algorithm is designed to accommodate inherent graph imperfections from both the graph creation stage and the participant's hand sketch. For example, if the algorithm checks for a horizontal edge, there we allow a slope between 0.3 and -0.3 since
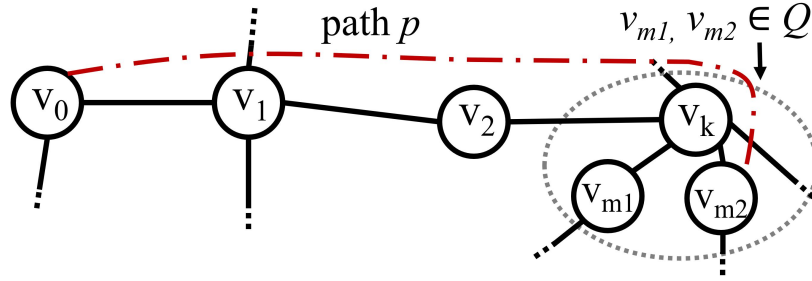
Figure 5.7: Finding path p for the top horizontal side of detail 2's rectangle. Dotted area on right indicates $dist$ radius. In this example $v_{m2} = c_1$, $dir = Right$ and $nextdir = Down$ (See Algorithms 1 and 2). This also serves as an example for any agent-based graph crawling recognizer.

participants are not expected to produce a perfectly horizontal line. The algorithms were designed to strike a balance between leniency to accommodate the imperfect nature of a hand-drawn shape and precision to find the expected shape if it exists.

5.1.4.1 High-Level Recognition Categories

Although the composition of the ROCF necessitates specific recognition requirements for each detail, all of our recognition approaches can be categorized in four separate techniques:

- **Agent-based graph crawling.** This technique places an agent in a specified vertex and instructs the agent to move in a specific direction, changing directions when it finds a corner or can no longer move in said direction (see Figure 5.7). This technique saves every edge and vertex that the agent crawls through, constructing the specified shape along the way. A recognizer using this technique would have parameters to construct the specific detail, with enough leniency to recognize hand-drawn shapes that are not geometrically perfect. This technique is the primary method used for details 2, 4, 5, 7, 13, 15, 16, and 18.

- **Bounded depth-first search.** This technique is employed for details where the graph created in Stage 1 is expected to vary too widely for the previous technique to be effective for hand-drawn ROCFs. It uses a slightly modified version of the common depth-first searches found in tree and graph traversal [305, 306], and it is intended to crawl through and return all
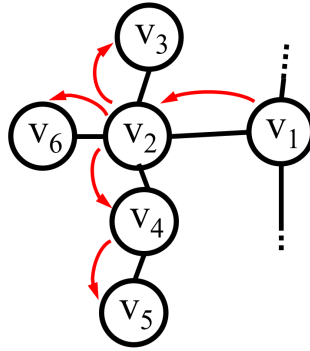
Figure 5.8: An example of a bounded depth-first search recognition algorithm. It defines an initial direction to begin DFS, then crawls finding any vertex and edge not part of a predetermined outer bound.

connected vertices and edges inside a bounded area of a graph (see Figure 5.8). For this technique we bound the area by labeling adjacent details that have already been identified, indicating that the depth-first search should go no further if it encounters edges or vertices associated with other shapes. If certain edges are meant to be shared (such as the left-hand side of detail 2 and the left-hand side of detial 6), the edges are added through a second-pass technique, usually a short version of agent-based graph crawling or Dijkstra's shortest-path. Bounded depth-first search is the primary technique used for recognizing details 1, 6, 17

- **Dijkstra's shortest-path for diagonal detection.** [307] Due to the imperfect nature of hand-drawn ROCFs, diagonal lines across shapes that connect several details would be difficult to identify via graph-crawling. This technique leverages the inherent geometric property that diagonals inside a rectangular shape are the shortest path between the rectangle's corners (see Figure 5.9). This technique is used twice, in details 3 and 18, where it is most applicable for diagonal lines embedded within rectangular shapes.

- **Area-based recognition**. This technique is used in rare instances where the detail in question is expected to vary too widely to accurately identify the shape via the previous two recognition techniques. This technique constructs an area where the detail is expected to be drawn in, and all vertices and edges situated inside this area are deferred to the Stage 3

33
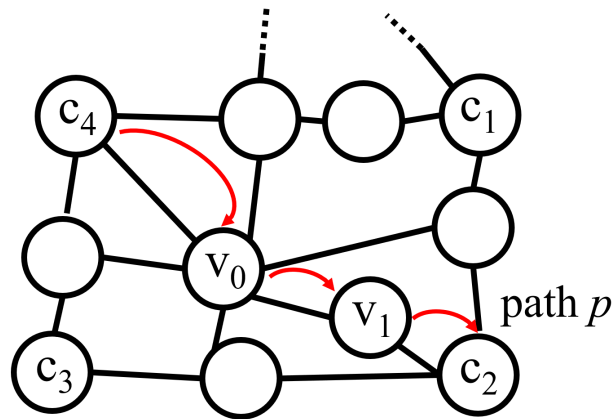
Figure 5.9: Dijkstra's shortest-path algorithm is used to find diagonals, since geometrically they are the shortest path between corners such as $c4$ and $c2$.
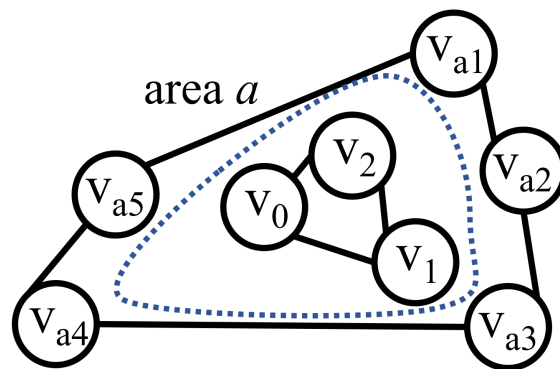
Figure 5.10: Area-based recognition defines a space a bounded by, for example, v_a1 to v_a5, constructs a polygon comprising these vertices, and returns all vertices such as v_0 to v_2 inside the bounds of said polygon area a.

template matcher to identify whether the sketch inside this area is that of the specified detail (see Figure 5.10). As a result, this offloads all recognition to the Stage 3 template matcher. This technique is only used for two details: 11 and 12.

The recognizers were implemented with the intent to be able to recognize a drawn detail that matches the spirit of the description as detailed in The Compendium of Neuropsychological Tests [2]. In order to effectively communicate the comparison between the written description of the recognizer with our internal implementation we will provide a quote for each from the Com-

pendium along with a description of how each of the recognizers were designed and implemented. Detail 2's description varies slightly, as we describe it as primarily a graph-crawling algorithm. For detail 2 we provide pseudo-code as an aid due to the complexity of the recognizer and to provide a clearer picture of the code implementation. The rest of the details 17 detail recognizers in the paper are described at a higher level and more briefly out of space considerations.

5.1.4.2 Detail 2's Recognizer

"The large rectangle. The horizontal dimensions of the rectangle must not be greater than twice the vertical dimensions of the rectangle, nor must the rectangle resemble a square. Because there are so many possibilities of distorting the rectangle and it is not possible to score for position, a score of 1/2 point is given if the rectangle is incomplete or distorted in any way"
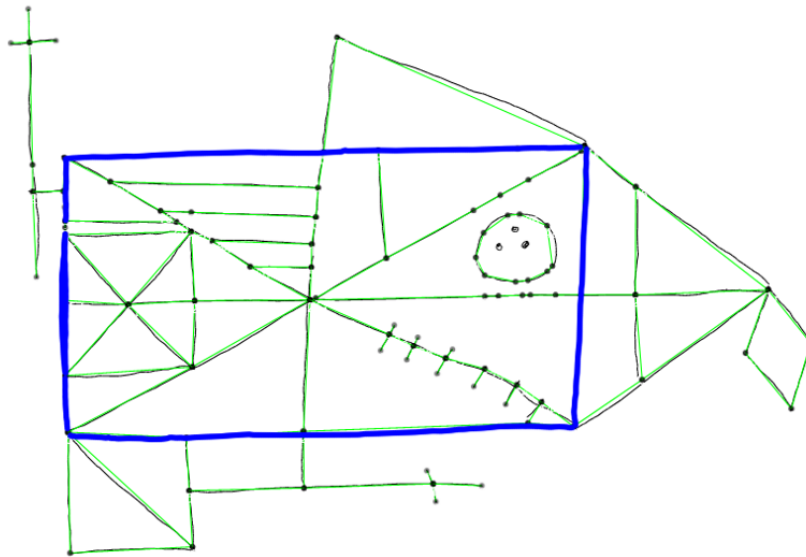
Figure 5.11: Detail 2, as detected by our recognizer.

Detail 2's recognition algorithm defined in Algorithm 1 and 2 is a greedy graph traversal algorithm tasked with finding the large rectangle that serves as the anchor for all other shapes in the

graph. Our pathfinder finds one rectangle side b and the corner at its end c at a time. For each side dir is the intended path direction, and $dirnext$ is the next path direction once we find our corner. Detail 2's recognizer consists of two main algorithms, algorithm 1 nested inside 2, to automatically determine the largest rectangle of the input sketch. The fuzziness and unpredictability of the input data necessitates that the algorithm consider every vertex as a possible candidate for the top-left corner of the detail 2 rectangle. Algorithm 1 then finds the four sides and corresponding corners and cralws as far as it can along every side. This ensures that not only will we receive the largest rectangle we can recognize, it also ensures that the starting vertex v_0, which automatically serves as corner c_0, is also corner c_4 at the end of crawling. If $c_0 = c_4$, it a closed shape and a potential candidate for detail 2. The algorithm then produces the largest such rectangle as detail 2, since the rectangle of detail 6 and some erroneous drawings of detail 18 can also return closed rectangles through algorithms 1 and 2.

Algorithm 1 SideAndCorner

Input: Start node n, directions dir, $dirnext$
Output: Side d of rectangle, corner c

1: push n to stack s
2: **while** s not empty **do**
3:     **pop** s to p, mark as visited
4:     **for all** adjacent pairs (p_1,p_2) of p **do**
5:         **if** direction of $(p_1, p_2) = dir$ **or** distance e (p_1,p_2)<25 **then**
6:             **push** p_2 to s
7:             p=p_2, repeat from line 5
8:         **end if**
9:         **if** dead end p_n reached **then**
10:            **add** shortest path as a stack from p to p_n to set c
11:         **end if**
12:     **end for**
13: **end while**
14: **for all** current longest path a in c **do**
15:     **for all** adjacency pairs (p_{a1},p_{a2}) of leaf p_{an} in a **do**
16:         **if** p_{a2} is $dirnext$ of p_{a1} **then**
17:             **return** $c = p_{a1}$, $d = a$
18:         **else**
19:             **pop** p_{an} from a
20:             **repeat** from line 15
21:         **end if**
22:     **end for**
23: **end for**

Algorithm 2 Detail 2: Largest Rectangle

Input: Sketch Graph's vertex adjacency list
Output: Largest rectangle vertices

1: **for all** node n in Graph **do**
2:　　corner c_1=SideAndCorner(n,$Right$,$Down$)
3:　　corner c_2=SideAndCorner(c_1,$Down$,$Left$)
4:　　corner c_3=SideAndCorner(c_2,$Left$,Up)
5:　　corner c_4=SideAndCorner(c_3,Up,$Right$)
6:　　**if** $n = c4$ **then**
7:　　　　add all SideAndCorner b sides to rectangle q
8:　　**end if**
9: **end for**
10: **return** largest q

We define N as single agents where every agent $n \in N$ has a start location $s_n \in G$ and a goal location $g_n \in G$. The path p of n consists of one side b of our rectangle where $g_n = c$. Path p is of length k that is a sequence of vertices $p = \{v_0, v_1, v_2, ..., v_k\}$ such that each consecutive vertex is either in a defined direction dir (up, down, left, right, or diagonals) or Eucledian distance $r < 25$. At the end of our sequence v_k one of two **conditions** is true:

1. v_k is connected to a vertex v_x such that direction of $(v_k, v_x) = dirnext$

2. v_k is connected to *other* vertices v_m such that r of $(vk, vx) < 25$.

The conditions describe that we have either (1) found a corner characterized by the start of the next side of the rectangle, or (2) the end of our sequence consists of various vertices very close together. This creates a set of "dead-end" nodes Q. For condition 1, $v_k \in Q$ and $c = v_k$. For any v_m that satisfies condition 2, $\{v_{m1}, v_{m2}, ..., v_{mn}\} \in Q$. We check all vertices in Q and return the vertex v_e that satisfies condition 1. The final sequence is $p = \{v_0, v_1, v_2, ..., v_e\}$, and $c = v_e$. This is repeated four times to find the four sides of our rectangle, and return the largest such rectangle as detail 2.

"The cross at the upper left corner, outside of the rectangle. The cross must come
down to the horizontal midline of the rectangle and must extend above the rectangle.
The line that joins the cross to the rectangle must be approximately in the middle of
the cross and must come between Detail 7 and the top of the rectangle."
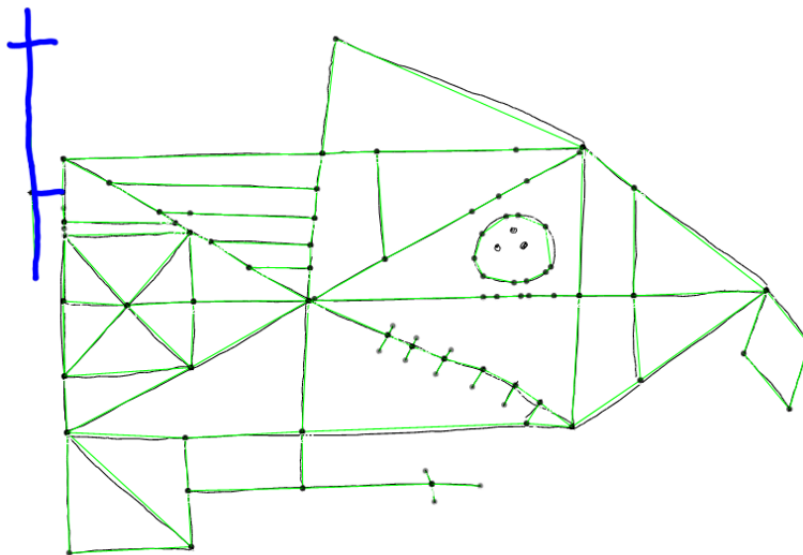
Figure 5.12: Detail 1, as detected by our recognizer.

Once detail 2 is recognized, the recognizer is able to use its sides and location and an anchor through which to recognize every other detail. From this point forth we follow the same order as presented by the Compendium's grading rubric.

Detail 1 uses a two-step combination of defining edge borders to isolate an area, then run a depth-first-search algorithm to identify all vertices and edges within said area. This method proves effective for several details that have two conditions: 1) the area in question is relatively straightfoward to isolate under normal drawing conditions, and 2) the sketch inside the area is

expected to vary widely enough that makes a more heuristics-based pathfinding recognizer used in Stage 2 to be inadequate.

When designing heuristics for a potential pathfinding algorithm for recognizing a cross, in practice we found that the Stage 1 graph generator would sometimes alter the edge and vertex composition to such a degree that it would require a high amount of alternate recognizers for crosses even if the sketch itself resembled a cross to a human eye. In these cases a template matcher would be considerably stronger in recognizing a cross and, more importantly, the proportions of the cross line lengths as described in the Compendium hand-graded rubric.

Detail 1 is recognized by taking the left-hand side of the detail 2 rectangle and its vertices as the border limit, then employing DFS to isolate all sketch data to the immediate left of the complex figure. The recognizer starts by iterating through the left-hand side of the detail 1 rectangle starting from the top and moving toward the bottom, then employs a directed DFS algorithm aimed toward the leftward direction when it detects that any of the iterated vertices have an edge to its left.

"The diagonal cross must touch each of the four corners of the rectangle and must intersect in the middle of the rectangle"
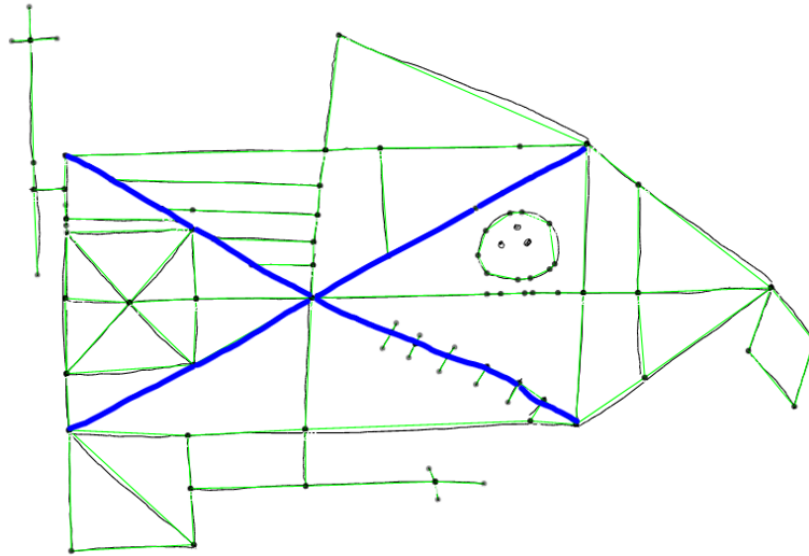
Figure 5.13: Detail 3, as detected by our recognizer.

Despite the wide variability of vertex composition and the tendency for participants to draw the diagonals in non-straight trajectories resulting in relatively unpredictable graphs, the algorithm for identifying detail 3 was straightfoward. The four corners of detail 2 are identified by extracting the last vertex of each returned rectangle side from detail 2 which will return the top-right corner, bottom-right corner, bottom-left corner and bottom-right corner in that order. These vertices are then saved and identified by their location. An implementation of Dijkstra's shortest-path algorithm [299] was employed twice, identifying shortest possible connecting path between the corners, where the starting and ending vertices were the opposing diagonal corners from detail 2. This cross of two diagonals intersecting in the middle are identified as detail 3 and passed to the template matching algorithm for identification.

We found this shortest-path algorithm the most effective at identifying diagonals within a rectangle given the geometric properties of said diagonals; by definition these diagonals constitute the shortest path between opposing corners of a rectangle. If the diagonals are incomplete, such as when a participant does not completely connect the diagonal lines to the corners, or when the lines themselves are incomplete, the template matcher would find a significant discrepancy between templates provided for the detail 3 diagonal cross and the sub-sketch returned from this recognizer.

Thus the advantage of this three-step process becomes emphasized. In some cases, such as detail 2, the recognizer itself is able to mostly verify the integrity of the shape within the recognition algorithm by itself. By contrast, some algorithms such as detail 3, or detail 1, the recognizer merely identifies and isolates a sub-shape where it expects the detail to exist if it is drawn correctly, and then passes whatever is returned to the Stage 3 template matcher. This allows us to create recognizers that are not just relatively flexible, it also keeps the recognizer algorithms feasible by not needing to account for every possible alternate graph composition from the provided sketches. Should those details be drawn too poorly, the Stage 3 template matcher would identify that and deduct points due to a distorted shape.

> *"The horizontal midline of the rectangle must go clearly across from the midpoint of the left side of the rectangle to the midpoint of the right side of the rectangle in one unbroken line"*
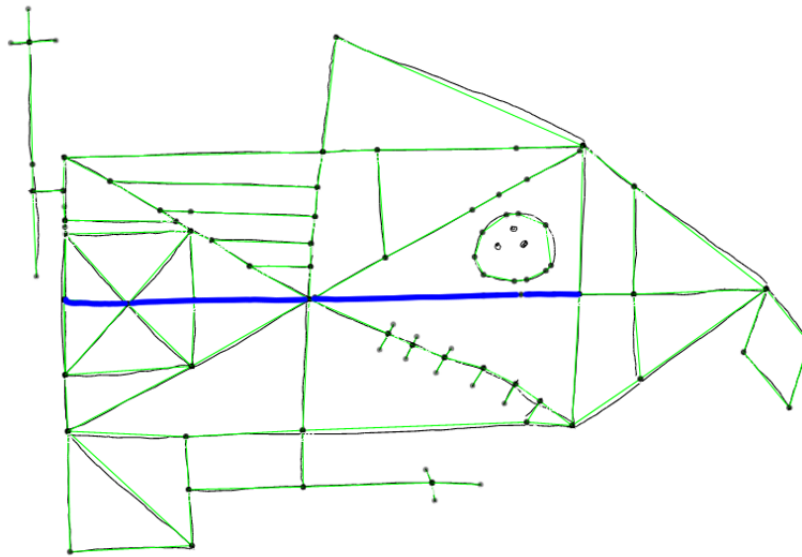
Figure 5.14: Detail 4, as detected by our recognizer.

The recognition of detail 4 exemplifies the sketch recognition problem of recognizing a simple shape in a relatively complex location. While recognizing a single horizontal line is trivial in digital sketch recognition, the fact that we seek a specific horizontal line embedded within several lines inside a complex figure is the challenge of this automated ROCF grader, and our solution shows why we believe a graph-traversal based approach is straightforward in concept and execution.

Detail 4 is a horizontal line situated across the width of detail 2. The key is differentiating this line, which runs along the middle of the rectangle of detail 2, from the horizontal lines of the rectangle of detail 2 itself. We exclude the horizontal lines of detail 2 by dropping the last and

the first nodes of the left and right sides of the detail 2 rectangle from consideration in the detail 4 recognizer.

With the first and last nodes of each side excluded from consideration, the agent's starting position index is $l - 1/2$, the middle of the path of the left-hand side of detail 2. A uni-directional DFS, the same implementation used for detail 2, is deployed to find a horizontal line from one side of the rectangle to the other. For the detail 4 line to be valid, the ending vertex must be part of the right-hand side of Detail 2, meaning the horizontal line must run the entire width of detail 2.

If the uni-directional DFS line is not found at the starting position of $l - 1/2$, the agent moves outward in increasing distance from the center with each new check. We introduce a counter i that increases in absolute distance of 1 with each new iteration. If no line is found in any of the iterations, the detail 4 recognizer will return "not found" and award no points.

"The vertical midline must start at the midpoint of the bottom of the rectangle and go through in one unbroken line to the midpoint at the top of the rectangle. In scoring of details 4, 5, and 6, they should intersect at the midpoint of the rectangle. Usually, if they do not , only one of them is scored as incorrect for position. Very seldom are all three scored incorrect for not being in position"
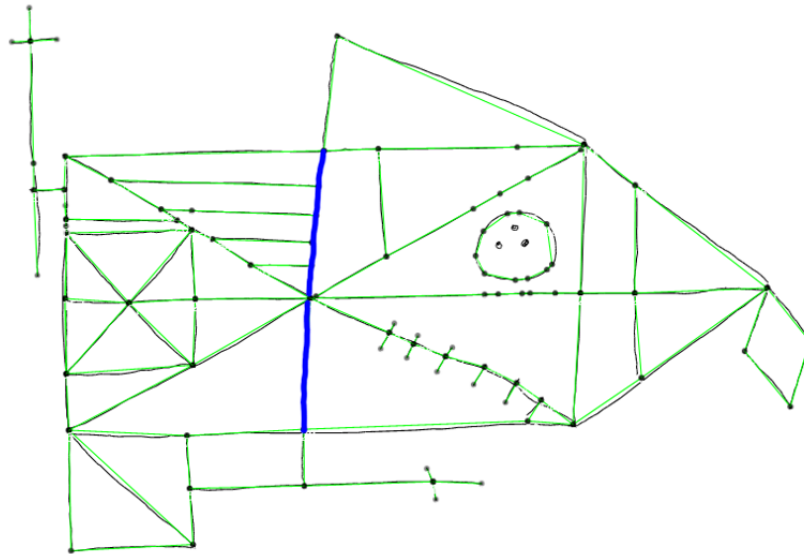
Figure 5.15: Detail 5, as detected by our recognizer.

Detail 5 is a very similar implementation to that of detail 4, with the exception that the starting and ending sides are the top and bottom sides of detail 2 and the uni-directional algorithm seeks to recognize a vertical line from top to bottom. The starting and ending points of the top side of detail 2 are still excluded, the starting position of the agent is still the middle of the top line of detail 2, and the agent still moves outwardly away from the center on both sides from the center in similar fashion as the detail 4 recognizer. Therefore, the implementation of detail 5's recognizer is largely the same with the exception of direction.

5.1.4.7   Detail 6's Recognizer

"The small rectangle within the large rectangle and to the left side of it. The boundaries of Detail 6 are defined by the top of the rectangle falling between lines 2 and 3 of the parallel lines that make up Detail 8, and the width of the small rectangle must be approximately one-quarter of the width of the large rectangle; that is, it should come to the midpoint between the left side of the large rectangle and the vertical midpoint of the rectangle. The cross within Detail 6 must come from the four corners of the rectangle and should intersect at the mid-point of the rectangle (i.e., words intersecting on Detail 4)"
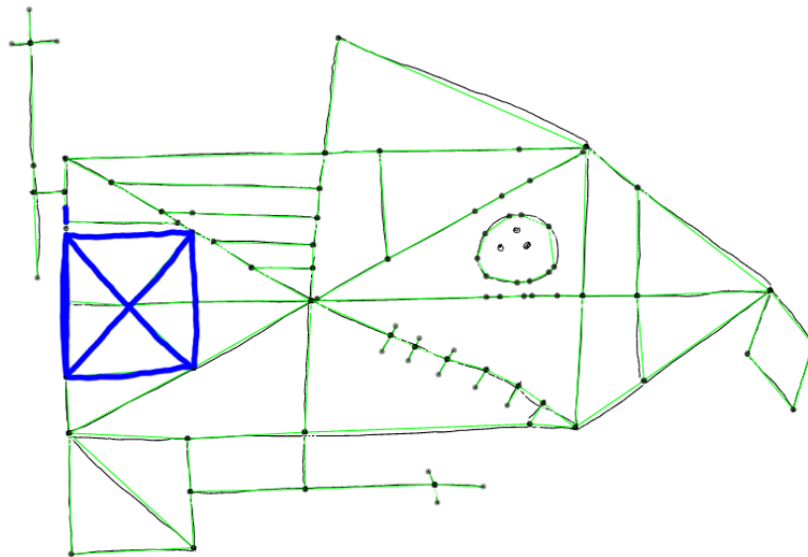
Figure 5.16: Detail 6, as detected by our recognizer.

Detail 6 recognizer has the most robust behavior of all features in this automated grader. An eye-level observation of the composition of detail 6 in a ROCF test allows one to intuit why; the shape consists of both a square shape as well as a cross inside it, and is embedded between the left-hand side of detail 2, and shares the remaining two corners with two lines from the cross of

detail 3. Additionally, detail 4's line runs through the middle of detail 6 but is actually excluded from detail 6 itself. All of these characteristics need to be taken into account in the recognizer's code, with enough leniency to still recognize the shape when it is drawn less than ideally.

The first step is in identifying the bounds of the shape. As previously described, the bounds consist of the left-hand side of the detail 2 rectangle and the two lines that form the cross of detail 3. Additionally, the vertex and edge sets for details 7 and 4 are saved only for the purposes of exclusion so that they are not recognized as part of detail 6.

Once these bound vertices are established, several omni-directional DFS algorithms are deployed. The reason why several DFS algorithms must be deployed is that the bounded vertices forces the algorithm to stop the search in that direction to prevent the recognizer from continuing past the vertex bounds. As such, the recognizer might fail to find every line of detail 6 if only one omni-direcitonal bounded DFS algorithm is used. An agent is placed on the left-hand side of detail 2 and the first and last vertices are excluded from consideration (see detail 4 for the reasoning as to why). The agent then iterates through each vertex and conducts an omni-directional DFS algorithm with a starting direction toward the right. This type of DFS differs from the uni-directional DFS algorithm described in previous details in that it is meant to search every direction but its very first search is aimed in a specific direction to help focus the search in a particular area. Otherwise, a true omni-directional DFS will jump beyond the bound vertices at the beginning of the search.

The final step is in excluding all the edges and vertices that are part of detail 6 and 7, and including the edges and vertices that are part of detail 2 but connect the top-left and bottom-left corners of detail 6. This needs to be done separately at the end of the process since the left hand side is used beforehand to construct the boundaries of the DFS searches.

As with the other DFS-based recognition approaches, the detail 6 recognizer uses stage 3 to verify the integrity of the drawn shape before awarding points. This shape in particular, with the challenges in dissecting the specific edges that part of detail 6 since not every line inside the rectangle is part of it, is one of the details that benefits the most from the Stage 3 recognition pass.

47

5.1.4.8 Detail 7's Recognizer

"The straight line above Detail 6 must be shorter than the horizontal aspect of Detail

6 and must fall between the top of Detail 6 and the second line of Detail 8"
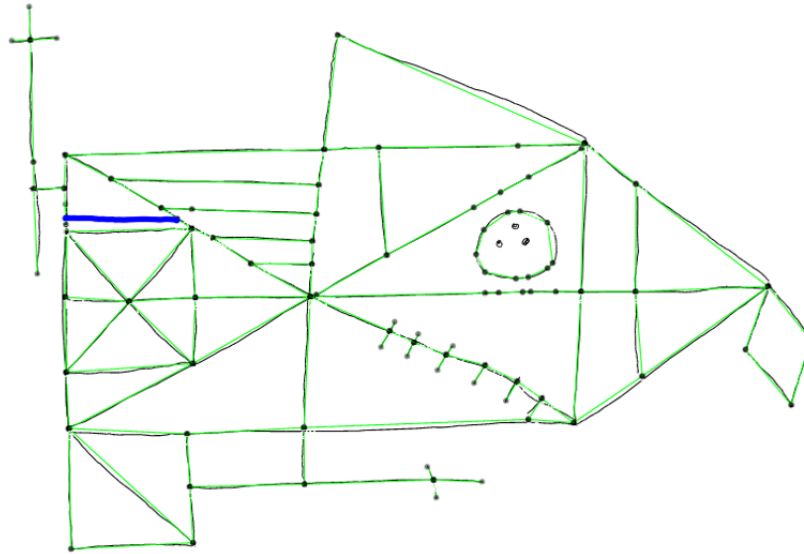
Figure 5.17: Detail 7, as detected by our recognizer.

Detail 7, along with detail 10, is among the most straightforward recognizers since they are both single lines with a common vertex with detail 2. In the case of detail 7, the line is horizontal and is situated just above detail 6. The algorithm places an agent along the left-hand side of detail 2, excluding the first and last vertices to avoid detecting the top and bottom lines of detail 2 as the horizontal line. The detail 4 edges and corresponding vertices are also excluded from consideration, since they are also horizontal lines and would otherwise be confused with detail 7 if the participant does not draw detail 7 in their complex figure.

The detail 7 recognizer returns a single edge and pair of vertices corresponding to the line just above detail 6 if it exists. The algorithm makes a single direction comparison per inner vertex v of the left-hand side of detail 2.

> *"The four parallel lines within the rectangle in the upper left corner should be parallel*
>
> *with the spaces between them approximately equal. If the lines are unduly slanted or,*
>
> *of course, if there are more or less than four of them, then the scoring is penalized."*
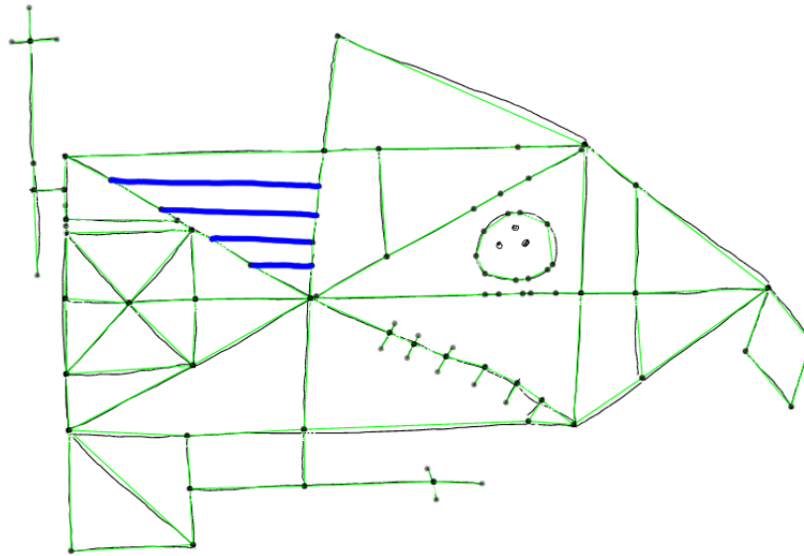
Figure 5.18: Detail 8, as detected by our recognizer.

The detail 8 recognizer also uses the Stage 3 template matching before awarding points, since it is the most robust way to check for number of lines and parallel line straightness. Detail 8 is contingent on the existence of detail 5 for the parallel lines to be attached to the right vertical line and detail 4 for the lines to be attached the diagonal cross of detail 4. We are not aware of a widespread amount of instances where detail 8 exists in a drawn ROCF but without the existence of a detail 4 and 5. For that reason we consider details 4 and 5 as pre-requisites in our hierarchy for recognizing detail 8.

The algorithm places two agents, one to crawl along the vertical line of detail 4 and one along the diagonal line of detail 3. The crawling for both agents iterates from the top of the detail 2

rectangle, excluding the vertices along the top line of detail 2. The crawlers stop at the intersections of details 3 and 5, since we do not expect parallel lines belonging to detail 8 to exist on the right half of detail 2, or the bottom half of detail 2.

The agent that moves along the vertical line of 4 checks for single lines to the left of the detail 4 line. The agent that moves along the diagonal line of 3 checks for single lines to its right. The recognizer checks for both directions since the stage 1 graph creator often inserts a vertex along the parallel lines of detail 8, and thus we hope to account for those instances.

"The triangle above the rectangle on the upper right, with the height less than the base"
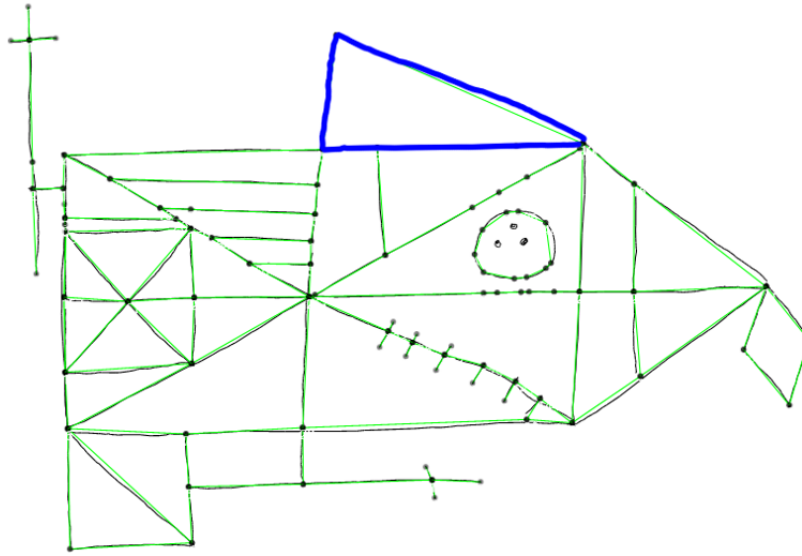
Figure 5.19: Detail 9 as detected by our recognizer.

Detail 9 is among the first details that were previously recognized in past efforts by Canham *et al.* [94] due to the simplicity of the shape along with the fact that it is isolated at the top of detail 2. In a similar fashion our recognizer is relatively straightforward and focuses on identifying the vertical and diagonal lines that make up the upper two lines of the right triangle.

Our detail 9 recognizer places an agent along the top of detail 2. The agent searches for a vertical line moving upward from the top of detail 2. When it finds one the agent moves to the top vertex of this line and then searches for a slanted line in the south-east direction. If it finds such a slanted line that intersects back into detail 2 then it has completed the closed right triangle and thus we have detected a potential detail 9 triangle. The algorithm performs two comparisons for each vertex (one for the vertical line and one for the slanted line).
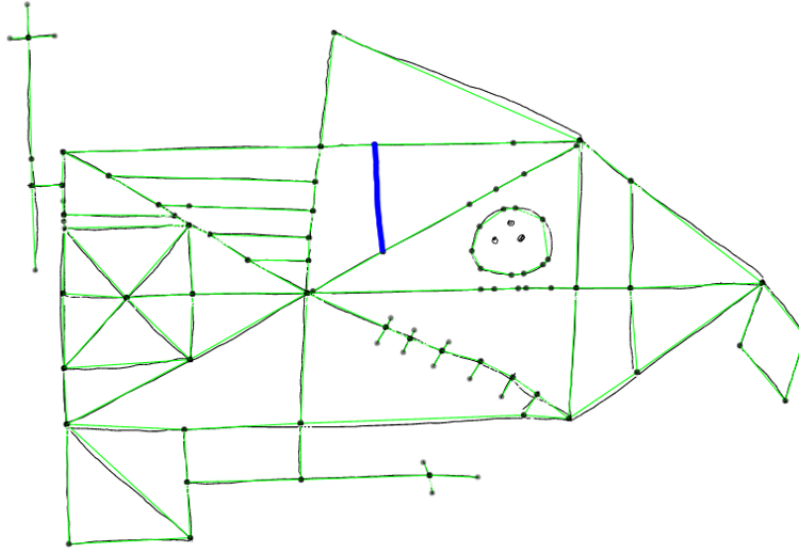
Figure 5.20: Detail 10, as detected by our recognizer.

5.1.4.11   Detail 10's Recognizer

"The small vertical line within the rectangle just below Detail 9. The line should be

clearly shifted to the left within the upper right quadrangle in the rectangle"

This is the vertical counterpart to detail 7, in that it is a single line pointing inwards from an outer edge of the detail 2 rectangle. Similar to detail 7, we wish to exclude recognition of other unrelated vertical lines such as the vertical lines of detail 2 on either side, as well as the line of detail 5.

This recognizer places a single agent along the top line of detail 2 and excludes the first and last vertices of the top line. It seeks the first downward line it finds that is not already part of detail 5. This recognizer performs single comparisons with each iterated vertex.

"The circle with three dots in the lower right half of the upper right quadrangle. It must not touch any of the three sides of the triangular area in which it is placed, and the positioning of the dots should be such that there are two above and one below, so that it resembles a face"
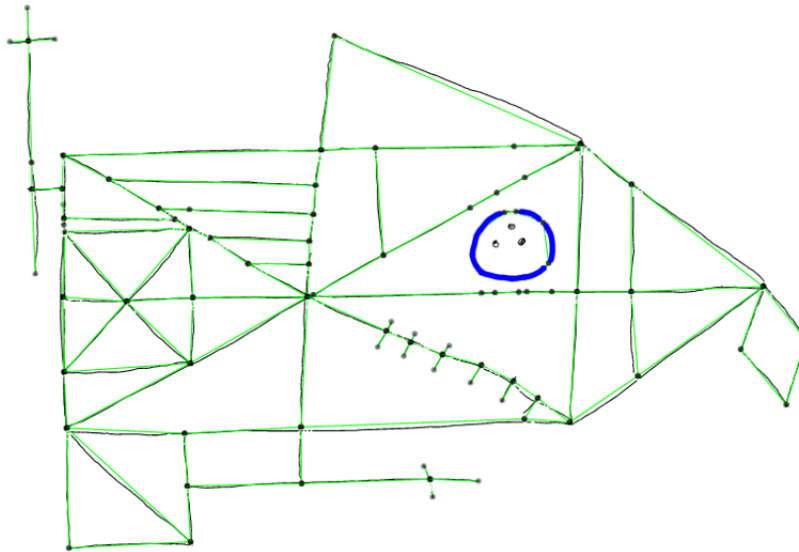
Figure 5.21: Detail 11, as detected by our recognizer.

Detail 11 was one of the primary reasons for the recognition technique of returning all vertices and edges within a certain location and having the stage 3 template matcher verify the integrity of the shape. It is important to note that even though stage 1 constructs straight-line edges and approximates corners into vertices, it associates the vertices and edges to the actual hand-drawn sketch. Therefore, we can use the vertices and edges to quickly identify an area of interest and return the actual sketch data to the stage 3 template matcher easily.

Detail 11 employs an area-based boundary that differs from vertex/edge boundaries found for BFS-based recognizers. This area-based boundary constructs a polygon from a set of vertices to

serve as an outer boundary, then returns every vertex and edge inside that area.

The boundary is defined by the right half of the detail 4 horizontal line, the upper-right half of the detail 3 cross, and the upper half of the detail 2 right side. This boundary information is then fed into a function that generates an n-sided polygon constructed from the set of vertices that make up the boundary.

This helper function, named "CreatePolygon," saves the coordinates of the generated polygon to identify which nodes in the graph are inside. To determine whether a vertex is inside of the generated polygon, we employ a horizontal line test, a technique frequently used to determine whether a point is inside of an n-sided polygon. In a horizontal line test, we use the y-coordinate of the point we want to check and a theoretical "infinite" horizontal line to the right of the he x-axis. We then how many times such horizontal line intersects with the sides of the n-sided polygon. If the number of times the horizontal line, which extends to a distance of infinity to the right of the vertex's coordinates, intersects an odd number of times, it means the coordinates lie inside the n-sided polygon since the horizontal line "entered" the polygon but did not leave. Conversely, an even number of intersections indicates the the horizontal line entered the polygon as many times as it left, meaning the coordinates are outside of the n-sided polygon. This process is repeated to find every vertex that lies inside the n-sided polygon, and by returning the edges associated with these vertices we are able to determine the shapes that are inside of the area we define in the n-sided polygon.

This is used to identify every vertex and edge that is situated inside the triangular area as seen in Fig. 5.21. The ROCF guidelines specify only the circle and three smaller dots inside said circle are expected to be found in this area. We not only determine that only this shape is inside the area, but we also determine the neatness of the shape of detail 11 using the template matching algorithm in Stage 3.

"The five parallel lines that are crossing the lower right aspect of Detail 3 must all be within the lower right quadrangle. They must not touch any sides of the quadrangle, and they should be approximately equidistant from one another"
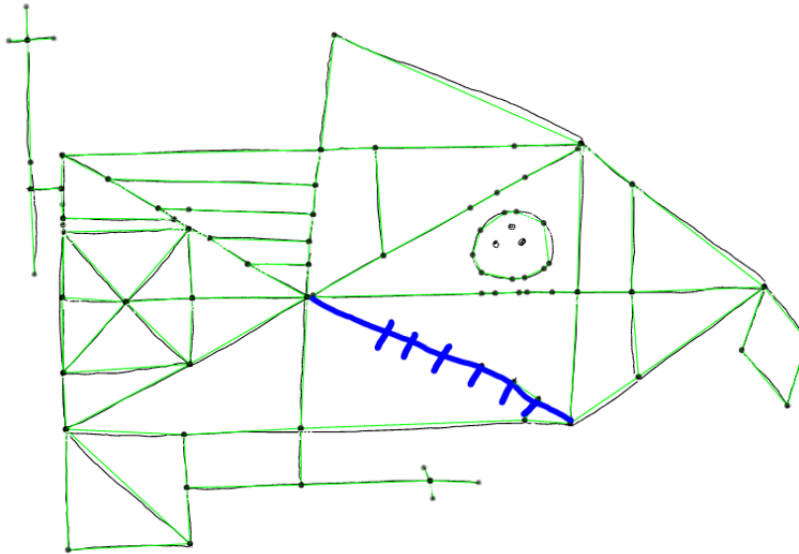
Figure 5.22: Detail 12, as detected by our recognizer.

The technique to recognize detail 12 is very similar to that of detail 11, with the main change being the area of comparison. In this case, because the diagonal line is itself a portion of the detail itself, we use a larger area of comparison.

The n-sided polygon consists of the area bounded by the lower portion of the right-hand side of the detail 2 rectangle, the right portion of the lower side of the detail 2 rectangle, the right-hand side of detail 4, and the bottom half of detail 5. Every line vertex and edge that is inside this defined area is considered detail 12 and checked against the template in Stage 3 to verify that this shape was drawn correctly.

"The triangle on the right end of the large rectangle. The height of the triangle must not be greater than half of the horizontal mid-line of the rectangle and, as already mentioned, the slope of the sides of the triangle must not be a continuation of the slope of Detail 9."
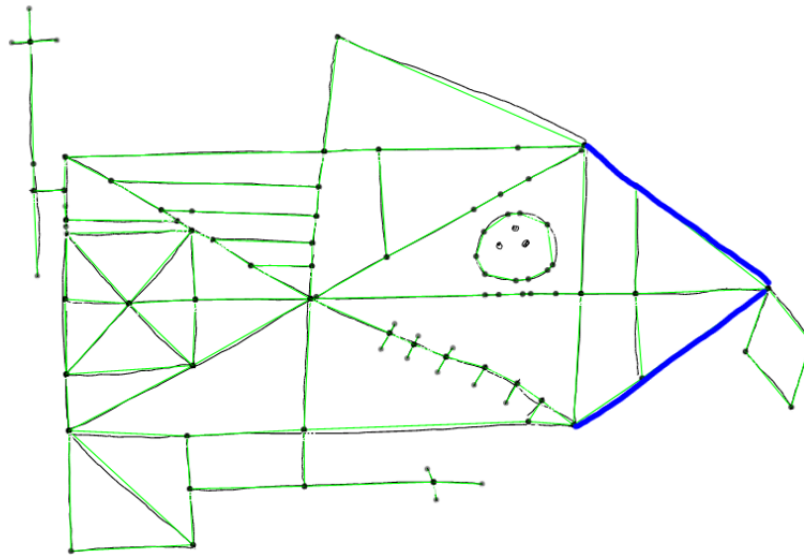
Figure 5.23: Detail 13, as detected by our recognizer.

The detail 13 triangle can be complicated by imperfect drawings of the attached diamond of detail 14, and adjacent details 15 and 16. The main consideration behind detail 13 is ensuring this triangle can be independently detected regardless of whether details 14, 15, and 16 are drawn correctly or at all.

Our recognition of detail 13 sees a return of edge crawling not too dissimilar from detail 2. This recognizer is more straightforward because we can estimate the agent's starting point; we place an agent at the bottom-right corner of the detail 2 rectangle, then crawl upwards as we search for the first slanted line with a slope that is upward and to the right. Once the first such line is detected (and

is ideally the agent's starting vertex at the bottom-right corner of the detail 2 rectangle), the crawler then continuously moves across the slanted line and stops when it cannot continue in the same up-and-right direction. If the ROCF is drawn correctly, this stopping point will place the agent at the right-most tip of the detail 13 triangle. The agent then changes direction and continues crawling along attached edges that have an up-and-left direction. To avoid crawling along the slanted line of detail 9 recognizer, the agent stops when it either encounters a vertex or edge associated with either detail 9, or detail 2. If this detail is drawn correctly, all edges and vertices that the agent has crawled through are those belonging to the two lines associated with the detail 13 triangle to the right of detail 2.

5.1.4.15 Detail 14's Recognizer

> *"The diamond attached to the end of Detail 13 should be diamond-shaped and must be attached to the end of Detail 13; it must not extend down below the bottom of the large rectangle, Detail 2."*
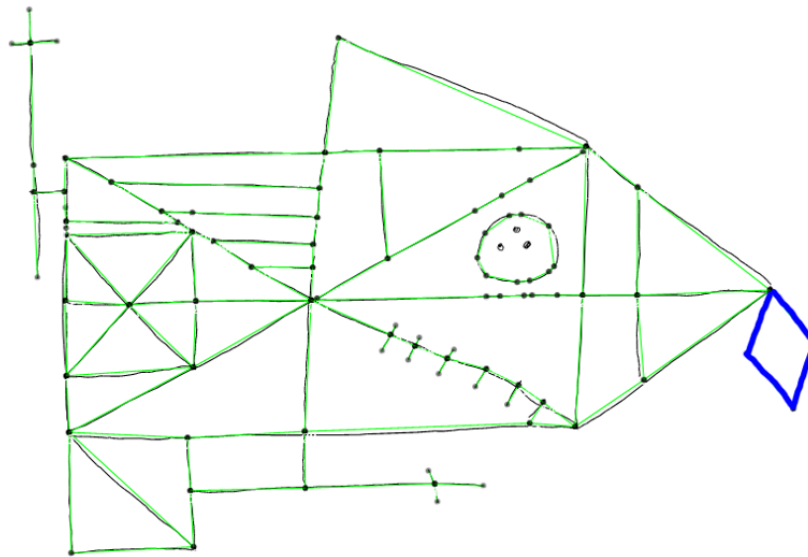
Figure 5.24: Detail 14, as detected by our recognizer.

Despite the diamond being simple in shape during testing and analysis of participant data we encountered that the diamond results in a fairly wide variety of vertices and edges when Step 1 of our process converts the ROCF to a graph. For that reason, rather than opting for an agent-based crawler, we have instead decided to employ a directed depth-first search somewhat similar to the bounded and directed DFS system found in detail 1 and detail 6.

The outer bounds of of this DFS-based recognizer is defined by the edges and vertices defined by detail 13, 15, and 16, and the DFS search begins at the tip of the detail 13 triangle. Because this requires details 15 and 16 to be included in the boundary, this means those two details are detected first (see Fig. 5.24).

"The vertical line within triangle 13 must be parallel to the right vertical of Detail 2,

the large rectangle, and it must be shifted to the left within Detail 13."
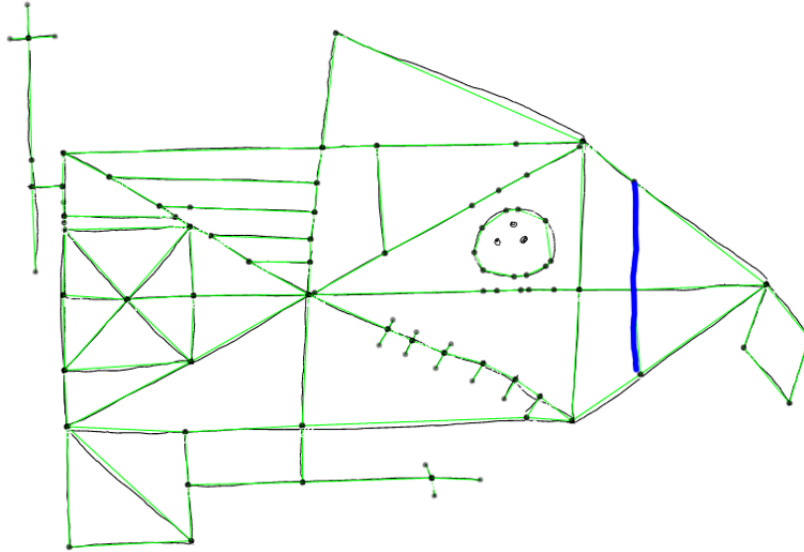
Figure 5.25: Detail 15, as detected by our recognizer.

Detail 15 is a relatively straightforward implementation of an crawler-based agent recognizer, working backwards from the default order of a detail 13 triangle and only crawling when it encounters a vertical line moving downward that is not already associated with the detail 2 rectangle. Once such a vertical edge is detected, the agent keeps moving downward and returns the vertical lines and vertices it crawls through until it encounters the bottom side of the detail 13 triangle.

This recognizer makes a single comparison for every n-node in the detail 13 triangle, and continues up to an m number of vertical edges in the downward direction, generally not more than 2 or 3 depending on the composition of the drawn ROCF and how the graph is created in Stage 1.

"The horizontal line within Detail 13, which is a continuation of Detail 4 to the right, must come from the midpoint of the right side of the large rectangle and extend to the top of triangle 13. If triangle 13 is slightly askew, or if Detail 4 does not meet the midpoint of the right side of the rectangle, Detail 16 should still be scored as a full 2 points if it went to the top of the triangle from the midpoint of the right side of the rectangle"
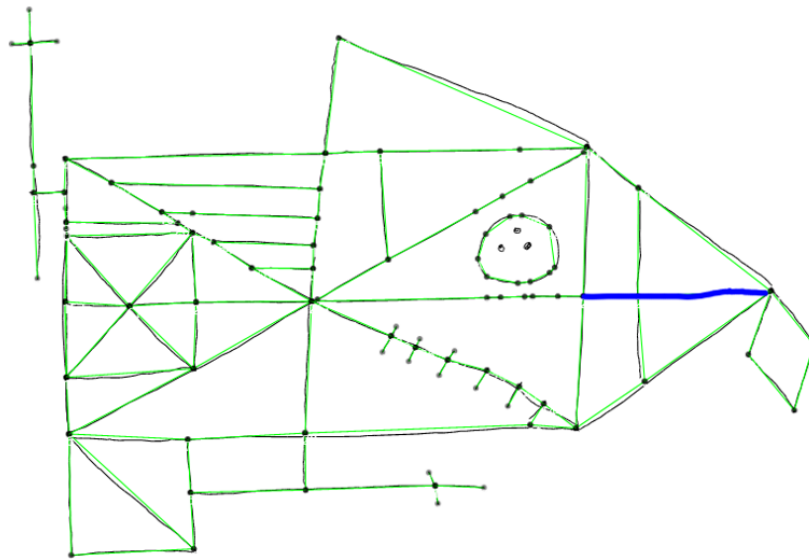
Figure 5.26: Detail 16, as detected by our recognizer.

Similar to detail 15, detail 16 is a crawling agent that moves along the right-hand vertical side of detail 2 and begins crawling in a rightward direction the first time that it can. It continues crawling until it arrives a vertex without a rightward edge attached.

The recognizer makes similar comparisons as detail 15, with n vertices of the right-hand side of detail 2 and m vertices of the horizontal line comprising detail 16.

5.1.4.18 Detail 17's Recognizer

> *"The cross attached to the lower center area of the rectangle. The right side of the cross must be clearly longer than the left side of the cross but must not extend beyond the right end of the large rectangle. It should also, at its left end, commerce at the midpoint of the right side of the square, which is Detail 18."*
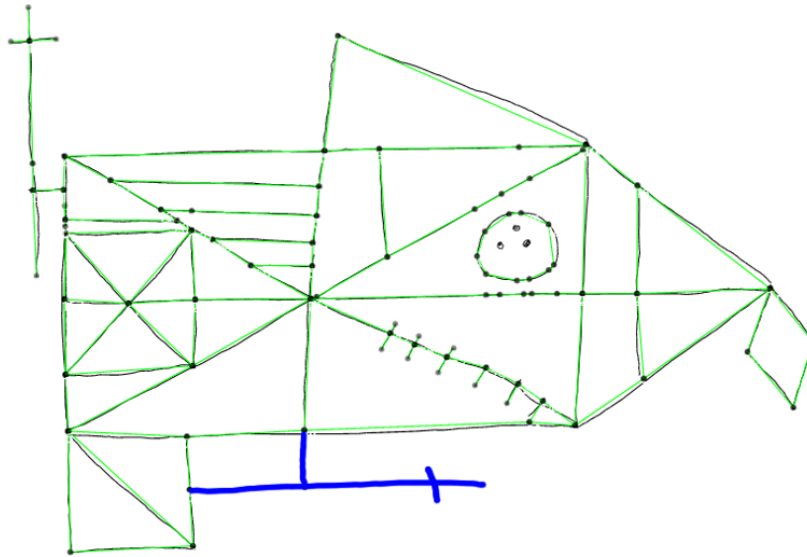
Figure 5.27: Detail 17, as detected by our recognizer.

Detail 17 is recognized in a similar fashion as the detail 14 diamond, which in itself is similar to the only other cross in an ROCF, detail 1. The reasoning is similar; all three of these details generate widely varying vertices and edges when the graph is created in Stage 1. Detail 17 further shares a similarity with detail 14 in that it this recognizer is deployed after the subsequent detail, in this case detail 18, since recognizing the latter detail significantly simplifies the generation of boundaries for the DFS-based recognition.

The boundaries for this DFS recognition is defined by vertices and edges associated with detail 18, and vertices and edges associated with detail 2. The DFS crawler finds detail 17.

> *"The square on the lower left corner of Detail 2. It must clearly be a square, as opposed to the rectangular shape of Detail 6, and its sides should be the same size as the vertical aspect of Detail 6, extending halfway between the left side of the rectangle and the vertical midline of the rectangle."*
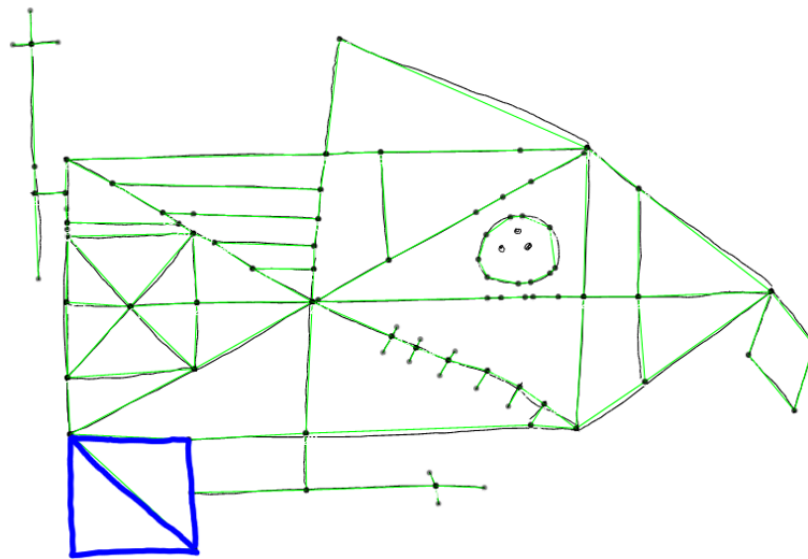
Figure 5.28: Detail 18, as detected by our recognizer.

The most effective version of the detail 18 recognizer is one that employs slightly modified versions of the detail 2 crawler and the detail 3 diagonal detector. The detail 18 recognizer benefits from a set starting point for the agent being placed at the bottom of the detail 2 rectangle. The agent is placed on the left and starts crawling to the right.

The agent then checks for the first edge that is in a downward direction, moving down along the first edge it detects. Once it moves down it continues its motion until it can detect an edge to its right, then repeats the process until it finds an edge to the top, and once it moves to the top it should encounter an edge that is part of the detail 2 rectangle again. The agent then crawls to the

Det.	Method of Recognition	Det.	Method of Recognition
1	Isolate region, then DFS to fill	**10**	Greedy single-direction pathfinding
2	Greedy pathfinding, repeated per side	**11**	Isolate triangle, then DFS to fill
3	Dijkstra's between diagonal corners of #2	**12**	Isolate lower-right region, DFS
4	Greedy single-direction pathfinding	**13**	Find upward, downward diagonals
5	Greedy single-direction pathfinding	**14**	DFS to find all edges on tip of #13
6	Direction path for top/bottom, Dijkstra's	**15**	Greedy single-direction pathfinding
7	Greedy single-direction pathfinding	**16**	Greedy single-direction pathfinding
8	Connect horizontal lines bet. #3 and #5	**17**	Isolate region, then DFS to fill
9	Find vertical, diagonal line above #2	**18**	#2's technique, then single diagonal

Table 5.1: General recognition method types for all 18 details. DFS is the Depth-First Search pathfinding algorithm. Dijkstra's is the Dijkstra Shortest-Path Algorithm

left where it is expect to come into contact with its starting vertex. If it does not find its starting vertex in this final step, it is not considered a closed rectangular shape situated below detail 2, and the agent is then placed further along the bottom of detail 2 in a rightward direction. However, a correctly drawn ROCF is expected to result in a closed rectangular shape of detail 18 the first time.

Detail 18 also consists of a single diagonal line connecting the top-left of the detail 18 square with the bottom-right corner. We employ Dijkstra's shortest path algorithm to detect this diagonal in the same way the diagonals are detected for detail 3. Because we observed several instances where participants would erroneously also draw another diagonal to form a cross like in detail 3, we have chosen to also run a second Dijkstra's shortest-path algorithm from the top-right to the bottom-left corners of the detail 18 square. If the ROCF is drawn correctly and no such diagonal exists, it will make no change to the shape that is given to Stage 3. If, however, the second diagonal is erroneously included, the additional line will be included and Stage 3 will detect this additional line, deducting points during grading.

5.1.5 Stage 3: Detail Validation

Stage 3 compares only the **isolated sample** of the recognized detail to a set of template details to score the sample's quality. With the exception of details that are a single straight line, every other detail in the ROCF test is passed through Stage 3 for validation. The system begins by centering,

scaling, and resampling the isolated sample points so that they lie in a $[-1, 1]$ range on the x and y plane. Re-sampling is done a second time to ensure that the isolated shapes can all be uniformly compared against the resampled template shapes.

Stage 3 uses a system loosely based on the $P recognizer [237] that turns gestures into point-clouds whose distances from provided templates serve as the basis for shape recognition. Point-cloud techniques have been used for fast recognition [308–311], and different implementations employ their own versions, with $P itself being loosely based off of Hidden Markov Models [312].

Figures 5.12 through 5.28 highlight in blue the line data that is passed to Stage 3, and is the sampled data from the pen that that pre-processed for conversion to a graph for Stage 1.In the process of recognition, however, small parts of the lines might get omitted from recognition during Stage 2 in order to facilitate the algorithm function. Certain corners and segments where lines intersect, particularly with lines that are not a part of the detail, might get omitted. By re sampling the isolated details we are able to restore the small "holes" created during the Stage 2 recognition process.

The resampled points are then used to create a dot density a map of an empirically provided resolution n such that a detail is represented as a $n \times n$ matrix where space in the figure is mapped to a cell of the matrix, each cell being some range $[x_i, x_j]$, $[y_i, y_j]$. Each cell is then given a p-value based on the number of points that lie within the range for each cell. A visualization of this is shown in figure 5.29. This is applied to each of the templates, then each template matrix is then averaged cell-wise to form a template mapping. This template point-density matrix data is calculated with shapes that have been isolated ahead of time. New templates can always be added and existing ones can be replaced without any change in the rest of the Stage 3 code. Although any number of templates can be added for any one particular detail, we have chosen 5 templates for each detail based on our testing during development.

The template matrix Q is compared sampled detail matrix P with equation 5.4 to determine how closely the two matched. The best match is then found by shifting the the sample matrix by row and column to find the best possible position when compared to the template, given that
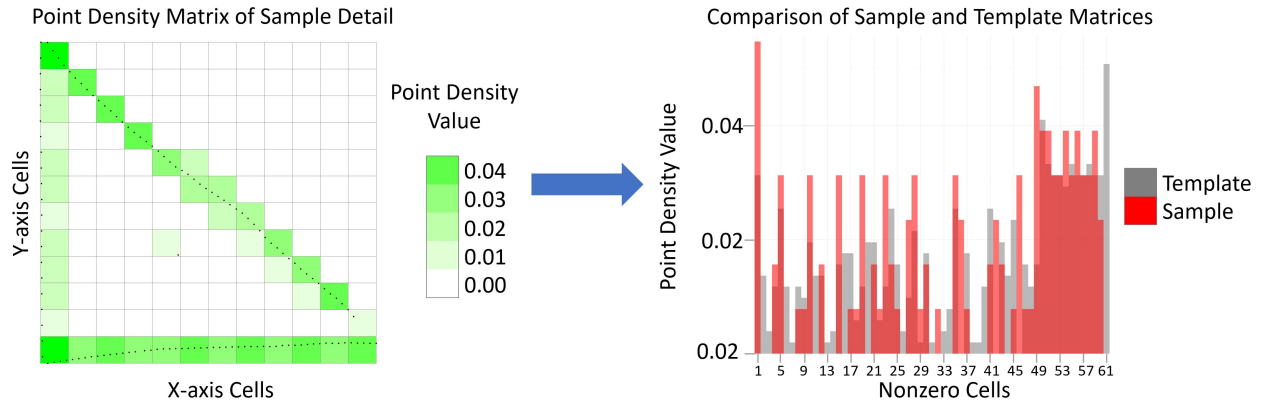
Figure 5.29: Comparing two point-density matrices to asses detail distortion: our sample, and a saved template. The process is repeated for all templates.

some details will match in terms of their stroke and dimension but have somewhat different centers relative to the the template.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \max\left(0, Q_{i,j} - P_{i,j}\right) \tag{5.4}$$

The value of "distance" between template and sample is between 0 and 1, with 0 being the best. A shape that receives full credit for neatness is characterized as how close the sample is to the templates. Any value below 0.5 assigned to a sample is given full credit of 2 points. A value between 0.5 and 0.9 is given partial credit of 1 point. A value above 0.9 is given 0 points. This is distinct from point-cloud recognizers such as $P that match input gestures to its closest shape category in what is considered a weighted-matching problem [313]. Our system, rather than matching to a type of shape, instead focuses on the distortion of the provided shape, since we can make a closer inference about the shape that is intended to be drawn. For example, if we compare against detail 2 templates, we know the user should have drawn a rectangle.

The design of this template-matching technique places emphasis on both the line composition of the proportion of the shape. Extra lines in the isolated sample or omitted lines will create a

65

difference in distance p when compared with the template. Proportion also works in a similar fashion; observing the point-density matrix on the left of Figure 5.29 shows that the horizontal line at the bottom of the matrix has higher point density value than the vertical line on the left of the matrix. This discrepancy is due to the fact that the triangle in question is that of detail 9 in an ROCF (see Figure 5.19), which is a right triangle where the horizontal line is considerably longer than the vertical line. The resolution of the matrix is a static $n \times n$ across the details, meaning shape dimensions will be compressed and the longer one dimension is, the more dense the cells are in that dimension when the matrix is created. This essentially preserves the proportion of length to height of each shape for both templates and samples. While an ROCF's shape is bounded by the size of the physical page, it is important to note that total size of an ROCF does vary in size. This "compression" of the shape proportions when creating the point density matrix allows us to grade for shape proportion regardless of shape's absolute size.

5.2 Data Collection and Results

5.2.1 Data Collection Protocol

We conducted a study with 68 cognitively healthy participants to complete a Rey-Osterrieth Complex Figure Test between the ages of 19-32, following the protocol as specified in the Compendium of Neuropsychological Tests [2]. Participants were recruited via email and classroom announcements at the local university. All participants took the test in a simulated neuropsychologist's test environment of an office setting free of distractions and below-average background noise. Every participant was tested one at a time and first completed the necessary consent form. Participants were provided the following: the prompt ROCF printed on a piece of paper, a Neo SmartPen already connected via Bluetooth to a laptop running our *Auto ReyO* app for data collection, and "blank" canvas page that tracks the pen's location and pressure in real-time. Participants were asked to re-create the ROCF they see into the blank canvas page using the SmartPen, informing them to pay as close attention as possible to the composition of the shape as they will be asked to re-create it from memory. They are asked to draw as cleanly as possible,

although they are allowed to draw over their own line once or twice to correct for mistakes. There is no time limit to this or any other condition on the test. This stage constitutes the "Copy Condition" of an ROCF test.

Immediately after the shape is copied, we collect the pre-printed ROCF and their drawn copy, as testing protocol does not allow participants to study the ROCF by memorizing it through sight alone. A second "blank" canvas page is given to them, this time with no pre-printed ROCF page along with it, and are asked to draw the same shape again from memory. They are informed they should construct as much of the shape as possible, and to try their best even if they cannot remember every shape. They can think aloud through their process but the test proctor does not provide hints answers. This constitutes the "Recall Condition."

The "Delayed Recall Condition" is the final part of the testing protocol. It is the same as the previous "Recall Condition" but is not started until after 30 minutes of other activity. For ROCF tests, this 30 minute wait time for "Delayed Recall" can vary in length: test proctors can wait anywhere between 20 minutes to 50 to conduct a "Delayed Recall," and in this time participants are asked to engage in an unrelated activity. In our protocol they were asked to engage with any activity of their choosing on their personal laptop or phone. Most commonly, participants browsed the internet, checked email, worked on homework, or any combination. At the conclusion of 30 minutes participants were asked to re-create the ROCF with a new "blank canvas" page one last time.

All three test conditions were recorded separately as their own sketches, and were graded by our *Auto ReyO* system for analysis at a later time.

5.2.2 Data Analysis

Although this test is meant to assess constructional ability and memory loss, healthy participants do not always score full marks on an ROCF [314], and indeed our testing corpus reflects a wide range of scores that conform to established normative data for our participants.

A total of 204 sketches from the 68 participants were collected. Of these, 5 perfect-score tests were set aside to be used as templates for Stage 3 validation. 14 were not gradable or their sketch

data was corrupted, bringing the total graded to 185. All tests were also graded by two field experts whose grades we consider "ground truth" in this context. The first grader is a practicing clinical neuropsychologist and the second is a professor specializing in cognitive and visual perceptual rehabilitation in older adults. We measure our system's success in two ways: the F1-Score of our recognition algorithm for each detail, and the comparison between our system's total grade and the expert graders' total grade. For the latter, both our system's and the grades are on the 36-point scale as defined in Section 3.1.4.

A key factor considered when calculating F1-score was the subjectivity of distortion thresholds. While we implemented our own thresholds for distortion in Stage 3, instructions for the ROCF in the literature leave the definition of "distortion" at the discretion of the grader [2]. For recognition purposes we are interested in gauging whether our system can successfully either find a detail or confirm its absence. The F1-score reports our system's ability to recognize the existence of a detail. Since we are still interested in comparing 36-point grades that also integrate distortion as partial credit, we also calculate Spearman's rank coefficient ($\rho = 0.767$) between our automatically-graded tests and those of our expert grader.

5.2.3 Results Discussion

In clinical neuropsychology, grading Rey Osterrieth Complex Figure tests has been the subject of constant iteration and is an active research topic, with numerous methods of interpretation being proposed and refined. As such, analyzing the process of grading these ROCFs automatically is not a trivial subject. Our analysis centered on simulating the perception of an detail since the granular differences in distortion are frequently attributed to grader subjectivity. Our aim, then, was the provide evidence the system perceived the details correctly, even if they might have been slightly distorted, or in the case of severe distortion the Stage 3 validation stage would be able to separate those clear cases. In terms of overall score comparisons, we sought to analyze how far apart individual test grades our system was from those of expert graders. Although the grades from individual graders were closer to each other than between each grader and *Auto ReyO*, our system compares favorably due to the high F1-score of the vast majority of details, and the average

68

Det. #	$\Delta_{a,g1}$	$\Delta_{a,g2}$	$\Delta_{g1,g2}$	F1-Score	Det. #	$\Delta_{a,g1}$	$\Delta_{a,g2}$	$\Delta_{g1,g2}$	F1-Score
1	0.69	0.68	0.37	0.787	10	0.21	0.20	0.10	0.962
2	0.18	0.44	0.33	0.857	11	0.47	0.45	0.21	0.878
3	0.07	0.11	0.16	0.978	12	0.46	0.44	0.11	0.872
4	0.27	0.30	0.12	0.927	13	0.13	0.11	0.11	0.966
5	0.08	0.49	0.49	0.978	14	0.42	0.41	0.15	0.881
6	0.31	0.56	0.26	0.958	15	0.50	0.49	0.09	0.770
7	0.47	0.26	0.13	0.855	16	0.21	0.19	0.08	0.919
8	0.28	0.33	0.13	0.966	17	0.57	0.66	0.34	0.788
9	0.35	0.20	0.07	0.904	18	0.45	0.49	0.21	0.925

Table 5.2: Classification results and average scoring differences for each detail across all graded tests. **n=141** for all details except for Detail 2, where **n=185**. $\Delta_{a,g1}$ denotes the **average point score difference** between Auto Rey-O and Grader 1, $\Delta_{a,g2}$ is the difference between Auto Rey-O and Grader 2, and $\Delta_{g1,g2}$ between Grader 1 and Grader 2.

difference between *Auto ReyO* and our system being around 3 points out of 36 possible points for an ROCF test. We believe these results are significant in light of the fact that a fully automated ROCF grader that grades all 18 details has yet been proposed.

Also of note is the performance of detail 2, the large rectangle that serves as the anchor for the rest of the sketch. The organizational strategy score of the Rey-Osterrieth Complex Figure test places the highest priority on the existence of Detail 2 in a sketch due to its importance to the overall figure structure [315, 316]. For the purposes of our system, this resulted in calculating F1-Score for recognition of the 18 details being conditional on whether detail 2 could be successfully recognized within a sketch. Exceptionally poor figures that lack a discernible detail 2 almost always result in very low or ungraded scores when hand-graded. Similarly, in very rare cases a poorly drawn ROCF could be graded by *Auto ReyO* if detail 2 could be recognized, while another ROCF that would score higher might not be graded due to a detail 2 that could not be recognized. For this reason, we have designed our recognition hierarchy such that the test is not graded if it cannot automatically recognize detail 2. This provides the most consistent application of grading requirements that is still consistent with the grading rubric as presented in the Compendium of Neuropsychological Tests [2].

A total score of 0, however, is not necessarily due to a true negative. For 44 sketches, our algorithm was unable to find detail 2 due to a sloppy or unconnected drawing, but other details would exist. If we flatly calculated F1-Score of all Details for every sketch included the ungraded ones, this would assign incorrect false negatives to the rest of the details. For that reason we tier F1-score calculations; for detail 2 we calculate it for all sketches (n=185), and for all other Details we calculate it where detail 2 was correctly detected (n=141).

The F1-Score results in Table 5.2 and the graph in Figure 5.30 demonstrate the effectiveness of Auto Rey-O. Our top-down system correctly identifies and validates the details with a high enough F1-score that shows the system working for typical test-takers. Table 5.2 also shows the average differences in scores (in points, maximum of 2) assigned between our Auto Rey-O system and our expert graders. All of the average differences in scores for each detail are well below 1 point and the vast majority below half a point, indicating a marginal difference in scoring between expert graders and our Auto Rey-O system. The system also works successfully for ROCFs with higher amounts of distortion. Such instances display the flexibility of the system in still identifying present details even if the participant has heavy lapses in memory.

The lowest-performing details are 1, 17 and 15. These had the highest amount of false negatives, although our manual review of these false negatives showed our system did recognize the details but chose to grant a score of 0 due to our threshold for distortion. Further refinements of distortion threshold values for these two details would improve their recognition quality.

Figure 5.30 compares the scores assigned by our automated grader and the two expert graders. Between the two expert graders, the correlation was $p = 0.948$, a Spearman's rank coefficient of $\rho = 0.942$, and an average difference in scores of $\Delta_{g_1,g_2} = 1.68$. Between our system and grader 1, $p = 0.799$, $\rho = 0.765$, and average $\Delta_{auto,g_1} = 3.21$. Between our system and grader 2, $p = 0.829$, $\rho = 0.802$, and average $\Delta_{auto,g_2} = 2.78$. Our automated system produced grades with a generally high correlation with those of the graders, although the grades from the experts were more similar to each other. In all three cases, low-scoring tests somewhat deviate across all graders, even between the expert graders. This is likely due to the aforementioned ambiguity

36-Point Grades For All Graded ROCF Tests
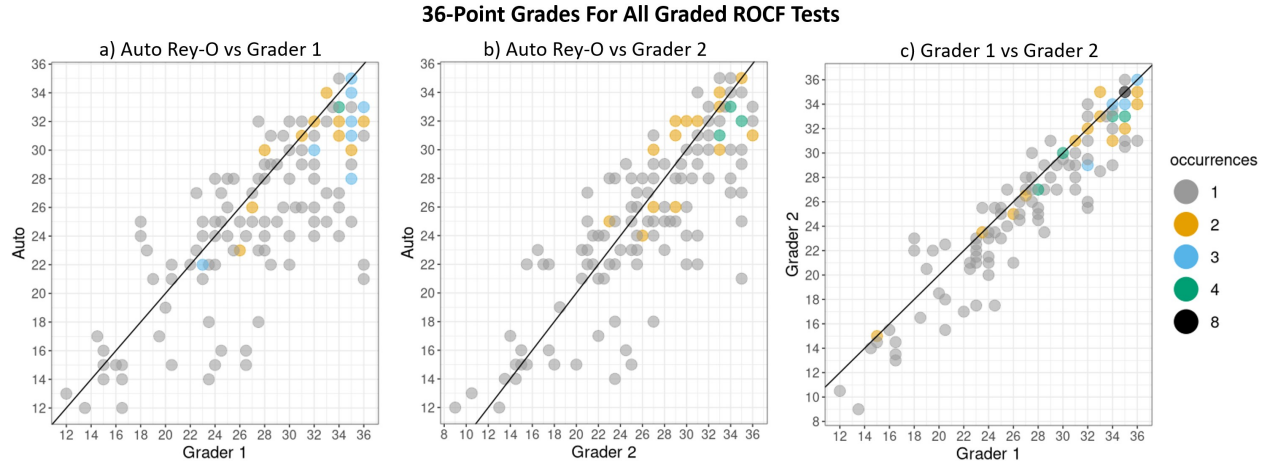
Figure 5.30: Grade plots for all 36-point scores, compared between Auto Rey-O and expert graders (**n=141**). (a) p=0.799, (b) p=0.829, (c) p=0.948

in interpreting detail distortion. Our automated grader can also be observed to be consistently too strict on grading that produces consistently lower scores, which is partially attributed to the fact that it does not recognize details that were placed in the wrong location. In addition, at the suggestion of the expert graders who also served as domain experts, we chose to prioritize consistency in grading over leniency when deciding on partial credit thresholds since consistency is one of the key advantages of an automated recognition system.

5.2.4 Insight Into Other Tasks

As a point of higher-level discussion, we also wish to specify the insights that we have gained while implementing this solution, particularly in how we might approach other tasks. Of immediate interest is that this approach has granted us the ability to create generalized sub-shape recognition for any scanned sketch that does not have timestamped per-stroke sketch information. Although our saved sketches contain within them timestamped coordinates, we do not use timing data for our Auto ReyO system. Our approach, on a high level, can also be applied to any ROCF shape provided we are able to digitize a completed examination via scanning, and would thus allow us to conduct automated grading for tests that have been completed in years prior.

To further expand to the topic of general sketch recognition, our solution drew inspiration from

71

the existing work provided by the Mechanix project [130–133, 135], which largely provided its solution for hand-drawn truss recognition by turning the small sketch into a graph and ensuring the resulting graph was a series of triangles with one common side on each. The same approach for truss recognition used in tandem with an application of our system, however, could be an asset in the task of recognizing other static-body systems that require the analysis of specific components. This would be especially useful in instances where a top-down recognition system is necessary, such as implementing an option to scan and grade sketches completed in pencil and paper. Such scans would lack the timestamped stroke data that Mechanix relies on, but the individual pieces of the system would still need to be individually recognized.

This approach in particular would provide us with a low-cost, computationally inexpensive alternative to sub-shape sketch recognition from scanned sketches. The more common deep-learning approaches require a far higher amount of training data (whereas we only used five well-drawn ROCFs for training), would require significant computational power to train a multi-layered network, and would be unable to exactly point to the absence or presence of individual lines. Our approach has shown us it is possible to build a graph-based approach for sketch recognition where individual details are of interest, usually any graded task such as in education or neuropsychological tests. The fact our system is able to produce strong recognition results without the need to analyze high granularity timestamped sketch data would allow scanned images to be brought into the corups of data without the need for computationally expensive deep-learning solutions.

5.2.5 Discussion on Grade Timing

Although an extensive study on the computational complexity and time performance of this system was out of scope for this project, we still would like to make some brief remarks on the improvements of time taken to complete this task that is inherent to an automated system. Canham *et al.* [94] reports that, on average, the time to fully grade an ROCF is between 5 and 15 minutes per test. Expanding this out to the three conditions of a full battery of ROCF, this translates to between 15 and 45 minutes for each subject.

By comparison, the full 3-stage pipeline we have created on average takes between 1.5 and 3 seconds per test to be completed, with the possibility of taking upwards of 10 seconds per test depending on the speed of the CPU of the computer in which this test is graded, and depending on the overall messiness of the sketch drawn. We further automated the process of automatically grading every sketch file in a folder containing all 185 sketches from our test corpus with a one-click solution. This included the reading of the file, the full 3-stage grading process, and the output of each test's automated grades into a file in the comma-separated value (CSV) format. This entire process took Auto ReyO to grade 185 tests in **5 minutes, 24 seconds** in our system consisting of an Intel i7-7700K CPU clocked at a frequency of 4.2GHz and 16 GB DDR4 RAM clocked at 4,100 MHz. Based on the estimates provided by Canham *et al.*, the same task of grading 185 tests by a human grader would have taken roughly **15 and 46 hours**. These remarks intend to highlight the inherent advantages of time savings for clinicians who wish to automate a process currently done by hand. While we did not request the expert graders of our study to time themselves, but will include some time-keeping mechanism for our expert graders in a future study to produce a more formal timing comparison.

5.2.6   Comparison with Normative Data

A comparison of interest is in analyzing how the scores reported by the Auto ReyO system comes close to reporting true-positive predictions of a subject's cognitive decline compared against the same scores reported by the expert graders. In practice this is a challenging comparison as cut-off scores in normative data depend on a variety of factors, particularly depending on the cognitive state or prior diagnosis of the participant themselves. Little research work exists for cut-offs among healthy populations, since poor performance of the ROCF alone is not necessarily indicative of an undiagnosed condition. Rather than score cut-offs, normative data for various populations that include reporting of mean and standard deviation are published, and clinicians make a comparison of scores for ROCF in addition to performance of other tests as well as their qualitative observations to complete a full behavioral profile. Mean and standard-deviation normative scores for aging populations for both healthy and those with cognitive decline exists, as does normative data exists

to help differentiate between healthy subjects and malingering subjects. The Compendium of Neuropsychological Tests defines "malingering subjects" as healthy participants who perform poorly "based on litigation status and failure of effort measures" [2].

For our dataset we perform an assumption of malingering subjects who fit the second category of "failure of effort measures." Although we did not conduct a survey to measure the participants' level of effort in an exact manner for each of the conditions, we still believe this would be the most accurate form of providing a score cut-off comparison since all recruited subjects were deemed to be cognitively healthy in our population. Strauss *et al.* cites a study that found that "examined patients with suspect effort" versus those who had "various neurological and psychiatric disorders" and found that "patients suspected of poor effort displayed significantly lower recognition scores than did patients with bona fide visual memory impairment," and thus cutoffs would indicate that a lower score among healthy populations below a specific cutoff could be signs of malingering subjects.

Along these lines, the Compendium of Neuropsychological Tests [2] reports that a cut-off score of 25 or less for the Copy condition resulted in a Predictive Positive Value (PPV) of 90%, which translates to a 90% possibility that the subjects **below** this cutoff fit in the malingering status. A cutoff of 9 or less had a PPV value of 78% for the Immediate Recall. In that study Delayed Recall was not reported, so cutoff data was unavailable. For the Copy condition our first expert grader had **no** tests that fell below our ≤ 25 cut-off, our second expert grader had **4** tests that fell below this cut-off, and Auto ReyO had **5** tests that fell below this cut-off. When examining the Immediate Recall condition, the first expert grader reported **1** test that scored below the ≤ 25 cut-off, the second expert grader reported **2** tests below the cutoff, and Auto ReyO reported **no** tests.

These cut-offs serve as a point of comparison of how the scores might be interpreted as belonging to different categorical classifications. While the majority of the time normative data only reports mean and standard deviation of the scores of a given population and the cut-off for what a clinician might consider signs of cognitive decline or malingering subjects, we have attempted to perform the closest point of score cut-off comparison that is identifying malingering

subjects. Clearly the expert graders did not fully agree with each other how many or which participants fell under this cut-off, but the number of tests were similar in magnitude and Auto ReyO makes a similar amount of comparisons. A more formal study would need to be performed for a full comparison of predicting malingering subjects. It is worth noting that according to Strauss *et al.*, "individual ROCF scores ... were not very sensitive in capturing suspect effort" [2], since it bears repeating that the ROCF by itself is unable to make such a determination.

5.2.7 Comparison With Off-the-Shelf Solutions

Although a common point of comparison when gauging the performance of an automated system is to compare it to existing state-of-the-art solutions, such a comparison is made challenging against this particular topic since there is currently no automated system for in existence for detecting the ROCF's 18 details. Still, there is value in the comparison between the closest equivalent of a baseline. Because there is no equivalent of a sub-shape recognizer, we have opted for the closest off-the-shelf system that exists, the creation of a machine learning classification system using common features created by Rubine *et al.* [103, 212].

Feature-based machine learning classification relies on the geometric properties of drawn shapes to help classify them to labels of various types. This technique is used extensively in our second project featured in Chapter 6. Most commonly, these labels in question are of the types of shapes that are drawn, such as numbers, letters, shapes, and other simple geometric symbols. This kind of classification can be effective at classifying input drawings on shape types since the Rubine features have been shown over the decades to provide a strong enough description of how a shape is drawn. These tend to be applied in bottom-up recognition systems but can also be applied top-down by treating an entire sketch as the input once and calculating the 12 Rubine features at once.

This technique is inadvisable to perform for complex figures since the features were created for gestures, not entire sketches (see section 3.2.1). Some amount of segmentation would have to be performed on-the-fly and a custom algorithm would have to be created for the ROCFT to repeatedly attempt to re-classify the sketch as the subject draws. However, for illustrative purposes, and for the purpose of demonstrating that existing shape classification techniques are not adequate for the

automation of grading ROCFTs, we have attempted to produce a classification algorithm on all 18 ROCF sub-shapes with the off-the-shelf solution using Rubine features to develop a machine learning classifier using Random Forest [317], the most commonly well-performing classifier.

All 185 sketches from the user study that were collected had their 12 Rubine features calculated (see [103] for a detailed explanation of each feature) on a per-sketch basis, as any more granular would require extensive custom code for the ROCF domain and thus would not be off-the-shelf. For labels we used the same expert grader's grades that were used as ground-truth scores to calculate the F1-scores seen in table 5.2. The grades for each of the 18 details were linked to the Rubine features of the corresponding sketch, along with the total grade of the entire sketch from the expert grader.

We produced two kinds of predictive algorithms: **(1)** a regression to attempt a prediction of the feature score itself (between 0 and 2 for each detail) based on the total score given by the expert grader, and **(2)** a binary classification that attempts to reproduce the "is there" vs. "not there" classes that we derive and describe in section 5.2.2. For the first predictive algorithm, we report on two measurements that report on the Relative Absolute Error (in percentage, lower is better) of the predicted score (between 0 and 2 per detail). For the second predictive algorithm, the off-the-shelf system is effectively tasked with attempting to classify the sketches as to whether or not each of the 18 details exist, based on a training set that contains a randomized sample of the provided sketches. The F1-score that we report, in a value between 0 and 1 (1 being best), represents how effective each classifier is at detecting which of the 18 details are missing in each sketch, if any. Because we aim to compare this with our system's task, which is in finding which individual details are present in an ROCF, we create 18 different predictive algorithms with one for each detail. The results are reported in Table 5.3, with the RAE for prediction **(1)** and the F1-Score for prediction **(2)**. Lastly, we report on a predictive algorithm to predict the **total score** between 0 and 36: an RAE of 81.708%. All predictive algorithms were tested with a standard 10-fold cross-validation method that prevents overfitting of data.

This method, regardless of whether we attempted to predict the whole ROCFT score, the score

Det.	RAE (%)	F1 Rubine	F1 Auto	Det.	RAE (%)	F1 Rubine	F1 Auto
1	91.394	0.096	0.787	10	83.410	0.561	0.962
2	87.396	0.019	0.857	11	92.964	0.041	0.878
3	92.390	0.011	0.978	12	89.654	0.126	0.872
4	92.479	0.016	0.927	13	77.863	0.062	0.966
5	91.456	0.016	0.978	14	97.263	0.038	0.881
6	84.173	0.082	0.958	15	82.763	0.189	0.770
7	87.223	0.186	0.855	16	91.721	0.135	0.919
8	87.227	0.039	0.966	17	88.230	0.155	0.788
9	67.708	0.119	0.904	18	89.496	0.073	0.925

Table 5.3: Comparison with baseline of off-the-shelf machine-learning based solution. RAE is Relative Absolute Error (in percentage) of raw score prediction. F1 Rubine is the F1-Score of the machine-learning solution that uses Rubine features for classification of the entire sketch. For comparison purposes, we repeat the classification F1-Scores of our own system in F1 Auto. For all classifications, **n=185**

of each of the 18 details, or whether the 18 details could be individually detected, we can see that the traditional methods of digital sketch recognition are inadequate for the task of sub-shape recognition. The RAE for the prediction of each detail and the overall sketch was very high, and similarly the F1-Score for all but one of the details was unable to achieve a value above 0.1. For reference, an indicator of an effective F1-Score in digital sketch recognition should be at least 0.8, preferably closer to 0.9 or above. This allows us to recognize the effectiveness of our Auto ReyO solution.

5.2.8 Limitations

The main limitation of this graph-based approach to top-down sketch recognition is the reliance on line connections. Our vertex-contraction algorithm in Step 1 of the system's process does connect lines with corners within a certain radius. We found this technique worked very well if sketches were drawn with reasonable neatness. If the lines are disconnected by more than half an inch, however, these lines will remain disconnected. This was a conscious design choice since vertex contraction cannot be too aggressive; otherwise, regions where any correct sketch would have high numbers of vertices would all get incorrectly contracted into one. This is the case such

as the area where detail 6, 7, 3, and 8 all converge—even neatly-drawn sketches have a high concentration of vertices here. We intend to improve refine the recognition system to "jump" gaps and close disconnected lines only where appropriate.

Additionally, another limitation of note is the custom nature of the individual recognizers. As per section 5.1.2, we discussed the careful balance that needed to be struck in designing a series of recognizers aimed at identifying not only geometric details, but specific ones in certain locations. The series of generalized approaches that we deployed on each detail's recognizer can also be applied in a similar manner as the other 7 complex-figure examinations found in the Compendium of Neuropsychological Tests [2]. However, we should point out the code itself would need to be written for each of the recognizers.

In that sense, of the three stages that we present in our system, the second stage that comprises of the individual recognizers would need to be written for the other 7 examinations. However, several of the details in other other complex figures would likely share a large proportion of similar coude. In the case of Detail 2, for example, the code to find the largest rectangle that serves as an anchor for the rest of the shapes would be largely the same as the solution we have presented. Other more specific shapes, especially geometric figures that do not appear in the ROCF, would need to be written in a similar sense. This work disseminates the approach to the graph-traversal approaches that we use to recognize sub-shapes as a framework from which grading other complex figure examinations could also be automated. It should also be emphasized that stages 1 and 3 can be applied to any complex figure examination with little additional work. Stage 1 would not require any changes when applied to other tests, and stage 3 would only need the templates for well-drawn examples of each detail. In that sense stages 1 and 3 do not use custom solutions that would only work for the ROCF.

5.3 Future Work and Conclusion

Refinements can be made to help recognize specific kinds of poorly-drawn details. As previously mentioned, most sources of grading inaccuracies for our system came from poorly connected graphs due to sketch sloppiness. For healthy participants taking this test, our expert graders

attributed sloppiness as a lack of effort rather than genuine memory loss if the patient has no hand motor issues. Still, there would be an interest in supplementing our graph traversal with connecting otherwise unconnected vertices to improve recognition performance.

Additionally, improvements to our Stage validation approach could be made to recognize finer details. Our validation method sometimes may not properly distinguish between small changes, such as an extra stray mark or one line missing. Identifying missing lines is important for details 8 and 12, where the number of parallel lines drawn is relevant to its grading. Our validation method is able to find these discrepancies somewhat frequently, but potential for improvement exists.

We also seek to further generalize this work by allowing a GUI-powered method of graph crawling. With such an approach a clinician would be able to produce new complex figure examinations by drawing the test in a GUI, letting the system convert it to a graph, and letting the clinician label with click-and-drag motions each of the details that need to be automatically recognized. The system would internally keep track of the order of the path such that this behavior can be re-created when input graphs from subjects.

Lastly, we aim to work with clinical neuropsychologists to administer their test to willing clients to evaluate system usability in a clinical setting. This would produce additional sketch data taken from actual patients, and would allow us to perform UI/UX usability studies for clinicians. The ultimate aim of the system is to aid diagnosis process by automating the grading of an ROCF, so evaluating the user experience of clinicians as they collect the digital data and use the Auto Rey-O application for themselves is the next step to further this project.

Our *Auto ReyO* automatic Rey-Osterrieth Complex Figure test grader demonstrates the validity of a top-down sketch recognition approach using graph traversal algorithms. This significantly simplifies the recognition process where a bottom-up approach would need to take into consideration a prohibitively wide array of possible shape interpretations and re-interpretations. By employing graph crawling, vertex search, and optimization algorithms we are able to identify key sub-shapes of geometric shapes and thus are able to present one of the first fully automated ROCF grader that analyzes all 18 details.

6. DEVELOPING PREDICTIVE ALGORITHMS TO DETECT MILD COGNITIVE IMPAIRMENT

With the number of Alzheimer's patients reaching 5 million in 2014 according to the Center for Disease Control and Prevention, increasing emphasis has been placed on identifying and understanding its precursor condition, Mild Cognitive Impairment (MCI). MCI is characterized by subtle but abnormal cognitive decline and is challenging to detect without formal testing. Neuropsychologists use paper-and-pencil tests such as the Trail-Making Test (TMT) for diagnosis, and ongoing research places importance on high-granularity sketch data from digital TMTs. We present SmartStrokes, a digital TMT app designed to simulate the paper-and-pencil testing experience on a tablet and stylus. Our contribution is a pair of classification models that identify MCI on an individual segmented line basis, which could provide localized sketching behavior indicative of MCI. We also present a refinement of line segmentation algorithms from previous research to better distinguish between the actions that a participant takes when completing the exam.

6.1 Applying Sketch Recognition Techniques to Detect Mild Cognitive Impairment in Digitized Trail-Making Tests

This chapter contains work currently under review for publication.*

6.2 Motivation

The Center for Disease Control and Prevention has reported 5 million Alzheimer's patients in 2014, with the expected number to more than double to 13.9 million by 2060. Due to advancements in interventions aimed at mild-to-moderate cases of Alzheimer's disease, neuropsychologists have placed an increasing emphasis on early detection of Mild Cognitive Impairment (MCI) to better preserve quality of life [318–320]. A clinical neuropsychologist typically conducts paper-and-pencil cognitive examinations with a patient to help detect MCI, and the process is laborious, re-

---

quires multiple rounds of testing, and frequently subjective analysis of a patient's subtle behavioral patterns. Digitizing these clinical examinations, specifically the Trail-Making Test among them, has allowed researchers to attempt to aid the diagnosis process by employing machine learning for behavioral analysis. Existing work in this space has not yet leveraged recognition techniques used in digital sketch recognition, particularly research that links sketching with cognition.

Our proposed contribution presents a classification model that can distinguish between MCI and healthy participants on a per-line basis. It builds on existing work from Dahmen *et al.* [321], which in its conclusion states the belief that the high-granularity digital sketch data from a digitized version of the TMT could provide higher-granularity analysis. Per-line classification would offer two advantages: 1) targeting individual lines for classification of MCI could give a more localized assessment of individual dots that challenged the participant, and 2) building a classification model that analyzes sketches on a per-line basis could allow generalizability for more layouts, since the TMT frequently needs a wide variety of dot layouts to avoid the Learning Effect.

Further, our proposed model compares uses the scores from the Montreal Cognitive Assessment (MoCA) to detect MCI, whereas the existing work from Dahmen uses the TICS [322] and FAB [323] to assess more advanced dementia. MCI is characterized as being much more subtle in nature, meaning milder cases of MCI frequently result in sketching behavior that only slightly deviates from a healthy participant. Our proposed solution achieves an accuracy in line with Dahmen's existing work, with the added advantage of offering classifications on a more granular per-line basis, and classifying for a more subtle degree of cognitive decline.

Our goal with this research was driven by the main research question:

R1—*Can we create a predictive algorithm for estimating cognitive decline using digital sketch recognition features?*

To answer R1, we conducted a study with 37 participants with varying degrees of MCI as reflected on associated MoCA scores. A portion of the segmented data was used for predictive algorithm training purposes while the testing data was used for verification and accuracy performance metrics.

Our contributions in this work include:

- An analysis of how existing digital sketch recognition techniques can be applied to the domain of clinical neuropsychological examinations.

- A predictive model that estimates MoCA scores based on how they complete a digital TMT.
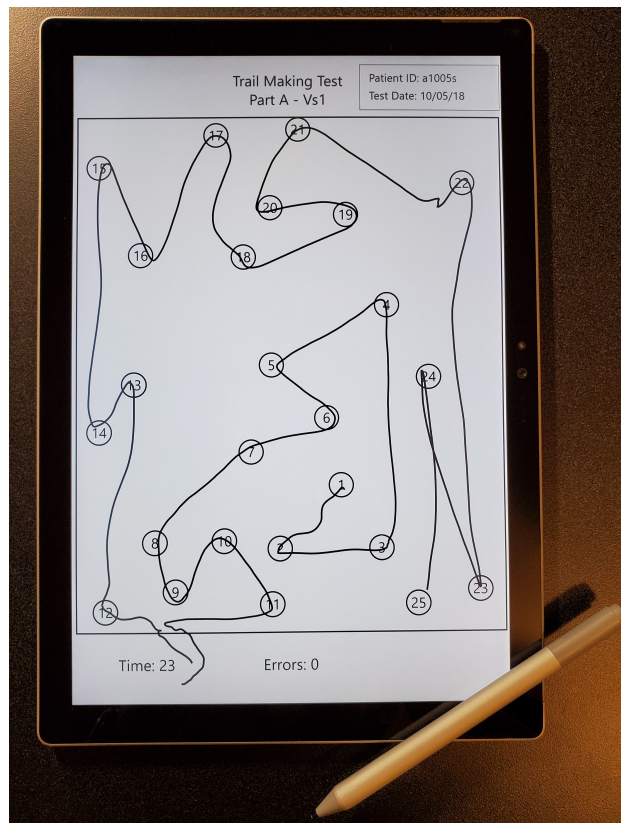
Figure 6.1: A sample completed test in our SmartStrokes app. The interface is designed to be as close as possible to an actual paper-and-pencil test

6.3 Design

SmartStrokes is a digital testing suite focused on re-creating commonly-used Trail-Making Test layouts for use on Microsoft Surface Pro 4 devices. The Universal Windows Platform (UWP) was chosen for reasons that include rapid development of a mobile-style application on Windows

devices, ease of exporting pen data for analysis, and its firmware-level digital pen integration with Surface Pen devices and allows us to easily extract pen pressure to supplement our feature set.

The system's users are intended to be test proctors; either medical or research professionals working with participants. Every participant is associated with individual users and every test is directly associated with each participant. All participant data including completed tests can only be accessed from within the app if the test proctor username and password is entered at the login screen. Proctors have the option to export digital images of the completed examinations at the conclusion of each test, at which point the proctor should ensure proper data anonymization.
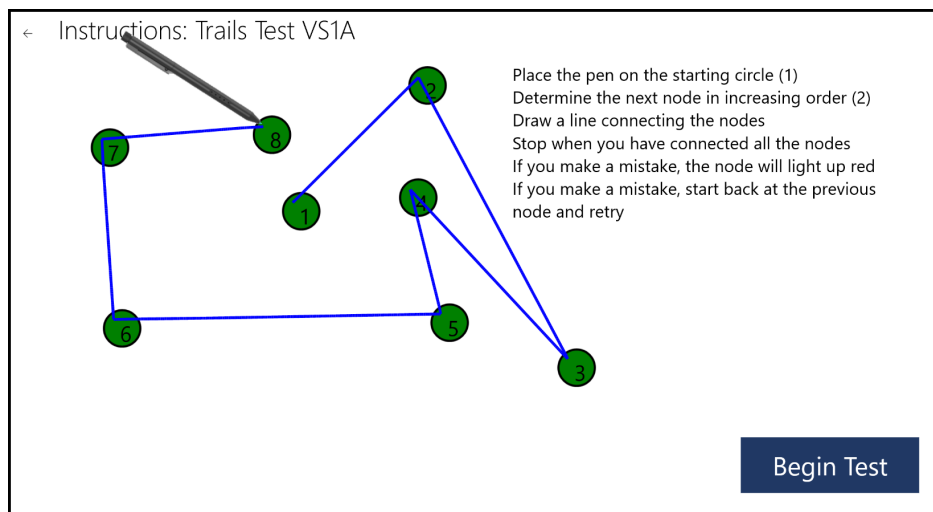
Figure 6.2: The *SmartStrokes* interface displaying the instructions for TMT-A for participants. The instructions are also delivered orally by the proctor.

A total of 8 separate Trail-Making Test layouts were converted into a digital format, comprising of 4 pairs of the "A" and "B" variations. These Trail-Making Test layouts are among those that are generally used by neuropsychologists when conducting these tests in their practice. Dimensions of the white space were cropped to account for the different aspect ratio between the Surface Pro 4 and a regular 8" 1/2 x 11" piece of paper, and the layout and size of the dots were scaled accordingly. The test interface itself resembles a paper-and-pen test as much as possible. This includes extending the drawing canvas across the entire screen, beyond the black large rectangle
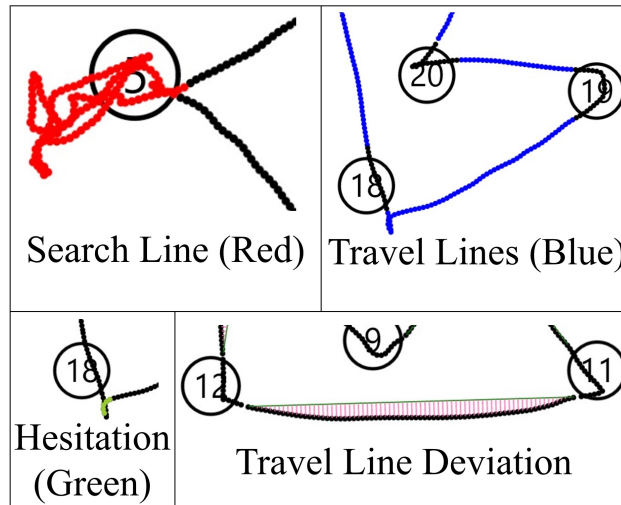
Figure 6.3: Four of the color-coded features and sketch properties that *SmartStrokes* can display. Search and travel Lines are also used to segment data for constructing the classificaton models

where the dots are placed, since in a real piece of paper some participants may draw outside of lines even when it is not advised to do so. Our intention with this interface is to capture the same types of mistakes a participant might make with a traditional pencil-and-paper modality. SmartStrokes also intentionally offers no indication of visual feedback given to participants when the next dot is connected in sequence. An earlier version of the test turned correctly connected dots green, but experts advised that feedback be only given in the case of a mistake since that is the only scenario in which a clinician would intervene. Although testing protocol dictates participants should only complete each pair once to avoid the Learning Effect, *SmartStrokes* has the ability to test each participant as many times as they wish on any arbitrary layout and order to accommodate for any testing procedure.

Completed tests can be viewed at any time if the application is signed into the proctor's profile. This allows proctors to review each participant's tests at their leisure and can also choose to "play back" the test in real-time to qualitatively review the participant's performance. Additionally, *SmartStrokes* can display color-coded visualizations of the sketch that include: separation of "travel" and "search" actions during the test, pen speed, pressure, location of "Hesitation" regions, and line straightness.

SmartStrokes also assists in data analysis by performing feature calculation of individual tests and outputting the anonymized data into a local Comma-Separated Value file (CSV). Additionally, the user can choose to automatically perform this calculation for every test associated with that proctor. This allows researchers to conduct data analytics by easily importing the CSV for rapid visualization and machine-learning analytics tasks.

6.4 Analyzing Digitized Trail-Making Tests

One of the significant challenges in analyzing the Trail-Making Test is in the proper segmentation of the data. Although the task is designed to result in simple straight lines, the ideal resulting sketch consists of a singular line making 25 stops that change direction each time. Analysis is further complicated by behaviors arising from cognitive decline, most commonly involving repeated mistakes and prolonged periods of searching for the next dot, hesitation, or doubt.

Complicated line drawings are frequently segmented in the digital sketch recognition domain in order to properly characterize key elements in the sketch. The most appropriate domain-specific method of line segmentation separates the lines in two different categories: **Search** lines, and **Travel** lines.

- **Search**: all lines drawn when the participant is looking for the next dot

- **Travel**: the line segment where the participant is actively moving from one dot to the next.

The following two subsections outline the differences between the two types of lines, what thresholds exist between the line segmentation, and which sketching characteristics we believed would be the most relevant to identifying MCI.

6.4.1 Search Lines

According to the protocol of Trail-Making tests as outlined by the Compendium of Neuropsychological Tests [2], participants are required to have their pen on the test at all times, even if they are not moving from one place to the next. This is done for two reasons. The first is that it is less likely that a participant loses their place if they do not lift their pen as they search for their next
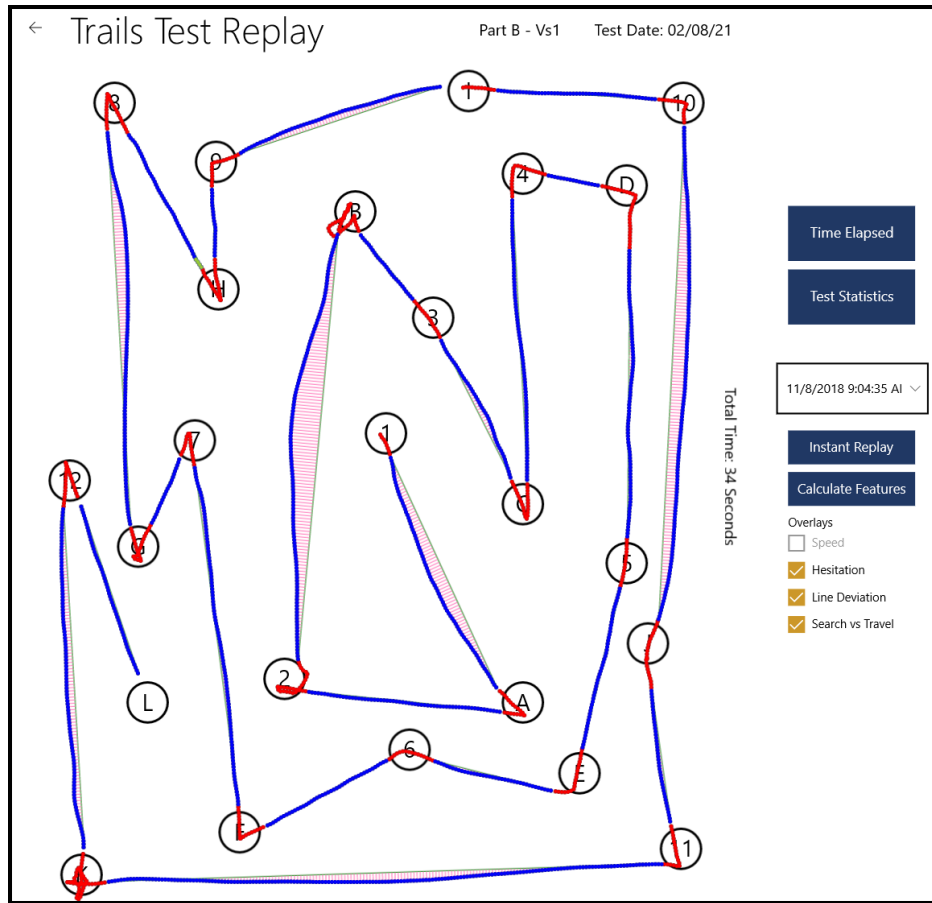
Figure 6.4: Test analysis interface of *SmartStrokes*, demonstrating line deviation and separation of search and travel lines on a completed test

dot. The second reason is that this maximizes the amount of data collection, since a participant who leaves their pen on the paper as they search for the next dot almost always results in random pen movements while they move their hand to see the rest of the test. This kind of behavior is typically characterized by noisy, erratic movement that tends to "hover" around the current dot as the participant searches for the next one. This is the kind of line that we identify as a Search line.

We define the beginning of a search line as the instant a participant enters the next correct dot in the TMT sequence. We define the end of the search line as the moment the participant identifies the next dot and moves out of the area of the current dot. We complicate the definition of the end of the search line beyond simply "outside of a dot," because of how participants behave when searching for a dot for a long time; participants who hover around a dot for a long time frequently
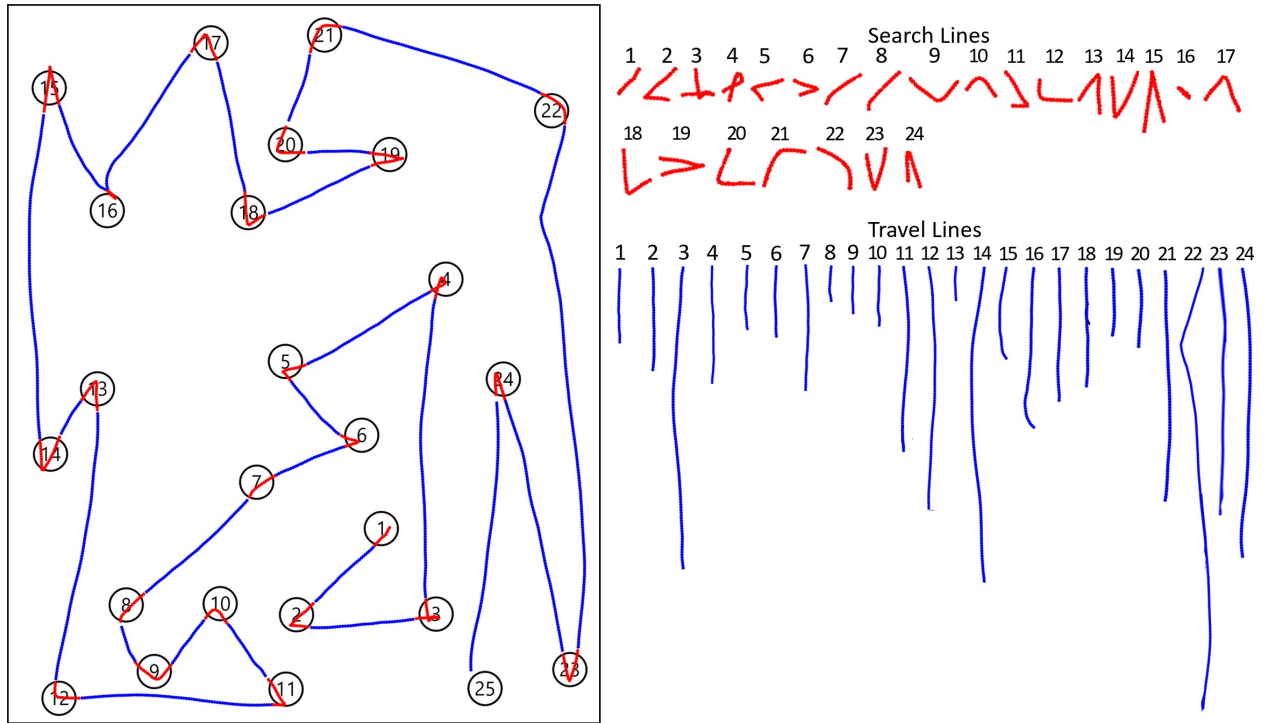
Figure 6.5: Separated travel and search lines. Travel lines are rotated to always face a top-to-bottom orientation.

move the pen inside and outside of the dot's area as they move the pen erratically. They may also stray away from the dot moments before identifying the next one. For these reasons we include an additional speed threshold outside of a dot's area as the end of the search line segment.

Healthy participants typically do not pause for long as they search for the next dot in the sequence, if at all. Indeed, search lines from typical healthy participants are usually short in length and are of a single curve clearly detailing the change in direction from the previous dot in the sequence to the next with very little or no erratic "hover" behavior. MCI participants or any other participants who find the TMT challenging typically remain in this search state for longer, resulting in longer and more erratic search line segments.

6.4.2 Travel Lines

Our parameters for defining travel lines are more straightforward, as we define travel lines as the moment the participant begins to move with intent to arrive at the next dot, and the travel

line segment ends when the next dot in the sequence is reached. When done correctly, the travel line will be a single straight line from the previous dot to the next. We implemented a pen speed threshold to identify this "intent to move" to help us clearly delineate between a search line outside of a dot's area and the moment the participant moves toward the next.

Every dot in a Trail-Making Test in sequence can be connected with a single straight line. For that reason, participants who perform well in the TMT usually have a series of travel lines drawn straighter and without turning to change direction while moving from one dot to the next. Participants who perform poorly sometimes stop in the middle of a "travel" line to either check their destination again or change direction as they realize they are going to the wrong dot.

Sometimes, such participants stop entirely in the middle of travel and begin a similar "search" behavior to find the next dot. We call all of these mid-travel stops or significant reduction in speed as "hesitation." While not every participant with MCI enters this state, several instances of these hesitation states in one test is likely to point to a poorly-performing examination

6.5 Experiment Design and Analysis

This subsection details the process by which TMT data collection was conducted, and the sketch recognition features that were selected and applied to a machine-learning classification model to detect MCI.

6.5.1 Data Collection

37 participants were recruited for data collection and classification purposes. Participants were screened and classified as MCI or healthy based on scores from the Montreal Cognitive Assessment (MoCA) [54]. The inclusion criteria for participants were the following:

- **Healthy subjects without MCI**: Healthy older adults, normal cognition. MoCA score is 26 or above. Subject group labeled as "Healthy" for model classification purposes.

- **Healthy subjects with MCI**: Healthy older adults. MoCA score is between 19 and 26. Subject group labeled as "MCI" for model classification purposes.

Table 6.1: Participant demographics for user study. 95% confidence interval for participant age is 0.94, for MoCA scores 2.49

Age Group	Male	Female	MCI	Non-MCI	Avg. Age	Avg. MoCA
55-59	1	1	0	2	57	26.5
60-64	1	5	1	5	63	26.5
65-69	5	2	5	2	67.7	23.9
70-74	3	5	4	4	71.5	25.5
75-79	4	5	7	2	76	23.3
80-84	2	0	1	1	82	25.5
85-89	2	1	3	0	85.7	21.3
Totals	18	19	21	16	71.43	24.54

- **Exclusion criteria**: Older adult subjects with a MoCA score below 19; any history of severe medical/neurological/psychiatric disease, including diabetes/hypertension; taking medication primarily targeting central nervous system; any other condition at investigator's judgment that clearly demonstrates severe cognitive decline

Additional demographic information is available on Table 6.1.

For the purposes of classification we refer to our subjects without MCI as "healthy," but we must highlight that our MCI participants are not considered "unhealthy" by contrast. All participants were given two sets of Trail-Making tests, the first in the morning and the second in the afternoon. Each test set consisted of a Trails A variant (numbers) and its accompanying Trails B variant (alternating numbers and letters). Each of the two sets used different standard dot layouts to eliminate a learning effect [324]. All participants were given the same Microsoft Surface Pro 3 device with accompanying Surface Pen to complete the digital tests.

6.5.2 Preprocessing

Several pre-processing steps are conducted on individual completed examinations. Each test's sketch data is separated by travel and search lines according to the description in Section 4. Sketch

data is then resampled to uniform interspace S using the formula:

$$S = \frac{\sqrt{(x_m - x_n)^2 + (y_m - y_n)^2}}{c} \qquad (6.1)$$

Where S is the new spacing between each sample, (x_m, y_m) is the lower-right corner of the sketch (x_n, y_n) is the upper-left corner of the sketch, and c is a constant $c = 40$.

Lastly, we implemented an additional key step in this process by normalizing individual line rotation for travel lines. The chosen features explained later in this section make significant use of sketch direction, either as a per-sample basis or the entire line. In more typical digital sketch recognition problems, features relating to direction inform a user's style of drawing, or are directly related to the type of shape that the user intends to draw. The Trail-Making Test, however, places all dots in pre-arranged locations that strongly influences the direction of a correct line. This would introduce a confounder, since differences between angles or sketch direction would not be attributed to MCI but rather the layout of the test's dots. We normalize travel lines by rotating every line such that the endpoint of the line is directly underneath the start point. This allows us to still be able to leverage direction-related sketch features to calculate characteristics like tremor, changes in direction due to mistakes made, and other types of directionality affected by the participant's performance rather than the layout of the Trail-Making Test. We are not aware of similar work in constructing a Trail-Making Test classification model that employs this segmented line direction normalization technique.

6.5.3 Feature Calculation

6.5.3.1 Rubine Features

We implemented a combination of digital sketch recognition features known to yield accurate models in similar research projects. The first set of 13 features introduced by Rubine *et al.*, abbreviated as "Rubine features" [103]. The 13 features were first introduced alongside a recognition technology named GRANDMA (Gesture Recognizers Automated in a Novel Direct Manipulation Architecture), a toolkit that sought to provide users with the ability to train any
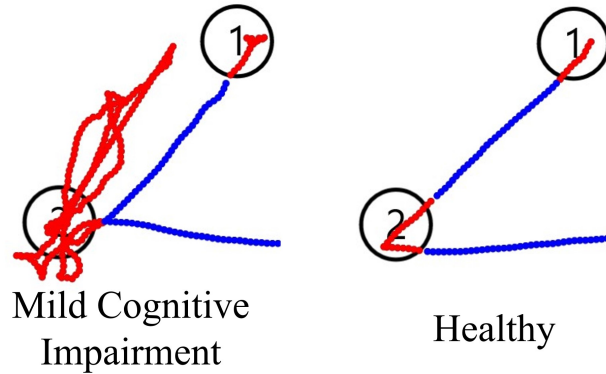
Figure 6.6: A clear example of the search Line difference between an MCI participant and a healthy one. This discrepancy is usually the result of the participant unable to locate the next dot in the sequence for an extended period of time. Although the discrepancy is obvious in this example, not all MCI participants exhibit this behavior, making diagnosis challenging.
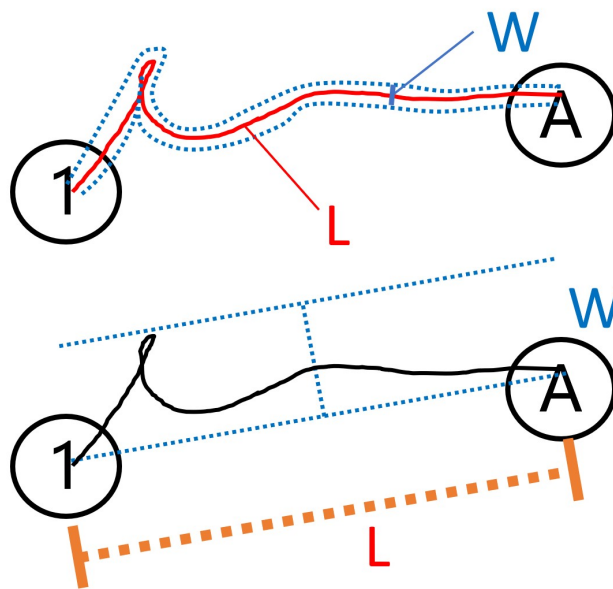
Figure 6.7: The traditional application of the Steering Law is on top, with *W* and *L* being predetermined. Our use of Steering Law, on bottom, creates a simple tunnel with *W* based on the total "width" of the pen trajectory.

gesture for recognition using a click-and-drag interface. The Rubine features themselves have since then been implemented in various sketch recognition projects that can gauge not only the type of shape that is drawn, but also the cognitive state of the participant who drew them. Rubine

Table 6.2: Rubine features f_1 through f_{13}. Let $\Delta x_p = x_{p+1} - x_p$, and $\Delta y_p = y_{p+1} - y_p$, and $\Delta t_p = t_{p+1} - t_p$

Rubine Features	
$f_1 = \dfrac{x_2-x_0}{\sqrt{(x_2-x_0)^2+(y_2-y_0)^2}}$	$f_8 = \sum_{p=1}^{P-2} \sqrt{\Delta x_p^2 + \Delta y_p^2}$
$f_2 = \dfrac{y_2-y_0}{\sqrt{(x_2-x_0)^2+(y_2-y_0)^2}}$	$f_9 = \sum_{p=1}^{P-2} \theta_p$
$f_3 = \sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2}$	$f_{10} = \sum_{p=1}^{P-2} \lvert \theta_p \rvert$
$f_4 = arctan\dfrac{y_{max}-y_{min}}{x_{max}-x_{min}}$	$f_{11} = \sum_{p=1}^{P-2} \theta_p^2$
$f_5 = \sqrt{(x_{p-1} - x_0)^2 + (y_{p-1} - y_0)^2}$	$f_{12} = \max_{p=0}^{P-2} \dfrac{\Delta x_p^2 + \Delta y_p^2}{\Delta t_p^2}$
$f_6 = \dfrac{(x_{p-1}-x_0)}{f_5}$	$f_{13} = t_{P-1} - t_0$
$f_7 = \dfrac{(y_{p-1}-y_0)}{f_5}$	$\theta_p = arctan\dfrac{\Delta x_p \Delta y_{p-1} - \Delta x_{p-1}\Delta y_p}{\Delta x_p\Delta x_{p-1}+\Delta y_p\Delta y_{p-1}}$

features f_1 and f_2 specify the cosine and sine features of the first few samples, usually limited to the first two samples as was done in our implementation. The bounding box diagonal of the entire gesture is analyzed as features f_3 and f_4. The distance in pixels between the first and the last point is specified in feature f_5. The difference between the first and last point of a gesture is analyzed through features f_6 cosine and f_7 sine between the start and end points, the total length of the gesture is calculated for f_8, and the total angle traversed is f_9. Three total summations are calculated, with f_9 being the total angle traversed over the course of the gesture, f_{10} being the sum of the absolute value of the angle per mouse point that does not take into account direction, and f_{11} being the sum of the square of the value of f_9. The square of the maximum speed achieved in the gesture is f_{12}, and the last feature f_{13} is the total duration of the gesture, measured in milliseconds. The calculations for the Rubine features are provided on Table 6.2.

The Rubine features represent the various geometric properties of any given gesture. They can measure speed, curvature, direction at the start and ends of the gesture, total time taken, and the properties of the total area (referred to by Rubine as the "bounding box") of any particular gesture. These features offer an alternative to template-matching recognition in that they do not require a point-for-point comparison, but rather are geometric calculations of the gestures themselves. Although these have been used mostly for recognizing gestures, their frequent use in recognizing

shapes provides us with an opportunity for analysis of cognitive impairment.
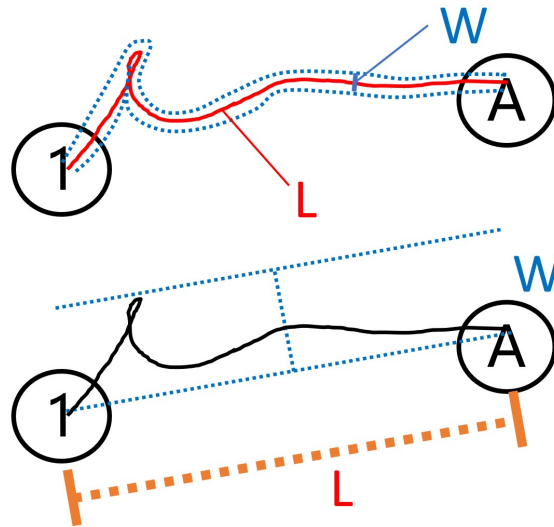
Figure 6.8: The traditional application of the Steering Law is on top, with *W* and *L* being predetermined. Our use of Steering Law, on bottom, creates a simple tunnel with *W* based on the total "width" of the pen trajectory.

6.5.3.2 *Fitts' and Steering Law Features*

We leverage principles from Fitts' Law by calculating that law's Index of Performance [325]:

$$ID_F = \log_2 \frac{2D}{W} \tag{6.2}$$

Fitts' Law was originally conceived as a method to quantify pointing task complexity [326–329] and has has been widely used in HCI research, particularly UI navigation tasks [330–336]. Because Fitts' Law is rooted in tracing lines across distances between targets and measure that task's complexity and a user's performance [337–339], we believed it could be leveraged to help identify task performance.

A related feature we use is the more recent variant, the Steering Law. The Steering Law assesses the difficulty of a user navigating a pointer through a path with a set width [340–344].

For a generic tunnel C, and a width $W(s)$ along the path, the Steering Law's Index of Difficulty ID_S is:

$$ID_S = \int_C \frac{ds}{W(s)} \tag{6.3}$$

For our purposes, we use a straight path of length L and a constant with W as defined by Pastel et al. [345], which reduces ID_S to:

$$ID_S = \frac{L}{W} \tag{6.4}$$

By using the user's input lines as the basis for calculating W, we essentially create a form of performance index using the Steering Law. For the Trail-Making Test, a narrower line width W is straighter and effectively more difficult to recreate. We integrated this metric as a feature for the classification model to test whether a participant with MCI would create lines with a generally lower ID_S. We also scaled and averaged ID_F and ID_S as a separate feature to explore a possible combination of the two.

6.5.3.3   Additional Behavioral Features

Hesitation is a feature that we briefly discussed in section – as a feature unique to travel lines. It characterizes the prevalence of stop-and-go motion for participants who start connecting a dot but stop or slow down significantly while inside a travel state. *Hesitation* begins when the pen slows down to an empirically-derived speed of 0.4 over five consecutive sampled points, and our calculated feature is distance the pen traveled while the pen remains in this state. The pen exits this state when at least five consecutive sampled points have a speed above 0.4. If the pen enters *Hesitation* state multiple times inside a single travel line, the total distance across all of these states is reported for the one travel line

Line Ratio is a feature meant to normalize the length of a participant's drawn line. We believe the length of the line is important to understand how confident and accurate the lines were connected, since meandering behaviors and course correction would naturally result in a longer line than a straight line drawn directly from dot to dot. However, a drawn line will also be longer if the

correct dots are placed further apart. The Trail-Making Test is explicitly designed to place dots a variety of distances from each other to measure a participant's ability to identify dots that might be further away from their immediate location. To take *relative* line length into account we divide the total distance drawn from one dot to the next by the theoretical "perfect" line drawn from one dot to the next. The closer the number is to 1, the closer to "perfect" this distance becomes and the better a participant performs. The formula for Line Ratio R_{ln} is found below, where (x_n, y_n) is the final sampled dot of the input line:

$$R_{ln} = \frac{\sqrt{(x_n - x_0)^2 + (y_n - y_0)^2}}{\sum_{i=0}^{n} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}} \tag{6.5}$$

Pen Lift Time is the amount of time during each segment that the user lifts their pen. Although participants are required to leave their pen on the tablet at all times as per the instructions of the Trail-Making Test, some participants still absent-mindedly lift the pen when searching for a dot or when correcting a mistake. This feature is intended to capture the behavior of both of these scenarios to explore a possible correlation with MCI.

Pen Pressure Average and *Pen Pressure Standard Deviation* are features pertaining to the pressure that a participant places on the pen as they complete the test. We wanted to explore the possibility that a participant places more pressure on the tablet if they are unsure of their trajectory or if the test is difficult for them to complete.

We complete the feature set by adding a few sets from existing sketch and gesture recognition literature. We implemented 11 features from Long *et al.* [215] as a supplement to the Rubine features for general-purpose sketch recognition. Alamudun *et al.* [218] applied Rubine and Long features and added two direction-based features to help with saccade detection in an eye-tracking task, but have we believe can also be implemented as general-purpose sketch recognition as well. Finally, Paulson *et al.* introduced two features, normalized distance between direction extremes (NDDE) and direction change ratio (DCR), as general-purpose sketch recognition features that we also included for this study [216, 346].

Table 6.3: Classification features. **Model** describes whether the feature was used for the model for classification of travel (T) or search (S) lines. Some features were excluded due to high collinearity and/or were inappropriate for a specific model.

Name	Model	Name	Model	Name	Model	Name	Model
rubine1	T	rubine10		avgPressure	T+S	openness	T+S
rubine2	T	rubine11		stdevPressure	T+S	boundBoxArea	
rubine3	T+S	rubine12	T+S	avgSpeed	T+S	logArea	T+S
rubine4	T+S	rubine13	T+S	stdevSpeed	T+S	rotRatio	T+S
rubine5	S	fitts		aspect		lengthLog	T+S
rubine6		steering	T+S	curviness	T+S	aspectLog	T+S
rubine7		lineRatio	T	relativeRot	T+S	fittsSteering	
rubine8	S	hesitation	T	densityMetric1		ndde	T
rubine9	T+S	penLiftTime	T+S	densityMetric2	T+S	dcr	

6.5.4 Model Construction

Because Trail-Making Test behavior is characterized by the distinct actions of travelling to the next line and searching for the next, we decided to produce two separate classification models to explore the possibility of either being more indicative of MCI and compare their performance. Additionally, because the actions yield different behaviors, not all features were applicable for both types of actions. For example, line direction is important for travel lines to identify incorrect line deviation after we normalize travel lines as shown in Fig. 6.5. However, search lines cannot be normalized since direction at entry and at exit of a dot, even for healthy participants, depends heavily on the test layout itself. Table 6.3 lists every feature initially integrated into the feature set, and the subscripts next to the feature names indicate which were chosen for the models.

Some features were also removed from search and travel classification models due to a high collinearity value (>0.90). Fig. 6.11 shows the collinearity heatmap of the remaining features that were used for both search and travel lines. Further, all values were normalized between 0 and 1.

Every segmented line from the 149 tests is included and is given the label according to the participant's cognitive state (MCI or healthy). 3,490 search lines and an equal number of travel lines were used for their respective classification models.

Models were constructed according to a 90/10 split for a 10-fold cross-validation. The models were trained and evaluated according to the two labels of MCI or healthy assigned during the screening phase of the study. We used 7 binary classification models to compare performance.

Recursive Feature Elimination vs. Accuracy for MoCA Prediction (Search)
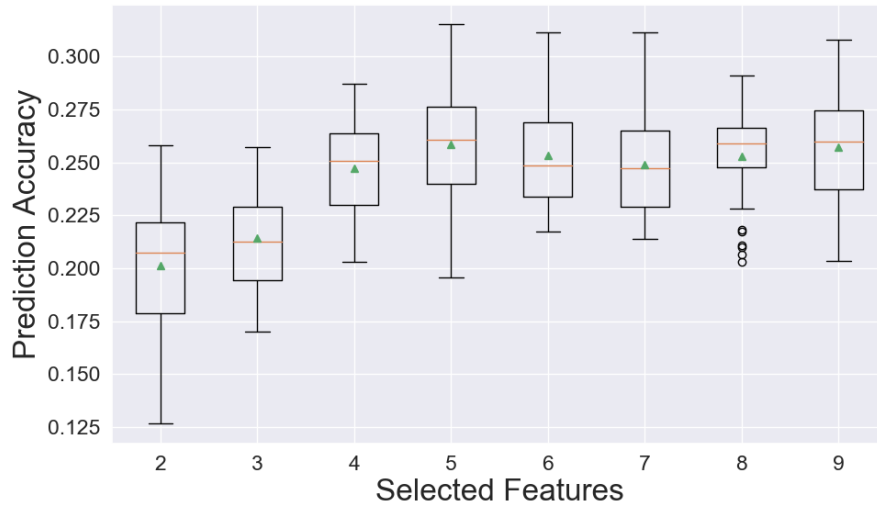
(a) Feature amounts chosen for Search.

Recursive Feature Elimination vs. Accuracy for MoCA Prediction (Travel)
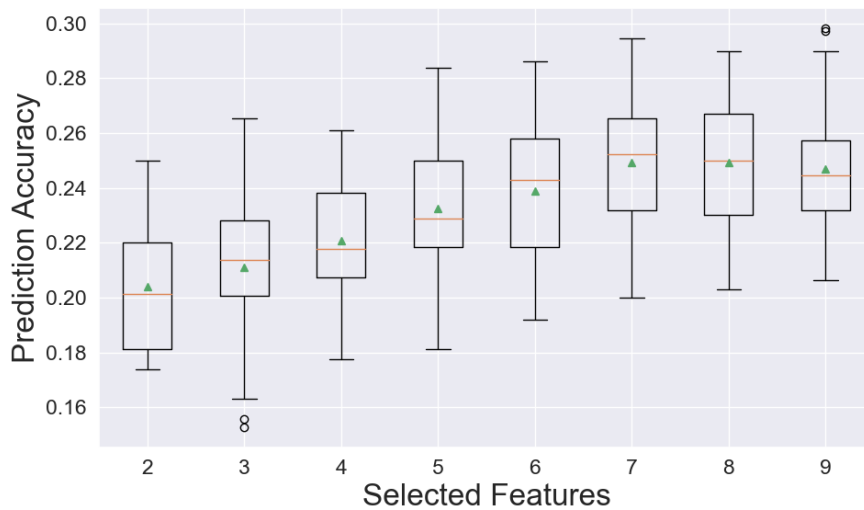
(b) Feature amounts chosen for Travel.

Figure 6.9: Box plot of different feature amounts chosen for Recursive Feature Elimination.

6.5.5 Prediction of MoCA Scores

Participant labels of "healthy, without MCI" and "healthy, with MCI" essentially is dividing participants between two broad categories of MoCA scores. Section 6.5.1 specifies the categories are a MoCA score of 26 and above for "healthy, without MCI" and between 19 and 26 for "healthy, with MCI." This in effect means our classification attempts to predict a wide range of the MoCA scores of the participants. However, we also sought to more directly predict the MoCA score in a more granular fashion as part of the data analysis of this study.

Table 6.4: Features chosen by Recursive Feature Elimination to directly predict MoCA scores.

Search Model RFE Features	Travel Model RFE Features
aspectLog	*avgPressure*
avgPressure	*avgSpeed*
avgSpeed	*rubine12*
logArea	*rubine13*
rubine11	*stdDevPressure*
stdDevPressure	*steering*

Our approach to MoCA score prediction is similar to the prediction of broad categories in that we are using the same training and classification features, and the same 90/10 split for 10-fold cross-validation. The similarity also extends to the training and classification being performed on the individual lines, and the F1-score and accuracy being calculated on how closely each **line** is being predicted to the actual MoCA score associated with that line's entire test. This is distinct from other methods of classification that seek to analyze the entire page and create a single prediction. We believe that participants of Trail-Making Test do not perform evenly through the entirety of the test; while they might perform well for a few dots, a single dot might prove to be difficult for participants to find. Indeed, several participants from our empirical observations who performed poorly would find certain dots easy while finding others significantly more difficult. Our wish to capture this particular type of behavior is the reason behind our use of the segmented

lines. We believe per-line sketch analysis and predictions might yield a novel insight into the participants' behavior. Because the original scoring system was conceived at a time when granular sketch analysis was not possible, we believe per-line analysis can provide a more granular and complete picture of a participant's behavior during the test. Prediction was trained and tested on both the Travel and the Search models.

We performed a Recursive Feature Elimination (RFE) on the Travel and Search Models to determine the top-ranking features that can be included in a logistic regression to predict the MoCA scores. The number of features ideal for both was determined to be 6 since that is the number of features where the accuracy plateaus for both Search and Travel models. The features selected for the Search and Travel models are listed in Table 6.4, and the box plot depicting the comparison of feature selection to accuracy is shown in Figure 6.9.

A logistic regressor with an iteration of limit of $n = 100000$ was employed for both models to predict the MoCA scores based on the features selected by the RFE. We then used a repeated K-Fold cross-validator, with a 90/10 split repeated 3 times for a total of 30 comparisons in the calculation of the predictors. To gauge the performance of the predictions being made, we calculated the average Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) of the predicted vs. the actual MoCa test scores for all of the lines. In this prediction algorithm all of the segmented lines of test and training participants have been labeled their respective MoCA scores. Although MoCA is not typically labeled on a per-line basis, our experiment is to determine whether such a prediction can be accurately made on per-line granularity. MAE and RSME were both used to help determine the mean error between the predictions of the logistic regressions. Both of the Travel and Search prediction algorithms had RSME and MAE as well as their standard deviation calculations and are shown in Table 6.6.

Table 6.5: Classification metrics. Acc is accuracy, F1 is F1-score, Prec is precision. For both the travel lines and search lines models, **n=3,490**

	Travel Lines				Search Lines			
Classifier	Acc	F1	Prec	Recall	Acc	F1	Prec	Recall
Majority	0.51	0.51	0.50	0.50	0.53	0.53	0.52	0.52
Gaussian Naive-Bayes	0.47	0.36	0.60	0.53	0.47	0.38	0.58	0.53
Decision Tree	0.59	0.59	0.58	0.58	0.60	0.60	0.59	0.59
K-Nearest Neighbor	0.60	0.60	0.59	0.59	0.58	0.58	0.57	0.57
Linear Regression	0.65	0.64	0.65	0.63	0.62	0.59	0.61	0.58
SVM	0.65	0.63	0.66	0.62	0.63	0.61	0.64	0.60
LDA	0.65	0.63	0.65	0.62	0.62	0.60	0.61	0.59
Random Forest*	**0.67**	**0.73**	**0.67**	**0.80**	**0.66**	**0.72**	**0.68**	**0.77**

6.6 Results

6.6.1 Accuracy Metrics of MCI Prediction

The main results of model performance are reported on Table 6.5, meant to report on how well a classification model trained and tested on travel lines and search lines independently is able to identify whether the author of those lines had MCI or was a healthy participant. A total of eight different classification models, listed in the **Classifier** column on the table, were trained with the features listed in section 5.3. These are classifiers that are most commonly used in either binary classification tasks or have historically performed well with digital handwriting recognition: Majority classifier [347] as a baseline, Gaussian Naive-Bayes [348], Decision Tree [349], K-Nearest Neighbor [350], Linear Regression [351], SVM [352, 353], LDA [353], and Random Forest [317, 354]. Results are reported for both the search line model and the travel line model, and we report the models' accuracy, F1-score, precision, and recall. For both travel and search lines, Table 6.5 shows that the best performing models were created using a Random Forest classifier. Additionally, pressure-related features had among the highest feature importances when analyzing drop-column importances for the random-forest classifiers.

As a point of comparison, we refer to the Specificity/Recall metrics of diagnostic classification

accuracy as provided by the study led by Ashendorf *et al.* reports a Specificity of TMT-B based on number of errors alone is 70.3% [47], whereas our per-line classifier achieves a specificity of 80% if classifying travel lines and 77% if classifying search lines. It is important to note that our model does not differentiate between TMT-A and TMT-B, which emphasizes the layout-agnostic element of these classification models. While this makes direct comparisons non-trivial with accuracy metrics that limit their accuracy statistical analysis to a subset of test layouts, we are still able to report on some level a comparison of specificity with non-machine-learning classification efforts.
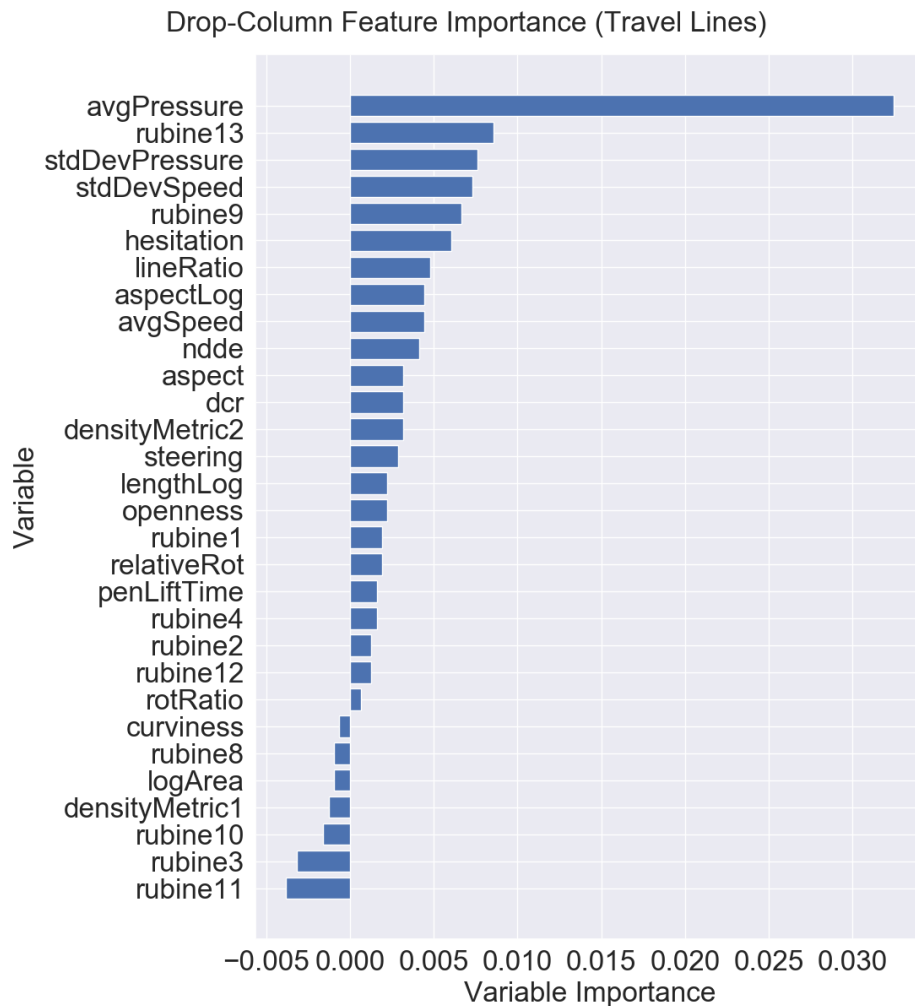
Figure 6.10: Feature importance metrics of the Random Forest classification model built for travel lines
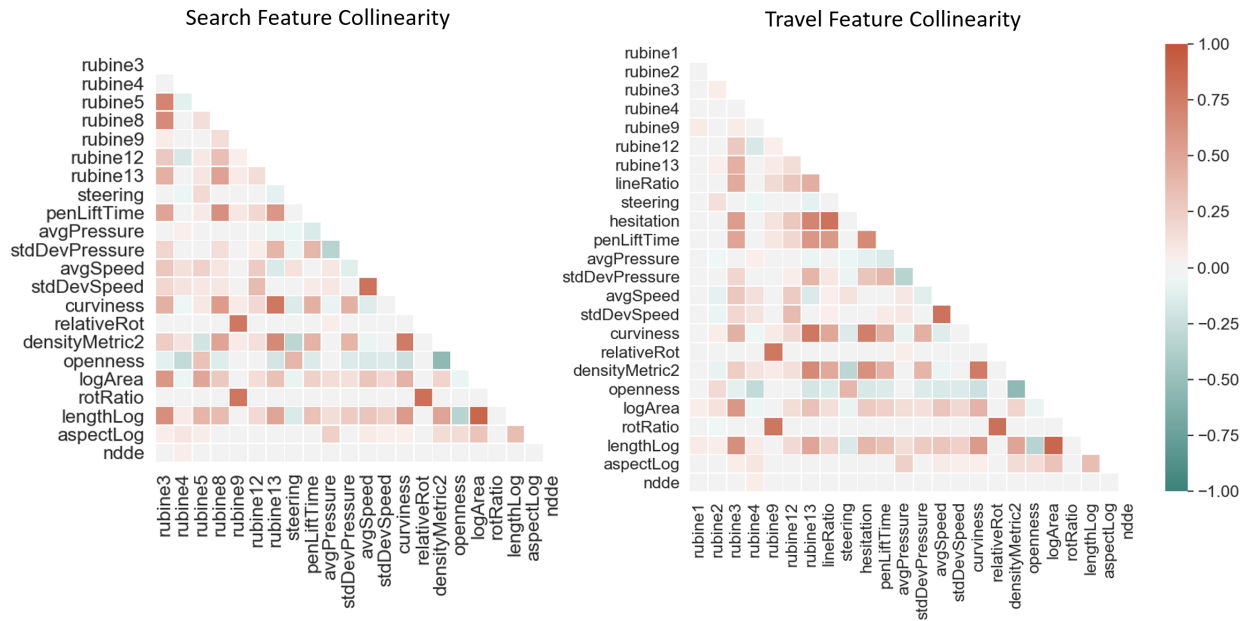
Figure 6.11: Feature collinearity for both search and travel lines. Features with collinearity above 0.9 were removed from the model

6.6.2 Accuracy Metrics of MoCA Score Prediction

Two sets of metrics can be reported for the MoCA score prediction: the results of Recursive Feature Elimination and how the number of features affects the prediction accuracy, and the results of the average MAE and RSME of the predictions made on the test data. Prediction of the MoCA, as opposed to the prediction of MCI, is non-binary and more of a continuous set of data in nature. For this exercise we allowed fractions of numbers to be predicted, since our chief method of comparison is the calculation of RSME and MAE. Small discrepancies in MoCA scoring due to the inclusion of non-whole fractions would be minor, if that were the chief difference between predicted and actual scores. The accuracy metrics are reported in Table 6.6.

6.6.3 Discussion

6.6.3.1 Mild Cognitive Impairment

One of the primary challenges in detecting MCI is the inherently subtle nature of changes. Research such as that of Zhang *et al.* [37] outline difficulties in formalizing behaviors that correlate

Table 6.6: Root Mean Squared Error (RSME) and Mean Absolute Error (MAE) of predicted points of MoCA scores.

Error Metric	Travel Lines		Search Lines	
	Average	Std. Dev	Average	Std. Dev
RSME	3.325	0.132	3.315	0.134
MAE	2.415	0.110	2.406	0.105

significantly with the manifestation of MCI in the Trail-Making Test. Depending on the severity of cognitive decline and multiple factors in how MCI affects each participant, they may not find the TMT specifically that challenging. For that reason, it is generally believed that the TMT, while proven sensitive to MCI in many cases, is not alone the only tool needed to reliably detect MCI.

The results from the accuracy metrics of the travel and search lines supports the notion that detecting subtle levels of MCI is inherently challenging if only analyzing one test. In several of our observed cases, participants who we classified as just under our MCI threshold based on their MoCA score completed the test in a similar manner as a typical healthy user. In these cases a clinical neuropsychologist would continue testing their patient with several other kinds of exams or use the Trail-Making Test to primarily to identify other conditions of cognitive decline. This differs from other digital sketch recognition problems where the exhibited behaviors are not subtle by nature, or if the goal is to differentiate between discrete shapes. Models for those problems typically result in much higher accuracy and F1-scores (closer to >0.9) since the labels are more cleanly delineated.

Overall, we believe the results present a meaningful contribution on the analysis of MCI through the TMT, largely due to the analysis and model construction on a per-line basis. Our implementation refined steps to segment the sketches by integrating speed thresholds to identify when the participant has found the next dot in the process. Whereas previous work in analyzing digitized TMT sketch data tends to average behaviors over an entire test, we sought to leverage the high-granularity nature of sketch data to provide analysis on individual lines. Our contribution also extends to the normalization of line direction and total length to avoid differences between lines

that are due to the TMT individual dot locations. The key is to eliminate potential confounders introduced by the fact that the TMT stimulates all participants to change line directions and total line length. We chose not to map a "perfect" line for each of the different segments to gauge performance, since Trail-Making Test layouts are numerous and clinicians frequently use modified versions for their own purposes. We sought to create a classification model that would work regardless of the dot layout to avoid creating a model that only works on that specific layout. Ultimately we sought to explore whether segmented lines could individually be labeled as MCI or healthy with at least similar performance as existing work.

A popular method for creating behavioral models is in the leveraging of deep learning techniques such as neural networks. These techniques are becoming more prevalent due to its ease of deployment for large data sets and higher efficacy in classification. However, we did not believe these techniques appropriate for this experiment for two primary reasons. The first is due to the necessity of collecting a considerably larger dataset for the creation of a classification algorithm using deep learning techniques. Challenges related to the proper collection of data for this experiment are explained in the following section. The second reason is due to the lack of explainability in deep learning techniques. While it would be possible to produce a more accurate behavioral model provided we acquire a considerably larger dataset, we would be unable to explain to a clinician which behaviors of the participant are responsible for the conclusion they are likely to have MCI. We believe that behavioral analysis in these types of domains should be usable to domain experts, thus motivating the manual creation of features to explain behavior.

6.6.3.2 *Montreal Cognitive Assessment Score Prediction*

Challenges of the MoCA score prediction were similar to those of predicting MCI, but were exacerbated by the labeling of a single score point to every line. Per-dot line segmentation resulted in likely an unbalanced training set, since a small subset of participants who performed fairly well the MoCA could skew the training and test sets considerably. This unevenness in MoCA distribution suggests that a much larger and wider range of MoCA scores is needed for accurate score prediction. As it stands, the Recursive Feature Elimination for both models as shown in

Figure 6.9 suggests that even the optimal amount of chosen features yields only an accuracy above 0.35 for the Search model and an accuracy of up to 0.30. For this current version of the calculated features and those chosen by the RFE, we believe that additional features and changes to the existing ones would be necessary to increase the prediction accuracy.

At present the results for predicting MoCA scores are inconclusive. The errors as reported in Table 6.6 might suggest an average error of about 10% given the range of the MoCA scores to be from 0 to 30 points. The demographic data shown in Table 6.1 and discussed in section 6.5.1 shows an average MoCA score of 24.54 for all participants as well as the total overall criteria for inclusion of participants from 19 onward. Due to limitations on protocol safety, we are unable at present to recruit and test for participants with more severe cognitive impairment for participants who score below 19. This is largely due to safety protocols requiring participants below that age to be accompanied by a guardian or healthcare official, since institutional review boards consider severely cognitive impaired individuals who would be unable to provide informed consent of their own volition. Following the safety protocols fortunately does not impair significantly the prediction of MCI vs. non-MCI populations since MCI participants are still able to provide informed consent, but this does reduce the efficacy of predicting MoCA performance as a continuous score. In order to create a more accurate MoCA predictor, we will require a larger corpus of data with a more even distribution of MoCA scores such that the scores are more evenly distributed as per established normative data. At present the inclusion of an MCI predictor did somewhat limit the performance of a MoCA score predictor.

6.7 Limitations and Future Work

One of the main challenges in building an accurate predictive behavioral model is the creation of a new dataset for that specific purpose. Despite the fact that the Trail-Making Test has been in use for several decades, the granularity of digital data and the requirement of a digital pen necessitated the creation of a new dataset. The prevalence of different Trails Test layouts and the small differences of protocol that vary from clinician to clinician also necessitated a unified testing protocol. Accompanying this challenge is the laborious recruitment process. Although the task is

simple, the administration of the MoCA and the proper administration of the Trail-Making Test resulted in a slower rate of data collection that is typical of sketch recognition tasks.

Currently, the age ranges of the Trail-Making Test's normative data, as found Tombaugh's stratified normative data for paper-and-pencil Trail Making Tests [31], divides the age range into 11 distinct categories. Our normative data covers the latter 7 bins with our participants ranging from 57 to 86 years of age. Since the focus of this experiment is in identifying MCI among middle aged and older individuals, the study focused on that age range. Future studies will continue the data collection process to build a more complete normative body of data across all age ranges. These might potentially result in differing behaviors between patients with MCI from different age ranges, but a solid body of data from those age ranges is necessary for verification. We also aim to further expand on localizing areas that were difficult for participants with MCI, reporting these lines on a UI-level in real time and evaluating a clinician's diagnosis experience with such an automated tool.

Also of note is the fact that a protocol of collecting data on an MCI population inherently removes a full range of ages and conditions for normative data. This impacts the ability for a prediction system to make a machine-learning-based prediction of the actual MoCA score. The prediction of the MoCA scores yielded relatively small error percentages but when taking into account the reduced range of MoCA scores that were available for testing and training, we conclude the results for direct prediction of exact MoCA scores are inconclusive despite being somewhat promising. We are considerably more confident about the binary classification between non-MCI and MCI populations precisely because the range of data and the collection protocol yielded the most appropriate data for that kind of classification. Future work will require a wider range of participants with normative data closer to that of Tombaugh *et al.* [31]. We are confident the digital sketch data from digital TMTs can be used to make much more accurate predictions about the participants' MoCA scores if said data were available.

Subtle changes in behavior due to Mild Cognitive Impairment continue to present significant challenges in identifying the earliest possible signs for conditions that may lead to dementia and

Alzheimer's disease. Existing efforts to aid in this challenge highlight the difficulty of finding the nuances in behavioral changes present in a Trail-Making Test. However, with significant improvements over previous efforts we present a solution that suggests individual lines, regardless of their direction, can distinguish between MCI and Healthy with noticeably higher levels of accuracy. We look forward to employing additional preprocessing methods, features, and a larger digital sketch dataset to further improve on this effort. We believe sketch data from the Trail-Making Test still has the potential to yield insights into behavioral changes that are yet to be discovered.

7.1   3D-Trail-Making Test: A Touch-Tablet Cognitive Test to Support Intelligent Behavioral Recognition

This chapter contains previously published work* [355].
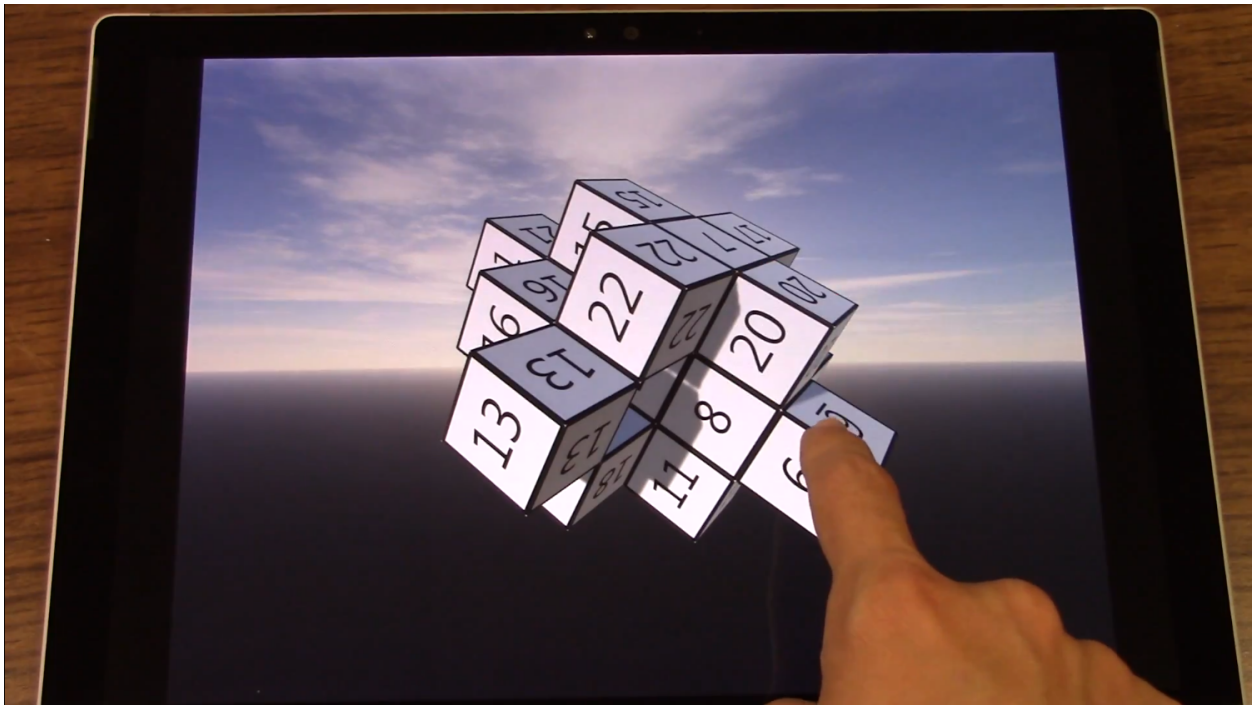
Figure 7.1: A participant completing the A variation of the 3D-Trail-Making Test.

7.2   Motivation

The decades-old Trail-Making Test has established itself as an effective and versatile cognitive testing tool. However, its reliance on a paper-and-pencil method of administration severely limits its capabilities for quantitative evaluation. We evaluate input data via sketch recognition algorithms

on a digitized version of the Trail-Making Test, as well as exploring the viability of a novel touch-based version designed around the basic concepts of the traditional paper-and-pencil exam.

For this project we were interested in exploring the creation of a novel cognitive test, one that would only be able to be performed and administered via a mobile device such as a tablet. Typical applications of neuropsychological tests in a digitized space adapt existing tests to work with current technology, but we believe a true intersection of human-computer interaction digital user experiences and neuropsychological testing have remained relatively unexplored.

One of the reasons for the lack of exploration is in the fact that for every novel test or variant of an existing one, new normative data must be constructed. In practice, neuropsychologists must have an established body of past data from which to draw comparisons to assess whether a patient's performance is within parameters of neurodivergence whether these may be inherited, a result of trauma, or as part of cognitive decline.

This work aims to examine the viability of a novel neuropsychological test by evaluating whether a proposed novel test is sensitive an aspect of cognition an existing test is already sensitive to. Existing work cited below shows the Trail-Making Test is sensitive to cultural and ethnic background, since behaviors informed by culture and heritage result in changes to normative data. We believe if a proposed novel test may also elicit a cognitive response that differs between cultural backgrounds from an adapted test, there may yet be a path to create novel digital tests adapted from existing paper-and-pencil tests. This would allow further research into creating tests that may examine certain parts of cognition currently not feasible in paper and pencil tests, mainly dynamic tests that change significantly on user input and touch-based examinations.

7.3 Neuropsychological Influence

Neuropsychologists employ a variety of cognitive examinations through which to observe a patient's behavior. Some of these involve manipulating physical objects, answering a series of oral questions, solving exercises that integrate drawing exercises. These tests typically involve paper and pencil, and the patient is required to complete an examination by sketching, drawing on objects, crossing out or circling shapes, and other forms of interaction using pen on paper.

The TMT was conceived and originally introduced as a testing tool in the field of clinical neuropsychology to assess general intelligence. Shortly thereafter, clinicians found utility in the test's ability to aid in the detection of cognitive abnormalities. Clinicians who administer the Trail-Making test observe the patient's behavior according to a predetermined checklist. Among the items include sitting posture, how well they maintain eye contact, reaction to mistakes, and their efforts in remembering the next item in the sequence among others [2]. Tests produce one quantitative performance metric in the form of the time taken to complete the test, rounded to the nearest second. A clinician compares the score against established normative data depending on which age classification the patient belongs to. The lengthy diagnosis process combined with the increased emphasis for early detection due to the climbing proliferation of Alzheimer's disease and other types of dementia[†] has maintained the TMT's relevance in modern diagnoses.

7.4 Design

Conceptually, our novel 3D-Trail-Making Test is an intelligent cognitive testing computer application designed to leverage modern mobile capabilities to ultimately facilitate behavioral recognition. Paper-based examinations are used due to the ease of administration and quick completion times, but they lack the inherent ability to test for more complex cognitive functions such as difficulties with depth perception and 3-dimensional spatial relationship of objects. Mendez *et al.* describes the ways in which neuropsychologists test for these difficulties, which still involves drawing and observing flat images on a paper to simulate 3-dimensional perception [356]. Other forms of testing visuospatial capabilities have included complex experiments, such as the work of Prvulovic *et al.* which involved custom made foam pads, projectors, frosted screens, and several different stimuli per participant [357]. The complexity of using custom-built objects and the fact that experts already test depth perception from simulating the perception of depth on a flat surface motivated the design of a touch-based mobile application on a tablet.

After careful consideration of the existing technology applications for early diagnosis, we identified five key components for a new examination tool:

[†]Centers for Disease Control and Prevention: www.cdc.gov/features/alzheimers-disease-deaths/index.html

- Leverage advantages of existing clinical neuropsychological diagnosis methods by **adapting an existing paper test** into a more modern examination

- **Preserve** the paper test's ease of administration and speed of test completion while testing **new aspects** of a participant's cognitive abilities that a paper test cannot.

- Carefully design the user experience to develop a **streamlined, intuitive interface** that captures a user's **behavior** rather than examine their proficiency with consumer electronics

- **Capture** usage data on a scale appropriate for machine learning classification

- Provide behavioral usage data that can be used to **intelligently** identify differing behaviors among participant groups

7.4.1 Design Overview

This 3-dimensional test contains both versions of the traditional Trail-Making test, wherein version A has all the boxes labeled in ascending numbers, and version B has the boxes alternating between numbers and letters. Like in traditional 2-dimensional paper-based exams, the correct sequence is different for both the A and B variations and participants complete both exams as a pair. This test was created on the Unity game engine platform and compiled to run on Windows 10.

7.4.1.1 Input

The classic Trail-Making test serves as the primary basis for this new examination with some differences to better leverage modern tablet technology in the diagnosis process. Rather than placing every labeled dot on a flat 2-dimensional surface with every dot visible to the participants at once, we have rendered every "dot" as interactive 3-dimensional boxes. The boxes are arranged in a spherical pattern and float in an empty virtual space, and with one finger participants drag anywhere on the screen to rotate the sphere on every axis at its center point. The rotation was designed specifically to emulate the rotation gestures of Google Earth [‡] due to its familiarity and

---
[‡]Google Earth: google.com/earth/

111

simplicity. Rotation speed is calculated by directly measuring how much the finger moved across the screen (X and Y, measured in pixels) each frame, multiplied by a constant of 0.5 after testing different rotation speeds in pilot studies. With the same finger, the participant taps on the next box in the sequence. If the correct box is tapped, it disappears with a shrinking animation to clearly indicate correct input. If the incorrect box is tapped, it changes its color to red and remains static to indicate a mistake has been made. When boxes disappear, they reveal more boxes hidden in deeper layers of the sphere. This design is intentional since our goal is to create a dynamic examination wherein not every number or letter is visible at once. This encourages participants to be closely engaged with the exam at all times, since the dynamic nature of the sphere geometry will require their constant re-evaluation of the state of the exam.
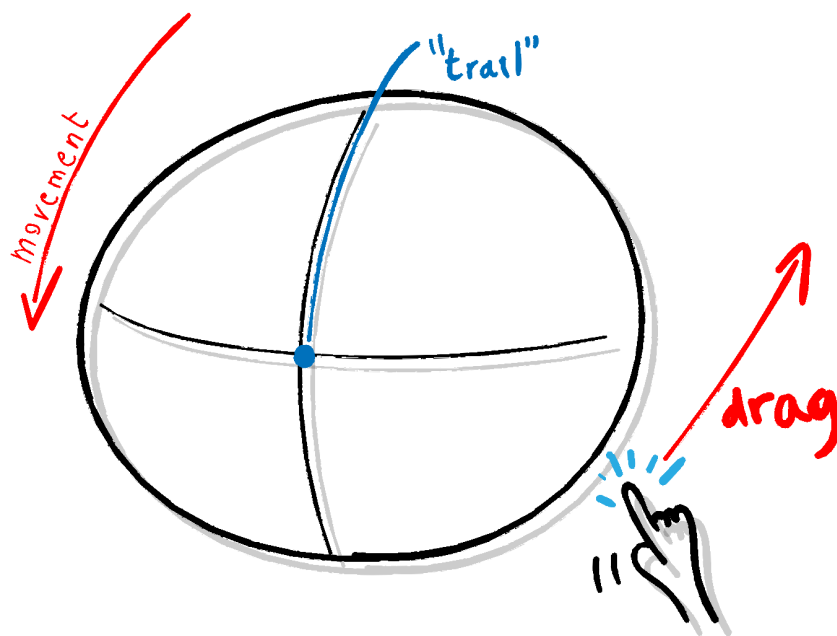
Figure 7.2: The participant taps and drags the screen to rotate the test sphere. The program samples the sphere's position dozens of times per second, serving as the movement "trail" data to be analyzed for classification.

The one-finger input modality is intended to mirror the TMT in its ease of use; participants of all age brackets, cultures, educational backgrounds and cognitive states are expected to be able to reasonably engage with the test [358]. When designing our test, the commitment to ergonomic design led us to avoid using Virtual Reality and Augmented Reality despite the more realistic depictions of depth; these relatively cutting-edge technologies are more expensive to deploy than a touch tablet, require more complex setups, and inherently demand a level of comfort with more complex technology. Our intention is to avoid testing a participant's acumen in consumer electronics, and instead to simply produce data directly sensitive to their cognitive state.

7.4.2 Visual Design

7.4.2.1 Depiction of Free Space

We have also made efforts to avoid possible confusion resulting from an interactive 3-dimensional object projected into a 2-dimensional screen. We took special consideration in several aesthetic design choices to ensure a participant of any age can clearly differentiate between objects of varying depths and angles. During the test, the sphere is surrounded by a realistic depiction of a sky in the distance in order to clearly indicate that the sphere is floating in a free, empty space and that the sphere's motions will not be interrupted by collision with other objects.

7.4.2.2 Depiction of Depth

The background depicting a clear afternoon sky also communicates to the user to expect daytime lighting, which we have implemented into the examination with a global illumination system. This heightens the depiction of depth as the sphere of boxes is rotated. The boxes are also capable of casting shadows on each other, with shadow lengths dynamically changing depending on the sphere's location relative to the source of global illumination. This produces very clear distinctions between boxes that differ in depth, as without shadows boxes might seem at the same depth relative to the screen. To further communicate the notion of depth of a 3-dimensional object, each box is rendered with a slight metallic material so that the box subtly shines as it is being

rotated. We carefully muted this shine effect so that the reflection does not impede a participant from properly reading the box's number or letter. To make rotation as intuitive as possible, the center box is always the last box of the examination since it is also the center point of rotation of the entire sphere. This serves as a visual anchor for the participant so that rotation is not disorienting even toward the end of the exam when the cluster of boxes no longer forms a spherical shape.
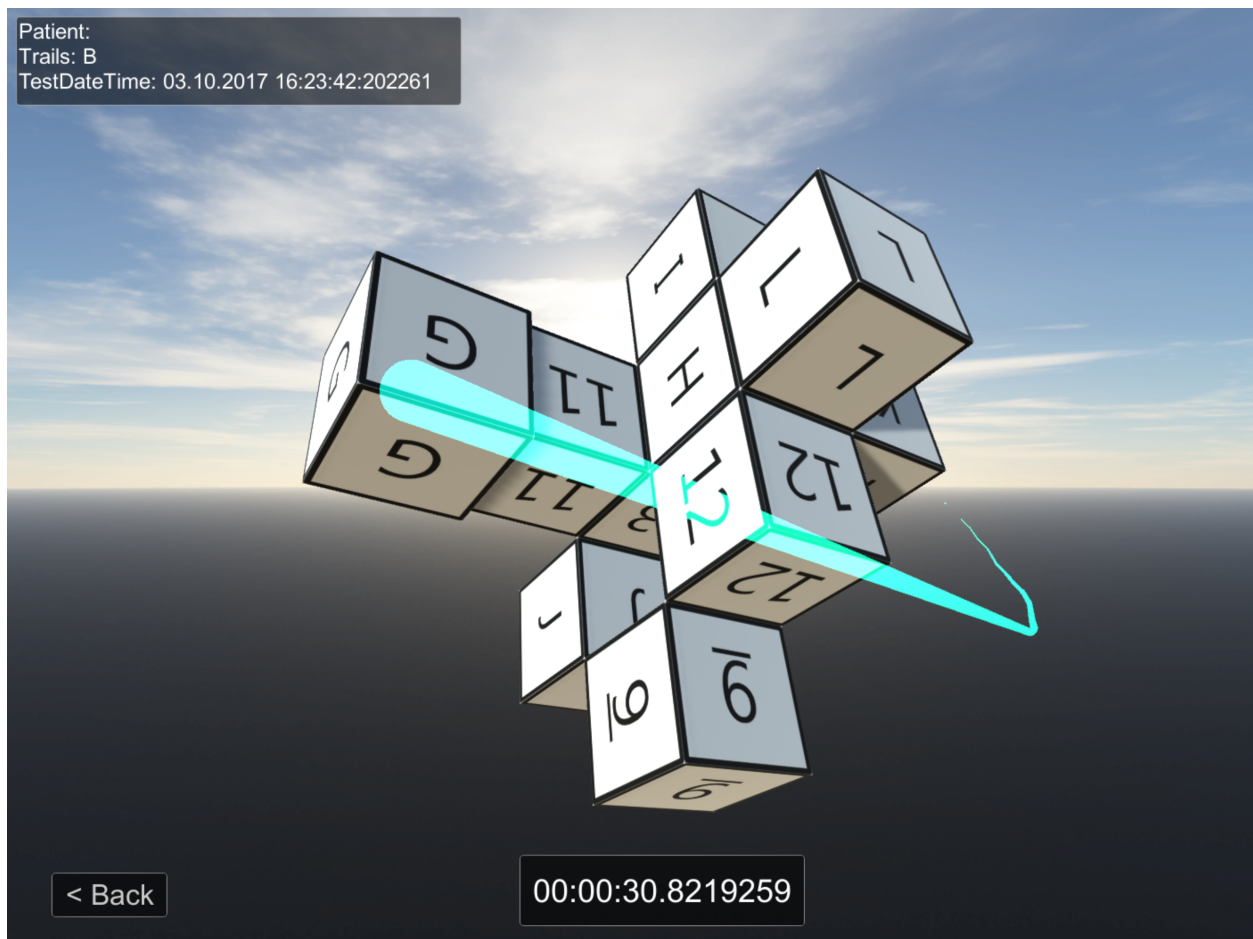
Figure 7.3: The interface of the Test Replay feature in the 3D-Trail-Making Test application. It provides a real-time playback of a participant's movements. Additional UI features include a running clock, participant information on the top left, and a real-time trail of captured finger movements synchronized with the sphere rotations and finger tap actions.

7.4.3 Replay of Completed Tests

The fine granularity of recorded input data allows us recreate the entirety of participant's exam by reading the file in sequence and advancing the position of the sphere and tapped-box actions with the proper timing and speed. As a result, we have created an additional "Replay" functionality so that a participant or researcher can view the completed test as many times as desired. This functionality also displays a translucent trail of the participant's finger drag across the screen to directly see their input. This "Replay" enhances the observer's ability to produce qualitative analysis, although it is not used for the quantitative evaluation we present in the following section.

7.4.4 Effects in Cognitive Load

The 3D-TMT deliberately introduces two cognitive tasks not found in the original paper TMT: performing mental rotation tasks, and manipulating a test that dynamically introduces and removes elements. The former is tested by box labels rotating naturally as the participant rotates the test sphere (resulting in boxes that may be skewed, angled, or upside down), while the latter is tested through the mechanic of disappearing cubes that reveal more cubes underneath. These two tasks have been explored in clinical neuropsychology previously, and frequently involve the introduction of new tests [290, 359, 360]. In leveraging the computing power of modern tablet technology, our interest is in carefully designing tests that intuitively combine tasks already known to be explored in the field of clinical neuropsychology. Traditional paper tests requires only one or two of these tasks being tested at one time [361]. As a result, the 3D-TMT is observed to have a higher cognitive load. We recognized this would likely mean a higher average time to test completion than the traditional TMT, a prediction supported by our evaluation results.

7.5 System Evaluation

Cognitive examinations, specifically the TMT, have been previously administered to overseas populations and compared to existing normative data, such as the work of Kim *et al.* [362], Seo *et al.* [363], and Cavaco *et al.* [99]. Our first of the two quantitative studies presented in this paper compares differences between the performance baseline of this examination with that of

the traditional paper-and-pencil version. The second quantitative study explores the recognition capabilities of the 3D-TMT, aiming to differentiate between two populations in a way that would be impossible to do with a typical paper-and-pencil examination. All participants in this study are cognitively healthy.

Table 7.1: Summary of the normative data collected for test variations A and B across paper, digital, and the new 3D examination. Tombaugh's normative data is the first row for reference.

TMT-A - Time (sec.)

	Mean	SD	Median	Min-Max	Skewness	Kurtosis
Tombaugh Normative Data	22.93	6.87	21.70	12-57	1.64	4.46
Paper-based TMT	22.57	5.69	21	15-42	1.36	3.24
Digital TMT	23.97	6.61	23	16-46	1.93	4.63
3D-TMT	53.5	15.24	50.5	35-79	0.45	-1.32

TMT-B - Time (sec.)

	Mean	SD	Median	Min-Max	Skewness	Kurtosis
Tombaugh Normative Data	48.97	12.69	47.0	29-95	0.91	0.92
Paper-based TMT	42.93	13.80	42.5	27-102	2.73	11.26
Digital TMT	50.86	15.97	47.5	30-99	2.01	4.24
3D-TMT	58.83	15.08	56.5	38-92	0.57	-0.61

7.5.1 Normative Studies

7.5.1.1 *Experiment Design*

In order to establish a performance baseline for the 3D-TMT, we conducted normative studies with cognitively healthy individuals with a methodology that mirrors that of the existing paper-and-pencil TMT. The 3D-TMT's performance score in this normative study mirrors that of the method of scoring the examination of the traditional TMT, which is "expressed in terms of the time in seconds required for completion of each of the two parts of the test" [2]. For instance, a test with a score of 26 indicates that a participant completed the test in 26 seconds. Existing normative data for the TMT includes dividing the test population into different age groups, one of which is the age group of participants between the ages of 18-24. Normative data can be established with the

data of around 30 participants. Following established protocols, our quantitative study involves 30 cognitively healthy participants between the ages of 18-24.

In order to explore the level of consistency between the paper-and-pencil TMT and the 3D-TMT, we conducted a within-subjects study in which each participant in our user study was tasked with completing three sets of Trail-Making tests. The first was a regular paper-based examination administered using the protocol as outlined in Compendium of Neuropsychological Tests. The second set of tests tasks the participant with completing a digital version of the same connect-the-dots exercise by drawing on a Microsoft Surface Pro 4 with a Surface Pen. This simulated the paper-based examination in every way, but with a different layout and positioning of the dots so as to avoid data skew through a Learning Effect. This second test also serves as an anchor point linking the traditional TMT and the touch-based 3D-TMT; establishing consistency in performance between using a paper-based exam and one with a touch-enabled tablet eases concerns about introducing a confounding effect by having to use a tablet for the completion of the 3D-TMT. The third set of tests is the 3D-Trail-Making test completed on the same Microsoft Surface Pro 4 with the participant using his or her finger as input. For each of these three tests, the participants completed both the A variation of the test (order 1, 2, 3, 4) as well as the B variation (order 1, A, 2, B).

7.5.1.2 Results

Table 1 shows the summary of the collected data from 30 cognitively healthy individuals between the ages of 18-24. For comparison, the established normative data Tombaugh's normative data that is stratified by age [31]. Skewness and kurtosis are two statistical measures typically reported in normative data studies. Skewness is the measure of symmetry where the data distribution is symmetric if it looks the same to the left and right of the center point, while kurtosis is a measure of how heavy the distribution tails are, where high kurtosis values indicate higher presence of outliers[§]. The general performance across the digital version falls in line with that of the paper version, although performance baselines change for the 3D-TMT. The 3D-TMT test A's mean and

---

[§]Measures of Skewness and Kurtosis: itl.nist.gov/div898/handbook/eda/section3/eda35b.htm

standard deviations are considerably higher than the either of the two 2-dimensional examinations (paper or digital), and while test B showed a noticeable increase as well in completion time, its increase is considerably less pronounced. Also of note is the fact that the performance across both tests A and B are remarkably close to each other, indicating that across all cognitively healthy participants, expected time to complete the test is more uniform for both tests A and B.

7.5.1.3 Discussion

Time to test completion has appreciably changed for both tests A and B for the 3D-TMT. Specifically, test A's completion time has nearly doubled in time. Despite the time increase, we do not consider this detrimental since the average completion time is still under a minute and within the normal range for neuropsychological examinations. Observing a difference in average time to completion alone is not a measure of test efficacy; every neuropsychological test has its own normative time to completion. A more accurate measure of efficacy is consistency among participants of the same classification.

Overall, observed participant behavior remained consistent in nature across all forms of the examination for tests A and B. The changes in test completion time can be attributed to the changes inherent to the different format and increased cognitive load of the examination. It is important to note that sknewness for the 3D-TMT is lower than even that of the skewness presented in Tombaugh's established paper-based normative data, suggesting that test completion times are more consistent between healthy participants. This results in observably consistent behavior among healthy participants, which provides us with strong results as our first set of normative data for the 3D-TMT.

7.5.2 Classification Capabilities of the 3D-TMT

7.5.2.1 Experiment Design

The 3-D Trail-Making Test was designed to produce rich volumes of data in order to facilitate machine learning classification. The simplified input modality results in two possible labeled actions: "tap" and "drag." Each of these actions is logged into the usage metrics data along with

a time-stamp of when these occurred, the cartesian coordinates of the center box of the sphere, and the x-y coordinates of where the participant's finger was touching the screen. New actions are logged as "Drag" any time the finger is placed and a change in finger position is detected. "Tap" actions get two additional labeled points: which box the participant tapped, and whether it was the next correct one in the sequence. New "Tap" or "Drag" data samples are created only as the participant completes said action. That is, there are no "no action" or "blank" action logs to conserve space, and participant pauses are still preserved by finding the difference between the samples' time-stamps. In total, participants log actions at an average magnitude of several dozen actions per second, typically resulting between 1,000 to 2,000 actions per test. Each usage metric log is saved into a text file along with a participant's anonymized ID number, age, which variant of the exam this log belongs to (A or B), and the date and time that this exam was taken.

This quantitative study is based in concept on the previously mentioned Trail-Making Test studies that analyze populations across different geographic regions [97–99, 362, 363]. A total of 52 participants completed both variations of test A and B. Of these 52, 32 of them live in the US and identify as having American nationality, and 20 were native citizens of Japan currently residing there. All members of both groups were cognitively healthy and both had an age range of 18-45 years of age.

7.5.2.2 *Feature Calculation*

Data features were calculated with the intention to capture participant's behavioral patterns. This is motivated largely due to multiple qualitative observations that we made during the user studies. For instance, members of the Japanese population would more frequently rotate in only one direction as they searched for the next box in the sequence, and they tended to rotate the sphere at a higher speed than that of the American participants. For this reason, our calculated features focused on speed and direction of movement in each of the sphere's three axes. Time to completion was included as well due to the test's relation to the paper-based TMT, as well as timing data in finer granularity. Along those same lines, time spent moving in each direction was also included as a feature. Every feature calculated is shown in Table 2.

Table 7.2: Total list of features used in this classification experiment. "[1–25]" indicates each feature was split into an additional 25, one per the data captured in between each of the boxes. "[X,Y,Z]" indicates each feature was further split into 3, one per the direction in which the movement is cumulatively captured.

Behavioral Features
TotalMistakes
TotalTestTime
TimeStartDelay
BoxTimes [1-25]
TimeDragPositive [X,Y,Z] [1-25]
TimeDragNegative [X,Y,Z] [1-25]
DeltaDragPositive [X,Y,Z] [1-25]
DeltaDragNegative [X,Y,Z] [1-25]
AverageTimeDragPositive [X,Y,Z]
AverageTimeDragNegative [X,Y,Z]
AverageDeltaDragPositive [X,Y,Z]
VelocitySamples [X,Y,Z] [1-25]
Magnitudes [1-25]

"TotalMistakes" is the total amount of mistakes that the participant made over the course of the test. "TimeStartDelay" is the amount of time that a participant took between starting the test and beginning to control the sphere to find the first box. "BoxTimes" is the time taken to correctly tap the next box in the sequence (e.g., the time taken between Box 1 and 2, box 2 and 3, etc.). "TimeDragPositive" is the amount of time that a participant spent rotating the test sphere in the direction that increased the coordinate value (X, Y, and Z were calculated separately). "TimeDragNegative" is, similarly, the amount of time a participant spent rotating in the opposite direction. "DeltaDragPositive" is the distance that the test sphere traveled while the participant rotated the sphere in the positive direction. This was calculated as a separate feature since at different speeds a participant may cover different distances. "DeltaDragNegative" similarly captures distance traveled in the negative direction. The averages of each of these described features across the entire tests were also calculated and included as features. Finally, the velocity vector of the X, Y and Z coordinates is captured in order to track the speed of the participant's movements. The velocity vector is calculated with the formula:

$$v = \frac{[x_t - x_{t-1}, y_t - y_{t-1}, z_t - z_{t-1}]}{timeDiff} \tag{7.1}$$

"Magnitudes" are the calculated magnitudes of the velocity vector using the formula:

$$v_m = \sqrt{v[x]^2 + v[y]^2 + v[z]^2} \tag{7.2}$$

Some features were further segmented into each box, indicated by "[1–25]" in Table 2. For example, "TimeDragPositiveX" was divided further into 25 segments, one per box, to indicate the time that the participant spent moving the test sphere in the positive X direction between box 1 and 2, 2 and 3, etc. This would allow us to analyze the participant's behavior on a per-box basis rather than just across the entirety of the test, since we anticipated that patterns might emerge between specific boxes.

This resulted in a total of 440 features, the vast majority of which we anticipated would be culled during feature analysis as we introduced this data into our classifier. We used the standard Naive Bayes classifier, as it is optimal for the Binary classification that we intended to perform between a Japanese and an American population. Classification was supported by 10-fold cross validation.

7.5.2.3 Results

Classification was performed with the Weka data mining software [364]. Table 4 shows the subset of features across both variations A and B of the 3D-TMT that yielded the highest accuracy in classifying examinations as belonging to either an American or Japanese participant. These features support our earlier intuition that differentiation may be more effective if we focused on tracking the specific direction of motion. For test A, moving the test sphere in the negative X direction across greater distances significantly contributed to the country of origin of the participant. A total of four features related to velocity support the observation that Japanese participants moved the sphere at greater speeds. For test B, a similar result is observed with respect to emphasis of

Table 7.3: Features yielding the best classification results for the A and B versions of the 3D-Trail-Making Test

Test A	Test B
DeltaDragPositiveY_23	BoxTimes_1
DeltaDragPositiveZ_12	BoxTimes_25
DeltaDragPositiveZ_18	TimeDragPositiveZ_2
DeltaDragNegativeX_6	TimeDragPositiveZ_21
DeltaDragNegativeX_10	TimeDragNegativeX_24
DeltaDragNegativeX_16	DeltaDragPositiveY_6
DeltaDragNegativeX_17	DeltaDragPositiveZ_4
DeltaDragNegativeX_25	DeltaDragPositiveZ_11
DeltaDragNegativeY_12	VelocitySamplesZ_7
DeltaDragNegativeZ_13	VelocitySamplesZ_8
DeltaDragNegativeZ_14	VelocitySamplesZ_25
DeltaDragNegativeZ_21	Magnitudes_8
VelocitySamplesZ_5	
VelocitySamplesZ_9	
VelocitySamplesZ_15	
VelocitySamplesZ_16	

movement velocity, albeit with fewer features needed for classification and the inclusion of timing data for two boxes.

Table 7.4: A small set of user test data (10 samples) as it is recorded on the text file. The format for each line is: Action Type, X, Y, Z coordinates of the center "anchor" box, X, Y coordinates of the finger touch for replay purposes, Cube ID of tapped box (only if Action Type is "Tap"), whether the tapped box was correct (only if Action Type is "Tap"), and the timestamp of the sample.

Line	Action	Pos X	Pos Y	Pos Z	Touch X	Touch Y	Time(s)	Cube ID	Correct?
...									
739	DRAG	325.72	348.95	205.25	600	108	22.80338		
740	DRAG	325.23	348.89	205.36	600	107	22.81962		
741	DRAG	325.23	349.39	205.36	599	107	22.86997		
742	DRAG	327.69	349.19	204.82	779	328	22.13622		
743	TAP	327.69	349.19	204.82	779	328	23.15491	Cube (15)	False
744	TAP	327.69	349.19	204.82	788	366	23.82257	Cube (7)	True
745	DRAG	322.29	344.94	206.12	713	682	24.78589		
746	DRAG	311.26	335.84	210.62	725	659	24.80566		
747	DRAG	305.86	329.57	214.81	731	647	24.81960		
...									

Table 5 shows the detailed accuracy statistics separated by class for both the A and B variation of the 3D-TMT. Specifically, for the A variation, the test was correctly classified as belonging to either an American or Japanese participant with an accuracy of 96.154% and an F-Measure of 0.962. For the B variation, the classification had an accuracy of 88.235% and F-Measure of 0.883.

7.5.2.4 Discussion

As previously mentioned, "DeltaDrag" refers to the total distance that the participant moved in any particular direction between boxes, split into either a positive or negative direction, and further split between X, Y, or Z direction of movement. For test A, this usage metric became important features in classifying the nationality of these participants. For test B, time played a slightly larger role, with the first and last time taken to tap on the correct box being an important feature while the total time taken in dragging in the X and Z directions being significant features also at the beginning and end of the examination.

Behavioral Observations. Observations on the participants while they were completing these tasks have yielded some interesting insights, which are supported by the most significant features chosen for classification. The Japanese participants visibly focused on completing the test as fast as possible. Although this did result in the participants moving the test sphere more rapidly, this did not result in faster time to completion, which is evidenced by the fact that total test completion time was not a significant feature for classification. However, this did result in a difference in behavior, which is evident in the classification features since velocity and the magnitude for the velocity vector were among the most significant features for classification. This may be attributed to a cultural difference due to the higher proliferation of mobile devices in Japan, and much higher rates of gaming mobile devices among the population [365]. Japanese participants tended to interpret this examination as a game, with the implicit objective being the fastest completion of time. This subtle cultural behavior was evident in the most important features, which in turn resulted in a feature set that yielded high accuracy and F-measure.

Test Accuracy Differences. A point of consideration are differences in classification features and accuracy between the tests A and B. We had already anticipated that the features for both

Table 7.5: Summary of accuracy statistics for both tests A and B of the 3D-Trail-Making Test. TP/FP Rates, Precision, Recall, and F-Measure shown are the weighted averages between the two classification labels.

	Accuracy	TP Rate	FP Rate	Precision	Recall	F-Measure
Test A	50 of 52 (96%)	0.96	0.04	0.96	0.96	0.96
Test B	45 of 51 (88%)	0.88	0.11	0.89	0.88	0.88

tests would be considerably different, since the correct box order across both tests are entirely different like on the paper-based TMT. The considerable decrease in accuracy for test variation B can be explained by its increased difficulty, since participants across all forms of this test, paper or otherwise, frequently take a few moments between each number or letter to think which is the next correct box or dot in the sequence. We can observe this as most participants think aloud, repeating numbers and letters in sequence to themselves to remember. On this 3D-TMT version of test B we noticed that as they were thinking they would frequently move the sphere, often times in random directions as they would not yet know what box they were looking for. This has a cumulative effect of a considerably larger number of "random" movements in unpredictable directions in total when considering all 25 boxes, making participants slightly more difficult to classify based on movement patterns.

Establishing Behavioral Recognition. Overall, the classification accuracy across both tests A and B is high, strongly supporting the notion that this new 3D-TMT test, much like its paper-based counterpart, is sensitive to behavioral patterns. In this case, we can observe that the 3D-TMT is noticeably sensitive to a participant's geographic location and nationality. This classification task would be impossible with the traditional paper-based TMT, since a completed, fully connected set of dots would yield no insight as to the country of origin of its participant, and the paper's existing metric of noting completion time would also provide limited information into country of origin.

Limitations. The two quantitative studies and their respective results also highlighted areas of improvement. Although both the paper-based TMT and the novel 3D-TMT appear reasonably

sensitive to geographic location and nationality, utility in broad behavioral recognition should firmly be established by conducting additional tests with patients with cognitive impairments. Additionally, we recognize the need for expanding the evaluation to determine whether the type of data gathered for this experiment may extend to general classification that also includes cognitive impairment.

7.6   Future Work

We are interested in continuing to explore the sensitivity of the 3D-TMT, testing for behavioral patterns such as those mentioned in Section 1.3. We seek to determine whether the test is sensitive in similar ways to the traditional TMT, such as age, level of education, sleepiness, inebriation, concussions, and Mild Cognitive Impairment among others. We also plan on integrating touch data into the classification features, including finger movement speed, location, acceleration, curvature, and motions indicating among others. This will introduce a new dimension of collected data which should even further increase the rich volumes of behavioral data that this examination can capture.

7.7   Conclusion

The 3-dimensional Trail-Making test is an examination heavily based on the principles of the established paper-based Trail-Making test with an added focus on collecting large volumes of behavioral data. We observe the participant's manipulation of the sphere object as a function of time with a high enough sample rate as to produce upwards of 2,000 data samples per test, which usually lasts less than one minute. The result is an ample set of features, 440 in total, and a rich set of data sensitive to subtle changes in behavior. To demonstrate this, we were able to classify completed tests as belonging to either Japanese or American participants with an accuracy of up to 96.154%. We have also begun building our own normative data set for this new examination, with an improved skewness than that of established paper-based examinations. Future integration of touch data into the classification algorithm can provide an even higher volume of data likely to be more sensitive to behavioral patterns.

8.   SUMMARY

8.1   Answers to Research Questions

8.1.1   R1. Can we create a predictive algorithm for estimating cognitive decline using digital sketch recognition features?

A1.   Yes.   A combination of existing and novel features were implemented onto various machine-learning behavior recognition models for training and prediction, and we found the best-performing F1-score for binary classification of whether participants are in the MCI or non-MCI group were $F1 = 0.73$ when building a model only on segmented "Travel" lines and $F1 = 0.72$ when building a model only on "Search" lines.   As discussed in Chapter 6, "Travel" and "Search" lines are automatically segmented based off of an estimate on what action the participant is performing. Based on the classification results, it is feasible to conclude that both kinds of lines yield similar amount of information that is meaningful enough to create classification algorithms of similar performance.

The second approach for estimating cognitive decline was through directly estimating MoCA scores based on the same digital sketch data that was analyzed for the binary classification experiment. This approach yielded somewhat positive results with an average root mean squared error of $RMSE = 3.325$ for Travel lines, $RMSE = 3.315$ for Search lines, a mean absolute error of $MAE = 2.415$ for Travel lines, and $MAE = 2.406$ for Search lines.   However, we believe that due to the necessary safety protocols in ensuring only MCI participants were recruited, we were unable to gather a true normative dataset for estimating MoCA scores directly due to no subject with a MoCA score under 19 being recruited.

8.1.2   R2. How closely can an automated grader for a complex figure examination mirror a human grader for ROCFTs?

A2. As discussed in Chapter 5, we have shown the closeness of an automated top-down graph-based grader in two ways: accuracy and F1-scores of whether a drawn detail exists and drawn

correctly, and how closely related the scores are to the two expert graders used in this study. With F1-scores ranging from 0.770 to as high as 0.966 and an average of 0.898, we believe that the recognition performance is considerably higher than state-of-the-art. Between the two expert graders, the correlation was $p = 0.948$, a Spearman's rank coefficient of $\rho = 0.942$, and an average difference in scores of $\Delta_{g_1,g_2} = 1.68$. Between our system and grader 1, $p = 0.799$, $\rho = 0.765$, and average $\Delta_{auto,g_1} = 3.21$. Between our system and grader 2, $p = 0.829$, $\rho = 0.802$, and average $\Delta_{auto,g_2} = 2.78$.

Additionally, the results mark the first time that a fully automated Rey-Osterrieth Complex Figure grader can assess all 18 details, whereas previous efforts have only been able to automatically grade a small subset of the shapes.

8.1.3 R3. How well can we differentiate between participants of different nationalities from a novel cognitive examination?

A3. We believe that our novel neuropsychological examination based on TMT can yield sensitivity to cultural background in the same way that the original paper-and-pencil test is. We are able to differentiate between participants with an accuracy and F1-score of 0.96 for Test A, and accuracy and F1-score of 0.88 for Test B, both of which are touch-based 3-dimensional exams with boxes placed in numerical order for Test A, and alternating alphanumerical order for Test B similar to the original TMT. This means that we were able to differentiate between participants from Japan and the United States with a high accuracy.

8.2 List of Contributions

- A digital Rey-Osterrieth Complex Figure Examination grader that is the first to fully automate the grading process of all 18 details.

- A 3-stage complex figure recognizer that is able to accurately discern specific lines or sets of lines based on a set of heuristics.

- A set of heuristics for sub-shape recognition that can be easily adapted to the other 6 complex figure examinations

- A template-matching algorithm that iterates on point-cloud system by reducing gestures to a empirically-defined $n \times n$ point-density matrix. Such iteration differs from the existing "Dollar family" of recognizers from which it is based by being able to provide a more accurate distortion metric rather than classifying the input gesture into generalized types of shapes.

- A set of features applied to Trail-Making Test digital sketch data that can perform as well in determining subject MCI as current state-of-the-art

- Novel sketch recognition features aimed at classifying participants based on whether or not they may have MCI

- An improved method over state-of-the-art for segmenting Trail-Making-Test sketches based on stop-and-go motion associated with searching for the next dot in a sequence and traveling to the next dot.

- A novel 3-dimensional touch-based neuropsychological examination based on the existing Trail-Making Test

- A set of novel behavioral features specific to determining whether touch input belongs to that of a Japanese or American test subject as they complete the novel touch-based neuropsychological examination

- Promising classification results that may yield intuitive insights into behavioral differences between Japanese and American test subjects when completing the novel neuropsychological examination

9. FUTURE WORK

The presented work served as a strong foundation in the implementation of digitizing neuropsychological examinations, automating or enhancing existing assessment methods, and introduces new examinations only possible in a digital domain. However, we believe there are several areas in which the presented work will be improved upon further development.

9.1 More Comprehensive Normative Data

Normative data presented by Tombaugh *et al.* [31] and Boone *et al.* [81] for the TMT and ROCF respectively provide a larger corpus of data from which to draw conclusions and normative data of cognitively healthy or neurodivergent behavior. We endeavored to gain a representative sample of the population as equally proportional as possible of the conditions and populations we wished to classify. However, more data from as many demographics and age groups as possible is always able to improve the existing recognition techniques, especially for the automated TMT classifier as it is based on machine-learning data.

9.2 Expansion to other Complex-Figure Tests

Auto ReyO is able to produce an accurate automated grader for the Rey-Osterrieth Complex Figure test. However, as explained in Chapter 5, six other complex figure tests are in frequent use as recognized by the Compendium of Neuropsychological Examinations [2]. The vast majority of the 18 details of the six other complex figures share most geometric properties in common, given that they are mostly combinations of straight lines, crosses, circles, triangles, and rectangles in a similar fashion as the ROCF. However, it is important to note that a few shapes and line types do not appear in the ROCF such as curved lines, tear-drop shapes, star shapes, etc. as can be seen in Figure 5.5. These are small in number but can be feasibly implemented with similar techniques as those as described in Chapter 5.

9.3 Exploration of UI/UX Studies for Both Clinicians and Patients

Every effort was made for the three presented works to provide participants with an intuitive and easy-to-use interface for proctors and subjects. Comprehensive user studies that explored the user experience on a UI level, however, proved to be out of scope for this dissertation. In future work we hope to conduct usability studies and informal interviews to determine key areas of improvement for both clinicians who proctor tests and patients who take them.

9.4 Sensitivity of Novel Examinations to Cognitive Impairment

Chapter 7 outlined the sensitivity of the novel neuropsychological examination to two different cultural backgrounds of the participants who took part in the study. However, since this was conceived and implemented as a neuropsychological examination, we wish to produce a study that collects data from participants who have MCI as well as participants who do not, such that we can determine whether the test is sensitive to various degrees of cognitive decline.

9.5 Data Visualization and Reporting for Both Clinicians and Patients

Reporting classification and analysis results of sketch information needs to be done intuitively such that clinicians are able to interpret the results in a way that is domain-appropriate. As such, we seek to produce a user study in which different visualizations of results are provided to clinicians to make a determination in how best to communicate this novel way of analyzing neuropsychological test data.

10. CONCLUSION

This dissertation explored the various ways in which digital sketch recognition and its associated techniques can be used to facilitate the practice of clinical neuropsychology. We sought to produce three projects that examined how the testing process can be improved in three major ways: grading tests in the same way that a clinician grades them, grading tests in a novel way based on machine-learning classification algorithms, and an entirely new examination only possible with modern touch tablets.

By these accounts the aforementioned research projects were successful. We were able to produce the first fully automated Rey-Osterrieth Complex Figure test based on a top-down graph-based digital sketch recognition pipeline. We produced a per-line examination of Trail-Making Tests that yielded comparable results with state-of-the-art based on much more granular examination, and we were able to provide a new examination that was sensitive to participant nationality in the same way a Trail-Making Test can be sensitive to different cultural backgrounds.

While the methods and presented projects contain within them much room for improvement, they still represent a strong contribution to research endeavors related to digitizing clinical neuropsychology. More specifically we are able to provide a strong foundation from which clinicians can automate and standardize the grading process of examinations that would otherwise be time-consuming and subjective, can provide the clinician with key new insights into a subject's cognitive state with analysis at a level of granularity not previously feasible, and the emergence of novel neuropsychological examinations that can be deployed on digital spaces. These exciting developments can further advance the practice of neuropsychology from the page into the digital domain and unlock a level of behavioral analysis not previously available.

REFERENCES

[1] MoCA Montreal–Cognitive Assment, "Montreal cognitive assessment," 2019. [Online; accessed May 18, 2019].

[2] E. Strauss, E. M. Sherman, O. Spreen, *et al.*, *A compendium of neuropsychological tests: Administration, norms, and commentary*. American Chemical Society, 2006.

[3] R. Brookmeyer, D. A. Evans, L. Hebert, K. M. Langa, S. G. Heeringa, B. L. Plassman, and W. A. Kukull, "National estimates of the prevalence of Alzheimer's disease in the United States," *Alzheimer's & Dementia*, vol. 7, pp. 61–73, Jan. 2011.

[4] D. A. Evans, H. H. Funkenstein, M. S. Albert, P. A. Scherr, N. R. Cook, M. J. Chown, L. E. Hebert, C. H. Hennekens, and J. O. Taylor, "Prevalence of Alzheimer's Disease in a Community Population of Older Persons: Higher Than Previously Reported," *JAMA*, vol. 262, pp. 2551–2556, Nov. 1989.

[5] D. A. Evans, "Estimated Prevalence of Alzheimer's Disease in the United States," *The Milbank Quarterly*, vol. 68, no. 2, pp. 267–289, 1990.

[6] L. Fratiglioni, M. Grut, Y. Forsell, M. Viitanen, M. Grafström, K. Holmen, K. Ericsson, L. Bäckman, A. Ahlbom, and B. Winblad, "Prevalence of Alzheimer's disease and other dementias in an elderly urban population: Relationship with age, sex, and education," *Neurology*, vol. 41, pp. 1886–1886, Dec. 1991.

[7] R. Katzman, "Education and the prevalence of dementia and Alzheimer's disease," *Neurology*, vol. 43, pp. 13–13, Jan. 1993.

[8] T. Polvikoski, R. Sulkava, L. Myllykangas, I. L. Notkola, L. Niinistö, A. Verkkoniemi, K. Kainulainen, K. Kontula, J. Pérez-Tur, J. Hardy, and M. Haltia, "Prevalence of Alzheimer's disease in very elderly people: A prospective neuropathological study," *Neurology*, vol. 56, pp. 1690–1696, June 2001.

[9] K. Trevisan, R. Cristina-Pereira, D. Silva-Amaral, and T. A. Aversi-Ferreira, "Theories of Aging and the Prevalence of Alzheimer's Disease," *BioMed Research International*, vol. 2019, p. e9171424, June 2019.

[10] R. S. Wilson, D. R. Weir, S. E. Leurgans, D. A. Evans, L. E. Hebert, K. M. Langa, B. L. Plassman, B. J. Small, and D. A. Bennett, "Sources of variability in estimates of the prevalence of Alzheimer's disease in the United States," *Alzheimer's & Dementia*, vol. 7, no. 1, pp. 74–79, 2011.

[11] C. for Disease Control and Prevention, "Deaths from alzheimer's disease," may 2017.

[12] N. Amoroso, M. La Rocca, S. Bruno, T. Maggipinto, A. Monaco, R. Bellotti, and S. Tangaro, "Brain structural connectivity atrophy in alzheimer's disease," *arXiv preprint arXiv:1709.02369*, 2017.

[13] M. T. Angelillo, D. Impedovo, G. Pirlo, L. Sarcinella, and G. Vessio, "Handwriting Dynamics as an Indicator of Cognitive Reserve: An Exploratory Study*," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 835–840, Oct. 2019.

[14] M. Barz, K. Altmeyer, S. Malone, L. Lauer, and D. Sonntag, "Digital Pen Features Predict Task Difficulty and User Performance of Cognitive Tests," in *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '20, (New York, NY, USA), pp. 23–32, Association for Computing Machinery, July 2020.

[15] F. Aşıcıoğlu and N. Turan, "Handwriting changes under the effect of alcohol," *Forensic science international*, vol. 132, no. 3, pp. 201–210, 2003.

[16] M. R. Patil and C. F. Mulimani, "Influence of alcohol on handwriting," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 17, no. 9, pp. 4579–4584, 2020.

[17] M. R. Patil and C. F. Mulimani, "Forensic examination of the handwriting characteristics of alcoholics," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 17, no. 11, pp. 271–278, 2020.

[18] C. A. Tripp, F. A. Fluckiger, and G. H. Weinberg, "Effects of alcohol on the graphomotor performances of normals and chronic alcoholics," *Perceptual and Motor Skills*, vol. 9, no. 2, pp. 227–236, 1959.

[19] A. C. Stevens, *The effects of typical and atypical grasps on endurance and fatigue in handwriting*. PhD thesis, Texas Woman's University, 2008.

[20] M. H. Tseng and S. A. Cermak, "The influence of ergonomic factors and perceptual–motor abilities on handwriting performance," *American Journal of Occupational Therapy*, vol. 47, no. 10, pp. 919–926, 1993.

[21] P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, and A. Korczyn, "Handwriting process variables discriminating mild alzheimer's disease and mild cognitive impairment," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, vol. 61, no. 4, pp. P228–P236, 2006.

[22] A. Schröter, R. Mergl, K. Bürger, H. Hampel, H.-J. Möller, and U. Hegerl, "Kinematic analysis of handwriting movements in patients with alzheimer's disease, mild cognitive impairment, depression and healthy subjects," *Dementia and geriatric cognitive disorders*, vol. 15, no. 3, pp. 132–142, 2003.

[23] P. Ghaderyan, A. Abbasi, and S. Saber, "A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of alzheimer's disease," *Expert Systems with Applications*, vol. 114, pp. 428–440, 2018.

[24] J. Behrendt, "Alzheimer's disease and its effect on handwriting," *Journal of Forensic Science*, vol. 29, no. 1, pp. 87–91, 1984.

[25] M. J. Slavin, J. G. Phillips, J. L. Bradshaw, K. A. Hall, and I. Presnell, "Consistency of handwriting movements in dementia of the alzheimer's type: A comparison with huntington's and parkinson's diseases," *Journal of the International Neuropsychological Society*, vol. 5, no. 1, pp. 20–25, 1999.

[26] M. A. El-Yacoubi, S. Garcia-Salicetti, C. Kahindo, A.-S. Rigaud, and V. Cristancho-Lacroix, "From aging to early-stage alzheimer's: Uncovering handwriting multimodal behaviors by semi-supervised learning and sequential representation learning," *Pattern Recognition*, vol. 86, pp. 112–133, 2019.

[27] D. E. Barnes and K. Yaffe, "The projected effect of risk factor reduction on Alzheimer's disease prevalence," *The Lancet Neurology*, vol. 10, pp. 819–828, Sept. 2011.

[28] P. J. Snyder, K. Kahle-Wrobleski, S. Brannan, D. S. Miller, R. J. Schindler, S. DeSanti, J. M. Ryan, G. Morrison, M. Grundman, J. Chandler, R. J. Caselli, M. Isaac, L. Bain, and M. C. Carrillo, "Assessing cognition and function in Alzheimer's disease clinical trials: Do we have the right tools?," *Alzheimer's & Dementia*, vol. 10, pp. 853–860, Nov. 2014.

[29] B. Leifer, "Early diagnosis of alzheimer's disease: Clinical and economic benefits," *Journal of the American Geriatrics Society*, vol. 51, no. 5s2, pp. S281–S288, 2003.

[30] C. M. Parsey and M. Schmitter-Edgecombe, "Applications of technology in neuropsychological assessment," *The Clinical Neuropsychologist*, vol. 27, no. 8, pp. 1328–1361, 2013.

[31] T. N. Tombaugh, "Trail making test a and b: Normative data stratified by age and education," *Archives of Clinical Neuropsychology*, vol. 19, no. 2, pp. 203–214, 2004.

[32] P. Osterrieth, "Le test de copie d'une figure complexe," *Archives de Psychologie*, vol. 30, pp. 205–550, 1944.

[33] B. Reisberg, S. H. Ferris, A. Kluger, E. Franssen, J. Wegiel, and M. J. De Leon, "Mild cognitive impairment (mci): a historical perspective," *International Psychogeriatrics*, vol. 20, no. 1, pp. 18–31, 2008.

[34] B. Reisberg and S. Gauthier, "Current evidence for subjective cognitive impairment (sci) as the pre-mild cognitive impairment (mci) stage of subsequently manifest alzheimer's disease," *International psychogeriatrics*, vol. 20, no. 1, pp. 1–16, 2008.

[35] C. P. Hughes, L. Berg, W. Danziger, L. A. Coben, and R. L. Martin, "A new clinical scale for the staging of dementia," *The British journal of psychiatry*, vol. 140, no. 6, pp. 566–572, 1982.

[36] L. S. Broster, S. L. Jenkins, S. D. Holmes, M. G. Edwards, G. A. Jicha, and Y. Jiang, "Electrophysiological repetition effects in persons with mild cognitive impairment depend upon working memory demand," *Neuropsychologia*, vol. 117, pp. 13–25, Aug. 2018.

[37] Y. Zhang, B. Han, P. Verhaeghen, and L.-G. Nilsson, "Executive functioning in older adults with mild cognitive impairment: Mci has effects on planning, but not on inhibition," *Aging, Neuropsychology, and Cognition*, vol. 14, no. 6, pp. 557–570, 2007.

[38] M. D. Greicius and D. L. Kimmel, "Neuroimaging insights into network-based neurodegeneration," *Current opinion in neurology*, vol. 25, no. 6, pp. 727–734, 2012.

[39] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, A. D. N. Initiative, *et al.*, "Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects," *Neuroimage*, vol. 104, pp. 398–412, 2015.

[40] E. E. Bron, M. Smits, W. M. Van Der Flier, H. Vrenken, F. Barkhof, P. Scheltens, J. M. Papma, R. M. Steketee, C. M. Orellana, R. Meijboom, *et al.*, "Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural mri: the caddementia challenge," *NeuroImage*, vol. 111, pp. 562–579, 2015.

[41] Z. S. Khachaturian, "Diagnosis of alzheimer's disease," *Archives of Neurology*, vol. 42, no. 11, pp. 1097–1105, 1985.

[42] K. M. Harrell, S. S. Wilkins, M. K. Connor, and J. Chodosh, "Telemedicine and the evaluation of cognitive impairment: The additive value of neuropsychological assessment," *Journal of the American Medical Directors Association*, vol. 15, no. 8, pp. 600–606, 2014.

[43] H. E. Feenstra, I. E. Vermeulen, J. M. Murre, and S. B. Schagen, "Online cognition: factors facilitating reliable online neuropsychological test results," *The Clinical Neuropsychologist*, vol. 31, no. 1, pp. 59–84, 2017.

[44] G. P. Lee, D. W. Loring, and J. L. Thompson, "Construct validity of material-specific memory measures following unilateral temporal lobe ablations," *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, vol. 1, no. 3, pp. 192–197, 1989.

[45] K. J. Meador, E. E. Moore, M. E. Nichols, O. L. Abney, H. S. Taylor, E. Y. Zamrini, and D. W. Loring, "The role of cholinergic systems in visuospatial processing and memory," *Journal of Clinical and Experimental Neuropsychology*, vol. 15, pp. 832–842, Sept. 1993.

[46] C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, and A. Scotto di Freca, "Handwriting analysis to support neurodegenerative diseases diagnosis: A review," *Pattern Recognition Letters*, vol. 121, pp. 37–45, Apr. 2019.

[47] L. Ashendorf, A. L. Jefferson, M. K. O'Connor, C. Chaisson, R. C. Green, and R. A. Stern, "Trail making test errors in normal aging, mild cognitive impairment, and dementia," *Archives of Clinical Neuropsychology*, vol. 23, no. 2, pp. 129–137, 2008.

[48] P. Julayanont and Z. S. Nasreddine, "Montreal cognitive assessment (moca): concept and clinical review," in *Cognitive screening instruments*, pp. 139–195, Springer, 2017.

[49] K. B. Kortte, M. D. Horner, and W. K. Windham, "The trail making test, part b: cognitive flexibility or ability to maintain set?," *Applied neuropsychology*, vol. 9, no. 2, pp. 106–109, 2002.

[50] I. . Sánchez-Cubillo, J. Periáñez, D. Adrover-Roig, J. Rodríguez-Sánchez, M. Rios-Lago, J. Tirapu, and F. Barcelo, "Construct validity of the trail making test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities," *Journal of the International Neuropsychological Society: JINS*, vol. 15, no. 3, p. 438, 2009.

[51] S. F. Crowe, "The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts a and b of the trail making test," *Journal of clinical psychology*, vol. 54, no. 5, pp. 585–591, 1998.

[52] J. J. O'Rourke, L. J. Beglinger, M. M. Smith, J. Mills, D. J. Moser, K. C. Rowe, D. R. Langbehn, K. Duff, J. C. Stout, D. L. Harrington, *et al.*, "The trail making test in prodromal

huntington disease: contributions of disease progression to test performance," *Journal of clinical and experimental neuropsychology*, vol. 33, no. 5, pp. 567–579, 2011.

[53] J. A. Arnett and S. S. Labovitz, "Effect of physical layout in performance of the trail making test.," *Psychological Assessment*, vol. 7, no. 2, p. 220, 1995.

[54] Z. S. Nasreddine, N. A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J. L. Cummings, and H. Chertkow, "The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, pp. 695–699, 2005.

[55] S. McLennan, J. Mathias, L. Brennan, and S. Stewart, "Validity of the Montreal Cognitive Assessment (MoCA) as a Screening Test for Mild Cognitive Impairment (MCI) in a Cardiovascular Population," *Journal of Geriatric Psychiatry and Neurology*, vol. 24, pp. 33–38, Mar. 2011.

[56] S. Freitas, M. R. Simões, L. Alves, and I. Santana, "Montreal Cognitive Assessment: Validation Study for Mild Cognitive Impairment and Alzheimer Disease," *Alzheimer Disease & Associated Disorders*, vol. 27, pp. 37–43, Jan. 2013.

[57] J. Hobson, "The montreal cognitive assessment (moca)," *Occupational Medicine*, vol. 65, no. 9, pp. 764–765, 2015.

[58] T. Smith, N. Gildeh, and C. Holmes, "The montreal cognitive assessment: validity and utility in a memory clinic setting," *The Canadian Journal of Psychiatry*, vol. 52, no. 5, pp. 329–332, 2007.

[59] Z. S. Nasreddine and B. B. Patel, "Validation of montreal cognitive assessment, moca, alternate french versions," *Canadian Journal of Neurological Sciences*, vol. 43, no. 5, pp. 665–671, 2016.

[60] G. Santangelo, M. Siciliano, R. Pedone, C. Vitale, F. Falco, R. Bisogno, P. Siano, P. Barone, D. Grossi, F. Santangelo, *et al.*, "Normative data for the montreal cognitive assessment in an italian population sample," *Neurological Sciences*, vol. 36, no. 4, pp. 585–591, 2015.

138

[61] C. Delgado, A. Araneda, and M. I. Behrens, "Validation of the Spanish-language version of the Montreal Cognitive Assessment test in adults older than 60 years," *Neurología (English Edition)*, vol. 34, pp. 376–385, July 2019.

[62] N. Ojeda, R. Del Pino, N. Ibarretxe-Bilbao, D. J. Schretlen, and J. Pena, "Montreal Cognitive Assessment Test: normalization and standardization for Spanish population," *Revista De Neurologia*, vol. 63, pp. 488–496, Dec. 2016.

[63] S. T. Pendlebury, F. C. Cuthbertson, S. J. Welch, Z. Mehta, and P. M. Rothwell, "Underestimation of cognitive impairment by mini-mental state examination versus the montreal cognitive assessment in patients with transient ischemic attack and stroke: a population-based study," *Stroke*, vol. 41, no. 6, pp. 1290–1293, 2010.

[64] G. Chiti and L. Pantoni, "Use of montreal cognitive assessment in patients with stroke," *Stroke*, vol. 45, no. 10, pp. 3135–3140, 2014.

[65] M. Janssen, M. Bosch, P. Koopmans, and R. Kessels, "Validity of the montreal cognitive assessment and the hiv dementia scale in the assessment of cognitive impairment in hiv-1 infected patients," *Journal of neurovirology*, vol. 21, no. 4, pp. 383–390, 2015.

[66] M. G. Borda, C. Reyes-Ortiz, M. U. Pérez-Zepeda, D. Patino-Hernandez, C. Gómez-Arteaga, and C. A. Cano-Gutiérrez, "Educational level and its Association with the domains of the Montreal Cognitive Assessment Test," *Aging & Mental Health*, vol. 23, pp. 1300–1306, Oct. 2019.

[67] J. Toglia, K. A. Fitzgerald, M. W. O'Dell, A. R. Mastrogiovanni, and C. D. Lin, "The mini-mental state examination and montreal cognitive assessment in persons with mild subacute stroke: relationship to functional outcome," *Archives of physical medicine and rehabilitation*, vol. 92, no. 5, pp. 792–798, 2011.

[68] O. Godefroy, A. Fickl, M. Roussel, C. Auribault, J. M. Bugnicourt, C. Lamy, S. Canaple, and G. Petitnicolas, "Is the montreal cognitive assessment superior to the mini-mental state

examination to detect poststroke cognitive impairment? a study with neuropsychological evaluation," *Stroke*, vol. 42, no. 6, pp. 1712–1716, 2011.

[69] J. Dalrymple-Alford, M. MacAskill, C. Nakas, L. Livingston, C. Graham, G. Crucian, T. Melzer, J. Kirwan, R. Keenan, S. Wells, *et al.*, "The moca: well-suited screen for cognitive impairment in parkinson disease," *Neurology*, vol. 75, no. 19, pp. 1717–1725, 2010.

[70] D. R. Roalf, P. J. Moberg, S. X. Xie, D. A. Wolk, S. T. Moelter, and S. E. Arnold, "Comparative accuracies of two common screening instruments for classification of alzheimer's disease, mild cognitive impairment, and healthy aging," *Alzheimer's & Dementia*, vol. 9, no. 5, pp. 529–537, 2013.

[71] A. Rey, "L'examen psychologique dans les cas d'encéphalopathie traumatique.(les problems.).," *Archives de psychologie*, 1941.

[72] M. D. Lezak, D. B. Howieson, D. W. Loring, J. S. Fischer, *et al.*, *Neuropsychological assessment*. Oxford University Press, USA, 2004.

[73] J. Knight, E. Kaplan, and L. Ireland, "Survey findings of the rey-osterrieth complex figure use among the ins membership," *22nd Annual Meeting of the International Neuropsychological Society*, 1994.

[74] D. T. R. Berry, R. S. Allen, and F. A. Schmitt, "Rey-Osterrieth complex figure: Psychometric characteristics in a geriatric sample," *Clinical Neuropsychologist*, vol. 5, pp. 143–153, Apr. 1991.

[75] M.-S. Shin, S.-Y. Park, S.-R. Park, S.-H. Seol, and J. S. Kwon, "Clinical and empirical applications of the rey–osterrieth complex figure test," *Nature protocols*, vol. 1, no. 2, p. 892, 2006.

[76] M.-S. Shin, Y.-H. Kim, S.-C. Cho, and B.-N. Kim, "Neuropsychologic characteristics of children with attention-deficit hyperactivity disorder (adhd), learning disorder, and tic

disorder on the rey-osterreith complex figure," *Journal of Child Neurology*, vol. 18, no. 12, pp. 835–844, 2003.

[77] M. M. Cherrier, M. F. Mendez, M. Dave, and K. M. Perryman, "Performance on the rey–osterrieth complex figure test in alzheimer disease and vascular dementia.," *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*, 1999.

[78] R. J. Melrose, D. Harwood, T. Khoo, M. Mandelkern, and D. L. Sultzer, "Association between cerebral metabolism and rey–osterrieth complex figure test performance in alzheimer's disease," *Journal of clinical and experimental neuropsychology*, vol. 35, no. 3, pp. 246–258, 2013.

[79] S. M. Silverstein, L. M. Osborn, and D. R. Palumbo, "Rey-osterrieth complex figure test performance in acute, chronic, and remitted schizophrenia patients," *Journal of Clinical Psychology*, vol. 54, no. 7, pp. 985–994, 1998.

[80] P. S. Fastenau, N. L. Denburg, and B. J. Hufford, "Adult norms for the rey-osterrieth complex figure test and for supplemental recognition and matching trials from the extended complex figure test," *The Clinical Neuropsychologist*, vol. 13, no. 1, pp. 30–47, 1999.

[81] K. B. Boone, I. M. Lesser, E. Hill-Gutierrez, N. G. Berman, and L. F. D'elia, "Rey-osterrieth complex figure performance in healthy, older adults: Relationship to age, education, sex, and iq," *The Clinical Neuropsychologist*, vol. 7, no. 1, pp. 22–28, 1993.

[82] N.-A. Wilson and J. Batchelor, "Examining rey complex figure test organization in healthy adults," *Journal of clinical and experimental neuropsychology*, vol. 37, no. 10, pp. 1052–1061, 2015.

[83] D. P. Waber and J. M. Holmes, "Assessing children's memory productions of the rey-osterrieth complex figure," *Journal of Clinical and Experimental Neuropsychology*, vol. 8, no. 5, pp. 563–580, 1986.

[84] K. Watanabe, T. Ogino, K. Nakano, J. Hattori, Y. Kado, S. Sanada, and Y. Ohtsuka, "The rey–osterrieth complex figure as a measure of executive function in childhood," *Brain and Development*, vol. 27, no. 8, pp. 564–569, 2005.

[85] D. P. Waber and J. M. Holmes, "Assessing children's copy productions of the rey-osterrieth complex figure," *Journal of clinical and experimental neuropsychology*, vol. 7, no. 3, pp. 264–280, 1985.

[86] M. W. Kirkwood, M. D. Weiler, J. H. Bernstein, P. W. Forbes, and D. P. Waber, "Sources of poor performance on the rey–osterrieth complex figure test among children with learning difficulties: A dynamic assessment approach," *The Clinical Neuropsychologist*, vol. 15, no. 3, pp. 345–356, 2001.

[87] K. Lang, M. Roberts, A. Harrison, C. Lopez, E. Goddard, M. Khondoker, J. Treasure, and K. Tchanturia, "Central coherence in eating disorders: a synthesis of studies using the rey osterrieth complex figure test," *PLoS One*, vol. 11, no. 11, p. e0165467, 2016.

[88] M. Rosselli and A. Ardila, "Effects of age, education, and gender on the rey-osterrieth complex figure," *The clinical neuropsychologist*, vol. 5, no. 4, pp. 370–376, 1991.

[89] J. Bennett-Levy, "Determinants of performance on the rey-osterrieth complex figure test: An analysis, and a new technique for single-case assessment," *British Journal of Clinical Psychology*, vol. 23, no. 2, pp. 109–119, 1984.

[90] R. A. Stern, E. A. Singer, L. M. Duke, N. G. Singer, C. E. Morey, E. W. Daughtrey, and E. Kaplan, "The boston qualitative scoring system for the rey-osterrieth complex figure: description and interrater reliability," *The Clinical Neuropsychologist*, vol. 8, no. 3, pp. 309–322, 1994.

[91] A. K. Troyer and H. A. Wishart, "A comparison of qualitative scoring systems for the rey-osterrieth complex figure test," *The Clinical Neuropsychologist*, vol. 11, no. 4, pp. 381–390, 1997.

[92] S. L. Hamby, J. W. Wilkins, and N. S. Barry, "Organizational quality on the rey-osterrieth and taylor complex figure tests: A new scoring system.," *Psychological assessment*, vol. 5, no. 1, p. 27, 1993.

[93] P. S. Fastenau, J. M. Bennett, and N. L. Denburg, "Application of psychometric standards to scoring system evaluation: is "new" necessarily "improved"?," *Journal of Clinical and Experimental Neuropsychology*, vol. 18, no. 3, pp. 462–472, 1996.

[94] R. Canham, S. L. Smith, and A. M. Tyrrell, "Automated scoring of a neuropsychological test: the rey osterrieth complex figure," in *Proceedings of the 26th Euromicro Conference. EUROMICRO 2000. Informatics: Inventing the Future*, vol. 2, pp. 406–413, IEEE, 2000.

[95] J. T. Barth, W. M. Alves, T. V. Ryan, S. N. Macciocchi, R. W. Rimel, J. A. Jane, and W. E. Nelson, "Mild head injury in sports: Neuropsychological sequelae and recovery of function," *Mild head injury*, pp. 257–275, 1989.

[96] R. J. Echemendia, M. Putukian, R. S. Mackin, L. Julian, and N. Shoss, "Neuropsychological test performance prior to and following sports-related mild traumatic brain injury," *Clinical Journal of Sport Medicine*, vol. 11, no. 1, pp. 23–31, 2001.

[97] R. Hashimoto, K. Meguro, E. Lee, M. Kasai, H. Ishii, and S. Yamaguchi, "Effect of age and education on the trail making test and determination of normative data for japanese elderly people: the tajiri project," *Psychiatry and Clinical Neurosciences*, vol. 60, no. 4, pp. 422–428, 2006.

[98] A. C. Hamdan and E. M. L. Hamdan, "Effects of age and education level on the trail making test in a healthy brazilian sample," *Psychology & Neuroscience*, vol. 2, no. 2, pp. 199–203, 2009.

[99] S. Cavaco, A. Gonçalves, C. Pinto, E. Almeida, F. Gomes, I. Moreira, J. Fernandes, and A. Teixeira-Pinto, "Trail making test: Regression-based norms for the portuguese population," *Archives of Clinical Neuropsychology*, vol. 28, no. 2, pp. 189–198, 2013.

[100] R. Z. Goldstein, A. C. Leskovjan, A. L. Hoff, R. Hitzemann, F. Bashan, S. S. Khalsa, G.-J. Wang, J. S. Fowler, and N. D. Volkow, "Severity of neuropsychological impairment in cocaine and alcohol addiction: association with metabolism in the prefrontal cortex," *Neuropsychologia*, vol. 42, no. 11, pp. 1447–1458, 2004.

[101] K. Cheshire, H. Engleman, I. Deary, C. Shapiro, and N. J. Douglas, "Factors impairing daytime performance in patients with sleep apnea/hypopnea syndrome," *Archives of internal medicine*, vol. 152, no. 3, pp. 538–541, 1992.

[102] M. Koso and S. Hansen, "Executive function and memory in posttraumatic stress disorder: a study of bosnian war veterans," *European Psychiatry*, vol. 21, no. 3, pp. 167–173, 2006.

[103] D. Rubine, "Specifying gestures by example," *SIGGRAPH Comput. Graph.*, vol. 25, p. 329–337, July 1991.

[104] L. B. Kara and T. F. Stahovich, "Hierarchical parsing and recognition of hand-sketched diagrams," in *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pp. 13–22, 2004.

[105] C. Calhoun, T. F. Stahovich, T. Kurtoglu, and L. B. Kara, "Recognizing multi-stroke symbols," in *AAAI Spring Symposium on Sketch Understanding*, pp. 15–23, Stanford University, AAAI Technical Report SS-02-08, AAAI Press, 2002.

[106] M. Bresler, D. Průša, and V. Hlaváč, "Detection of arrows in on-line sketched diagrams using relative stroke positioning," in *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 610–617, IEEE, 2015.

[107] T. Hammond and B. Paulson, "Recognizing sketched multistroke primitives," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 1, no. 1, pp. 1–34, 2011.

[108] W. Li and T. Hammond, "Using scribble gestures to enhance editing behaviors of sketch recognition systems," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, (New York, NY, USA), pp. 2213–2218, Association for Computing Machinery, May 2012.

[109] H. Choi and T. Hammond, "Sketch recognition based on manifold learning," in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI) Student Abstracts*, (Chicago, IL, USA), pp. 1786–1787, AAAI, July 13–17, 2008.

[110] H. Choi, B. Paulson, and T. Hammond, "Gesture recognition based on manifold learning," in *Structural, Syntactic, and Statistical Pattern Recognition (SSPR), Lecture Notes in Computer Science*, vol. 5342, pp. 247–256, Springer Berlin Heidelberg, 2008. ISBN: 978-3-540-89688-3.

[111] K. Dahmen and T. Hammond, "Distinguishing between sketched scribble look alikes," in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI) Student Abstracts*, (Chicago, IL, USA), pp. 1790–1791, AAAI, July 13–17, 2008.

[112] T. Hammond and R. Davis, "Ladder: A language to describe drawing, display, and editing in sketch recognition," in *Proceedings of the International Joint Conference on Aritificial Intelligence (IJCAI)*, (Alcapulco, Mexico, Mexico), pp. 461–467, AAAI, 2003.

[113] T. Hammond and R. Davis, "Ladder, a sketching language for user interface developers," *Comput. Graph.*, vol. 29, p. 518–532, Aug. 2005.

[114] T. Hammond and R. Davis, "Automatically transforming symbolic shape descriptions for use in sketch recognition," in *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI)*, (San Jose, CA, USA), pp. 450–456, AAAI, July 25-29, 2004.

[115] T. Hammond, "Ladder: A perceptually-based language to simplify sketch recognition user interface development," PhD Doctoral Dissertation, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, February 2007. Advisor: Randall Davis.

[116] T. Hammond and R. Davis, "Interactive learning of structural shape descriptions from automatically generated near-miss examples," in *Proceedings of the 11th International Conference on Intelligent User Interfaces*, (Sydney, Australia, Austalia), pp. 210–217, ACM, January 29 - February 1n 2006. ISBN: 1-59593-287-9.

[117] T. Hammond and R. Davis, "Debugging shape definitions for use in sketch recognition," in *MIT Lab Abstract. Computer Science and Artificial Intelligence Laboratory (CSAIL)*, (Cambridge, MA, USA), p. 2 pages, MIT, September 2004.

[118] T. Hammond, "Automatically generating sketch interfaces from shape descriptions," in *Proceedings of the 4th Annual MIT Student Oxygen Workshop*, (Gloucester, Massachusetts, USA), p. 4 pages, MIT, July 2004.

[119] T. Hammond and R. Davis, "Ladder: A sketch recognition language," in *MIT Lab Abstract. Computer Science and Artificial Intelligence Laboratory (CSAIL)*, (Cambridge, MA, USA), p. 2 pages, MIT, September 2004.

[120] T. Hammond, M. Sezgin, , A. Adler, M. Oltmans, C. Alvarado, and R. Hitchcock, "Multi-domain sketch recognition," in *Proceedings of the 2nd Annual MIT Student Oxygen Workshop*, (Cambridge, MA, USA), p. 2 pages, MIT, Artificial Intelligence Laboratory, September 2002.

[121] T. Hammond and R. Davis, "Testing shape descrpitions by automatically translating them for use in sketch recognition," in *MIT Lab Abstract. Computer Science and Artificial Intelligence Laboratory (CSAIL)*, (Cambridge, MA, USA), p. 2 pages, MIT, September 2004.

[122] C. Alvarado, M. Sezgin, D. Scott, T. Hammond, Z. Kasheff, M. Oltmans, and R. Davis, "A framework for multi-domain sketch recognition," in *MIT Lab Abstract. Artificial Intelligence Laboratory*, (Cambridge, MA, USA), p. 3 pages, MIT, September 2001.

[123] T. Hammond and R. Davis, "Recognizing interspersed sketches quickly," in *Proceedings of Graphics Interface (GI). Kelowna, Canada. Canadian Information Processing Society, May 25-27, 2009*, (Kelowna, Canada, Canada), pp. 157–166, Canadian Information Processing Society, May 25–27, 2009. ISBN: 978-1-56881-470-4.

[124] J. Johnston and T. Hammond, "Computing confidence values for geometric constraints for use in sketch recognition," in *Proceedings of the Seventh Sketch-Based Interfaces*

and Modeling Symposium (SBIM), (Annecy, France, France), pp. 71–78, Eurographics Association, June 7–10, 2010. ISBN: 978-3-905674-25-5.

[125] T. Hammond, "Enabling instructors to develop sketch recognition applications for the classroom," in *37th Annual Frontiers In Education Conference (FIE)*, (Milwaukee, WI, USA), pp. S3J11–S3J16, IEEE, October 10-13, 2007.

[126] T. Hammond and R. Davis, "Shady: A shape description debugger for use in sketch recognition," in *AAAI Fall Symposium on Making Pen-Based Interaction Intelligent and Natural (AAAI)*, (Arlington, VA, USA), p. 7 pages, AAAI, October 22-24, 2004. ISBN: 978-1-57735-217-4.

[127] T. Hammond and R. Davis, "Creating the perception-based ladder sketch recognition language," in *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS)*, (Aarhus, Denmark, Denmarkl), pp. 141–150, ACM, August 16-20, 2010. ISBN: 978-1-4503-0103-9.

[128] C. Alvarado and R. Davis, "Sketchread: a multi-domain sketch recognition engine," in *ACM SIGGRAPH 2007 courses*, p. 34, ACM, 2007.

[129] T. Y. Ouyang and R. Davis, "Chemink: a natural real-time recognition system for chemical drawings," in *Proceedings of the 16th international conference on Intelligent user interfaces*, pp. 267–276, ACM, 2011.

[130] M. Field, S. Valentine, J. Linsey, and T. Hammond, "Sketch recognition algorithms for comparing complex and unpredictable shapes," in *Proceedings of the Twenty-Second international Joint Conference on Artificial Intelligence (IJCAI)*, vol. 3, (Barcelona, Spain, Spain), pp. 2436–2441, AAAI Press, July 16-22, 2011.

[131] S. Valentine, F. Vides, G. Lucchese, D. Turner, H.-h. Kim, W. Li, J. Linsey, and T. Hammond, "Mechanix: a sketch-based tutoring and grading system for free-body diagrams," *AI magazine*, vol. 34, no. 1, pp. 55–55, 2013.

[132] T. Nelligan, S. Polsley, J. Ray, M. Helms, J. Linsey, and T. Hammond, "Mechanix: A sketch-based educational interface," in *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*, pp. 53–56, 2015.

[133] O. Atilola, M. Field, E. McTigue, T. Hammond, and J. Linsey, "Mechanix: a sketch recognition truss tutoring system," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 54846, pp. 645–654, 2011.

[134] S. Valentine, M. Field, A. Smith, and T. Hammond, "A shape comparison technique for use in sketch-based tutoring systems," in *Proceedings of the 2011 Intelligent User Interfaces Workshop on Sketch Recognition (Palo Alto, CA, USA, 2011)*, vol. 11-5, (Palo Alto, CA, USA), p. 4 pages, IUI, February 13, 2011.

[135] S. Valentine, R. Lara-Garduno, J. Linsey, and T. Hammond, "Mechanix: A sketch-based tutoring system that automatically corrects hand-sketched statics homework," in *The impact of pen and touch technology on education*, pp. 91–103, Springer, 2015.

[136] S. Valentine, F. Vides, G. Lucchese, D. Turner, H.-H. A. Kim, W. Li, J. Linsey, and T. Hammond, "Mechanix: A sketch-based tutoring system for statics courses," in *Proceedings of the Twenty-Fourth Innovative Applications of Artificial Intelligence Conference (IAAI)*, (Toronto, Canada, Canada), pp. 2253–2260, AAAI, July 22-26, 2012.

[137] O. Atilola, S. Valentine, H.-H. A. Kim, D. Turner, E. McTigue, T. Hammond, and J. Linsey, "Mechanix: A natural sketch interface tool for teaching truss analysis and free-body diagrams," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AIEDAM)*, vol. 28-2, pp. 169–192, May 2014. ISSN: 1469-1760, `https://doi.org/10.1017/S0890060414000079`.

[138] J. Peschel and T. Hammond, "Strat: A sketched-truss recognition and analysis tool," in *2008 International Workshop on Visual Languages and Computing (VLC) at the 14th*

International Conference on distributed Multimedia Systems (DMS), (Boston, MA, USA), pp. 282–287, Knowledge Systems Instistute, September 4-6, 2008.

[139] K. Kebodeaux, M. Field, and T. Hammond, "Defining precise measurements with sketched annotations," in *Proceedings of the Eighth Eurographics Symposium on Sketch-Based Interfaces and Modeling (SBIM)*, (Vancouver, Canada, Canada), pp. 79–86, ACM, August 5–7, 2011. ISBN: 978-1-4503-0906-6.

[140] O. Atilola, M. Field, E. McTigue, T. Hammond, and J. Linsey, "Evaluation of a natural sketch interface for truss fbds and analysis," in *Frontiers in Education Conference (FIE)*, (Rapid City, SD, USA), pp. S2E–1 – S2E–6, IEEE, October 2011. ISBN: 978-1-61284-468-8.

[141] T. Hammond, "Workshop - integrating sketch recognition technologies into your classroom," in *38th Annual Frontiers in Education Conference*, (Saratoga Springs, NY, USA), pp. W2C–1, IEEE, October 22-25, 2008. ISBN: 978-1-4244-1969-2.

[142] S. Polsley, J. Ray, T. Nelligan, M. Helms, J. Linsey, and T. Hammond, "Leveraging trends in student interaction to enhance the effectiveness of sketch-based educational software," in *9th Workshop on Pen and Touch Technology in Education. WPTTE 2015*, (Microsoft Campus, Redmond, WA, USA), April 28–30, 2015.

[143] S. Polsley, J. Ray, T. Nelligan, M. Helms, J. Linsey, and T. Hammond, "Leveraging trends in student interaction to enhance the effectiveness of sketch-based educational software," in *Revolutionizing Education with Digital Ink: The Impact of Pen and Touch Technology on Education* (T. Hammond, S. Valentine, and A. Adler, eds.), Human-Computer Interaction Series, ch. 7, pp. 103–114, Switzerland: Springer, 2016. ISBN: 978-3-319-31193-7, `http://dx.doi.org/10.1007/978-3-319-31193-7_7`.

[144] M. Green, B. Caldwell, M. Helms, J. Linsey, and T. Hammond, "Using natural sketch recognition software to provide instant feedback on statics homework (truss free body diagrams): Assessment of a classroom pilot," in *2015 ASEE Annual Conference and*

Exposition, (Seattle, WA, USA), pp. 26.1671.1 – 26.1671.12, ASEE, June 14, 2015. ISBN: 978-0-692-50180-1.

[145] M. Bresler, T. Van Phan, D. Průša, M. Nakagawa, and V. Hlaváč, "Recognition system for on-line sketched diagrams," in *2014 14th International Conference on Frontiers in Handwriting Recognition*, pp. 563–568, IEEE, 2014.

[146] M. Bresler, D. Průša, and V. Hlaváč, "Online recognition of sketched arrow-connected diagrams," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 19, no. 3, pp. 253–267, 2016.

[147] C. Wang, H. Mouchere, C. Viard-Gaudin, and L. Jin, "Combined segmentation and recognition of online handwritten diagrams with high order markov random field," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 252–257, IEEE, 2016.

[148] O. Altun and O. Nooruldeen, "Sketrack: stroke-based recognition of online hand-drawn sketches of arrow-connected diagrams and digital logic circuit diagrams," *Scientific Programming*, vol. 2019, 2019.

[149] J.-I. Herrera-Camara and T. Hammond, "Flow2code: From hand-drawn flowcharts to code execution," in *Proceedings of the Symposium on Sketch-Based Interfaces and Modeling*, pp. 1–13, 2017.

[150] M. Agrawal, A. Zotov, M. Ye, and S. Raghupathy, "Context aware on-line diagramming recognition," in *2010 12th International Conference on Frontiers in Handwriting Recognition*, pp. 682–687, IEEE, 2010.

[151] A.-M. Awal, G. Feng, H. Mouchere, and C. Viard-Gaudin, "First experiments on a new online handwritten flowchart database," in *Document Recognition and Retrieval XVIII*, vol. 7874, p. 78740A, International Society for Optics and Photonics, 2011.

[152] R. A. Sottilare, A. C. Graesser, X. Hu, A. Olney, B. Nye, and A. M. Sinatra, *Design Recommendations for Intelligent Tutoring Systems: Volume 4 - Domain Modeling*. US Army Research Laboratory, July 2016.

[153] P. Taele, J. Peschel, and T. Hammond, "A Sketch Interactive Approach to Computer-Assisted Biology Instruction," *Intelligent User Interfaces Workshop on Sketch Recognition*, Feb. 2009.

[154] B. Williford, P. Taele, T. Nelligan, W. Li, J. Linsey, and T. Hammond, "PerSketchTivity: An Intelligent Pen-Based Educational Application for Design Sketching Instruction," in *Revolutionizing Education with Digital Ink: The Impact of Pen and Touch Technology on Education* (T. Hammond, S. Valentine, and A. Adler, eds.), Human–Computer Interaction Series, pp. 115–127, Cham: Springer International Publishing, 2016.

[155] B. Williford, "SketchTivity: Improving Creativity by Learning Sketching with an Intelligent Tutoring System," in *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, C&C '17, (New York, NY, USA), pp. 477–483, Association for Computing Machinery, June 2017.

[156] B. Williford, M. Runyon, A. Malla, W. Li, J. Linsey, and T. Hammond, "Zensketch: A sketch-based game for improving line work," in *The ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY), Student Game Design Competition*, (Amsterdam, The Netherlands), pp. 591–598, ACM, October 15-18, 2017. ISBN: 978-1-4503-5111-9, DOI: 10.1145/3130859.3130861, `https://dl.acm.org/citation.cfm?id=3130861`.

[157] E. Hilton, T. Gamble, W. Li, T. Hammond, and J. Linsey, "Back to basics: Sketching, not cad, is the key to improving essential engineering design skills," in *ASME IDETC 2018-DTM*, (Quebec City, Canada), August 2018. DOI: 10.1115/DETC2018-86325, `http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?articleid=2713706&resultClick=3`.

[158] E. Hilton, B. Williford, W. Li, T. Hammond, and J. Linsey, "Teaching engineering students free-hand sketching with an intelligent tutoring system," in *11th Conference on Pen and Touch Technology in Education. CPTTE 2017*, (Northwestern University, Evanston, IL, USA), October 12-14, 2017.

[159] E. Hilton, M. Paige, B. Williford, W. Li, T. Hammond, and J. Linsey, "Improving the sketching ability of engineering design students," in *International Conference on Engineering Design (ICED)*, vol. 9: Design Education, (Vancouver, BC, Canada), pp. 217–224, The Design Society, August 21–25 2017. ISBN: 978-1-904670-97-1.

[160] E. Hilton, M. Paige, B. Williford, W. Li, T. Hammond, and J. Linsey, "Engineering drawing for the next generation: Students gaining additional skills in the same timeframe," in *American Society for Engineering Education Annual Conference (ASEE)*, vol. Board # 52: Engineering Drawing for the Next Generation: Students Gaining Additional Skills in the Same Timeframe, (Columbus, OH, USA), pp. 1–12, ASEE, June 24-28 2017. `https://peer.asee.org/27874`.

[161] W. Li, E. Hilton, T. Hammond, and J. Linsey, "Persketchtivity: An intelligent pen-based online education platform for sketching instruction," in *Electronic Visualisation and the Arts (EVA 2016)*, (London, England), pp. 133–141, BCS, July 12–14 2016. `http://dx.doi.org/10.14236/ewic/EVA2016.28`.

[162] E. Hilton, B. Williford, W. Li, E. McTigue, T. Hammond, and J. Linsey, "Consistently evaluating sketching ability in engineering curriculum," in *4th ICDC, International Conference on Design and creativity*, (Atlanta, GA, USA), The Design Society, November 4, 2016. ISBN: 978-1-904670-82-7.

[163] B. Williford, M. Runyon, J. Cherian, W. Li, J. Linsey, and T. Hammond, "A Framework for Motivating Sketching Practice with Sketch-based Gameplay," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '19, (New York, NY, USA), pp. 533–544, Association for Computing Machinery, Oct. 2019.

[164] S. Keshavabhotla, B. Williford, E. Hilton, P. Taele, W. Li, J. Linsey, and T. Hammond, "Conquering the cube: Learning to sketch primitives in perspective with an intelligent tutoring system," in *SBIM '17 Proceedings of the Symposium on Sketch-Based Interfaces and Modeling*, vol. 2, (Los Angeles, CA, USA), pp. 1–11, ACM, July 29–30 2017. ISBN: 978-1-4503-5079-2, DOI: 10.1145/3092907.3092911.

[165] T. Hammond, S. P. A. Kumar, M. Runyon, J. Cherian, B. Williford, S. Keshavabhotla, S. Valentine, W. Li, and J. Linsey, "It's not just about accuracy: Metrics that matter when modeling expert sketching ability," *ACM Trans. Interact. Intell. Syst.*, vol. 8, July 2018.

[166] P. Taele, J. Peschel, and T. Hammond, "A sketch interactive approach to computer-assisted biology instruction," in *Proceedings of the Workshop on Sketch Recognition at the 14th International Conference of Intelligent User Interfaces Posters (IUI)*, (Sanibel, FL, USA), p. 2 pages, ACM, February 8-11, 2009.

[167] B. Chao, X. Zhao, D. Shi, G. Feng, and B. Luo, "Eyes Understand the Sketch! Gaze-Aided Stroke Grouping of Hand-Drawn Flowcharts," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17, (New York, NY, USA), pp. 79–83, Association for Computing Machinery, Mar. 2017.

[168] D. Dixon, M. Prasad, and T. Hammond, "iCanDraw: Using sketch recognition and corrective feedback to assist a user in drawing human faces," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, (New York, NY, USA), pp. 897–906, Association for Computing Machinery, Apr. 2010.

[169] D. Cummings, F. Vides, and T. Hammond, "I don't believe my eyes!: Geometric sketch recognition for a computer art tutorial," in *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling (SBIM)*, (Annecy, France, France), pp. 97–106, Eurographics Association, June 4-6, 2012. ISBN: 978-3-905674-42-2.

[170] T. Hammond, B. Eoff, B. Paulson, A. Wolin, K. Dahmen, J. Johnston, and P. Rajan, "Free-sketch recognition: Putting the chi in sketching," in *CHI '08 Extended Abstracts on*

Human Factors in Computing Systems, CHI EA '08, (New York, NY, USA), pp. 3027–3032, Association for Computing Machinery, Apr. 2008.

[171] S. Singh and M. Sinha, "Forensic sketch recognition using user specific facial region," *International Journal of Biometrics*, vol. 8, pp. 134–144, Jan. 2016.

[172] T. Hammond, M. Prasad, and D. Dixon, "Art 101: Learning to draw through sketch recognition," in *Proceedings of 10th International Symposium on Smart Graphics, Lecture Notes In Computer Science 6133*, vol. 633, (Banff, Canada, Canada), pp. 277–280, Springer Berlin Heidelberg, June 24-26, 2010. ISBN: 978-3-642-13543-9.

[173] A. Bhat and T. Hammond, "Using entropy to distinguish shape versus text in hand-drawn diagrams," in *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI)*, (Pasadena, CA, USA), pp. 1395–1400, AAAI, July 11-17, 2009. ISBN: 978-1-57735-426-0, `http://www.aaai.org/ocs/index.php/IJCAI/IJCAI-09/paper/download/592/906`.

[174] T. Chu, P. Taele, and T. Hammond, "Supporting Chinese Character Educational Interfaces with Richer Assessment Feedback through Sketch Recognition," in *Proceedings of the 44th Graphics Interface Conference*, GI '18, (Waterloo, CAN), pp. 50–57, Canadian Human-Computer Communications Society, June 2018.

[175] S.-L. Leung, P.-C. Chee, C. Chan, and Q. Huo, "Contextual vector quantization modeling of hand-printed Chinese character recognition," in *Proceedings of the 1995 International Conference on Image Processing (Vol. 3)-Volume 3 - Volume 3*, ICIP '95, (USA), p. 3432, IEEE Computer Society, Oct. 1995.

[176] P. Taele and T. Hammond, "LAMPS: A sketch recognition-based teaching tool for Mandarin Phonetic Symbols I," *Journal of Visual Languages & Computing*, vol. 21, pp. 109–120, Apr. 2010.

[177] P. Taele and T. Hammond, "Using a geometric-based sketch recognition approach to sketch chinese radicals," in *Proceedings of the Twenty-Third AAAI Conference on Artificial*

Intelligence (AAAI) Student Abstracts, (Chicago, IL, USA), pp. 1832–1833, AAAI, July 13-17, 2008.

[178] P. Taele and T. Hammond, "Enhancing instruction of written east asian languages with sketch recognition-based intelligent language workbook interfaces," in *The Impact of Pen and Touch Technology on Education* (T. Hammond, S. Valentine, A. Adler, and M. Payton, eds.), Human-Computer Interaction Series, ch. 13, pp. 119–126, Switzerland: Springer, 2015. ISBN: 978-3-319-15594-4, `http://dx.doi.org/10.1007/978-3-319-15594-4_12`.

[179] P. Taele and T. Hammond, "Boponoto: An intelligent sketch education application for learning zhuyin phonetic script," in *Proceedings - DMS 2015: 21st International Conference on Distributed Multimedia Systems*, (Vancouver, Canada), September 2015. `https://doi.org/10.18293/DMS2015-031`.

[180] P. Taele and T. Hammond, "Hashigo: A Next-Generation Sketch Interactive System for Japanese Kanji," *Twenty-First IAAI Conference*, p. 6, 2000.

[181] P. Taele, J. I. Koh, and T. Hammond, "Kanji Workbook: A Writing-Based Intelligent Tutoring System for Learning Proper Japanese Kanji Writing Technique with Instructor-Emulated Assessment," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 13382–13389, Apr. 2020.

[182] S. Nabeel, B. Paulson, and T. Hammond, "Urdu qaeda: Recognition system for isolated urdu characters," in *Proceedings of the Workshop on Sketch Recognition at the 14th International Conference of Intelligent User Interfaces (IUI)*, (Sanibel, FL, USA), p. 4 pages, ACM, February 8-11, 2009.

[183] M. T. Parvez and S. A. Mahmoud, "Offline arabic handwritten text recognition: A Survey," *ACM Computing Surveys*, vol. 45, pp. 23:1–23:35, Mar. 2013.

[184] M. H. Shirali-Shahreza, K. Faez, and A. Khotanzad, "Recognition of handwritten Persian/Arabic numerals by shadow coding and an edited probabilistic neural network,"

in *Proceedings of the 1995 International Conference on Image Processing (Vol. 3)-Volume 3 - Volume 3*, ICIP '95, (USA), p. 3436, IEEE Computer Society, Oct. 1995.

[185] E. G. Miller and P. A. Viola, "Ambiguity and constraint in mathematical expression recognition," in *AAAI/IAAI*, 1998.

[186] S. Lavirotte and L. Pottier, "Mathematical formula recognition using graph grammar," in *Document Recognition V*, vol. 3305, pp. 44–52, International Society for Optics and Photonics, 1998.

[187] Y. Eto and M. Suzuki, "Mathematical formula recognition using virtual link network," in *Proceedings of sixth international conference on document analysis and recognition*, pp. 762–767, IEEE, 2001.

[188] T. Hammond, D. Logsdon, B. Paulson, J. Johnston, J. Peschel, A. Wolin, and P. Taele, "A sketch recognition system for recognizing free-hand course of action diagrams," in *Proceedings of the Twenty-Second Innovative Applications of Artificial Intelligence Conference (IAAI)*, (Atlanta, GA, USA), pp. 1781–1786, AAAI, July 11-15, 2010.

[189] T. Hammond, D. Logsdon, J. Peschel, J. Johnston, P. Taele, A. Wolin, and B. Paulson, "A sketch recognition interface that recognizes hundreds of shapes in course-of-action diagrams," in *CHI'10 Extended Abstracts on Human Factors in Computing Systems (CHI)*, (Atlanta, GA, USA), pp. 4213–4218, ACM, April 10-15, 2010. ISBN: 978-1-60558-930-5.

[190] D. Cummings, S. Fymat, and T. Hammond, "Sketch-based interface for interaction with unmanned air vehicles," in *CHI'12 Extended Abstracts on Human Factors in Computing Systems (CHI)*, (Austin, TX, USA), pp. 1511–1516, ACM, May 5–10, 2012. ISBN: 978-1-4503-1016-1.

[191] D. Cummings, S. Fymat, and T. Hammond, "Reddog: A smart sketch interface for autonomous aerial aystems," in *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling (SBIM)*, (Annecy, France, France), pp. 21–28, Eurographics Association, June 4–6, 2012. ISBN: 978-3-905674-42-2.

[192] P. Taele, L. Barreto, and T. Hammond, "Maestoso: An intelligent educational sketching tool for learning music theory," in *The Twenty-Seventh Annual Conference on Innovative Applications of Artificial Intelligence at AAAI (IAAI 2015)*, (Austin, Texas, USA), p. 7 pages, AAAI, January 27-29, 2015.

[193] L. Barreto, P. Taele, and T. Hammond, "A stylus-driven intelligent tutoring system for music education instruction," in *Revolutionizing Education with Digital Ink: The Impact of Pen and Touch Technology on Education* (T. Hammond, S. Valentine, and A. Adler, eds.), Human-Computer Interaction Series, ch. 10, pp. 141–161, Switzerland: Springer, 2016. ISBN: 978-3-319-31193-7, `http://dx.doi.org/10.1007/978-3-319-31193-7_10`.

[194] T. Van Phan and M. Nakagawa, "Combination of global and local contexts for text/non-text classification in heterogeneous online handwritten documents," *Pattern Recognition*, vol. 51, pp. 112–124, 2016.

[195] A. Delaye and C.-L. Liu, "Multi-class segmentation of free-form online documents with tree conditional random fields," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 17, no. 4, pp. 313–329, 2014.

[196] P. Rajan, P. Taele, and T. Hammond, "Evaluation of paper-pen based sketching interface," in *Proceedings of the 16th International Conference on Distributed Multimedia Systems (DMS)*, (Oak Brook, IL, USA), pp. 321–326, SRL, October 14–16, 2010. ISBN: 1-891706-28-4.

[197] P. Rajan and T. Hammond, "Applying online sketch recognition algorithms to a scanned-in sketch," in *Proceedings of the Workshop on Sketch Recognition at the 14th International Conference of Intelligent User Interfaces Posters (IUI)*, (Sanibel, FL, USA), p. 3 pages, ACM, February 8–11, 2009.

[198] C. M. Bishop, M. Svensen, and G. E. Hinton, "Distinguishing text from graphics in on-line handwritten ink," in *Ninth International Workshop on Frontiers in Handwriting Recognition*,

pp. 142–147, Oct. 2004.

[199] T. Hammond and R. Davis, "Tahuti: A geometrical sketch recognition system for uml class diagrams," in *Technical Report SS-02-08: Papers from the 2002 Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium on Sketch Understanding*, (Menlo Park, California, USA), p. 8 pages, AAAI, July 28-August 1, 2002.

[200] T. Hammond, "Natural sketch recognition in uml class diagrams," in *Proceedings of the MIT Student Oxygen Workshop*, (Gloucester, Massachusetts., USA), p. 3 pages, MIT Aritifial Intelligence Laboratory, July 16, 2001.

[201] T. Hammond, K. Oshiro, and R. Davis, "Natural editing and recognition of uml class diagrams," in *MIT Lab Abstract. Artificial Intelligence Laboratory*, (Cambridge, MA, USA), p. 2 pages, MIT, September 2002.

[202] T. Hammond and B. O'Sullivan, "Recognizing free-form hand-sketched constraint network diagrams by combining geometry and context," in *Proceedings of the Eurographics Ireland*, (Dublin, Ireland, Ireland), pp. 67–74, Eurographics, December 17, 2007.

[203] T. Hammond, D. Logsdon, B. Paulson, J. Johnston, J. Peschel, A. Wolin, and P. Taele, "A Sketch Recognition System for Recognizing Free-Hand Course of Action Diagrams," *Innovative Applications of Artificial Intelligence*, p. 6, 2010.

[204] F. Han and S.-C. Zhu, "Bottom-up/top-down image parsing by attribute graph grammar," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, pp. 1778–1785, IEEE, 2005.

[205] A. Wolin, B. Paulson, and T. Hammond, "Sort, merge, repeat: An algorithm for effectively finding corners in hand-sketched strokes," in *Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling (SBIM)*, (New Orleans, LA, USA), pp. 93–100, ACM, August 1-2, 2009. ISBN: 978-1-60558-602-1.

[206] A. Wolin, M. Field, and T. Hammond, "Combining corners from multiple segmenters," in *Proceedings of the Eighth Eurographics Symposium on Sketch-Based Interfaces and*

Modeling (SBIM), (Vancouver, Canada, Canada), pp. 117–124, ACM, August 5-7, 2011. ISBN: 978-1-4503-0906-6.

[207] A. Wolin, B. Paulson, and T. Hammond, "Eliminating false positives during corner finding by merging similar segments," in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI) Student Abstracts*, (Chicago, IL, USA), pp. 1836–1837, AAAI, July 13-17, 2008.

[208] A. Wolin, B. Eoff, and T. Hammond, "Shortstraw: A simple and effective corner finder for polylines.," in *SBM*, pp. 33–40, 2008.

[209] Y. Xiong and J. J. LaViola Jr, "Revisiting shortstraw: improving corner finding in sketch-based interfaces," in *Proceedings of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling*, pp. 101–108, ACM, 2009.

[210] L. Lin, T. Wu, J. Porway, and Z. Xu, "A stochastic graph grammar for compositional object representation and recognition," *Pattern Recognition*, vol. 42, no. 7, pp. 1297–1307, 2009.

[211] L. Lin, X. Liu, and S.-C. Zhu, "Layered graph matching with composite cluster sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1426–1442, 2009.

[212] D. Rubine, "Combining gestures and direct manipulation," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 659–660, 1992.

[213] T. M. Sezgin, T. Stahovich, and R. Davis, "Sketch based interfaces: early processing for sketch understanding," in *ACM SIGGRAPH 2007 courses*, pp. 37–es, ACM, 2007.

[214] L. B. Kara and T. F. Stahovich, "An image-based, trainable symbol recognizer for hand-drawn sketches," *Computers & Graphics*, vol. 29, no. 4, pp. 501–517, 2005.

[215] A. C. Long Jr, J. A. Landay, L. A. Rowe, and J. Michiels, "Visual similarity of pen gestures," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 360–367, 2000.

[216] B. Paulson and T. Hammond, "Paleosketch: accurate primitive sketch recognition and beautification," in *Proceedings of the 13th international conference on Intelligent user interfaces*, pp. 1–10, 2008.

[217] B. Paulson, P. Rajan, P. Davalos, R. Gutierrez-Osuna, and T. Hammond, "What!?! no rubine features?: Using geometric-based features to produce normalized confidence values for sketch recognition," in *HCC Workshop: Sketch Tools for Diagramming*, pp. 57–63, 2008.

[218] F. T. Alamudun, T. Hammond, H.-J. Yoon, and G. D. Tourassi, "Geometry and gesture-based features from saccadic eye-movement as a biometric in radiology," in *International Conference on Augmented Cognition*, pp. 123–138, Springer, 2017.

[219] G. Lucchese, M. Field, J. Ho, R. Gutierrez-Osuna, and T. Hammond, "Gesturecommander: Continuous touch-based gesture prediction," in *CHI'12 Extended Abstracts on Human Factors in Computing Systems (CHI)*, (Austin, TX, USA), pp. 1925–1930, ACM, May 5–10, 2012. ISBN: 978-1-4503-1016-1.

[220] A. Schlapbach, M. Liwicki, and H. Bunke, "A writer identification system for on-line whiteboard data," *Pattern Recognition*, vol. 41, pp. 2381–2397, July 2008.

[221] B. D. Eoff and T. Hammond, "Who dotted that'i'? context free user differentiation through pressure and tilt pen data," in *Proceedings of Graphics Interface 2009*, pp. 149–156, Citeseer, 2009.

[222] H.-h. Kim, P. Taele, J. Seo, J. Liew, and T. Hammond, "Easysketch2: A novel sketch-based interface for improving children's fine motor skills and school readiness," in *Proceedings of the Joint Symposium on Computational Aesthetics and Sketch Based Interfaces and Modeling and Non-Photorealistic Animation and Rendering*, pp. 69–78, Eurographics Association, 2016.

[223] H.-H. A. Kim, P. Taele, S. Valentine, and T. Hammond, "Developing intelligent sketch-based applications to support children's self-regulation and school readiness," in *2014*

International Conference on Intelligent User Interfaces (IUI) Workshop on Sketch: Pen and Touch Recognition, (Haifa, Israel, Israel), pp. 1–8, ACM, February 24, 2014.

[224] H.-H. Kim, *Analysis of Children's Sketches to Improve Recognition Accuracy in Sketch-Based Applications*. Thesis, Texas A&M University, Dec. 2012.

[225] H.-h. Kim, P. Taele, S. Valentine, E. McTigue, and T. Hammond, "KimCHI: A sketch-based developmental skill classifier to enhance pen-driven educational interfaces for children," in *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling*, SBIM '13, (New York, NY, USA), pp. 33–42, Association for Computing Machinery, July 2013.

[226] H.-h. Kim, S. Valentine, P. Taele, and T. Hammond, "EasySketch: A Sketch-based Educational Interface to Support Children's Self-regulation and School Readiness," in *The Impact of Pen and Touch Technology on Education* (T. Hammond, S. Valentine, A. Adler, and M. Payton, eds.), Human–Computer Interaction Series, pp. 35–46, Cham: Springer International Publishing, 2015.

[227] B. Paulson, B. Eoff, A. Wolin, J. Johnston, and T. Hammond, "Sketch-based educational games: "drawing" kids away from traditional interfaces," in *Proceedings of the 7th International Conference on Interaction Design and Children*, IDC '08, (New York, NY, USA), pp. 133–136, Association for Computing Machinery, June 2008.

[228] P. Zham, D. K. Kumar, P. Dabnichki, S. Poosapadi Arjunan, and S. Raghav, "Distinguishing different stages of parkinson's disease using composite index of speed and pen-pressure of sketching a spiral," *Frontiers in neurology*, vol. 8, p. 435, 2017.

[229] S. Müller, O. Preische, P. Heymann, U. Elbing, and C. Laske, "Increased diagnostic accuracy of digital vs. conventional clock drawing test for discrimination of patients in the early course of alzheimer's disease from cognitively healthy individuals," *Frontiers in aging neuroscience*, vol. 9, p. 101, 2017.

[230] W. Souillard-Mandar, R. Davis, C. Rudin, R. Au, D. J. Libon, R. Swenson, C. C. Price, M. Lamar, and D. L. Penney, "Learning classification models of cognitive conditions from subtle behaviors in the digital clock drawing test," *Machine learning*, vol. 102, no. 3, pp. 393–441, 2016.

[231] J. B. Tornatore, E. Hill, J. A. Laboff, and M. E. McGann, "Self-administered screening for mild cognitive impairment: initial validation of a computerized test battery," *The Journal of neuropsychiatry and clinical neurosciences*, vol. 17, no. 1, pp. 98–105, 2005.

[232] M. N. Sabbagh, M. Boada, S. Borson, M. Chilukuri, P. Doraiswamy, B. Dubois, J. Ingram, A. Iwata, A. Porsteinsson, K. Possin, *et al.*, "Rationale for early diagnosis of mild cognitive impairment (mci) supported by emerging digital technologies," *The Journal of Prevention of Alzheimer's Disease*, pp. 1–7, 2020.

[233] R. P. Fellows, J. Dahmen, D. Cook, and M. Schmitter-Edgecombe, "Multicomponent analysis of a digital trail making test," *The Clinical Neuropsychologist*, vol. 31, no. 1, pp. 154–167, 2017.

[234] R. M. Bauer, G. L. Iverson, A. N. Cernich, L. M. Binder, R. M. Ruff, and R. I. Naugle, "Computerized neuropsychological assessment devices: joint position paper of the american academy of clinical neuropsychology and the national academy of neuropsychology," *Archives of Clinical Neuropsychology*, vol. 27, no. 3, pp. 362–373, 2012.

[235] J. O. Wobbrock, A. D. Wilson, and Y. Li, "Gestures without libraries, toolkits or training: a \$1 recognizer for user interface prototypes," in *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pp. 159–168, ACM, 2007.

[236] R.-D. Vatavu, L. Anthony, and J. O. Wobbrock, "\$ q: A super-quick, articulation-invariant stroke-gesture recognizer for low-resource devices," in *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–12, 2018.

[237] R.-D. Vatavu, L. Anthony, and J. O. Wobbrock, "Gestures as point clouds: a $ p recognizer for user interface prototypes," in *Proceedings of the 14th ACM international conference on Multimodal interaction*, pp. 273–280, 2012.

[238] L. Anthony and J. O. Wobbrock, "$ n-protractor: A fast and accurate multistroke recognizer," in *Proceedings of Graphics Interface 2012*, pp. 117–120, Assocation for Computing Machinery, 2012.

[239] L. Anthony and J. O. Wobbrock, "A lightweight multistroke recognizer for user interface prototypes," in *Proceedings of Graphics Interface 2010*, pp. 245–252, Assocation for Computing Machinery, 2010.

[240] R.-D. Vatavu, "Improving gesture recognition accuracy on touch screens for users with low vision," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 4667–4679, ACM, 2017.

[241] H. Chen, Z. J. Xu, Z. Q. Liu, and S. C. Zhu, "Composite templates for cloth modeling and sketching," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 943–950, IEEE, 2006.

[242] L. Lin, S. Peng, J. Porway, S.-C. Zhu, and Y. Wang, "An empirical study of object category recognition: Sequential testing with generalized samples," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.

[243] L. Zhu, Y. Chen, and A. Yuille, "Unsupervised learning of a probabilistic grammar for object detection and parsing," *Advances in Neural Information Processing Systems*, 2007.

[244] A. Blackler, D. Mahar, and V. Popovic, "Older adults, interface experience and cognitive decline," in *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI '10, (New York, NY, USA), pp. 172–175, Association for Computing Machinery, Nov. 2010.

[245] P. Werner, S. Rosenblum, G. Bar-On, J. Heinik, and A. Korczyn, "Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment," *The*

Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, vol. 61, pp. P228–236, July 2006.

[246] H. B. Jimison, M. Pavel, K. Wild, P. Bissell, J. McKanna, D. Blaker, and D. Williams, "A neural informatics approach to cognitive assessment and monitoring," in *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on*, pp. 696–699, IEEE, 2007.

[247] 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.

[248] A. Kluger, J. G. Gianutsos, J. Golomb, A. Wagner, D. Wagner, and S. Scheurich, "Clinical features of MCI: Motor changes," *International Psychogeriatrics*, vol. 20, pp. 32–39, Feb. 2008.

[249] A. Vuono, M. Luperto, J. Banfi, N. Basilico, N. A. Borghese, M. Sioutis, J. Renoux, and A. Loufti, "Seeking Prevention of Cognitive Decline in Elders via Activity Suggestion by A Virtual Caregiver," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '18, (Richland, SC), pp. 1835–1837, International Foundation for Autonomous Agents and Multiagent Systems, July 2018.

[250] P. Wang, X. Zhang, Y. Liu, S. Liu, B. Zhou, Z. Zhang, H. Yao, X. Zhang, and T. Jiang, "Perceptual and response interference in Alzheimer's disease and mild cognitive impairment," *Clinical Neurophysiology*, vol. 124, pp. 2389–2396, Dec. 2013.

[251] J. H. Yan, S. Rountree, P. Massman, R. S. Doody, and H. Li, "Alzheimer's disease and mild cognitive impairment deteriorate fine movement control," *Journal of Psychiatric Research*, vol. 42, pp. 1203–1212, Oct. 2008.

[252] D. Riboni, C. Bettini, G. Civitarese, Z. H. Janjua, and R. Helaoui, "Smartfaber: Recognizing fine-grained abnormal behaviors for early detection of mild cognitive impairment," *Artificial intelligence in medicine*, vol. 67, pp. 57–74, 2016.

[253] N. Josman, R. Kizony, E. Hof, K. Goldenberg, P. L. Weiss, and E. Klinger, "Using the Virtual Action Planning-Supermarket for Evaluating Executive Functions in People with Stroke," *Journal of Stroke and Cerebrovascular Diseases*, vol. 23, pp. 879–887, May 2014.

[254] C.-F. Jiang and Y.-S. Li, "Development and verification of a vr platform to evaluate wayfinding abilities," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 2396–2399, IEEE, 2009.

[255] B. Drew and J. Waters, "Video games: Utilization of a novel strategy to improve perceptual motor skills and cognitive functioning in the non-institutionalized elderly.," *Cognitive Rehabilitation*, 1986.

[256] S. Giroux, J. Bauchet, H. Pigot, D. Lussier-Desrochers, and Y. Lachappelle, "Pervasive behavior tracking for cognitive assistance," in *Proceedings of the 1st International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '08, (New York, NY, USA), pp. 1–7, Association for Computing Machinery, July 2008.

[257] H. Kim, C.-P. Hsiao, and E. Y.-L. Do, "Home-based computerized cognitive assessment tool for dementia screening," *Journal of Ambient Intelligence and Smart Environments*, vol. 4, pp. 429–442, Sept. 2012.

[258] M. M. Mielke, S. D. Weigand, H. J. Wiste, P. Vemuri, M. M. Machulda, D. S. Knopman, V. Lowe, R. O. Roberts, K. Kantarci, W. A. Rocca, C. R. Jack, and R. C. Petersen, "Independent comparison of CogState computerized testing and a standard cognitive battery with neuroimaging," *Alzheimer's & Dementia*, vol. 10, pp. 779–789, Nov. 2014.

[259] V. Mylonakis, J. Soldatos, A. Pnevmatikakis, L. Polymenakos, A. Sorin, and H. Aronowitz, "Using robust audio and video processing technologies to alleviate the elderly cognitive decline," in *Proceedings of the 1st International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '08, (New York, NY, USA), pp. 1–8, Association for Computing Machinery, July 2008.

[260] M. E. Pollack, "Intelligent Technology for an Aging Population: The Use of AI to Assist Elders with Cognitive Impairment," *AI Magazine*, vol. 26, pp. 9–9, June 2005.

[261] R. Saha, A. Mukherjee, A. Sadhukhan, A. Roy, and M. De, "Handwriting Analysis for Early Detection of Alzheimer's Disease," in *Intelligent Data Analysis*, ch. 18, pp. 369–385, John Wiley & Sons, Ltd, 2020.

[262] I. Tarnanas, M. Tsolaki, T. Nef, R. M. Müri, and U. P. Mosimann, "Can a novel computerized cognitive screening test provide additional information for early detection of Alzheimer's disease?," *Alzheimer's & Dementia*, vol. 10, pp. 790–798, Nov. 2014.

[263] C. A. de Jager, A.-C. M. Schrijnemaekers, T. E. Honey, and M. M. Budge, "Detection of mci in the clinic: evaluation of the sensitivity and specificity of a computerised test battery, the hopkins verbal learning test and the mmse," *Age and ageing*, vol. 38, no. 4, pp. 455–460, 2009.

[264] H. Jimison, M. Pavel, J. McKanna, and J. Pavel, "Unobtrusive monitoring of computer interactions to detect cognitive status in elders," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, no. 3, pp. 248–252, 2004,.

[265] C. E. Drapeau, M. Bastien-Toniazzo, C. Rous, and M. Carlier, "Nonequivalence of computerized and paper-and-pencil versions of trail making test," *Perceptual and motor skills*, vol. 104, no. 3, pp. 785–791, 2007.

[266] C. T. Gualtieri, "Dementia screening using computerized tests," *JOURNAL OF INSURANCE MEDICINE-NEW YORK THEN DENVER–*, vol. 36, pp. 213–227, 2004.

[267] C.-C. Lin, M.-J. Chiu, C.-C. Hsiao, R.-G. Lee, and Y.-S. Tsai, "Wireless health care service system for elderly with dementia," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 4, pp. 696–704, 2006.

[268] U. Varshney, "A framework for wireless monitoring of mental health conditions," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 5219–5222, IEEE, 2009.

166

[269] Y.-J. Lin, H.-S. Chen, and M.-J. Su, "A cloud based bluetooth low energy tracking system for dementia patients," in *Mobile Computing and Ubiquitous Networking (ICMU), 2015 Eighth International Conference on*, pp. 88–89, IEEE, 2015.

[270] B. Xiao, M. Z. Asghar, T. Jamsa, and P. Pulii, "Canderoid: A mobile system to remotely monitor travelling status of the elderly with dementia," in *Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA), 2013 International Joint Conference on*, pp. 648–654, IEEE, 2013.

[271] I. Zavala-Ibarra and F. Jesús, "Ambient videogames for health monitoring in older adults," in *Intelligent Environments (IE), 2012 8th International Conference on*, pp. 27–33, IEEE, 2012.

[272] F. Gong, W. Xu, J.-Y. Lee, L. He, and M. Sarrafzadeh, "Neuroglasses: A neural sensing healthcare system for 3-d vision technology," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 2, pp. 198–204, 2012.

[273] G. Gowans, J. Campbell, N. Alm, R. Dye, A. Astell, and M. Ellis, "Designing a multimedia conversation aid for reminiscence therapy in dementia care environments," in *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, pp. 825–836, ACM, 2004.

[274] M. Moetesum, O. Zeeshan, and I. Siddiqi, "Multi-object sketch segmentation using convolutional object detectors," in *Tenth International Conference on Graphics and Image Processing (ICGIP 2018)*, vol. 11069, International Society for Optics and Photonics, 2019.

[275] M. Moetesum, I. Siddiqi, U. Masroor, and C. Djeddi, "Automated scoring of bender gestalt test using image analysis techniques," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pp. 666–670, IEEE, 2015.

[276] M. C. Fairhurst, T. Linnell, S. Glenat, R. Guest, L. Heutte, and T. Paquet, "Developing a generic approach to online automated analysis of writing and drawing tests in clinical patient profiling," *Behavior Research Methods*, vol. 40, no. 1, pp. 290–303, 2008.

[277] M. Bennasar, R. Setchi, Y. Hicks, and A. Bayer, "Cascade classification for diagnosing dementia," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2535–2540, IEEE, 2014.

[278] P. Ghaderyan, A. Abbasi, and S. Saber, "A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of alzheimer's disease," *Expert Systems with Applications*, vol. 114, pp. 428–440, Dec. 2018.

[279] K. K. Tsoi, M. W. Lam, C. T. Chu, M. P. Wong, and H. M. Meng, "Machine Learning on Drawing Behavior for Dementia Screening," in *Proceedings of the 2018 International Conference on Digital Health*, DH '18, (New York, NY, USA), pp. 131–132, Association for Computing Machinery, Apr. 2018.

[280] Z. Harbi, Y. Hicks, and R. Setchi, "Clock drawing test digit recognition using static and dynamic features," *Procedia Computer Science*, vol. 96, pp. 1221–1230, 2016.

[281] Z. Harbi, Y. Hicks, and R. Setchi, "Clock drawing test interpretation system," *Procedia computer science*, vol. 112, pp. 1641–1650, 2017.

[282] C. R. Pereira, D. R. Pereira, F. A. Da Silva, C. Hook, S. A. Weber, L. A. Pereira, and J. P. Papa, "A step towards the automated diagnosis of parkinson's disease: Analyzing handwriting movements," in *2015 IEEE 28th international symposium on computer-based medical systems*, pp. 171–176, IEEE, 2015.

[283] S. Glenat, L. Heutte, T. Paquet, R. Guest, M. Fairhurst, and T. Linnell, "The development of a computer-assisted tool for the assessment of neuropsychological drawing tasks," *International Journal of Information Technology & Decision Making*, vol. 7, no. 04, pp. 751–767, 2008.

[284] D. Impedovo, G. Pirlo, G. Vessio, and M. T. Angelillo, "A Handwriting-Based Protocol for Assessing Neurodegenerative Dementia," *Cognitive Computation*, vol. 11, pp. 576–586, Aug. 2019.

[285] C. Kahindo, M. A. El-Yacoubi, S. Garcia-Salicetti, V. Cristancho-Lacroix, H. Kerhervé, and A. S. Rigaud, "Semi-global Parameterization of Online Handwriting Features for Characterizing Early-Stage Alzheimer and Mild Cognitive Impairment," *IRBM*, vol. 39, pp. 421–429, Dec. 2018.

[286] J. Kawa, A. Bednorz, P. Stępień, J. Derejczyk, and M. Bugdol, "Spatial and dynamical handwriting analysis in mild cognitive impairment," *Computers in Biology and Medicine*, vol. 82, pp. 21–28, Mar. 2017.

[287] A. Prange, M. Barz, and D. Sonntag, "A categorisation and implementation of digital pen features for behaviour characterisation," *arXiv preprint arXiv:1810.03970*, 2018.

[288] R. Canham, S. Smith, and A. Tyrrell, "Location of structural sections from within a highly distorted complex line drawing," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 152, no. 6, pp. 741–749, 2005.

[289] J. Vogt, H. Kloosterman, S. Vermeent, G. Van Elswijk, R. Dotsch, and B. Schmand, "Automated scoring of the rey-osterrieth complex figure test using a deep-learning algorithm," *Archives of Clinical Neuropsychology*, vol. 34, no. 6, pp. 836–836, 2019.

[290] J. M. Thornbury, "The use of piaget's theory in alzheimer's disease," *American Journal of Alzheimer's Care and Related Disorders & Research*, vol. 8, no. 4, pp. 16–21, 1993.

[291] M. Moetesum, I. Siddiqi, U. Masroor, N. Vincent, and F. Cloppet, "Segmentation and classification of offline hand drawn images for the bgt neuropsychological screening test," in *Eighth International Conference on Digital Image Processing (ICDIP 2016)*, vol. 10033, p. 100334N, International Society for Optics and Photonics, 2016.

[292] H. B. Nazar, M. Moetesum, S. Ehsan, I. Siddiqi, K. Khurshid, N. Vincent, and K. D. McDonald-Maier, "Classification of graphomotor impressions using convolutional neural networks: An application to automated neuro-psychological screening tests," in *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, vol. 1, pp. 432–437, IEEE, 2017.

[293] R. Lara-Garduno, N. Leslie, and T. Hammond, "Smartstrokes: digitizing paper-based neuropsychological tests," in *Revolutionizing Education with Digital Ink*, pp. 163–175, Springer, 2016.

[294] L. A. Tupler, K. A. Welsh, Y. Asare-Aboagye, and D. V. Dawson, "Reliability of the rey-osterrieth complex figure in use with memory-impaired patients," *Journal of clinical and experimental neuropsychology*, vol. 17, no. 4, pp. 566–579, 1995.

[295] D. W. Loring, R. C. Martin, K. J. Meador, and G. P. Lee, "Psychometric construction of the rey-osterrieth complex figure: methodological considerations and interrater reliability," *Archives of Clinical Neuropsychology*, vol. 5, no. 1, pp. 1–14, 1990.

[296] J. E. Meyers and K. R. Meyers, "Rey complex figure test under four different administration procedures," *The Clinical Neuropsychologist*, vol. 9, no. 1, pp. 63–67, 1995.

[297] M. Moetesum, I. Siddiqi, S. Ehsan, and N. Vincent, "Deformation modeling and classification using deep convolutional neural networks for computerized analysis of neuropsychological drawings," *Neural Computing and Applications*, pp. 1–25, 2020.

[298] H. Imai and T. Asano, "Efficient Algorithms for Geometric Graph Search Problems," *SIAM Journal on Computing*, vol. 15, pp. 478–494, May 1986.

[299] E. W. Dijkstra *et al.*, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[300] J.-c. Chen, *Dijkstra's Shortest Path Algorithm*. Journal of Formalized Mathematics, 2003.

[301] A. Crauser, K. Mehlhorn, U. Meyer, and P. Sanders, "A parallelization of Dijkstra's shortest path algorithm," in *Mathematical Foundations of Computer Science 1998* (L. Brim, J. Gruska, and J. Zlatuška, eds.), Lecture Notes in Computer Science, (Berlin, Heidelberg), pp. 722–731, Springer, 1998.

[302] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM journal on computing*, vol. 1, no. 2, pp. 146–160, 1972.

[303] C. Alvarado, "Sketch Recognition User Interfaces: Guidelines for Design and Development," *Proceedings of AAAI Fall Symposium on Intelligent Pen-based Interfaces*, vol. 1, no. 2, p. 7, 2004.

[304] F. Ingram, V. M. Soukup, and P. T. Ingram, "The Medical College of Georgia Complex Figures: Reliability and preliminary normative data using an intentional learning paradigm in older adults," *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, vol. 10, pp. 144–146, Apr. 1997.

[305] A. Aggarwal and R. Anderson, "Parallel depth-first search in general directed graphs | Proceedings of the twenty-first annual ACM symposium on Theory of computing," Feb. 2021.

[306] G. J. Holzmann, D. Peled, and M. Yannakakis, "On Nested DepthFirst Search," in *Proc. Second SPIN Workshop, Rutgers Univ*, 1996.

[307] D. B. Johnson, "A Note on Dijkstra's Shortest Path Algorithm," *Journal of the ACM*, vol. 20, pp. 385–388, July 1973.

[308] Y. Li, "Protractor: a fast and accurate gesture recognizer," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 2169–2172, 2010.

[309] D. Willems, R. Niels, M. van Gerven, and L. Vuurpijl, "Iconic and multi-stroke gesture recognition," *Pattern recognition*, vol. 42, no. 12, pp. 3303–3312, 2009.

[310] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*. Courier Corporation, 1998.

[311] 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.

[312] T. M. Sezgin and R. Davis, "Hmm-based efficient sketch recognition," in *Proceedings of the 10th international conference on Intelligent user interfaces*, pp. 281–283, 2005.

[313] D. Avis, "A survey of heuristics for the weighted matching problem," *Networks*, vol. 13, no. 4, pp. 475–493, 1983.

[314] C. Gallagher and T. Burke, "Age, gender and iq effects on the rey-osterrieth complex figure test," *British Journal of Clinical Psychology*, vol. 46, no. 1, pp. 35–45, 2007.

[315] C. R. Savage, L. Baer, N. J. Keuthen, H. D. Brown, S. L. Rauch, and M. A. Jenike, "Organizational strategies mediate nonverbal memory impairment in obsessive–compulsive disorder," *Biological psychiatry*, vol. 45, no. 7, pp. 905–916, 1999.

[316] P. Anderson, V. Anderson, and J. Garth, "Assessment and Development of Organizational Ability: The Rey Complex Figure Organizational Strategy Score (RCF-OSS)*," *The Clinical Neuropsychologist*, vol. 15, pp. 81–94, Feb. 2001.

[317] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert Systems with Applications*, vol. 134, pp. 93–101, Nov. 2019.

[318] M. S. Albert, S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, *et al.*, "The diagnosis of mild cognitive impairment due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease," *Alzheimer's & dementia*, vol. 7, no. 3, pp. 270–279, 2011.

[319] D. M. Masur, M. Sliwinski, R. Lipton, A. Blau, and H. Crystal, "Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons," *Neurology*, vol. 44, no. 8, pp. 1427–1427, 1994.

[320] C. Flicker, S. H. Ferris, and B. Reisberg, "A two-year longitudinal study of cognitive function in normal aging and alzheimer's disease," *Journal of Geriatric Psychiatry and Neurology*, vol. 6, no. 2, pp. 84–96, 1993.

[321] J. Dahmen, D. Cook, R. Fellows, and M. Schmitter-Edgecombe, "An analysis of a digital variant of the trail making test using machine learning techniques," *Technology and Health Care*, vol. 25, no. 2, pp. 251–264, 2017.

[322] J. Brandt, M. Spencer, M. Folstein, *et al.*, "The telephone interview for cognitive status," *Neuropsychiatry Neuropsychol Behav Neurol*, vol. 1, no. 2, pp. 111–117, 1988.

[323] B. Dubois, A. Slachevsky, I. Litvan, and B. Pillon, "The fab: a frontal assessment battery at bedside," *Neurology*, vol. 55, no. 11, pp. 1621–1626, 2000.

[324] T. Miner and F. R. Ferraro, "The role of speed of processing, inhibitory mechanisms, and presentation order in trail-making test performance," *Brain and cognition*, vol. 38, no. 2, pp. 246–253, 1998.

[325] I. S. MacKenzie, "Fitts' law as a research and design tool in human-computer interaction," *Human-computer interaction*, vol. 7, no. 1, pp. 91–139, 1992.

[326] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement.," *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.

[327] J. Accot and S. Zhai, "Refining Fitts' law models for bivariate pointing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, (New York, NY, USA), pp. 193–200, Association for Computing Machinery, Apr. 2003.

[328] Y. Guiard, "The problem of consistency in the design of Fitts' law experiments: Consider either target distance and width or movement form and scale," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, (New York, NY, USA), pp. 1809–1818, Association for Computing Machinery, Apr. 2009.

[329] M. S. Raghu Prasad, S. Purswani, and M. Manivannan, "Modeling of Human Hand Force Based Tasks Using Fitts's Law," in *ICoRD'13* (A. Chakrabarti and R. V. Prakash, eds.), Lecture Notes in Mechanical Engineering, (India), pp. 377–386, Springer, 2013.

[330] I. S. MacKenzie and W. Buxton, "Extending fitts' law to two-dimensional tasks," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 219–226, 1992.

[331] J. Accot and S. Zhai, "Beyond Fitts' law: Models for trajectory-based HCI tasks," in *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '97, (New York, NY, USA), p. 250, Association for Computing Machinery, Mar. 1997.

[332] H. Drewes, "Only one Fitts' law formula please!," in *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, (New York, NY, USA), pp. 2813–2822, Association for Computing Machinery, Apr. 2010.

[333] Y. Guiard and M. Beaudouin-Lafon, "Fitts' law 50 years later: Applications and contributions from human-computer interaction," *International Journal of Human-Computer Studies*, vol. 61, no. 6, pp. 747–750, 2004.

[334] I. S. MacKenzie, "Fitts' Law as a Research and Design Tool in Human-Computer Interaction," *Human–Computer Interaction*, vol. 7, pp. 91–139, Mar. 1992.

[335] R. W. Soukoreff and I. S. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI," *International Journal of Human-Computer Studies*, vol. 61, pp. 751–789, Dec. 2004.

[336] C. E. Wright and F. Lee, "Issues Related to HCI Application of Fitts's Law," *Human–Computer Interaction*, vol. 28, pp. 548–578, Nov. 2013.

[337] D. Ahlström, "Modeling and improving selection in cascading pull-down menus using Fitts' law, the steering law and force fields," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, (New York, NY, USA), pp. 61–70, Association for Computing Machinery, Apr. 2005.

[338] Y. Guiard and H. H. Olafsdottir, "What is a Zero-Difficulty Movement? A Scale of Measurement Issue in Fitts' Law Research." Available: `https://hal.archives-ouvertes.fr/hal-00545567/document`, Oct. 2010.

[339] R. W. Soukoreff and I. S. MacKenzie, "Generalized Fitts' law model builder," in *Conference Companion on Human Factors in Computing Systems - CHI '95*, (Denver, Colorado, United States), pp. 113–114, ACM Press, 1995.

[340] J. Accot and S. Zhai, "Performance evaluation of input devices in trajectory-based tasks: an application of the steering law," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 466–472, 1999.

[341] J. Accot and S. Zhai, "Scale effects in steering law tasks," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1–8, 2001.

[342] J. T. Dennerlein, D. B. Martin, and C. Hasser, "Force-feedback improves performance for steering and combined steering-targeting tasks," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '00, (New York, NY, USA), pp. 423–429, Association for Computing Machinery, Apr. 2000.

[343] S. Kulikov, I. S. MacKenzie, and W. Stuerzlinger, "Measuring the effective parameters of steering motions," in *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, (New York, NY, USA), pp. 1569–1572, Association for Computing Machinery, Apr. 2005.

[344] R. Pastel, "Measuring the difficulty of steering through corners," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, (New York, NY, USA), pp. 1087–1096, Association for Computing Machinery, Apr. 2006.

[345] R. Pastel, "Measuring the difficulty of steering through corners," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1087–1096, 2006.

[346] B. Paulson and T. Hammond, "A system for recognizing and beautifying low-level sketch shapes using ndde and dcr," in *20th Annual ACM Symposium on User Interface Software and Technology Posters*, 2007.

[347] D. Ruta and B. Gabrys, "Classifier selection for majority voting," *Information Fusion*, vol. 6, pp. 63–81, Mar. 2005.

[348] I. Rish, "An empirical study of the naive Bayes classifier," *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, pp. 41–46, 2001.

[349] A. Priyam, Abhijeet, R. Gupta, A. Rathee, and S. Srivastava, "Comparative Analysis of Decision Tree Classification Algorithms," *International Journal of current engineering and technology*, vol. 3, no. 2, pp. 334–337, 2013.

[350] J. C. Bezdek, S. K. Chuah, and D. Leep, "Generalized k-nearest neighbor rules," *Fuzzy Sets and Systems*, vol. 18, pp. 237–256, Apr. 1986.

[351] I. Naseem, R. Togneri, and M. Bennamoun, "Linear Regression for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 2106–2112, Nov. 2010.

[352] G. Fung and O. L. Mangasarian, "Incremental Support Vector Machine Classification," in *Proceedings of the 2002 SIAM International Conference on Data Mining*, Proceedings, pp. 247–260, Society for Industrial and Applied Mathematics, Apr. 2002.

[353] Tao Xiong and V. Cherkassky, "A combined SVM and LDA approach for classification," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 3, pp. 1455–1459 vol. 3, July 2005.

[354] A. Palczewska, J. Palczewski, R. Marchese Robinson, and D. Neagu, "Interpreting Random Forest Classification Models Using a Feature Contribution Method," in *Integration of Reusable Systems* (T. Bouabana-Tebibel and S. H. Rubin, eds.), Advances in Intelligent Systems and Computing, pp. 193–218, Cham: Springer International Publishing, 2014.

[355] R. Lara-Garduno, T. Igarashi, and T. Hammond, "3d-trail-making test: A touch-tablet cognitive test to support intelligent behavioral recognition," in *Proceedings of Graphics Interface 2019*, GI 2019, Canadian Information Processing Society, 2019.

[356] M. F. Mendez, R. L. Tomsak, and B. Remler, "Disorders of the visual system in alzheimer's disease.," *Journal of Neuro-Ophthalmology*, vol. 10, no. 1, pp. 62–69, 1990.

[357] D. Prvulovic, D. Hubl, A. Sack, L. Melillo, K. Maurer, L. Frölich, H. Lanfermann, F. Zanella, R. Goebel, D. Linden, *et al.*, "Functional imaging of visuospatial processing in alzheimer's disease," *Neuroimage*, vol. 17, no. 3, pp. 1403–1414, 2002.

[358] M. Goebel, "Ergonomic design of computerized devices for elderly persons-the challenge of matching antagonistic requirements," *Universal Acess in Human Computer Interaction. Coping with Diversity*, pp. 894–903, 2007.

[359] J. Davidoff and E. K. Warrington, "The bare bones of object recognition: Implications from a case of object recognition impairment," *Neuropsychologia*, vol. 37, no. 3, pp. 279–292, 1999.

[360] O. H. Turnbull, D. P. Carey, and R. A. McCARTHY, "The neuropsychology of object constancy," *Journal of the International Neuropsychological Society*, vol. 3, no. 3, pp. 288–298, 1997.

[361] O. H. Turnbull, N. Beschin, and S. Della Sala, "Agnosia for object orientation: Implications for theories of object recognition," *Neuropsychologia*, vol. 35, no. 2, pp. 153–163, 1997.

[362] S.-W. Kim, J.-M. Kim, R. Stewart, K.-L. Bae, S.-J. Yang, I.-S. Shin, H.-Y. Shin, and J.-S. Yoon, "Correlates of caregiver burden for korean elders according to cognitive and functional status," *International journal of geriatric psychiatry*, vol. 21, no. 9, pp. 853–861, 2006.

[363] E. H. Seo, D. Y. Lee, K. W. Kim, J. H. Lee, J. H. Jhoo, J. C. Youn, I. Choo, J. Ha, and J. I. Woo, "A normative study of the trail making test in korean elders," *International journal of geriatric psychiatry*, vol. 21, no. 9, pp. 844–852, 2006.

[364] E. Frank, M. A. Hall, and I. H. Witten, *The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, fourth ed., 2016.

[365] S. Okazaki, R. Skapa, and I. Grande, "Global youth and mobile games: applying the extended technology acceptance model in the usa, japan, spain, and the czech republic," in *Cross-Cultural Buyer Behavior*, pp. 253–270, Emerald Group Publishing Limited, 2007.

[366] P. Corey and T. Hammond, "Gladder: Combining gesture and geometric sketch recognition," in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI) Student Abstracts*, (Chicago, IL, USA), pp. 1788–1789, AAAI, July 13–17, 2008.