Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test

by

William Souillard-Mandar

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2015

Author .....
Department of Electrical Engineering and Computer Science
January 30, 2015

Certified by .....
Randall Davis
Professor
Thesis Supervisor

Accepted by .....
Albert R. Meyer
Chairman, Masters of Engineering Thesis Committee

Learning Classification Models of Cognitive Conditions from Subtle Behaviors in the Digital Clock Drawing Test

by

William Souillard-Mandar

Submitted to the Department of Electrical Engineering and Computer Science
on January 30, 2015, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

The Clock Drawing Test – a simple pencil and paper test – has been used for more than 50 years as a screening tool to differentiate normal elderly individuals from those with cognitive impairment, and has proven useful in helping to diagnose dementias, such as Alzheimer's disease, Parkinson's disease, and other conditions.

A group of hospitals and clinics have been administering the test using a digitizing ballpoint pen that reports its position with considerable spatial and temporal precision, making available far more detailed data about the subject's performance. Using categorized stroke data from these drawings, we designed and computed a large collection of features, then explored the tradeoffs in performance and interpretability in classifiers built using a number of different subsets of these features and a variety of different machine learning techniques. We used traditional machine learning methods to build prediction models that achieve high accuracy. We operationalized widely used existing scoring algorithms so that we could use them as benchmarks for our models. We worked with clinicians to define guidelines for model interpretability, and constructed sparse linear models and decision lists designed to be as easy to use as scoring algorithms currently used by clinicians, but more accurate. We also extract insights from the data about the behavioral aspect of these conditions on patients.

While our models will require additional testing with subjects for validation, they offer the possibility of substantial improvement in detecting cognitive impairment earlier than currently possible, a development with considerable potential impact in practice.

Thesis Supervisor: Randall Davis
Title: Professor

Acknowledgments

This thesis allows me to contribute to what I believe to be an incredibly exciting and meaningful project. I learnt a lot about machine learning and and its application to the analysis of human behavior, as well as medical aspects of cognitive conditions. It also allowed me to continuously reinforce my interest in technology applied to medicine.

None of this would have been possible without the caring advising of Professor Randall Davis. Besides teaching and guiding me throughout the project, he inspired in me many soft skills, from the importance of paying attention to details to the usefulness of being able to clearly express concepts and ideas, as I was lucky to observe him do during our weekly meetings. I thank him most sincerely for the experience.

I would also like to thank Professor Cynthia Rudin for her advice and feedback, as well as sincere dedication and effort to the project.

Finally, I'd like to thank Dana Penney, Rhoda Au, David Libon, Rod Swenson, Catherine Price, and Melissa Lamar, for all the data collection work and medical feedback that made this project possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Cognitive Conditions and the Clock Drawing Test

With increasing progress of medicine extending life expectancy, populations world wide are "greying," producing a new set of healthcare challenges. As one example, recent estimates suggest that 13.9 percent of people above the age of 70 currently have some form of dementia [37], while the Alzheimer's Association projects that by 2050 the number of Americans with Alzheimer's will grow to some 13.8 million and the number worldwide will grow to 135 million [39]. As populations age there will clearly be huge financial, caregiver, and social burdens on our healthcare system and on society as we work to care for patients with cognitive impairments.

Research is underway on many fronts, including pharmaceuticals, but there is as yet no cure for cognitive impairments such as Alzheimer's and Parkinson's disease, and drugs often take 12 years from discovery to clinical approval. There is however the potential to slow the progress of some forms of cognitive decline, if caught early enough. Hence one important focus of research is early detection.

A variety of tests are used to screen for and assist with differential diagnosis of cognitive decline. One of the simplest and most widely used is called the Clock Drawing Test (CDT). In use for more than 50 years, it has been a widely accepted cognitive screening tool used in subjects with various dementias and other neurological disorders. The test asks the subject to draw a clock showing 10 minutes after 11 (called the Command clock), then asks them to copy a pre-drawn clock showing that time (the Copy clock).

As a simple paper and pencil test, it is quick and easy to administer, non-invasive and inexpensive, yet provides valuable clinical and diagnostic information. It is, for example, useful as a screening tool to differentiate normal elderly individuals from those with cognitive impairment, and has proven useful in helping to diagnose dementias, such as Alzheimer's disease, Parkinson's disease, and other conditions [14, 16]. The CDT is also used by neuropsychologists, neurologists and primary care physicians as part of a general screening for cognitive change in addition to dementia [46].

But there are drawbacks in the current use of the test. While there are a variety of well-regarded scoring systems designed for use by clinicians, these systems often rely on a clinician's subjective judgment of under-specified properties of the drawing. One current scoring system [29], for instance, calls for judging whether the clock circle has "only minor distortion," and whether the hour hand is "clearly shorter" than the minute hand, without providing quantitative definitions of those terms, leading to variability in scoring and analysis [38]. Other scoring system [30] specify more precise measures but are far too labor-intensive for routine use.

1.2 Digital Clock Drawing Test (dCDT)

For the past 7 years, a group of hospitals and clinics (The ClockSketch Consortium) has been administering the Clock Drawing Test using a digitizing pen (the DP-201 from Anoto, Inc.) that while functioning as an ordinary ballpoint also records its position on the page with considerable spatial (+- 0.005 cm) and temporal (12ms) accuracy. We then analyze the data using novel software developed for this task [10, 7, 11]. The resulting test is called the digital Clock Drawing Test (dCDT).

Administering the test using a digitizing pen enables a number of unique capabilities. The pen's spatial precision permits the software to do an unprecedented level of geometric analysis of the drawing, with no effort by the user. Because the data points are time-stamped, they capture the entire sequence of behaviors (every stroke, pause or hesitation), rather than just the final result (the drawing). Having time-stamped data means that the software can measure that behavior as well, including informative time intervals, as for example the delay between finishing numbering the clock and starting to draw the hands.

Processing raw data from the pen starts with sketch interpretation, i.e., classifying each

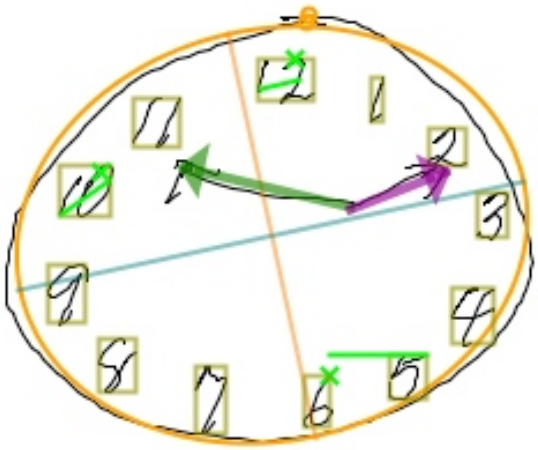


Figure 1-1: Example of a classified clock output by the ClockSketch software. An ellipse is fit to the clockface, with the major and minor axis shown; bounding box for each digits are computed; arrows show the overall direction of the hands; the lines on digits 5, 10, and 12 show the direction of hooklets, and the "x"s the start of the next stroke after each hooklet. The system adds the colored overlays as a way of making stroke classification visually obvious.

pen stroke as one or another component of the clock, e.g., as a minute hand, hour hand, as a specific digit, etc. This process is described by [10], and resulted in the Clocksketch software. The Clocksketch software analyzes the raw data from the pen, automatically classifying strokes as part of the clock face circle, hands, numbers, etc. It also permits assistance from the user, needed in difficult cases (e.g., for clocks by more impaired subjects). Figure 1-1 below shows a screenshot of the system after the strokes in a clock have been classified, showing the starting point for the work reported here.

Stroke classification is a key first step, as it enables appropriate measurement of clock features, e.g., the average size of the numerals, how accurately the hands show the time, the latency between finishing drawing the numerals and starting to draw the hands, etc. [8]. The spatial and temporal accuracy of the pen data permits the system to make precise measurements that are implausibly difficult with ordinary ink on paper.

From a user perspective, a significant advantage of the program is that all measurements are operationally defined in the software, hence free of user bias. They are also carried out in real time, with no additional burden to the user, a significant advantage in a clinical setting, where time is limited.

The next step is clinical interpretation: what does the drawing and the behavior that produced it tell us about the subject's cognitive state? This thesis reports on what light a

variety of machine learning techniques shed on answering this question. We describe our work on constructing features that are informative diagnostically, on building classifiers that predict a subject's condition, and on creating classifiers that are both accurate and comprehensible to clinical users.

1.3 Contributions

As a focus for this thesis we chose three categories of cognitive impairment based on their importance clinically and because they represent three of the most common diagnoses in our data: memory impairment disorders (MID) consisting of Alzheimer's and amnestic mild cognitive impairment (aMCI); vascular cognitive disorders (VCD) consisting of vascular dementia, mixed MCI and vascular/cognitive MCI; and Parkinson's disease (PD).

There are two forms of prediction we want to make. *Screening* distinguishes between healthy and one of the three categories. For each cognitive impairment category we built models that make a binary-choice prediction: whether someone has that condition or is healthy. We also do a group screening for these three conditions together, i.e., whether a subject has any one of the three condition or is healthy. The second task is the diagnosis-like task of *clinical group classification* – distinguishing one of the three conditions from every other of the 43 conditions in our data set, including healthy.

We define the following types of features, detailed in Section 3, on which our work is based:

- *Digital-pen features* are the features that we create from the data of the digital pen.
- *Clinician features* are the features used in the existing manual scoring systems created by and used by clinicians;
- *Operationalized clinician features (op-clinician features)* are rigorously defined and computed versions of the clinician features.
- *Simplest features* is a subset of all the features chosen because we believe they are particularly easy to evaluate by hand, hence less subject to inter-rater variance and usable in the pen-and-paper version of the test.

- *The set of all features* is the union of the digital-pen features, op-clinician features, and simplest features.

- *MRMR subset of all features* is the first 200 selected by Minimum-Redundancy-Maximum-Relevance [31] filtering from the set of all the features.

We began by using off-the-shelf machine learning methods for their ability to produce accurate predictive models when trained on large amounts of data. Section 4 describes the work and reports on the performance of six classification methods - Gaussian SVM, random forests, CART, C4.5, boosted decision trees, and regularized logistic regression - each of which had access to all features.

These classifiers performed very well in absolute terms, but determining the significance of their performance results requires a baseline to use as a point of comparison. While data are available on the performance of some of the scoring systems used by clinicians to score the traditional clock drawing test [49, 44, 45, 25], these are imperfect measures due to variations in the way the test is given (e.g., whether only one clock is to be drawn, whether the clock face circle is pre-drawn, etc.) and variations in the clinical populations used in evaluation.

To provide a more closely comparable measure of performance, we evaluated our clock data using seven of the most widely used existing manual scoring systems, selected after a careful review of the literature. Prompted by the impossibility of applying so many manual systems to our large collection of test, we created automated versions of these scoring systems. One challenge in doing this is that the scoring systems are designed for use by people, and often contain under-specified measures (e.g. deciding whether a clock circle has "only minor distortions.") We thus had to operationalize these algorithms, i.e., specify the computations to be done in enough detail that they could be expressed unambiguously in code. We refer to these as *operationalized scoring systems.*

One disadvantage of off-the-shelf machine learning classifiers is that they produce black box predictive models that may be impossible to understand as anything other than a numerical calculation. In response, another focus of our work has been on exploring the tradeoff between accuracy and interpretability. In Section 6.2, we provide a definition of interpretability for our problem. We then use a recently developed framework, Supersparse Linear Integer Models (SLIM) [51], and introduce a simple metric to prioritize more understandable features, enabling

us to build interpretable linear models.

In Section 6.3, we move to a second class of models consisting of rules and decision lists. We mine association rules from the data. Some of these rules turned out to confirm existing knowledge about correlations between pen-based features and diagnoses, while others appear novel, revealing correlations that had not been reported previously. In a further step in this direction, we constructed decision lists by employing a recently-developed machine learning technique called Bayesian Rule Lists (BRL) [22], which combines associations to create accurate-yet-interpretable predictive models.

Based on the framework outlined above, we carried out a number of experiments which resulted in the eight primary contributions of this thesis:

(i) Starting from a collection of novel clock test features created over the years (see e.g. [33, 34, 21, 32]), we create additional single-clock features, as well as features taking advantage of aggregate properties of the clocks and differences between the command and copy clocks. In addition, we operationalize the features used in existing scoring systems, producing the operationalized clinician features, and select a set of simplest features that we believe to be most easily and reliably measured by clinicians by eye.

(ii) We show that six state-of-the-art machine learning methods applied to the set of all features produce classifiers with AUC performance ranging from 0.89 to 0.93 for screening and 0.79 to 0.83 for clinical group classification. Published AUCs of existing clinician scoring systems [49, 44, 45, 25], which typically only attempt to distinguish healthy vs. cognitively impaired, range from 0.66 to 0.79 depending on the dataset. Our methods are thus not only significantly more accurate on this task, they are also capable of detecting more fine-grained classes of cognitive impairments for both screening and cognitive impairment classification.

(iii) We created operationalizations of seven widely used CDT scoring systems, to provide the most direct baseline for evaluating our models. Any free parameters in our operationalized scoring systems were chosen so as to maximize performance of the system, providing an upper bound on the performance of these systems on our data.

(iv) The classifiers produced by the state-of-the-art machine learning methods greatly outper-

formed the optimized operationalized scoring algorithms for both screening and clinical group classification. Where the machine learning methods produced AUCs from 0.89 to 0.93 for screening and 0.79 to 0.83 for group classification, the best operationalized scoring algorithms have AUCs of between 0.70 and 0.73 for screening and 0.65 and 0.69 for group classification. Thus, using the digital version of the CDT with our machine learning models would lead to more accurate predictions.

(v) We show that applying the machine learning methods to the clinician features leads to models with AUCs from 0.82 to 0.86 for screening and 0.70 to 0.73 for group classification, which is more accurate than the operationalized scoring algorithms. We also show that using the simplest features results in better performance than the operationalized scoring algorithms, with AUCs from 0.82 to 0.83 for screening and 0.72 to 0.73 for group classification. This opens up the possibility of clinicians recording these features and inputting them into our machine learning models, producing more accurate predictions of their patients' conditions, without changing what they attend to in evaluating the test.

(vi) We created Supersparse Linear Integer Models using simplest features, op-clinician features, and the MRMR subset of all features, that are all more accurate than existing scoring systems on the screening task, with AUCs from 0.73 to 0.83 depending on the feature set, and at least as accurate (and often better) on the group classification task, with AUCs from 0.66 to 0.77. These models contain very few features and prioritize understandable ones, leading to models that are at least as interpretable as existing algorithms and can be used reliably by clinicians.

(vii) We mined association rules and found many that were consistent with existing knowledge about connections between clock features and cognitive conditions, and found some that suggested plausible but previously unknown connections.

(viii) We created highly interpretable decision lists using simplest features, op-clinician features, and the MRMR subset of all features, resulting in classifiers with AUCs ranging from 0.78 to 0.85 for screening and 0.69 to 0.74 for clinical group classification, depending on the feature set and condition. As above, these models might be usable by clinicians at least as easily, and possibly more reliably and accurately, than existing scoring systems.

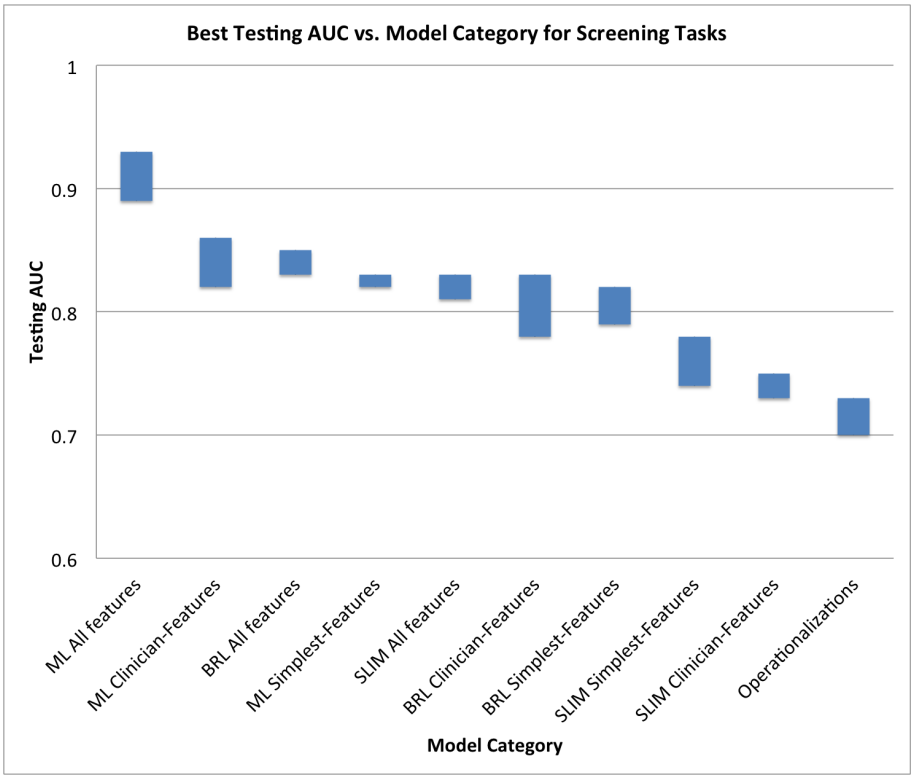


Figure 1-2: Summary of results for screening.

Figures 1-2 and 1-3 summarizes the results described above, showing the range of accuracies achieved by our different models for screening and clinical group classification, respectively, ordered left to right by decreasing upper bound. Each model category is a pairing of a class of model (traditional machine learning models, Supersparse Linear Integer Models, or Bayesian Rule Lists) with a feature set (simplest features, clinician features, or all features/MRMR subset of all features). Each bar shows the range of the AUC's across test folds for each condition, for the best algorithm in each category. For example, on the screening plot, "ML All features" indicates the range of accuracies of the best machine learning algorithms using all features, over the four possible screening tasks.

1.4 Related work

Machine learning, and more generally, artificial intelligence, have been a subject of interest to medicine as they offer potential for improving the detection and diagnosis of medical conditions and diseases, while at the same time increasing objectivity of the decision-making process.

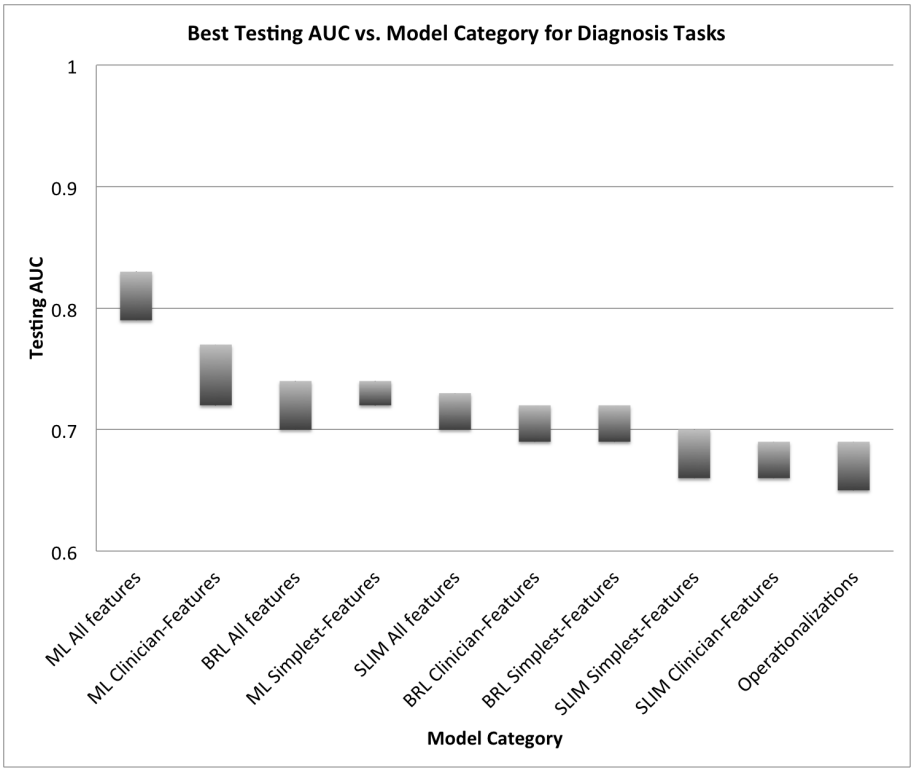


Figure 1-3: Summary of results for clinical group classification.

Successful applications date back to expert systems, for example tasked with selecting antibiotic therapy for bacteremia [9]. More recently, examples of successful machine learning medical applications include the detection and classification of tumors via X-ray and CRT images [36, 3]; the classification of malignancies from proteomic and genomic (microarray) assays [59, 12, 53]; heartbeat classification from electrocardiogram [56]; the creation of patient risk stratification models for Clostridium difficile [55].

At the same time, however, the goal of creating interpretable models – which was emphasized during the expert system years – has had considerably less attention over the last two decades. Recent research in statistics and machine learning has primarily focused on the accuracy of the models and the scalability of the algorithms, often sacrificing interpretability in the optimization process (e.g. by using proxy measures, such as the ℓ_1-norm instead of ℓ_0-norm). Some methods can be tailored to be more interpretable, such as decisions tree algorithms by restricting the height of possible trees. This is often at a high cost to accuracy, as those methods were not designed with interpretability in mind.

In many applications, but particularly medicine, the ease with which a model is used and

understood is very important. As we observed when reviewing existing scoring systems for the clock drawing test, the majority are short and use simple features combined linearly to arrive at a decision, a common pattern in medical scoring systems. Some work in applied machine learning has aimed to create such interpretable models. For example, a novel decision list construction algorithm [22] was used to generate a new scoring criterion to predict stroke in patients with atrial fibrillation, significantly outperforming the $CHADS_2$ [15] scoring system currently in use by doctors, while remaining equally interpretable; another work focuses on creating interpretable clinical decision support systems for gynecology [52].

There have been some attempts to create novel versions of the clock drawing test. The closest work [19, 18] builds a tablet-based clock drawing test that allows the collection of data along with some statistics about user behavior. However, the work focuses primarily on the user-interface aspects of the application, trying to ensure that it is usable by both subjects and clinicians, but not on automatically assessing the test or creating new data-driven machine learning classifiers to detect cognitive conditions.

Numerous papers in the clinical literature describe a variety of scoring systems for the clock test, but no work that we know of - and certainly none used in practice - has used state-of-the-art machine learning methods to create these systems or has reported levels of accuracy comparable to those obtained in this work. In addition, no work that we know of has tackled the problem of understanding the tradeoff between accuracy of prediction and interpretability for the clock drawing test.

Chapter 2

Data

Over the past seven years, the ClockSketch Consortium has accumulated a collection of 3541 digital clock tests whose strokes have been carefully classified and independently reviewed for accuracy. Some subjects have been tested multiple times over the years; to avoid issues that might arise from repeated exposure to the test, we selected only the first test for each subject, leaving us with 2169 tests (each of which has both a Command and Copy clock, yielding 4338 distinct drawings). The anonymized IDs of all the tests used can be found in Appendix B.

Each test has up to three diagnoses associated with it, each with a level of certainty ranging from 1 to 3. These diagnoses are often subclasses of major conditions. We decided to focus on four groups due to their importance clinically and because they represent three of the most common groups of diagnoses in our data, along with a group of healthy subjects.

- The memory impairment disorders (MID) group consists of 206 subjects diagnosed by consensus to have Alzheimer's or amnestic MCI. Alzheimer's is the most common form of dementia, accounting for 60 to 70 percent of dementia cases [35]. MCI (mild cognitive impairment) can present with a variety of symptoms; when memory loss is the predominant symptom it is termed "amnestic MCI" and is frequently seen as a transitional stage between normal aging Alzheimer's disease [1]. We would expect memory problems on the clock test but do not expect significant motor slowing during the early stages of the disease. In our sample, subjects with amnestic MCI meet criteria established by [35] and have circumscribed memory loss in the context of otherwise intact cognition and no report of functional problems. Our subjects with Alzheimer's disease are primarily at an early

stage of the disease.

- The vascular cognitive disorders (VCD) consisted of 121 subjects diagnosed with vascular dementia, mixed MCI, or vascular MCI. Vascular dementia is widely considered the second most common cause of dementia after Alzheimer's disease, accounting for 10 percent of cases [2], and is often preceded by vascular MCI. Many experts believe that vascular dementia and MCI remain under-diagnosed – like Alzheimer's disease – even though they are recognized as common. Early detection and accurate diagnosis are important, as risk factors for vascular dementia are important targets for medical intervention. We expect motor and cognitive slowing effects on the test performance.

- Parkinson's Disease (PD). There were 126 subjects diagnosed with Parkinson's disease. Early in the course of the disease the most obvious symptoms are movement-related and may include tremor, rigidity, slowness of movement and difficulty with gait. Later, thinking and behavioral problems may arise, with dementia most often occurring in the advanced stages of the disease. There is no cure yet, but medical and surgical treatments are effective at managing the motor symptoms of the disease.

  The tests that we use for the healthy set consist of "Healthy Control" subjects, who are non-patients (e.g. family members of patients who agree to be tested), as well as "Not demented" subjects, who have been longitudinally studied and evaluated by the Framingham Heart Study [1] and are judged with high confidence to be healthy. In order to ensure appropriate age matching in comparisons, the following sets of healthy patients were defined for each condition group:

  - Memory impairment disorders and vascular cognitive disorders subjects: "Healthy Control" and "Not demented" with confidence of 3 and age greater than 55, for a total of 406 healthy tests.
  - Parkinson's disease: "Healthy Control" and "Not demented" with confidence of 3 and age greater than 45, for a total of 587 healthy tests.

[1] The Framingham Heart Study began in 1948 with the goal of life-long physical examinations and lifestyle interviews of their participants every two to four years, to look for patterns related to heart disease. Their focus has since broadened to other diseases, but their methodology – recruiting and life-long examination of a large cohort of subjects – means that many of their subjects are healthy. The Study has been using the dCDT as part of its test suite for the past two years.

The remainder of the tests have other neurological, psychiatric, and medical conditions; their distribution is shown in Figure 2-1.

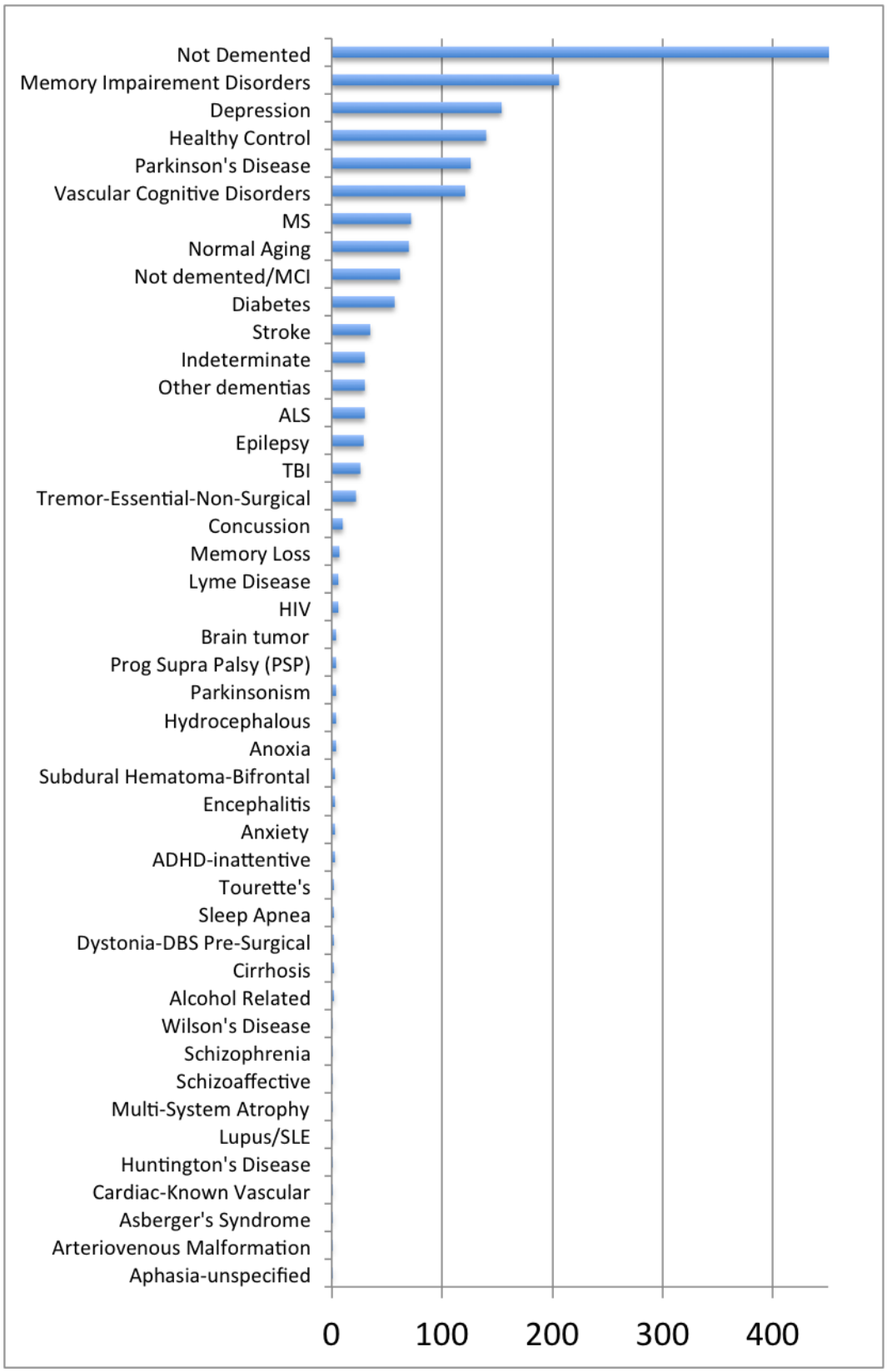


Figure 2-1: Frequency of each condition in the dataset.

Chapter 3

Feature construction

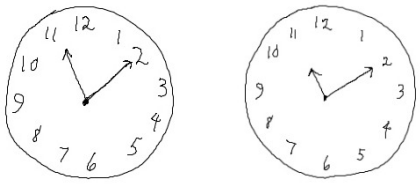
This section describes the clock drawing features we constructed, a process that involved conferencing with clinicians and examining intermediate results.

3.1 Example clocks

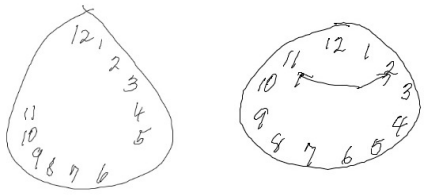
Figure 3-1 illustrates representative clock drawings from our dataset from a healthy control, a subject in the memory group impairment and a subject diagnosed with Parkinson's. As the figure suggests, clocks by healthy subjects are typically reasonably round, have all 12 digits present and spaced regularly around the clock, and have hands pointing towards digit 11 and digit 2. Hands often have arrowheads, and the minute hand is often but not invariably longer than the hour hand, following the traditional clock format. A center dot is also common.

There are many possible variations found in both healthy and impaired subjects.

- Clocks vary significantly in size, with some subjects drawing them much smaller (Figure 3-1c).
- There may be a gap between the start and the stop of the clockface (Figure 3-1c).
- Digits maybe be missing, crossed-out, repeated, or with poor angular spacing (Figure 3-1b).
- Digits greater than 12 are sometimes drawn.



(a) Healthy



(b) Alzheimer's

(c) Parkinson's

Figure 3-1: Example clocks from our data set for healthy, Alzheimer's disease, and Parkinson's disease, with command clock on the left and copy clock on the right

- Hands can be missing (Figure 3-1b), crossed-out, or repeated, with arrowheads sometime pointing toward the clock center.

- Some clocks contain stokes used by subjects for spatial arrangement, and tickmarks used as replacement for digits.

- Subjects sometime use additional text in their drawings, for example to write the time as a memory aid or in lieu of a number.

- We have defined "noise" as strokes that are not part of the representation of defined clock elements (e.g. hand, digit) but are clearly produced during the drawing process and are intentional (i.e. not random pen drops) [33]. They vary from tiny dots to longer lines (Figure 3-1c).

- A more subtle feature, hooklets [21, 32], can also be observed. These are abrupt changes in the direction at the end of a stroke that head toward the beginning of the next stroke. For example, when drawing the numbers on a clock, subjects may leave a "tail" on the end of one digit stroke that points toward the start of the first stroke of the next digit.

Starting from these observations, and iterating using results and doctor feedback, we constructed five sets of features to be used with our algorithms to obtain models that have specific characteristics. We detail the feature sets and their objectives below.

3.2 Digital-pen features

These are the features that we create from the data of the digital pen. They fall into the following four categories below:

3.2.1 Single-Clock-Measurements

These are measurements of geometric or temporal properties of components of a single clock. For example:

- For each component (e.g. the clockface, all digits, and all hands), the number of strokes, the total ink length, the time it took to draw, and the speed.

- The length of the major and minor axis of the fitted ellipse as well as the distance and angular difference between starting and ending points of the clock face (Figure 3-2A).
- Digits that are missing or repeated, the height and width of their bounding boxes (Figure 3-2B).
- Hands are checked for omissions or repetitions, the size ratio of hour hand to minute hand, the presence and direction of arrowheads, and angular error from their correct angle (Figure 3-2C).
- Whether the minute hand points to digit 10 instead of digit 2, which can happen as a consequence of the instruction to set the time to "10 past 11".
- The presence, length, and direction of hooklets are measured (Figure 3-2D).
- The presence of tick marks, spokes, any text (often used as a reminder of the time), or a center dot for the hands.
- The number and length of noise strokes.
- Timing information is used to measure how quickly different parts of the clock were drawn. One particularly interesting latency feature is one called the pre-firsthand latency, the time that elapsed between the first stroke of the first hand drawn and whatever was drawn previously [34].

3.2.2 Single-Clock-Aggregates

These are aggregates of geometric or temporal properties of a single clock. For example:

- The total time to draw the clock and the total number of strokes used.
- The average height, average width, and average length of all digits present.
- The number of digits missing or repeated.
- Measures of the distribution of digits around the clock. For example, one feature counts the number of digits in the clock that have fewer than 3 other digits within 45^o on either

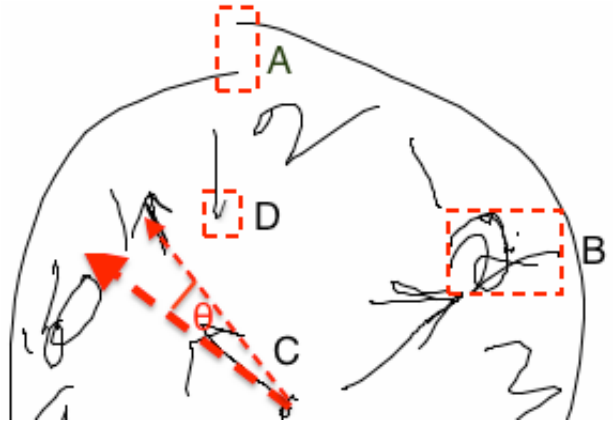


Figure 3-2: Example features. A: the distance between starting and ending point of the clockface, as well as the angular difference; B: digit repetition; width and height of the bounding box; C: the difference in angle between a hand and its correct angle; D: hooklet presence, length, and direction.

side; another feature reports whether all non-anchor digits are in their correct eighth; yet another the variance in the distance of digits from the clockface.

- The percentage of time spent drawing vs. thinking (holding the pen off the paper) for one clock.

3.2.3 Both-Clock-Aggregates

These are aggregates over both the command and the copy clock. For example:

- The total time to draw both clocks.
- The total number of strokes used.
- The average height, average width, and average length of all digits present in both clocks.
- The number of digits missing in both clocks.
- The percentage of time spent drawing vs. thinking for both clocks.

3.2.4 Clock Differences

We compute the difference in value of a feature across the command clock and the copy clock e.g, difference in the total time to draw each clock. This follows the intuition that because the command and copy clocks engage different cognitive functions, differences between them may be revealing.

3.3 Clinician features and operationalized-clinician features (op-clinician features)

These are computable versions of features found in existing manual scoring systems. Some of the clinician features are quantitative, such as checking for the presence of a digit or a hand. Others are less well defined: for example, one feature calls for determining whether the minute hand is “obviously longer" than the hour hand, while another checks whether there are “slight errors in the placement of the hands." These can be estimated by a clinician, but it is not immediately obvious how to compute them in a program in a way that captures the original intent. Section 5 describes our efforts to create the operationalized versions of these features.

The operationalized features then allow us to create operationalizations of the existing scoring systems, providing a baseline against which to compare the classifiers we build. In addition, we use these features with the machine learning algorithms in order to measure how predictive these features can be in models of other forms.

3.4 Simplest features

This is a subset of the features available in the traditional pen-and-paper version of the test, selecting those for which we believe there would be little variance in their measurement across clinicians. We expect, for example, that there would be wide agreement on whether a number is present, whether hands have arrowheads on them, whether there are easily noticeable noise strokes, etc.

Models created using this set of features would be applicable to the traditional pen-and-paper version of the test (i.e. without the digitizing pen), with clinicians able to measure the features more easily and consistently than for existing scoring systems.

3.5 All features

This set is a union of the three feature sets described above (i.e. digital-pen features, op-clinician features, simplest features). Our intent here is to build the best model possible, without regard to the number of features, their interpretability, etc., in order to get the maximum benefit from

the precise data by the digital pen. This is discussed in Section 4. Models built using this set of features can of course only be used with tests administered with the digital pen.

3.6 MRMR subset of all features

From among all of the features, we created a subset from the first 200 selected by Minimum-Redundancy-Maximum-Relevance [31] filtering. We use this set of features when using the set of all features is computationally too expensive.

Chapter 4

Traditional Machine Learning Models

Our aim in this section is to establish a measurement of the the highest accuracy attainable from our data, by applying state-of-the-art machine learning methods to the set of all features.

4.1 Machine Learning Methods

We constructed classifiers using a variety of machine learning methods, starting with filtering methods to get an initial idea of which features have high predictive power, and then moving on to decision trees, logistic regression, support vector machines, bagged decision trees, and boosted decision trees.

For each algorithm, we used stratified cross-validation to divide the data into 5 folds to obtain training and testing sets. For algorithms that had parameters to optimize, we further cross-validated each training set into 5 folds to optimize the parameters of each algorithm using grid search over a set of ranges.

4.1.1 Feature filtering

We ran different filters on our data to rank our features and get an idea of which ones had higher predictive power. Using [58], we ranked them according to Gini index and Information Gain. We also used Minimum-Redundancy-Maximum-Relevance (MRMR) [31] filtering to select features that are mutually far away from each other, while still highly correlated with the classification variable, to create a subset of all features as described in Section 3.

4.1.2 Decision Trees

We used the two most popular decision tree algorithms, CART and C4.5, to create classification models. For CART, we used the R library “rpart" with default parameters; for C4.5, we used the R library “RWeka" with default settings.

4.1.3 Random Forest

We used the MATLAB class TreeBagger with parameter “NVarToSample" set to the square root of the total number of variables and the variable “NTrees" for the number of trees set to 1000. This gave sufficient trees for the accuracy to converge to its asymptote for all of our classification problems.

4.1.4 Regularized logistic regression

We used the LIBLINEAR [13] implementation of logistic regression with l_1 regularization. We selected the regularization parameter $C_L R$ from $\{2^{-8}, 2^{-6}, ..., 2^8\}$, choosing the one with the best 5-fold cross-validation performance.

4.1.5 Support vector machines

We used three different algorithms.

- SVM with linear kernel: it performed significantly worse than the others.

- SVM with recursive feature elimination: it allowed us to construct linear SVM models with fewer features than the SVM with a linear kernel. However, we ended up using other models tailored for more interpretability, as detailed in Section 6.

- SVM with Gaussian kernel: performed very well on our data. We used SVMlight [17] with a radial basis function kernel. We selected the slack parameter C_{SVM} and the kernel parameter γ using a grid search over the ranges $C_{SVM} \in \{2^{-4}, 2^{-2}, ..., 2^{14}\}$ and $\gamma \in \{2^{-6}, 2^{-1}, ..., 2^{10}\}$

4.1.6 Boosting

- Boosted splines: considerably lower performance than boosted decision trees.

- Boosted decision trees: performed very well across the different classification problems. We used the MATLAB class "fitensemble" with 500 trees and parameter "LearnRate" set to 0.05.

4.2 Results for Machine Learning Methods

We present results for the machine learning methods described above that performed best: CART, C4.5, SVM with gaussian kernels, random forests, boosted decision trees, and regularized logistic regression.

We began with the screening task, seeking to develop classifiers able to distinguish healthy subjects from those with one of the conditions listed earlier: memory impairment disorders, vascular cognitive disorders, and Parkinson's, as well as whether the subject is healthy or has any of the three conditions.

Table 4.1 shows the prediction quality for all of the machine learning algorithms we used, reported as the mean and standard deviation of performance over the test folds. We chose to measure quality using area under the receiver operator characteristic curve (AUC) as a single, concise statistic; we display full ROC curves in Figure 4-1. Each curve is a mean over the 5 folds, with 95% confidence intervals displayed as bars along the curves. We assessed statistical significance for the experiments in Table 4.1 using matched pairs t-tests; bold indicates algorithms whose result was not statistically significantly different from the best algorithm.[1]. Note that no single machine learning method can be declared the winner across all experiments.

The best classifiers achieve AUC measures from the high 80s to the low 90s. With this level of prediction quality, these methods can be immediately helpful as decision aids for physicians.

For our sample of subjects, these results are superior to published accuracies of existing scoring systems, even where those scoring systems focused on the simpler screening task of distinguishing demented from non-demented subjects instead of the more fine-grained categories.

[1]These hypothesis tests are problematic because experiments between folds are not independent, but there is apparently no good alternative for testing (see, for instance, [27])

The published results reported AUC levels ranging from 0.60 to 0.79 [49, 44, 45, 25], with variance in the performance across reports. As an example of the ranges and variance, AUC accuracy for two widely used scoring systems have been reported from 0.66 [45] to 0.79 [44] for Shulman [42], and from 0.7 [45] to 0.78 [44] for Mendez [28].

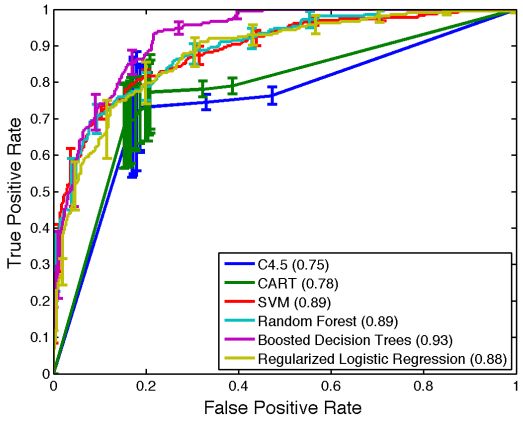
Full ROC curves are show in Figure 4-1. To produce these curves for a particular model (machine learning model, or scoring system), we rank subjects according to their score in the model and build the curve from the left (subjects with the highest score) to right (subjects with the lowest score). This way, the left part of the curve represents subjects most likely to have an impairment.

Algorithm	MID vs. Healthy	VCD vs. Healthy	Parkinson's vs. Healthy	All three vs. Healthy
C4.5	0.75 (0.08)	0.72 (0.07)	0.75 (0.06)	0.78 (0.08)
CART	0.78 (0.07)	0.75 (0.13)	0.76 (0.10)	0.76 (0.10)
SVM Gaussian	**0.89 (0.06)**	**0.84 (0.08)**	**0.86 (0.08)**	**0.91 (0.09)**
Random Forest	**0.89 (0.10)**	**0.88 (0.09)**	**0.91 (0.11)**	**0.89 (0.06)**
Boosted Decision Trees	**0.93 (0.09)**	**0.88 (0.11)**	**0.87 (0.08)**	**0.90 (0.12)**
Regularized Logistic Regression	**0.88 (0.11)**	**0.85 (0.07)**	**0.91 (0.08)**	**0.89 (0.09)**

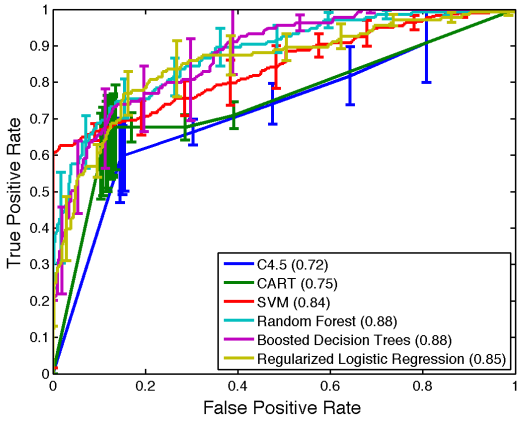
Table 4.1: Classification results for the screening task: distinguishing clinical group from healthy. Each entry in the table shows the mean and standard deviation AUC of a machine learning algorithm across 5 folds. The first column is for the task of distinguishing memory impairment disorders vs. healthy, the second column is for vascular cognitive disorders vs. healthy, the third column is for Parkinson's vs. healthy, and the last column is for any of the three cognitive impairments vs. healthy.

The second set of experiments aimed at clinical group classification, i.e., distinguishing subjects in one of our clinical groups from subjects who have any other medical, neurological, or psychological condition. Table 4.2 shows comparative accuracy results; Figure 4-2 shows the associated ROC curves. As expected, clinical group classification is a more difficult task, leading to the best algorithms having AUC's within the high 70's to low 80's.

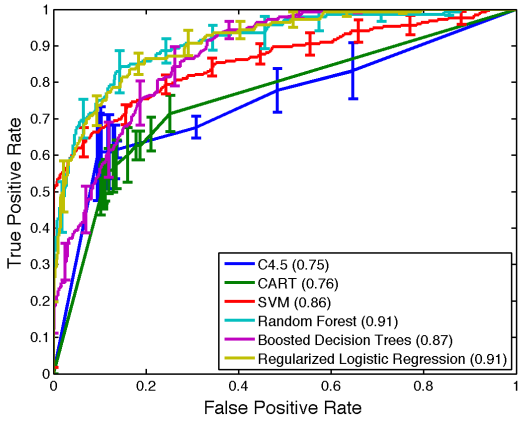
Having established performance for machine learning classifiers, we would like to know how they compare to the models currently in use by clinicians. Ideally, we would determine this by having a large number of our clock tests manually evaluated by clinicians using the scoring systems in current use, but this was not pragmatically possible. We were, however, able to establish a useful baseline by creating computational models of the existing scoring systems, resulting in models which we call *operationalized scoring systems.*



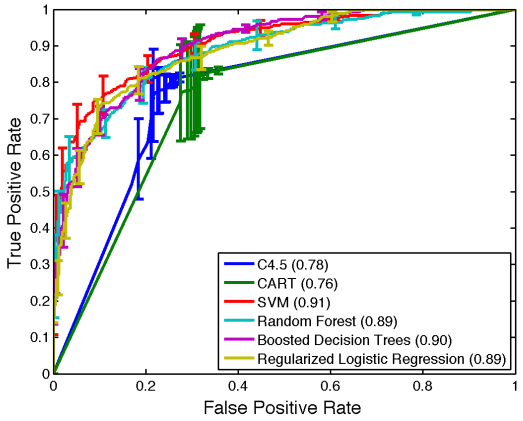
(a) Memory impairment disorders vs. Healthy



(b) Vascular cognitive disorders vs. Healthy



(c) Parkinson's vs. Healthy

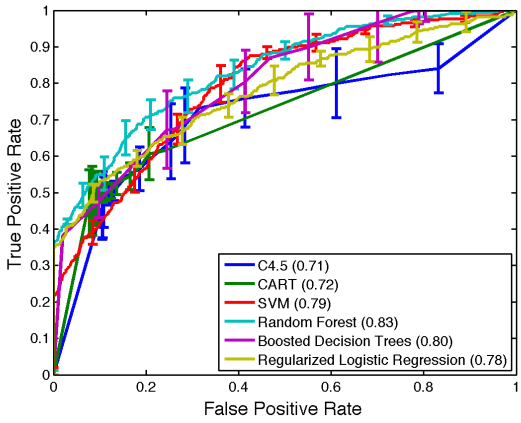


(d) All three vs. Healthy

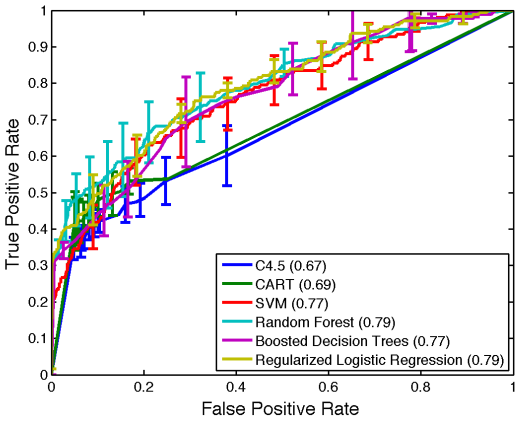
Figure 4-1: ROC curves for screening task (Table 4.1).

Algorithm	MID vs. All Others	VCD vs. All Others	Parkinson's vs. All Others	All three vs. All Others
C4.5	0.71 (0.10)	0.67 (0.06)	0.71 (0.09)	0.66 (0.09)
CART	0.72 (0.06)	0.69 (0.09)	0.68 (0.09)	0.64 (0.04)
SVM Gaussian	**0.79 (0.07)**	**0.77 (0.13)**	**0.81 (0.11)**	0.72 (0.06)
Random Forest	**0.83 (0.06)**	**0.79 (0.10)**	**0.81 (0.07)**	0.73 (0.04)
Boosted Decision Trees	**0.80 (0.09)**	**0.77 (0.08)**	**0.77 (0.09)**	**0.82 (0.05)**
Regularized Logistic Regression	**0.78 (0.06)**	**0.79 (0.05)**	**0.82 (0.05)**	**0.79 (0.07)**

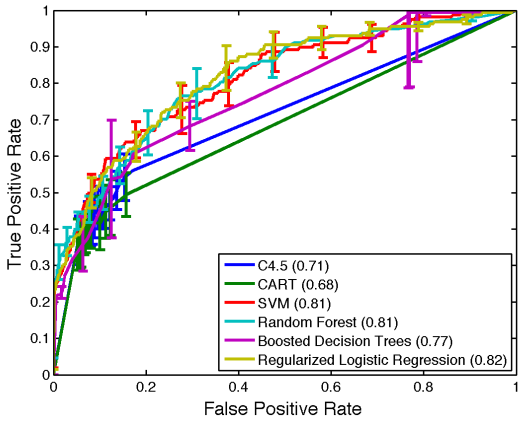
Table 4.2: Classification results for the clinical group classification task: distinguishing one cognitive impairment group from all other diagnoses. Each entry in the table shows the AUC and standard deviation of a machine learning algorithm for distinguishing one disease from the others. For instance, the entry in the table corresponding to memory impairment disorders and C4.5 indicates the accuracy in distinguishing memory impairment disorders from every other condition.



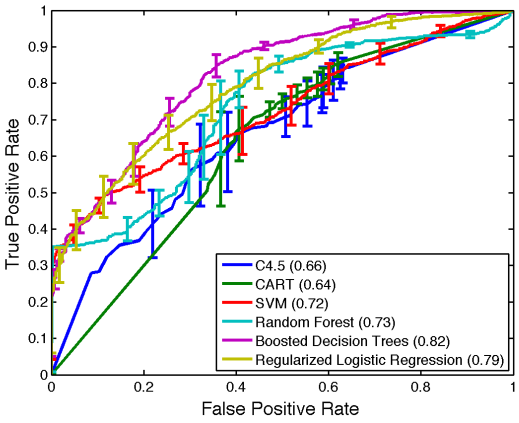
(a) Memory impairment disorders vs. All others



(b) Vascular cognitive disorders vs. All others



(c) Parkinson's vs. All others



(d) All three vs. All others

Figure 4-2: ROC curves for the clinical group classification task (Table 4.2).

Chapter 5

Operationalized scoring system

To evaluate the quality of our results with respect to the manual scoring systems currently in use for the CDT, we worked to reproduce the judgements made by clinicians when they apply one of the current scoring systems, creating fully automated versions of the scoring systems.

5.1 Existing scoring algorithms

There are a variety of scoring systems for the clock test, varying in complexity and the types of features they use. In each of the systems, points are added and subtracted based on features of the clock, such as whether clock hands are present, digits are missing, or the correct time is shown. A threshold is then used to decide whether the test gives evidence of impairment.

We reviewed the literature of existing scoring system for the clock test and worked with doctors to decide which to operationalize. The result is shown in Table 5.1.

5.2 Operationalizations

We were left with eight scoring algorithms to operationalize: Manos [26], Royall [41], Shulman [42], Libon [23], Rouleau [40], Mendez [28], MiniCog [6], and Watson [54]. Table 5.2 shows the Rouleau scoring criterion; we focus on it as an example of the operationalization process.

To operationalize these systems, we had to transform relatively vague terms, such as “slight errors in the placement of the hands" and “clockface present without gross distortion", into

Existing scoring system	Decision
Manos [26]	operationalized.
Royall [41]	operationalized.
Shulman (1986) [43]	not operationalized due to very similar newer version available; Shulman (1993) used instead.
Shulman (1993) [42]	operationalized.
Libon [23]	operationalized.
Rouleau [40]	operationalized.
Mendez [28]	operationalized.
MiniCog [6]	operationalized.
Watson [54]	operationalized.
Sunderland [47]	not operationalized; very similar to Libon.
Cahn [5]	not operationalized; very similar to Rouleau.
Wolf-Klein [57]	not operationalized; very similar to Watson and Shulman.
Tuokko [48]	not operationalized; not widely used according to doctors.
Lam [20]	not operationalized; not widely used according to doctors.
Lin [24]	not operationalized; almost identical to Royall and less widely used according to doctors.

Table 5.1: Summary of decisions for each existing scoring system found in the literature

precise rules that can be programmed. We conferred with clinicians for guidance on what was meant by the vague terms.

As one example, we translated “slight errors in the placement of the hands" to “exactly two hands present AND at most one hand with a pointing error of between ϵ_1 and ϵ_2 degrees", where the ϵ_i are parameters in the form of thresholds. Similarly, “clock face present without gross distortion" became “eccentricity of the clockface $\leq \epsilon_3$ AND clock face closed percentage $\geq \epsilon_4$".

Table 5.3 shows the non-obvious features used in the Rouleau scoring system (e.g. “digit missing" is obvious), while Table 5.4 shows the resulting operationalized scoring system. Operationalized scoring systems for all the other manual scoring systems are given in Appendix A.

The clinicians on our team confirmed the form and content of these operationalized scoring systems and provided initial values for the thresholds which they believed made the operationalizations capture the intent of the original manual scoring systems. For instance, the initial hand pointing thresholds were 15^o and 30^o.

maximum: 10 points
1. Integrity of the clockface (maximum: 2 points)
2: Present without gross distortion 1: Incomplete or some distortion 0: Absent or totally inappropriate
2. Presence and sequencing of the numbers (maximum: 4 points)
4: All present in the right order and at most minimal error in the spatial arrangement 3: All present but errors in spatial arrangement 2: Numbers missing or added but no gross distortions of the remaining numbers Numbers placed in counterclockwise direction Numbers all present but gross distortion in spatial layout 1: Missing or added numbers and gross spatial distortions 0: Absence or poor representation of numbers
3. Presence and placement of the hands (maximum: 4 points)
4: Hands are in correct position and the size difference is respected 3: Sight errors in the placement of the hands or no representation of size difference between the hands 2: Major errors in the placement of the hands (significantly out of course including 10 to 11) 1: Only one hand or poor representation of two hands 0: No hands or perseveration on hands

Table 5.2: Original Rouleau scoring system [40]

Variable	Description
Eccentricity of fitted ellipse	$\sqrt{(1-(\frac{b}{a})^2)}$ where a and b are half the major and minor axes respectively. A perfect circle has value 0, the value increases toward 1 as it gets flatter.
ClockfaceClosedPercentage	The percentage of the angle of the clockface that is closed.
DigitsAngleError	The average angle error of digits from their correct angle. A measure of the distribution of digits angularly.
DigitNeighborsTest	A count of the number of digits in the clock with fewer than 3 other digits within $\pm 45^o$. A second measure of the distribution of the digits angularly.
HandAngleError	The difference in angle between the hand and the digit it should point to.
HandRatio	The ratio: length of the hour hand / length of minute hand.

Table 5.3: Operationalized non-obvious features for Rouleau.

maximum: 10 points
1. Integrity of the clockface (maximum: 2 points)
2: eccentricity $\leq \epsilon_1$ AND clockface closed percentage $\geq \epsilon_2$ 1: eccentricity $> \epsilon_1$ OR clockface closed percentage $< \epsilon_2$ No clockface strokes OR normed residual $> \epsilon_3$
2. Presence and sequencing of the numbers (maximum: 4 points)
4: If all numbers present AND correct angular sequence AND DigitsAngleError $\leq \epsilon_4$ 3: If all numbers present AND correct angular sequence AND $\epsilon_4 \leq$ DigitsAngleError $\leq \epsilon_5$ 2: (At least one number missing OR at least one number repeated OR digits greater than 12 present) AND DigitNeighborsTest $= 0$) OR numbers counterclockwise OR All number present AND (at least one number outside the clock OR DigitNeighborsTest $\geq \epsilon_6$) 1: At least one number missing OR at least one number repeated OR digits greater than 12 present) AND DigitNeighborsTest $\geq \epsilon_6$) 0: No numbers
3. Presence and placement of the hands (maximum: 4 points)
4: Exactly two hands AND both HandAngleError $\leq \epsilon_7$ AND HandRatio $\leq \epsilon_8$ 3: Exactly two hands AND (at least one hand has $\epsilon_7 <$ HandAngleError $\leq \epsilon_9$ OR HandRatio $> \epsilon_8$) 2: Exactly two hands AND at least one hand has HandAngleError $> \epsilon_9$ OR minute hand within $\epsilon_1 0$ of “10" 1: One hand or more than two hands present 0: No hands present

Table 5.4: Operationalization of Rouleau scoring system

5.3 Results of Operationalizations

Starting from these initial values, we created a range of possible values for each parameter (Table 5.5), such as $\{0^o, 3^o, ..., 42^o, 45^o\}$ for the hand pointing error. We then selected parameter values via a 5-fold stratified cross-validation that maximized AUC. This maximization of the AUC ensures that our operationalized versions of the manual scoring systems provide an upper bound on the performance the scoring system is capable of.

Variable	Threshold values
Eccentricity of fitted ellipse	$\{0.45, 0.5, ..., 0.8, 0.85\}$
ClockfaceClosedPercentage	$\{70, 73..., 97, 100\}$
ClockfaceGap	$\{0, 2, ...28, 30\}$
DigitsAngleError	$\{0, 3, ..., 27, 30\}$
DigitNeighborsTest	$\{0, 1, 2, 3\}$
DigitClockfaceDistanceVariance	$\{0, 2, ...28, 30\}$
HandAngleError	$\{6, 9, ..., 27, 30\}$
HandRatio	$\{0.71, 0.73, ..., 1.03, 1.05\}$

Table 5.5: Parameter search values for operationalizations

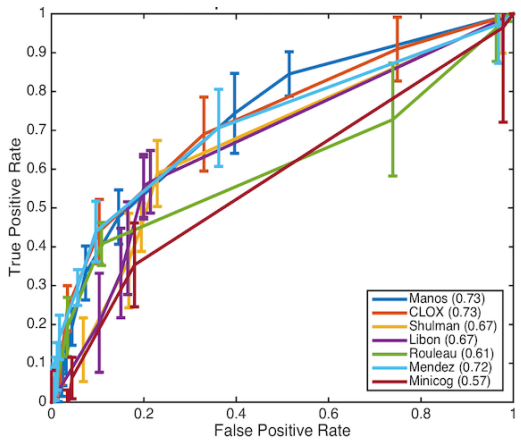
Table 5.6 and Figure 5-1 show the performance for each operationalized scoring system on the screening task.

Some of the manual scoring systems we operationalized have been evaluated on the task of screening for general dementia. Results reported for Shulman ranged from 0.66 [45] to 0.79 [44], while our operationalization of Shulman yielded 0.67 on memory impairment disorders and 0.71 on vascular cognitive disorders. Results reported for Mendez ranged from 0.70 [45] to 0.78 [44], while our operationalization of Mendez gave us 0.72 on memory impairment disorders and 0.70 on vascular cognitive disorders. Manos achieved 0.67 [25], while our operationalization gave us 0.73 on memory impairment disorders and 0.69 on vascular cognitive disorders. Thus, while there is a range of accuracies reported for these algorithms due in part to their being evaluated on different datasets and for different groupings of conditions (general dementia vs. memory impairment disorders/vascular cognitive disorders), our operationalized scoring systems achieve similar accuracies, providing a check on our operationalization process.

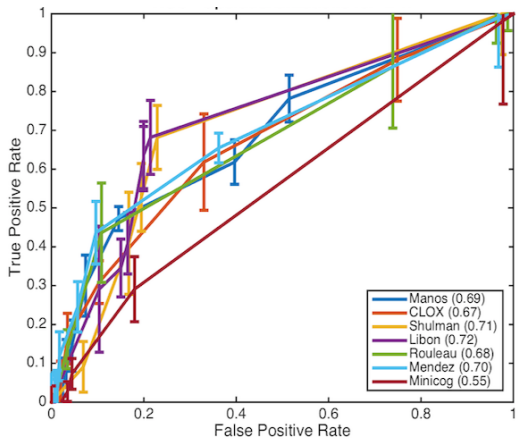
We then used a variety of machine learning methods on the op-clinician features and the simplest features. The lower part of Table 5.6 shows AUCs for the best machine learning algorithm on these two feature sets, followed by the AUCs of the best machine learning algorithm

Algorithm	MID vs. Healthy	VCD vs. Healthy	Parkinson's vs. Healthy	All three vs. Healthy
Manos	0.73 (0.08)	0.69 (0.13)	0.70 (0.11)	0.70 (0.07)
Royall	0.73 (0.14)	0.67 (0.13)	0.73 (0.09)	0.70 (0.06)
Shulman	0.67 (0.05)	0.71 (0.07)	0.66 (0.07)	0.67 (0.05)
Libon	0.67 (0.09)	0.72 (0.09)	0.68 (0.10)	0.68 (0.12)
Rouleau	0.61 (0.16)	0.68 (0.15)	0.59 (0.13)	0.61 (0.08)
Mendez	0.72 (0.11)	0.70 (0.12)	0.69 (0.07)	0.69 (0.06)
MiniCog	0.57 (0.08)	0.55 (0.13)	0.54 (0.15)	0.58 (0.12)
Best ML with op-clinician features	0.83 (0.09)	0.83 (0.11)	0.86 (0.08)	0.82 (0.10)
Best ML with simplest features	0.83 (0.06)	0.82 (0.07)	0.83 (0.08)	0.83 (0.07)
Best ML with all features	0.93 (0.09)	0.88 (0.11)	0.91 (0.11)	0.91 (0.09)

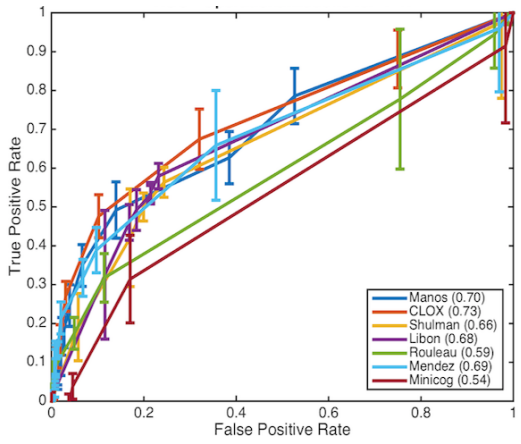
Table 5.6: Operationalized scoring system AUCs for screening task, together with AUCs of the best machine learning model on the op-clinician features, simplest features, and the set of all features.



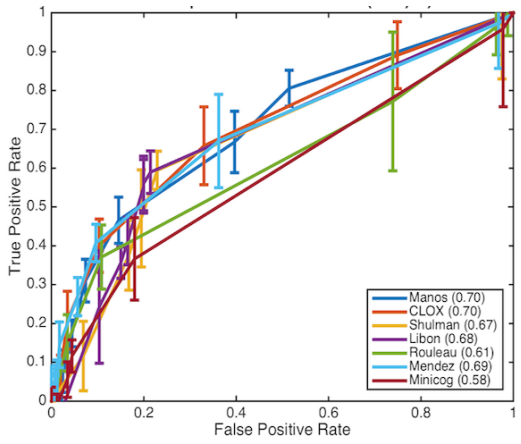
(a) Memory impairment disorders vs. Healthy



(b) Vascular cognitive disorders vs. Healthy



(c) Parkinson's vs. Healthy



(d) All three vs. Healthy

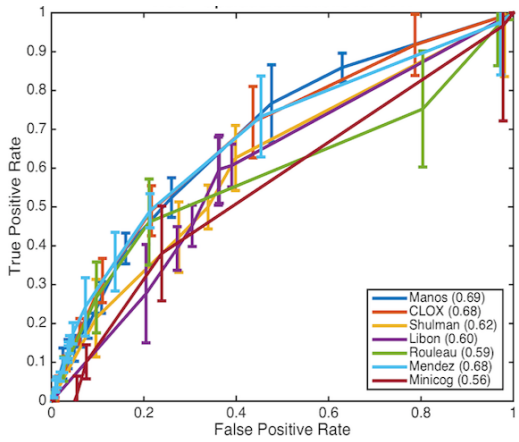
Figure 5-1: ROC curves for the experiments in Table 5.6.

Algorithm	MID vs. All Others	VCD vs. All Others	Parkinson's vs. All Others	All three vs. All Others
Manos	0.69 (0.07)	0.63 (0.08)	0.62 (0.07)	0.64 (0.06)
Royall	0.68 (0.08)	0.62 (0.07)	0.65 (0.07)	0.63 (0.09)
Shulman	0.62 (0.07)	0.65 (0.05)	0.59 (0.06)	0.63 (0.04)
Libon	0.60 (0.08)	0.65 (0.12)	0.60 (0.14)	0.64 (0.05)
Rouleau	0.59 (0.13)	0.64 (0.09)	0.53 (0.09)	0.60 (0.06)
Mendez	0.68 (0.06)	0.65 (0.05)	0.61 (0.07)	0.61 (0.07)
MiniCog	0.55 (0.07)	0.56 (0.07)	0.53 (0.05)	0.54 (0.07)
Best ML with op-clinician features	0.73 (0.06)	0.71 (0.08)	0.71 (0.05)	0.70 (0.06)
Best ML with simplest features	0.72 (0.05)	0.73 (0.07)	0.74 (0.08)	0.72 (0.05)
Best ML with all features	0.83 (0.06)	0.79 (0.05)	0.82 (0.05)	0.82 (0.05)

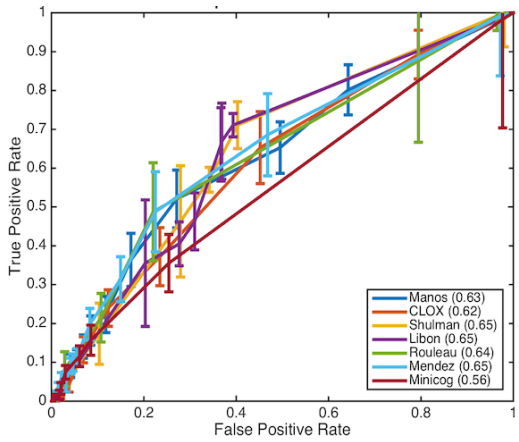
Table 5.7: Operationalized scoring system AUCs for clinical group classification task, together with AUCs of the best machine learning model on the op-clinician features, simplest features, and the set of all features.

on all features (reproduced from Section 4 for comparison). We can see that all three machine learning models are much more accurate than the operationalized scoring systems, even when using identical features (the op-clinician features), or ones that are even easier to measure (the simplest features).

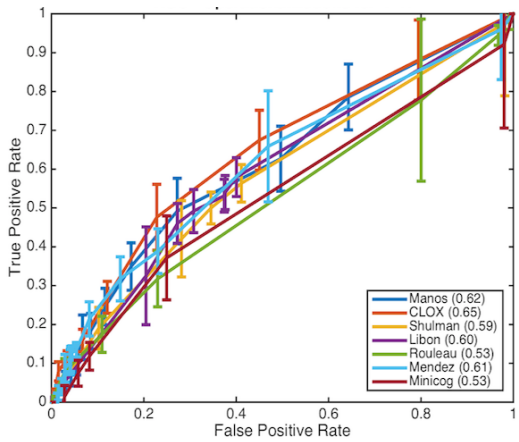
Table 5.7 and Figure 5-2 show corresponding accuracy results for the operationalized scoring systems on the clinical group classification task. Again, the machine learning classifiers created from all three feature sets are much more accurate than the operationalized scoring systems, which scored mostly in the low 60s. We were unable to find any published accuracies for these existing scoring systems on a comparable clinical group classification task. We can see that with these higher accuracies from the machine learning models, the dCDT could be considered not only as a general screening tool, but might also be able to guide diagnosis.



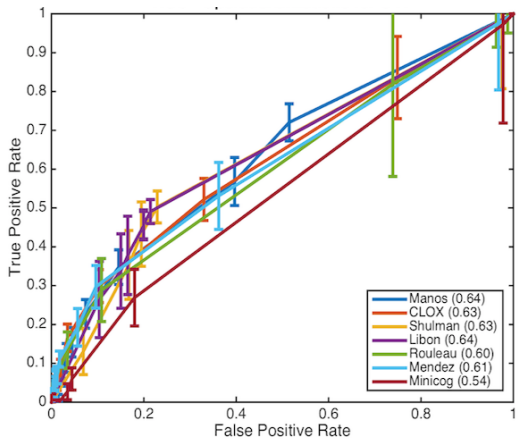
(a) Memory impairment disorders vs. All others



(b) Vascular cognitive disorders vs. All others



(c) Parkinson's vs. All others



(d) All three vs. All others

Figure 5-2: ROC curves for the experiments in Table 5.7.

Chapter 6

Interpretable Models

We have found that state-of-the-art machine learning methods on simplest features, clinician features, and the set of all features outperform existing scoring criteria; however, the existing scoring systems remain more interpretable. Interpretability is crucial if domain experts are to accept and use the model. We turn next to finding models that are more transparent and hence more likely to be accepted in practice, yet still outperform existing models.

6.1 Defining Interpretability for our task

The interpretability of a model is domain specific. To ensure that we produce models that can be used and accepted in a clinical context, we obtained guidelines from clinicians. This lead us to focus on three components: ease of feature measurements and their reliability, model computational complexity, and model understandability.

1. Ease of feature measurements and reliability: Some features can be measured quickly by eye (e.g. is there a minute hand present) while others would require a digital pen (time to draw the hand). In addition, some have a greater inter-clinician variance in measurements. This led us to focus on features that we believed would have the lowest variance, which, as noted, we call the simplest features. Models produced using these features could easily be used without a digital pen or other digitizing mechanism.

2. Computational complexity: the models should be relatively easy to compute, requiring a number of simple operations similar to the existing manual scoring systems. The existing

scoring systems discussed above have on average 8 to 15 rules, with each rule containing on average one or two features. We thus focus on models that use fewer than 20 features, and have a simple form, which in our case means either addition or subtraction of feature scores (i.e., a linear model), or an ordered sequence of if-then statements (a decision list). Clinicians should be able to evaluate these types of models rapidly.

3. Understandability: the rationale for a decision made by the model should be easily understandable, so that the user can understand why the prediction was made and can easily explain it. Thus if several features are roughly equally useful in the model, the most understandable one should be used. As one example of what we mean by "understandable," note that our feature set includes 3 measures of test taking time: the total time to draw the command clock, the total time to draw the copy clock, and the aggregate of the two, the total time to draw both. If using total time to draw both clocks produces the most accurate model, but almost all of the predictive power comes from only one of the components, say the total time to draw the command clock, it would be reasonable to trade some small amount of accuracy in order to use the simpler feature, the command clock drawing time. The form of the model is also important for understandability, leading us to focus on linear models and decision lists.

Our goal in the remainder of this thesis is to build classifiers that are at least as interpretable as existing scoring systems (according to the criteria mentioned above), but that are more accurate. While our focus will be on using the simplest features, we will also create interpretable models using op-clinician features and the MRMR subset of all features. These latter two might not be as practical to use manually, and may not be as interpretable, but exploring them allows us to test the predictive power of these more complex features. In addition, if these models achieve high accuracy, they could also be used for automatic scoring while providing interpretability for each prediction.

6.2 Interpretable Linear Models

We begin by using a recently developed framework, Supersparse Linear Interpretable Models (SLIM) [50, 51], designed to create sparse linear models that have integer coefficients and

constraints on the range of coefficients. To improve model understandability, we added feature preferences, where certain features would be preferred over others if performance is similar.

Given a dataset of N examples $D_N = \{(\boldsymbol{x_i}, y_i)\}_{i=1}^N$, each with F features, and a constant term, we want to build models of the form $y = \text{sign}(\boldsymbol{\lambda}^T \boldsymbol{x})$, where $\boldsymbol{\lambda} \subseteq \mathbb{Z}^{F+1}$ is a vector of integer coefficients. The framework determines the coefficients of the models by solving an optimization problem of the form:

$$\min_{\boldsymbol{\lambda}} \qquad \text{Loss}(\boldsymbol{\lambda}; D_n) + C \cdot \Phi(\boldsymbol{\lambda})$$

$$\text{s.t.} \quad \boldsymbol{\lambda} \in \mathcal{L}.$$

The Loss function $\text{Loss}(\boldsymbol{\lambda}; D_n)$ penalizes misclassifications. The interpretability penalty function $\Phi(\boldsymbol{\lambda})$: $\mathbb{R}^{F+1} \to \mathbb{R}$ allows for a tradeoff between desired interpretability qualities and accuracy, with the regularization parameter C controlling the balance. The framework also allows interpretability constraints by limiting $\boldsymbol{\lambda}$ to a user-defined set $\mathcal{L}$, to restrict coefficients to a particular set of values.

The framework allows for many of our interpretability goals. Integer coefficients allow for models that are more easily computable, have greater expository power, and have the same form as the scoring systems already in use; hard constraints on the coefficients allow us to set a hard limit on the number of variables used in the model, thus reducing computational complexity for evaluation of the model on a new patient.

We defined our own interpretability penalty function $\Phi(\boldsymbol{\lambda})$ to allow us to prioritize certain features, to ensure that the most understandable features appear in the model. We defined an understandability penalty u_i for each feature i by organizing our features into trees such that the children of each feature are those it depends on. For instance “total time to draw both clocks" has as children “total time to draw command clock" and “total time to draw copy clock." We define

$$u_i = \text{height(i)} \quad \forall i$$

which produces a bias toward simpler features, i.e., those lower in the tree.

Given that we want to regulate both the model complexity and the model understandability,

Features	MID vs. Healthy	VCD vs. Healthy	Parkinson's vs. Healthy	All three vs. Healthy
SLIM with simplest features	0.78 (0.08)	0.75 (0.05)	0.78 (0.07)	0.74 (0.05)
SLIM with op-clinician features	0.75 (0.10)	0.74 (0.07)	0.73 (0.11)	0.74 (0.06)
SLIM with MRMR subset	0.83 (0.09)	0.81 (0.13)	0.81 (0.10)	0.83 (0.09)
Best operationalized scoring system	0.73 (0.08)	0.72 (0.09)	0.73 (0.09)	0.70 (0.06)
Best ML with all features	0.93 (0.09)	0.88 (0.11)	0.91 (0.11)	0.91 (0.09)
Best ML with op-clinician features	0.83 (0.09)	0.83 (0.11)	0.86 (0.08)	0.82 (0.10)
Best ML with simplest features	0.83 (0.06)	0.82 (0.07)	0.83 (0.08)	0.83 (0.07)

Table 6.1: Results for Supersparse Linear Integer Models on screening task

we define our interpretability penalty function $\Phi(\boldsymbol{\lambda})$ as

$$\begin{aligned}\Phi(\boldsymbol{\lambda}) &= \text{computational complexity penalty} + \text{understandability penalty} \\ &= C_0 \sum_{i=1}^{F} \mathbf{1}[\boldsymbol{\lambda}_i \neq 0] + C_1 \sum_{i=1}^{F} u_i \cdot \mathbf{1}[\boldsymbol{\lambda}_i \neq 0].\end{aligned} \tag{6.1}$$

The first term simply computes the ℓ_0 semi-norm of $\Phi(\boldsymbol{\lambda})$, which is the count of the number of nonzero features. This term encourages the model to use fewer features. The second term adds our feature-based understandability penalty u_i for each feature used, which, while very simple, allows the optimization to potentially sacrifice a little accuracy for features lower in the tree, which we believe will be more understandable.

We ran our optimization problem on the set of simplest features and the clinician features, with a hard upper bound of 10 features, to keep them interpretable, and on the MRMR subset of all features with an upper bound of 20 features. Tables 6.1 and 6.2 present the AUCs for screening and clinical group classification, respectively. For screening, all the SLIM models outperformed the operationalized scoring systems, the best of which performed in the 0.70 to 0.73 range (Table 5.6). For clinical group classification, only the SLIM models with the MRMR subset of all features significantly outperforms the operationalized scoring systems, while the others perform similarly, the best of which performed in the 0.64 to 0.69 range (Table 5.7).

Table 6.3 shows a SLIM model containing only 9 binary features, yet achieving an AUC score of 0.78. Pushed by the understandability penalty, the model uses mostly simple features composed of a single property, except for the first line which consists of an aggregate of multiple simpler features, chosen by the optimization despite its complexity because of its high screening power. This model contains only elements from the simplest feature set, which means they do

Features	MID vs. All others	VCD vs. All others	Parkinson's vs. All others	All three vs. All others
SLIM with simplest features	0.68 (0.12)	0.66 (0.10)	0.66 (0.07)	0.69 (0.05)
SLIM with op-clinician features	0.67 (0.09)	0.66 (0.07)	0.66 (0.10)	0.70 (0.04)
SLIM with MRMR subset	0.75 (0.04)	0.72 (0.06)	0.77 (0.06)	0.76 (0.08)
Best operationalized scoring system	0.69 (0.07)	0.65 (0.05)	0.65 (0.07)	0.64 (0.05)
Best ML with all features	0.83 (0.06)	0.79 (0.05)	0.82 (0.05)	0.82 (0.05)
Best ML with op-clinician features	0.73 (0.06)	0.71 (0.08)	0.71 (0.05)	0.70 (0.06)
Best ML with simplest features	0.72 (0.05)	0.73 (0.07)	0.74 (0.08)	0.72 (0.05)

Table 6.2: Results for Supersparse Linear Integer Models on clinical group classification task

PREDICT MEMORY IMPAIREMENT DISORDER IF SCORE < 10	
Command clock:	
1. All digits are present, not repeated, and in the correct angular order	+5
2. Hour hand is present	+5
3. All of the non-anchor digits are in the correct eighth	+1
4. Crossed-out digits present	-3
5. Two hands not present	-1
6. More than 60 seconds to draw	-1
7. Minute hand points to digit 10	-6
Copy clock:	
8. All of the non-anchor digits are in the correct eighth	+4
9. Numbers are repeated	-3

Table 6.3: Supersparse Linear Integer Model for screening of memory impairment disorders

not have the problems present in many existing scoring systems; in particular, the features used in the model are not as subjective, producing a scoring criterion likely to be more reliable.

6.3 Rules and Decision Lists

We mined association rules from our data and used these rules to build interpretable decision lists. The rules allow us to gain insights about how different cognitive impairments influence behavior on the test. By constraining the width and length of our decision lists to levels similar to existing scoring systems, and by using simple features, we created decision lists that we believe can be easily interpreted by clinicians. Unlike the linear models above, rules and decision lists also allow us to use non-linear relationships in the data.

6.3.1 Mining Association Rules

The first step was to discretize all of our features into equal-frequency bins, using 2 and 5 bins per feature. We then mined globally for all IF-THEN rules in the data that obeyed certain conditions on the quality of the rule. In particular, we wanted the rules with sufficiently high support (i.e the number of subjects that obeyed the IF condition). We also wanted rules with high confidence (i.e. the empirical probability of the THEN condition to be true, given that the IF condition is true). We used FPGrowth [4] to extract decision rules from our data that predict each of our conditions (memory impairment disorders, vascular cognitive disorders, Parkinson's). We set a minimum support threshold of 40 tests, and required confidence to be greater than chance, where chance is simply the proportion of total patients who had the condition. Figure 6-1 shows the distribution of confidence and support for rules for each condition in the screening task.

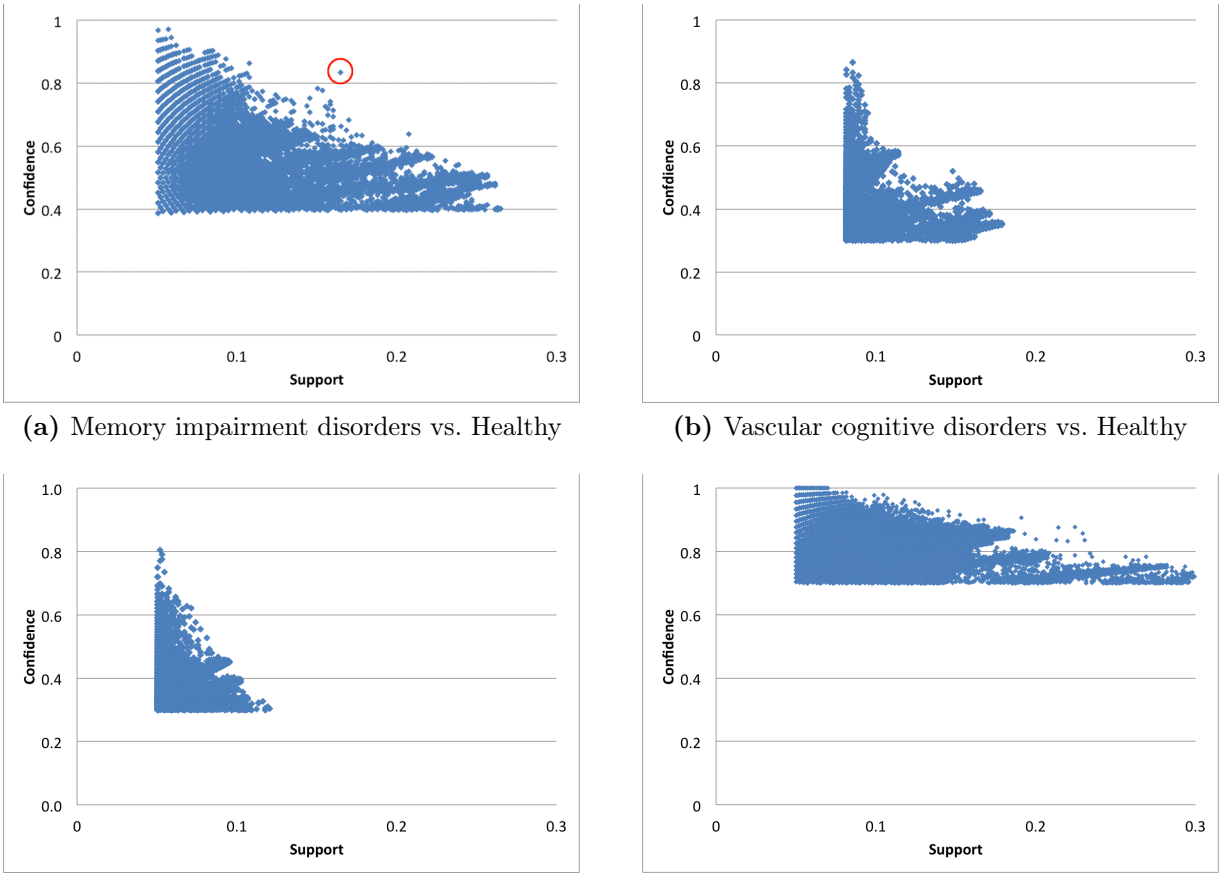
These graphs show us that some of these rules can be very accurate. For memory impairment disorders for example, we have a rule that, for our data, can be applied to 15% of the tests and can accurately predict memory impairment disorders over 80% of the time (circled in Figure 6-1(a)). This rule is: Pre-first-hand latency on copy clock is greater than 2.3 seconds, AND at least one hand missing on the command clock. This rule is consistent with what is known about memory impairment disorders, as we discuss below.

6.3.2 Interesting patterns

Some of the association rules confirm existing knowledge about correlations between pen-based features and clinical groups. Others appear to be novel, possibly providing insight into correlations not reported previously. Tables 6.4, 6.5, and 6.6 present a set of rules that focus on the screening task for memory impairment disorders, vascular cognitive disorders, and Parkinson's.

Memory impairment disorders

The first two rules in Table 6.4 show that, when compared to healthy subjects, the memory impairment group subjects tend to spend a greater percentage of the test-taking time thinking (i.e., with pen off the paper) and a smaller percentage of their test-taking time inking (with



(a) Memory impairment disorders vs. Healthy

(b) Vascular cognitive disorders vs. Healthy

(c) Parkinson's vs. Healthy

(d) All three vs. Healthy

Figure 6-1: Scatter plot of Confidence vs. Support for rules for each condition vs. healthy. Each dot on the plot represents an IF-THEN rule, where the condition is the THEN part of the rule. The right angle at the bottom left of each of these clusters shows the minimum confidence and support cutoffs used when mining the rules.

	Rule	Support	Confidence
1	Percentage thinking time is high, $> 65\%$ (alternative phrasing: Percentage inking time is low, $< 35\%$)	0.2	0.57
2	Pre-first-hand latency on copy clock is high, > 2.3 seconds	0.2	0.64
3	Pre-first-hand latency on copy clock is high, > 2.3 seconds, AND at least one hand missing on the command clock	0.14	0.84
4	There is at least one digit missing on command clock and none missing on copy clock	0.04	0.78
5	The minute hand is pointing more than 15^o away from digit 2 on command clock but points within 15^o degrees on copy clock	0.06	0.75

Table 6.4: Screening conditions implying memory impairment disorders

	Rule	Support	Confidence
1	Minute hand points within 15^o of digit 10 on command clock	0.04	0.79
2	One or more digits fail the quadrant test on command clock	0.19	0.52
3	Average time to draw digits on both clocks is high, > 2.5 seconds	0.2	0.52

Table 6.5: Screening conditions implying vascular cognitive disorders

	Rule	Support	Confidence
1	Average inking time over both clocks is high, > 17 seconds	0.2	0.38
2	Average angle gap of clock face is high, > 1	0.2	0.43
3	The average pen speed is low for both clocks	0.19	0.41
4	Average digit width is low, < 3mm	0.2	0.33
5	Average digit height is low, < 5mm	0.2	0.34
6	Average number of strokes per clock is high, > 27	0.16	0.34
7	Average number of noise strokes per clock is high, > 1.5	0.2	0.38
8	Average number of noise strokes smaller than 0.3mm per clock is high, > 0.5	0.2	0.49

Table 6.6: Screening conditions implying Parkinson's disease

pen on the paper). This is consistent with what's known about Alzheimer's and amnestic MCI.

The third rule indicates that memory impairment group subjects make a longer than normal pause between the first stroke of the hands and the last stroke that was drawn before the hands on the copy clock. This may result from decision-making difficulty, or from trouble recalling the instructions given (e.g., what time to set the clock to). Combining this third rule with the requirement that both hands be present on the command clock gives the fourth rule, which has a very high confidence.

Memory impairment patients tend to display signs of improvement from the command clock to the copy clock. Consistent with this, the fifth rule finds in the data that there is a significant chance someone belongs in the memory impairment group if they have one or more digits missing on their command clock but none missing on their copy clock. Similarly, the sixth rule tells us that this group is very likely if the minute hand is not aimed accurately in the command clock but is aimed accurately in the copy clock.

Vascular cognitive disorders

The patterns that distinguish the vascular-related cognitive disorders subjects from our healthy subjects are similar to those of the memory impairment group. These subjects also tend to spend more time thinking, less time inking, and show signs of improvements between the two clocks.

We highlight a few additional rules in Table 6.5. The first rule shows a particularly interesting phenomenon: some patients draw the minute hand pointing towards the 10 digit instead of towards the 2 digit, presumably driven by the words "ten" and "eleven" (as in the instructions

to set the time to "ten past eleven"). Almost 80% of people who do this fall in our vascular cognitive disorders group, making it a very accurate rule for screening. The second rule measures the angular distribution of the digits around the clock using the quadrant test, and if one or more digits fail the quadrant test, there is a high chance the subject belongs in our vascular cognitive disorders group. These subjects also tend to spend a long time drawing digits, as shown in the third rule.

Parkinson's Disease

The patterns for the Parkinon's group are very different. As expected, given the motor slowing and increased incidence of tremor characteristic of this disorder, instead of having low inking time like the memory group and the cognitive disorders group, subjects in the Parkinson's group tend to have high inking time over both clocks, likely due to motor impairment, as shown in the first rule of Table 6.6. The second rule shows that they tend to leave a larger angular gap in their clock face, possibly a consequence of their difficulty in starting, stopping, and persisting in motions, which might contribute to premature stopping, producing the gaps. They also tend to display signs of bradykensia, drawing slower than healthy patients, a common symptom of ParkinsonÕs, as shown in the third rule. The fourth and fifth rule show that the digits tend to be both shorter and narrower than those of healthy subjects, suggestive of micrographia, also common among ParkinsonÕs patients. Both their command and copy clocks also tend to have more total strokes (rule 6), and they also have a larger number of noise strokes (rule 7), particularly small strokes (rule 8), possibly due to tremors, or a pull to stimulus (i.e. the subject is resting the pen on a target of attention in the clock).

While these rules provide some interesting insights when considered individually, we also want to combine them to produce a classifier in the form of a decision list, yielding a classifier with a high degree of accuracy that remains interpretable. We turn next to this.

6.3.3 Decision Lists

To construct scoring systems for the CDT that are both accurate and interpretable using the rules mined above, we chose a recently developed machine learning algorithm called *Bayesian Rule Lists* (BRL) [22]. Its intent is to create classifiers that are accurate but more interpretable

Features	MID vs. Healthy	VCD vs. Healthy	Parkinson's vs. Healthy	All three vs. Healthy
BRL with simplest features	0.82 (0.06)	0.79 (0.08)	0.81 (0.05)	0.82 (0.06)
BRL with op-clinician features	0.82 (0.07)	0.78 (0.07)	0.83 (0.09)	0.78 (0.10)
BRL with MRMR subset	0.83 (0.10)	0.82 (0.07)	0.79 (0.09)	0.85 (0.09)
Best operationalized scoring system	0.73 (0.08)	0.72 (0.09)	0.73 (0.09)	0.70 (0.06)
Best ML with all features	0.93 (0.09)	0.88 (0.11)	0.91 (0.11)	0.91 (0.09)
Best ML with op-clinician features	0.83 (0.09)	0.83 (0.11)	0.86 (0.08)	0.82 (0.10)
Best ML with simplest features	0.83 (0.06)	0.82 (0.07)	0.83 (0.08)	0.83 (0.07)

Table 6.7: Results for BRL on screening task

than traditional machine learning models like CART, and thus more likely to be used by clinicians.

BRL derives from the data an ordered list of IF-THEN rules, known as a *decision list*. Table 6.9 shows an example. There are two main steps to the BRL algorithm:

- Find all of the feature combinations that occur sufficiently often (e.g., copy clock is missing numbers AND there is a missing hour hand on the command clock).

- Choose and order the feature combinations to form the left hand sides of rules for the decision list. This is done using a Bayesian modeling approach. BRL has two user-defined parameters that enter into its Bayesian prior over rule lists, allowing the user to specify the desired number of rules in the rule list, λ, and the desired number of conditions within each rule, η.

BRL's Bayesian modeling approach creates a posterior distribution of decision lists. The Bayesian prior encourages it to favor lists with approximately λ rules and η conditions per rule, as specified by the user.

We ran BRL on our three sets of features: simplest features, op-clinician features, and the MRMR subset of all features. AUCs for screening are shown in Table 6.7, and range from 0.79 to 0.85. These are significantly more accurate than the operationalized scoring systems (the best of which performed in the 0.70 to 0.73 range, Table 5.6). Clinical group classification AUCs, shown in Table 6.8, display a range from 0.69 to 0.74, only slightly better than the operationalized scoring systems (the best of which performed in the 0.64 to 0.69 range, Table 5.7).

Features	MID vs. All others	VCD vs. All others	Parkinson's vs. All others	All three vs. All others
BRL with simplest features	0.72 (0.08)	0.71 (0.05)	0.70 (0.08)	0.69 (0.06)
BRL with op-clinician features	0.70 (0.11)	0.72 (0.08)	0.69 (0.07)	0.72 (0.11)
BRL with MRMR subset	0.73 (0.08)	0.70 (0.05)	0.73 (0.06)	0.74 (0.08)
Best operationalized scoring system	0.69 (0.07)	0.65 (0.05)	0.65 (0.07)	0.64 (0.05)
Best ML with all features	0.83 (0.06)	0.79 (0.05)	0.82 (0.05)	0.82 (0.05)
Best ML with op-clinician features	0.73 (0.06)	0.71 (0.08)	0.71 (0.05)	0.70 (0.06)
Best ML with simplest features	0.72 (0.05)	0.73 (0.07)	0.74 (0.08)	0.72 (0.05)

Table 6.8: Results for BRL on clinical group classification task

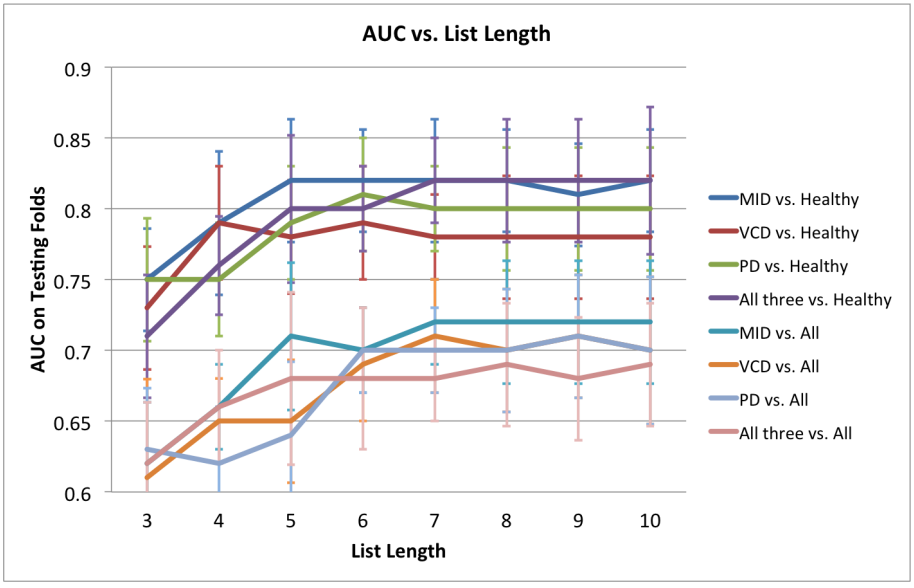


Figure 6-2: Plot of AUC on testing folds vs. list length for simplest features, for both screening and clinical group classification.

IF the command clock minute hand points within 15^o of digit 10	**THEN** 94% (88%−100%)
ELSE IF the command clock minute hand is present and drawn outwards from the center **AND** all of the non-anchor digits in the command clock are in the correct eighth	**THEN** 16% (12% − 20%)
ELSE IF all hands are present with arrowheads pointing outwards **AND** more than 5 of the non-anchor digits in the copy clock are in the correct eighth	**THEN** 24% (17% − 32%)
ELSE IF the total time to draw the command clock is greater than 40 seconds	**THEN** 92% (84% − 98%)
ELSE IF the total time to draw the copy clock is less than 20 seconds	**THEN** 12% (0% − 21%)
ELSE	33% (12% − 45%)

Table 6.9: BRL for screening of memory impairment disorders. Percentages are the probability of memory impairment disorders, with the 95% confidence interval in parentheses.

There is a tradeoff between accuracy and the size of the list (both width and length). Adding more rules and allowing them to have more antecedents will increase the accuracy, up to a limit. To make the decision list as interpretable as existing scoring systems, we restricted the width to at most 2, then maximized the accuracy over possible lengths. Figure 6-2 shows the tradeoff between testing AUC and list length for the simplest features. For the screening task, between 4 and 7 rules leads to the maximum AUC, while 5 to 8 is enough for clinical group classification. These models are both more concise and more accurate than existing scoring algorithms.

Table 6.9 presents a decision list obtained for the screening of memory impairment disorders; it was derived using the simplest features to allow the resulting decision list to be used with the pen-and-paper test, and allow clinicians to measure these features quickly and reliably by eye. Containing only 5 rules, each of similar complexity to a line from the existing scoring systems, it is shorter than most of the existing scoring systems, yet it achieves an AUC of 0.82, higher than the upper bound of 0.73 on the best existing scoring system that we examined.

Chapter 7

Conclusion

Traditional scoring systems created by clinicians are typically based on obvious features and thus have a transparency and face validity that is readily understood by the user population. A potential lack of transparency in machine learning-derived classifiers could be a barrier to clinical use.

Our goal was to have the best of both worlds: create an automated system based on new technology (the digital pen), state-of-the-art machine learning methods, and large amounts of patient data, but ensure the same interpretability qualities as the existing scoring systems. There are several important challenges we faced when trying to create our assessment models, in addition to the usual challenges of applying machine learning in practice for knowledge discovery applications.

The first challenge is *interpretability*. A major theme of this work is how to walk the line between interpretability and accuracy. We started with traditional (black box) machine learning methods to establish the highest accuracy baselines, then went to the other end of the spectrum by mining association rules, which provided accuracy baselines for the most interpretable methods. We then aimed to find the right balance of interpretability and accuracy using new machine learning techniques designed for this particular tradeoff. The models we learned have major advantages in accuracy over the traditional scoring systems for the clock drawing test, and even some advantages in interpretability because the traditional pen-and-paper scoring systems require subjective judgment and are not consistent across clinicians. Interpretability is notoriously difficult to quantify for a particular domain, but in this case, we

were able to use the new machine learning techniques to create models that mimic the form of model that the clinicians currently use. These techniques allowed us to optimize directly for interpretability as we chose to define it. The resulting models are potentially directly actionable. Our results indicate that some of our models are more robust, just as interpretable, more accurate than some widely used scoring systems, and require less computation on the part of the clinicians to compute the result, even without the benefit of the detailed data from the digital pen.

Another challenge we faced is how to create a reasonable assessment of the quality of our predictions, which required us to encode subjective human judgments in a way that captured the intent of those judgments. This led to our strategy of creating an optimized version of each of the existing scoring systems (the operationalized scoring systems). We were then able to show that even fully optimized versions of widely used scoring methods were not as accurate as a machine learning methods trained on data – even when that machine learning method was trained on the same features used in the existing scoring systems. This shows the power of combining machine learning with clinical knowledge.

This project brings together many important pieces: a new sensor (the digital pen), new techniques for handwritten stroke classification, techniques for optimizing calculations made using human judgment, new machine learning techniques for interpretability, and data created from many subjects' clock drawings and their subsequent clinical classifications. While our classifiers now need to be tested in actual clinical use, the results presented here suggest the potential of this work to make significant improvements in both screening and diagnosis of cognitive conditions.

Appendix A

All operationalized scoring systems

A.1 Additional features

We define two additional features that appear within the operationalized scoring systems, in Table A.1. The following subsections each provide an existing scoring system and our operationalization of it.

Variable	Description
ClockfaceGap	The distance between the start and end of the clock face
DigitClockfaceDistanceVariance	The variance in the distance of digits from the clockface.

Table A.1: Additional operationalized clinician features.

A.2 Manos

Table A.2 provides the original Manos scoring system, and Table A.3 shows our operationalization.

maximum: 10 points
1. Digit placement errors (maximum: 8 points)
The clock is divided into eighths, beginning with a line through “12" and the center of the circle (if “12" is missing the position is assumed to be counterclockwise from the “1" at a distance equal to that between the “1" and “2") For each eighths, add one point if the expected anchor digit is missing
2. Presence and placement of the hands (maximum: 2 points)
One point each is given for an obvious short hand pointing at the “11" and an obvious long hand pointing to the “2" The difference in the length of the hands must be obvious at a glance

Table A.2: Original Manos scoring system [26]

maximum: 10 points
1. Digit placement errors (maximum: 8 points)
Get angle of digit 12 If “12" present, go to step 2. Else if “12" not present but “1" and “2" present, get angle of “1" and “2", compute difference in angle, and add difference to angle of “1" to get approximate angle of “12". Else if “12" not present but “10" and “11"" present, get angle of “10" and “11", compute difference in angle, and subtract difference to angle of “11" to get approximate angle of “12". Else, bring up error. $\forall$ step $\in [-15, -14, ..., 0, ..., 14, 15]$ Break up clock into eighths using angle of “12" + step and adding multiples of 45^o to obtain eighths For each eighth, add one point if the expected anchor digit is missing Pick the the minimum score over all step values.
2. Presence and placement of the hands (maximum: 2 points)
If exactly two hands are present AND handRatio $\leq \epsilon_1$ If minute hand has handAngleError $\leq \epsilon_2$, add 1 If hour hand has handAngleError $\leq \epsilon_2$, add 1

Table A.3: Operationalization of Manos scoring system

A.3 Royall

Table A.4 provides the original Royall scoring system, and Table A.5 shows our operationalization.

maximum: 15 points; one point for each line satisfied
1. Does figure resemble a clock?
2. Circular face present?
3. Dimensions $>$ 1 inch ?
4. All numbers inside the perimeter?
5. “12", “6", “3" and “9" placed first?
6. Spacing intact? (symmetry on either side of “12" and “6" o’clock)
7. No sectoring or tic marks?
8. Only Arab numerals?
9. Only numbers 1-12 among the numerals present?
10. Sequence 1-12 intact? (no omissions or intrusions)
11. Only two hands present? (ignore sectoring/tic marks)
12. All hands represented as arrows?
13. Hour hand between 1 and 2 o’clock?
14. Minute hand longer than hour hand?
15. None of the following
(1) hand pointing to 10 o’clock
(2) “11:10" present?
(3) intrusions from “hand" or “face" present?
(4) any letters, words or pictures?
(5) any intrusion from circle below?

Table A.4: Original Royall scoring system [41]

maximum: 15 points; one point for each line satisfied
1. Clockface closed percentage $\geq \epsilon_1$ AND at least 4 digits present AND at least 1 hand present
2. Clockface present
3. Major axis of fitted ellipse to clockface greater than 1 inch
4. All numbers inside the clockface
5. “12", “6", “3", “9" all anchor digits
6. DigitsAngleError $\leq \epsilon_2$
7. No spokes or tick marks present
8. Always 1 (we do not have any clocks with other numerals in our dataset so assume it is very rare)
9. No digit greater than 12 present
10. All numbers present in correct order by angle, no repetitions, no numbers greater than 12, no text, crossed-out digits allowed
11. Two hands present, no repetitions of hands but allow crossed-out hands
12. Arrows present on both hands, direction must be correct
13. Angle of hour hand between angle of “11" and angle of “12". If either digits or hand missing, 0
14. HandRatio $\leq \epsilon_3$
15. None of the following
(1) Minute hand pointing within ϵ_4 of “10"
(2) Any text present
(3) Always false. Very hard to measure, and no example in dataset so assume it is very rare
(4) Any text present
(5) Always false.

Table A.5: Operationalization of Royall scoring system

A.4 Shulman

Table A.6 provides the original Schulman scoring system, and Table A.7 shows our operationalization.

maximum: 6 points

1. Perfect
2. Minor visuospatial errors

    Examples

    (a) Mildly impaired spacing of times

    (b) Draws times outside circle

    (c) Turns page while writing numbers so that some numbers appear upside down

    (d) Draws in lines (spokes) to orient spacing

3. Inaccurate representation of “10 after 11" when visuospatial organization is perfect or shows only minor deviations

    Examples

    (a) Minute hand points to “10"

    (b) Writes “10 after 11"

    (c) Unable to make any denotation of time

4. Moderate visuospatial disorganization of times such that accurate denotation of “10 after 11" is impossible

    Example

    (a) Moderately poor spacing

    (b) Omits numbers

    (c) Perseveration: repeats circle or continues on past 12 to 13, 14, 15 etc.

    (d) Right-left reversal: numbers drawn counterclockwise

    (e) Dysgraphia: unable to write numbers accurately

5. Severe level of disorganization as described in 4
6. No reasonable representation of a clock

Exclude severe depression or other psychotic states

    Example

    (a) No attempt at all

    (b) No semblance of a clock at all

    (c) Writes a word or name

Table A.6: Original Shulman scoring system [42]

maximum: 6 points
1. MOCAPtScore = 3
2. Minor visuospatial errors
(a) $\epsilon_1 <$ DigitsAngleError $\leq \epsilon_2$
(b) At least one digit outside the circle
(c) No way to measure automatically given our data, and very rare according to doctors
(d) At least one spoke present
3. Inaccurate representation of “10 after 11" when visuospatial organization is perfect or shows only minor deviations
(a) Minute hand points within ϵ_3 of “10"
(b) Any text present
(c) both hands have HandAngleError $> \epsilon_4$
4. Moderate visuospatial disorganization of times such that accurate denotation of “10 after 11" is impossible
(a) DigitNeighborsTest $\geq \epsilon_5$
(b) At least one digit missing
(c) More than one clockface OR at least one digit repeated OR digits greater than 12 present
(d) Numbers drawn counterclockwise
(e) At least one digit missing
5. Severe level of disorganization as described in 4
Severely poor spacing: DigitNeighborsTest $\geq \epsilon_6$
6. No reasonable representation of a clock
Clockface closed percentage $< \epsilon_7$ OR fewer than four digits present OR no hands present

Table A.7: Operationalization of Shulman scoring system

A.5 Libon

Table A.8 provides the original Libon scoring system, and Table A.9 shows our operationalization.

maximum: 10 points
Scores 10 to 6: Circle and Hands are basically intact, some impairment in hand placement.
10: Hands, numbers and circle are totally intact 9: Slight error(s) in hand number placement; hands of equal length; any self-correction 8: More noticeable errors in hand/number placement; hand length correct but shifted to one side or top/bottom 7: Significant errors in hand placement; hand placement intact with some numbers deleted; minor perseveration in number placement 6: Inappropriate use of clock hands i.e., digital display; circling numbers to indicate hand placement; connecting the numbers 10 and 11 or 11 and 2.
Scores 5 to 1: Circle, numbers and/or hand placement are grossly impaired.
5: Crowding numbers to one side; numbers reversed; significant perseveration of numbers within circle boundary 4: Loss of clock face integrity, numbers outside circle boundary, further distortion of number placement 3: Numbers and clock face no longer connected 2: Vague representations of a clock; clock face absent but numbers present 1: Either no attempt or response is made; scattered bits or fragments are produced

Table A.8: Original Libon scoring system [23]

maximum: 10 points
Scores 10 to 6: Circle and Hands are basically intact, some impairment in hand placement.
10: Both hands have HandAngleError $\leq \epsilon_1$ 9: At least one hand has $\epsilon_1 <$ HandAngleError $\leq \epsilon_2$ 8: Both hands hand have $\epsilon_1 <$ HandAngleError $\leq \epsilon_2$ 7: At least one hand not in correct quadrant 6: Ignore: Hard to measure automatically, and very rare in our data
Scores 5 to 1: Circle, numbers and/or hand placement are grossly impaired.
5: DigitNeighborsTest $\geq \epsilon_3$ 4: Any number missing or any number placed outside the clockface 3: DigitNeighborsTest $\geq \epsilon_4$ 2: Clockface closed percentage $\geq \epsilon_5$ OR at least four digits present OR at least one hand present 1: Clockface closed percentage $< \epsilon_5$ AND fewer than four digits present AND less than one hand present

Table A.9: Operationalization of Libon scoring system

A.6 Mendez

Table A.10 provides the original Mendez scoring system, and Table A.11 shows our operationalization.

maximum: 20 points; one point for each line satisfied
1. There is an attempt to indicate a time in any way
2. All marks or items can be classified as either part of a closure figure, a hand, or a symbol for clock numbers
3. There is a totally closed figure without gaps (closure figure).
Score only if Symbols for Clock Numbers Are Present
4. A 2 is present and is pointed out in some way for the time
5. Most symbols are distributed as a circle without major gaps
6. Three or more clock quadrants have one or more appropriate numbers: 12 to 3, 3 to 6, 6 to 9, 9 to 12 per respective clockwise quadrant.
7. Most symbols are ordered in a clockwise or rightward direction
8. All symbols are totally within a closure figure.
9. An 11 is present and is pointed out in some way for the time
10. All numbers 1-12 are indicated
11. There are not repeated or duplicated number symbols
12. There are no substitutions for Arabic or Roman numerals
13. The numbers do not go beyond the number 12
14. All symbols lie about equally adjacent to a closure figure edge
15. Seven or more of the same symbol type are ordered sequentially.
Score Only if One or More Hands Are Present:
16. All hands radiate from the direction of a closure figure center
17. One hand is visibly longer than another hand
18. There are exactly two distinct and separable hands
19. All hands are totally within a closure figure
20. There is an attempt to indicate a time with one or more hands.

Table A.10: Original Mendez scoring system [28]

maximum: 20 points; one point for each line satisfied
1. At least one hand present
2. No noise, no ticks, no spokes, no text
3. ClockfaceGap $\leq \epsilon_1$
Score only if Symbols for Clock Numbers Are Present
4. "2" is present, minute hand has handAngleError $\leq \epsilon_2$
5. DigitsAngleError $< \epsilon_3$
6. Break clock into quadrants, and at least three correct digits within each quadrant
7. More than half of digits present are in clockwise direction
8. No digit or hands present outside clockface
9. "11" is present, hour hand has handAngleError $\leq \epsilon_2$
10. All digits present
11. No repeated digits (cross-outs allowed)
12. Always 1 (hard to measure and very rare in our data)
13. No digits greater than 12 present
14. DigitClockfaceDistanceVariance $< \epsilon_4$
15. At least 7 digits are in correct order by angle
Score Only if One or More Hands Are Present:
16. Both hands are drawn in an outwards direction
17. HandRatio $\leq \epsilon_5$
18. Only two hands present (cross-outs allowed)
19. Hands drawn within the clockface
20. At least one hand present

Table A.11: Operationalization of Mendez scoring system

A.7 MiniCog

Table A.14 provides the original Mendez scoring system, and Table A.13 shows our operationalization.

maximum: 1 point	
If all numbers approximately in the correct position AND there are two hands pointing properly	+1

Table A.12: Original MiniCog scoring system [6]

maximum: 1 point	
If DigitsAngleError $< \epsilon_1$ AND both hands have HandAngleError $< \epsilon_2$	+1

Table A.13: Operationalization of MiniCog scoring system

A.8 Watson

maximum: 7 points	
1. The first quadrant (12-3) has less than three digits	+1
2. The second quadrant (3-6) has less than three digits	+1
3. The third quadrant (6-9) has less than three digits	+1
4. The fourth quadrant (9-12) has less than three digits	+4

Table A.14: Original Watson scoring system

maximum: 7 points
1. Get angle of “12"
If “12" present, go to step 2.
Else if “12" not present but “1" and “2" present, get angle of “1" and “2", compute difference in angle,
and add difference to angle of “1" to get approximate angle of “12".
Else if “12" not present but “10" and “11"" present, get angle of “10" and “11", compute difference
in angle, and subtract difference to angle of “11" to get approximate angle of “12".
Else, bring up error.
2. $\forall$ step $\in [-45, -44, ..., 0, ..., 44, 45]$
Break up clock into quadrants using angle of “12" + step
and adding multiples of 90$\circ$ to obtain quadrants
For each quadrant that does not have 3 digits,
add 1 if it’s one of the first three quadrants, and 4 if it’s the fourth
3. Pick the the minimum score over all step values.

Table A.15: Operationalization of Watson scoring system

Appendix B

ID of clock tests used

ADI0008256248 ADI0019362138 ADI0075002223 ADI0075005607 ADI0075245712
ADI0102773856 ADI0136018171 ADI0139079400 ADI0154339576 ADI0216580475
ADI0223663476 ADI0227360375 ADI0295762579 ADI0314259971 ADI0317654307
ADI0317754394 ADI0335808004 ADI0366611095 ADI0396704045 ADI0416346673
ADI0435757687 ADI0481548090 ADI0532797685 ADI0629818118 ADI0635491091
ADI0647537257 ADI0650804890 ADI0666531196 ADI0686390434 ADI0698564728
ADI0727437491 ADI0731431217 ADI0744278184 ADI0745745588 ADI0756774543
ADI0764806176 ADI0808144331 ADI0827276492 ADI0856368928 ADI0863722010
ADI0888617511 ADI0933504019 ADI0939369232 ADI0959223173 ADI0976513858
ADI0980424484 ADI1002420659 ADI1027293504 ADI1081210139 ADI1142049050
ADI1147871291 ADI1168630452 ADI1199876434 ADI1231588876 ADI1244167831
ADI1311590981 ADI1321165697 ADI1331086070 ADI1386733327 ADI1400319714
ADI1432494541 ADI1453617021 ADI1485720460 ADI1502112757 ADI1571025542
ADI1726887733 ADI1744985764 ADI1761571913 ADI1787209007 ADI1880235858
ADI1888630530 ADI1940442556 ADI1980332443 ADI2042488697 ADI2067299361
ADI2073369577 ADI2092543576 ADI2118006160 ADI2128501311 ADI2136876313
CIN0012554857 CIN0021048289 CIN0021464378 CIN0022817477 CIN0041466804
CIN0047736485 CIN0052221862 CIN0053407618 CIN0055911696 CIN0057065125
CIN0059530813 CIN0064170993 CIN0068752107 CIN0079657991 CIN0083950870
CIN0083967214 CIN0086289408 CIN0087127573 CIN0090074055 CIN0098487336

CIN0100446024 CIN0100603242 CIN0102618210 CIN0104871835 CIN0105093392
CIN0105350865 CIN0105520977 CIN0105982484 CIN0116966184 CIN0123862361
CIN0124525147 CIN0125579522 CIN0132590717 CIN0133238244 CIN0134665159
CIN0140686130 CIN0146564148 CIN0147692407 CIN0154369662 CIN0164742787
CIN0167812185 CIN0168920098 CIN0173101364 CIN0176652109 CIN0184968821
CIN0187365029 CIN0193304772 CIN0197977380 CIN0210382014 CIN0211552676
CIN0212916689 CIN0219479435 CIN0221953493 CIN0227530469 CIN0231776495
CIN0248655962 CIN0249837095 CIN0252802715 CIN0253153890 CIN0254203855
CIN0258862275 CIN0260717676 CIN0262387140 CIN0263633495 CIN0263688266
CIN0269492078 CIN0273483843 CIN0273928773 CIN0276692760 CIN0278615980
CIN0283371394 CIN0288132746 CIN0302941166 CIN0304569211 CIN0309311347
CIN0309972167 CIN0312155466 CIN0319786108 CIN0323991053 CIN0324244069
CIN0325886721 CIN0327850012 CIN0328353984 CIN0329765729 CIN0338325221
CIN0341880680 CIN0344007458 CIN0351611797 CIN0355693955 CIN0357138794
CIN0364224432 CIN0367938843 CIN0375598863 CIN0376235667 CIN0379627862
CIN0380389393 CIN0381774299 CIN0381839798 CIN0387958652 CIN0388064539
CIN0392789743 CIN0392916689 CIN0398505892 CIN0400188137 CIN0400930771
CIN0408536824 CIN0410358680 CIN0412588832 CIN0413002535 CIN0413582377
CIN0415564888 CIN0416841585 CIN0420909454 CIN0428936289 CIN0428982561
CIN0430675678 CIN0431929809 CIN0432253546 CIN0438467910 CIN0442155792
CIN0450280399 CIN0451985477 CIN0453287782 CIN0455388352 CIN0460213468
CIN0469013109 CIN0471230137 CIN0471926272 CIN0479180710 CIN0482970969
CIN0484415202 CIN0487468552 CIN0488921745 CIN0489803364 CIN0492585281
CIN0500127282 CIN0506813451 CIN0509887985 CIN0510660217 CIN0517439397
CIN0532840254 CIN0538240262 CIN0545159554 CIN0550602888 CIN0550945481
CIN0560040151 CIN0564655929 CIN0568585115 CIN0574546767 CIN0575818107
CIN0578666524 CIN0578707061 CIN0578857653 CIN0583937164 CIN0584487598
CIN0585527464 CIN0591539882 CIN0594516500 CIN0595908914 CIN0598760430
CIN0601776251 CIN0605679142 CIN0606445963 CIN0611403556 CIN0612040585
CIN0612721737 CIN0614448972 CIN0626337772 CIN0636412677 CIN0637371952
CIN0638156814 CIN0639609890 CIN0649466009 CIN0650029116 CIN0653161572

CIN0665001820 CIN0668974316 CIN0670471511 CIN0671250734 CIN0672793737
CIN0673918715 CIN0677771050 CIN0678537159 CIN0678593951 CIN0680822397
CIN0683990948 CIN0684048327 CIN0684591786 CIN0690199065 CIN0694525168
CIN0694636158 CIN0698323704 CIN0700659550 CIN0702630279 CIN0706851426
CIN0710043159 CIN0715562809 CIN0723152914 CIN0728847298 CIN0731389076
CIN0736560387 CIN0741037621 CIN0741257347 CIN0742533336 CIN0743332986
CIN0745286634 CIN0752776689 CIN0758179468 CIN0760212440 CIN0762170906
CIN0764536926 CIN0766003158 CIN0767033276 CIN0772192473 CIN0773050006
CIN0780355753 CIN0781482116 CIN0782521053 CIN0788977646 CIN0793119678
CIN0793690264 CIN0794004576 CIN0797570370 CIN0797612368 CIN0803820439
CIN0806005027 CIN0808780731 CIN0809634535 CIN0815906007 CIN0816762614
CIN0817443455 CIN0818723200 CIN0819407799 CIN0822210251 CIN0822866441
CIN0829359923 CIN0832677687 CIN0832904372 CIN0836717582 CIN0838333020
CIN0840815587 CIN0846222755 CIN0851145015 CIN0854992145 CIN0855218712
CIN0860855232 CIN0864490712 CIN0866222349 CIN0873632655 CIN0875849784
CIN0876102883 CIN0880825141 CIN0880940110 CIN0885753910 CIN0888150022
CIN0889674526 CIN0896053636 CIN0896384340 CIN0900030148 CIN0903144133
CIN0906909659 CIN0909373883 CIN0920378440 CIN0922540310 CIN0926258744
CIN0928346331 CIN0929223317 CIN0939054722 CIN0943834199 CIN0945986383
CIN0956001420 CIN0958482690 CIN0965876830 CIN0980194111 CIN0980285858
CIN0983401482 CIN0984378197 CIN0990384836 CIN0993769749 CIN0998325398
CIN0999089989 CIN1002367642 CIN1002545116 CIN1002547684 CIN1002796093
CIN1004647539 CIN1004681732 CIN1018227802 CIN1020097416 CIN1020906993
CIN1025370378 CIN1029212004 CIN1031695010 CIN1032333235 CIN1040066317
CIN1040239746 CIN1047290664 CIN1047511993 CIN1054910602 CIN1058930159
CIN1061551666 CIN1068353997 CIN1076069864 CIN1077285320 CIN1079793304
CIN1088914040 CIN1095883363 CIN1096699940 CIN1097468304 CIN1105631229
CIN1107967550 CIN1118890191 CIN1123731973 CIN1129292656 CIN1130636350
CIN1131508255 CIN1134532484 CIN1135689197 CIN1136301330 CIN1136641253
CIN1139003678 CIN1144547255 CIN1145514502 CIN1145606819 CIN1148569125
CIN1151887790 CIN1152189056 CIN1162209701 CIN1164661265 CIN1170361290

CIN1172871596 CIN1175052931 CIN1178387165 CIN1180313729 CIN1180502947
CIN1183088827 CIN1183788714 CIN1185390433 CIN1193918104 CIN1204798366
CIN1206532380 CIN1211875458 CIN1216824403 CIN1219856373 CIN1220912665
CIN1222437637 CIN1224429209 CIN1225679313 CIN1228679487 CIN1229225404
CIN1236929156 CIN1237389173 CIN1239163249 CIN1243687999 CIN1245861924
CIN1252905851 CIN1253884343 CIN1253903474 CIN1265744205 CIN1266613070
CIN1273579188 CIN1274701843 CIN1275013047 CIN1278874453 CIN1279534509
CIN1279745510 CIN1292926857 CIN1294207338 CIN1294891404 CIN1296931004
CIN1297300762 CIN1297384782 CIN1305012740 CIN1312564029 CIN1313182310
CIN1316808391 CIN1319180741 CIN1324121380 CIN1335400759 CIN1339875634
CIN1340059288 CIN1347970133 CIN1348028999 CIN1348645032 CIN1353387013
CIN1367052777 CIN1367127301 CIN1369389823 CIN1369614699 CIN1370460189
CIN1371523966 CIN1377038857 CIN1381742086 CIN1384247465 CIN1387021482
CIN1390405521 CIN1392067704 CIN1392849442 CIN1393233740 CIN1395920741
CIN1397286283 CIN1397380605 CIN1398100870 CIN1400972887 CIN1408340045
CIN1408521859 CIN1420950292 CIN1422940326 CIN1426253847 CIN1428072395
CIN1430331196 CIN1430474458 CIN1430555625 CIN1431384319 CIN1431510849
CIN1436063098 CIN1439741535 CIN1440212394 CIN1441194436 CIN1441223030
CIN1441566889 CIN1452870911 CIN1454909560 CIN1455936430 CIN1458421117
CIN1459287457 CIN1460847500 CIN1464730678 CIN1465436195 CIN1469615084
CIN1470707530 CIN1471768746 CIN1481185361 CIN1482047207 CIN1487380041
CIN1489174410 CIN1491059590 CIN1496615692 CIN1501863653 CIN1503842164
CIN1504062693 CIN1505288632 CIN1506613715 CIN1507887534 CIN1513621878
CIN1516213853 CIN1520349377 CIN1523269884 CIN1530958939 CIN1530974164
CIN1532674130 CIN1535924518 CIN1538240524 CIN1539469736 CIN1540131583
CIN1550057937 CIN1566979916 CIN1568009842 CIN1569282405 CIN1573515168
CIN1575363401 CIN1575689941 CIN1578339852 CIN1579029831 CIN1582152442
CIN1582724493 CIN1585055379 CIN1589355945 CIN1593283173 CIN1607345900
CIN1607370268 CIN1607419267 CIN1614329181 CIN1619000239 CIN1626877896
CIN1630012730 CIN1634537358 CIN1652821717 CIN1653986498 CIN1655231138
CIN1657782012 CIN1659964930 CIN1661876793 CIN1663888975 CIN1665583706

CIN1666607938 CIN1670814028 CIN1671055220 CIN1671233146 CIN1678376476
CIN1680619393 CIN1685239080 CIN1687365461 CIN1688813715 CIN1692371108
CIN1697275343 CIN1701712050 CIN1709050449 CIN1712773189 CIN1718216577
CIN1727768249 CIN1733790982 CIN1735040052 CIN1735536427 CIN1737373792
CIN1739530392 CIN1744559777 CIN1747864838 CIN1747870687 CIN1752153237
CIN1757731144 CIN1758561753 CIN1759690886 CIN1760264257 CIN1760929815
CIN1763208168 CIN1766642865 CIN1768017805 CIN1768861648 CIN1769392878
CIN1769615697 CIN1776225148 CIN1777673498 CIN1783746660 CIN1786411571
CIN1787372015 CIN1787816975 CIN1788079352 CIN1788425118 CIN1792641094
CIN1793304415 CIN1800007620 CIN1807983379 CIN1814114464 CIN1816461137
CIN1820078637 CIN1827632452 CIN1830331306 CIN1831363922 CIN1831995296
CIN1841368545 CIN1841775235 CIN1845199887 CIN1845623100 CIN1848110058
CIN1848188725 CIN1849101830 CIN1856129260 CIN1862250539 CIN1864733719
CIN1868194358 CIN1868861524 CIN1869969099 CIN1881432278 CIN1882327356
CIN1886581307 CIN1887040604 CIN1888464750 CIN1890265273 CIN1891149026
CIN1892122559 CIN1893992940 CIN1897577607 CIN1899384355 CIN1905170544
CIN1908734515 CIN1909521969 CIN1909600964 CIN1913591915 CIN1914662806
CIN1917974906 CIN1918085473 CIN1918089285 CIN1918811480 CIN1925705500
CIN1928613727 CIN1928686896 CIN1928783471 CIN1930780503 CIN1934340138
CIN1934430120 CIN1936459786 CIN1942918516 CIN1943645286 CIN1948577106
CIN1957141861 CIN1969571523 CIN1969694309 CIN1969787414 CIN1977335049
CIN1979025987 CIN1979449919 CIN1979450492 CIN1980144457 CIN1980597825
CIN1981306725 CIN1983534704 CIN1983985131 CIN1986393573 CIN1988368470
CIN1992775528 CIN1993089748 CIN1994388749 CIN1996429532 CIN1999330564
CIN2004016886 CIN2006670482 CIN2006928793 CIN2007039393 CIN2012822029
CIN2013617100 CIN2015222127 CIN2016349273 CIN2029699715 CIN2034954515
CIN2037335507 CIN2040680452 CIN2047216560 CIN2049827175 CIN2050302313
CIN2050460514 CIN2056483383 CIN2057090687 CIN2065300116 CIN2068560265
CIN2070842051 CIN2072315286 CIN2076705627 CIN2080545981 CIN2080763298
CIN2087014604 CIN2088492276 CIN2094549127 CIN2096243533 CIN2099425708
CIN2102762143 CIN2103879823 CIN2108247672 CIN2109964091 CIN2111583867

CIN2114802665 CIN2124810133 CIN2124841396 CIN2142434511 CIN2146016349
CIN2146196895 CIN2146741235 CIN2147142001 CIU0005486174 CIU0017622890
CIU0024206738 CIU0046081236 CIU0053722794 CIU0061793568 CIU0066457197
CIU0068842289 CIU0079703918 CIU0082207563 CIU0096160840 CIU0102246802
CIU0103195854 CIU0104748943 CIU0107435524 CIU0111920429 CIU0119497602
CIU0124645489 CIU0127099221 CIU0131890396 CIU0149301375 CIU0180915278
CIU0182779253 CIU0191046308 CIU0239612146 CIU0242808349 CIU0243557598
CIU0246900263 CIU0250532099 CIU0254593945 CIU0260523291 CIU0288026327
CIU0310647667 CIU0315713704 CIU0316818651 CIU0326567661 CIU0352050448
CIU0353216303 CIU0398141918 CIU0417725922 CIU0430121389 CIU0440019387
CIU0442363321 CIU0531924436 CIU0532303856 CIU0537529226 CIU0537878236
CIU0548573982 CIU0549197535 CIU0555381131 CIU0571042779 CIU0581395448
CIU0598926729 CIU0621757291 CIU0626994089 CIU0627168325 CIU0631823260
CIU0657961717 CIU0665356781 CIU0671099093 CIU0671168913 CIU0674087700
CIU0674760449 CIU0704488820 CIU0706949522 CIU0741381093 CIU0742443950
CIU0761996293 CIU0767129698 CIU0770682680 CIU0774261019 CIU0775198123
CIU0777098261 CIU0795311667 CIU0816972891 CIU0846291863 CIU0850202076
CIU0856293915 CIU0866526305 CIU0889161814 CIU0898738577 CIU0902712583
CIU0910967317 CIU0914197709 CIU0917521172 CIU0933702777 CIU0970572152
CIU0973378160 CIU1015022888 CIU1021799980 CIU1025987127 CIU1062624084
CIU1066180526 CIU1071085405 CIU1086930867 CIU1088956579 CIU1091674056
CIU1093137561 CIU1099068193 CIU1109603030 CIU1115939941 CIU1116621506
CIU1117495628 CIU1153290503 CIU1168872894 CIU1188729923 CIU1191221108
CIU1239254123 CIU1241827398 CIU1266990146 CIU1285844264 CIU1324389181
CIU1340967210 CIU1363604262 CIU1364012448 CIU1364222395 CIU1367274223
CIU1376568818 CIU1388319646 CIU1389708368 CIU1391272275 CIU1407432484
CIU1424128298 CIU1425492718 CIU1444860531 CIU1446664843 CIU1452312314
CIU1464204075 CIU1472879808 CIU1485690589 CIU1500115584 CIU1505689976
CIU1510709131 CIU1515285922 CIU1522021811 CIU1524538009 CIU1537040188
CIU1565546674 CIU1569337259 CIU1585029778 CIU1586251114 CIU1606392884
CIU1616463717 CIU1635413023 CIU1637829744 CIU1643693318 CIU1670019021

CIU1674727377 CIU1705630072 CIU1707691323 CIU1716540425 CIU1737092819
CIU1739040295 CIU1755286706 CIU1768209562 CIU1769416462 CIU1782960317
CIU1799629617 CIU1801225326 CIU1823972891 CIU1834101511 CIU1848319523
CIU1852397041 CIU1859562066 CIU1860878789 CIU1862526098 CIU1889489567
CIU1891410857 CIU1899494459 CIU1936049898 CIU1940265462 CIU1945551524
CIU1966246447 CIU1991059739 CIU1994588455 CIU1997850316 CIU2000716439
CIU2079212770 CIU2092895667 CIU2097017403 CIU2112529139 CIU2134257892
CIU2134391113 EGE0048723202 EGE0056676390 EGE0062352022 EGE0074582654
EGE0091443850 EGE0092099221 EGE0143851472 EGE0167599527 EGE0270005059
EGE0271665761 EGE0348579752 EGE0375558151 EGE0430106867 EGE0458758770
EGE0468185867 EGE0477008340 EGE0477340547 EGE0488148287 EGE0489394608
EGE0495314770 EGE0519424751 EGE0528412617 EGE0556887562 EGE0603076589
EGE0628057102 EGE0706902923 EGE0719336199 EGE0720764651 EGE0744818560
EGE0766567854 EGE0813752608 EGE0838358279 EGE0882081772 EGE0887135619
EGE0922540745 EGE0960117541 EGE0969601126 EGE1059111135 EGE1077000824
EGE1152023406 EGE1165700469 EGE1168812095 EGE1174719586 EGE1194738423
EGE1199219204 EGE1210175927 EGE1227383573 EGE1229216930 EGE1230811213
EGE1238003390 EGE1240346762 EGE1252519043 EGE1298297446 EGE1318066790
EGE1331665504 EGE1331724899 EGE1370827434 EGE1377705409 EGE1418164226
EGE1466972375 EGE1477402218 EGE1516019668 EGE1528314806 EGE1551806340
EGE1555519969 EGE1564463898 EGE1569302432 EGE1579069590 EGE1595650202
EGE1630941806 EGE1635278100 EGE1663533090 EGE1703572234 EGE1724348272
EGE1729607156 EGE1731929614 EGE1765900491 EGE1782096001 EGE1813478853
EGE1865516148 EGE1923413357 EGE1940454417 EGE1962770219 EGE1969144806
EGE2030406051 EGE2038384853 EGE2089890253 EGE2100566174 EGE2117649573
ORU0022621348 ORU0033873510 ORU0040221516 ORU0067812077 ORU0101280521
ORU0141409120 ORU0214002724 ORU0234651310 ORU0241277053 ORU0242642663
ORU0260103282 ORU0263782536 ORU0271454768 ORU0278498759 ORU0288347298
ORU0338895011 ORU0351166563 ORU0355087110 ORU0371179841 ORU0435765892
ORU0454992946 ORU0456648861 ORU0478357206 ORU0526107987 ORU0563716628
ORU0566807646 ORU0577384976 ORU0640807570 ORU0681309129 ORU0684255875

ORU0684461700 ORU0691777419 ORU0702838292 ORU0748619075 ORU0796572994
ORU0832245419 ORU0840094047 ORU0882917589 ORU0891269308 ORU0933676648
ORU1010706926 ORU1030612082 ORU1141527401 ORU1184517177 ORU1282217042
ORU1298744135 ORU1405425692 ORU1457366995 ORU1472034496 ORU1497888622
ORU1513723722 ORU1523209376 ORU1545747048 ORU1572369380 ORU1579143314
ORU1594504278 ORU1614808862 ORU1644707772 ORU1651237257 ORU1709324812
ORU1817010287 ORU1823106156 ORU1852069863 ORU1872703283 ORU1938503383
ORU1952963752 ORU1967217631 ORU1977877628 ORU1979847017 ORU1984069152
ORU2000037553 ORU2086234709 ORU2120290419 ORU2123368665 ORU2124971720
VIN0002546833 VIN0002752711 VIN0011253644 VIN0037442139 VIN0063271152
VIN0078085048 VIN0079536327 VIN0084450104 VIN0086266413 VIN0091872821
VIN0098771936 VIN0102603075 VIN0102736346 VIN0122024543 VIN0147882920
VIN0150986070 VIN0159227304 VIN0176770694 VIN0211110230 VIN0212724106
VIN0214894135 VIN0216801333 VIN0222487600 VIN0223245274 VIN0250487960
VIN0275434639 VIN0277293416 VIN0288160313 VIN0292653489 VIN0320646684
VIN0353132207 VIN0360962039 VIN0373688483 VIN0381902123 VIN0382623641
VIN0387262421 VIN0397194519 VIN0398883252 VIN0405636253 VIN0408418879
VIN0440705291 VIN0441352371 VIN0475791367 VIN0487443598 VIN0497097589
VIN0500200077 VIN0513420101 VIN0516652623 VIN0517074792 VIN0525310967
VIN0528492924 VIN0531061126 VIN0536748850 VIN0542469514 VIN0558833393
VIN0576293986 VIN0588654376 VIN0591255927 VIN0624107039 VIN0630798716
VIN0630915307 VIN0644921205 VIN0647261580 VIN0663836195 VIN0676446680
VIN0689239745 VIN0711833317 VIN0713714003 VIN0721968100 VIN0762360910
VIN0775888172 VIN0785980059 VIN0787138990 VIN0787678398 VIN0788428366
VIN0800441387 VIN0828088000 VIN0828812527 VIN0829757788 VIN0831798596
VIN0833140908 VIN0834666119 VIN0842680555 VIN0847029653 VIN0847870313
VIN0853226752 VIN0863786444 VIN0873509087 VIN0880699204 VIN0897425793
VIN0902133667 VIN0903667079 VIN0922826245 VIN0924591893 VIN0927265292
VIN0929937545 VIN0934868080 VIN0938575033 VIN0953299173 VIN0969671570
VIN0983535487 VIN0984345114 VIN0997183320 VIN1019479827 VIN1021052330
VIN1021133203 VIN1030668968 VIN1040347531 VIN1052780286 VIN1073197592

VIN1082466239 VIN1106145152 VIN1117825264 VIN1136939465 VIN1152904937
VIN1158639961 VIN1170676553 VIN1173894107 VIN1179048218 VIN1179560000
VIN1181803228 VIN1185197344 VIN1192862333 VIN1202464077 VIN1221788157
VIN1232857867 VIN1242373154 VIN1245859164 VIN1246286886 VIN1246459525
VIN1247964890 VIN1250546032 VIN1250612126 VIN1252719001 VIN1255156352
VIN1263579883 VIN1266811704 VIN1270987934 VIN1290845432 VIN1298002396
VIN1300160199 VIN1303699403 VIN1313750347 VIN1335376767 VIN1341672470
VIN1355430577 VIN1381455645 VIN1406164481 VIN1407793617 VIN1418679258
VIN1431673895 VIN1439528750 VIN1445890720 VIN1463925987 VIN1467304713
VIN1475529716 VIN1487156944 VIN1520402976 VIN1533730546 VIN1542794379
VIN1543094987 VIN1570034958 VIN1582668235 VIN1591138123 VIN1634366975
VIN1638973340 VIN1639731805 VIN1644104725 VIN1652062116 VIN1657771398
VIN1660017801 VIN1662416892 VIN1687212776 VIN1693665651 VIN1699673802
VIN1706069152 VIN1708306654 VIN1715687822 VIN1720450681 VIN1724539676
VIN1724927675 VIN1725370023 VIN1726811630 VIN1727767885 VIN1728476201
VIN1729760908 VIN1737987862 VIN1745613379 VIN1762161879 VIN1786441936
VIN1790499685 VIN1792676811 VIN1794164525 VIN1833967705 VIN1842240487
VIN1851737753 VIN1861900076 VIN1877439688 VIN1882970623 VIN1892070344
VIN1908439122 VIN1926831987 VIN1931807909 VIN1949946426 VIN1962208556
VIN1973134116 VIN1982923303 VIN1989294807 VIN1991121621 VIN1993325106
VIN1996974147 VIN1998384886 VIN2009985186 VIN2015542200 VIN2016358114
VIN2028485314 VIN2037668217 VIN2047727573 VIN2070830288 VIN2070879711
VIN2099133980 VIN2100335296 VIN2100703695 VIN2103774766 VIN2107627400
VIN2109342881 VIN2109969259 VIN2118764148 VIN2121743464 VIN2123733155
VIN2129103436 VIN2131858992 YDU0001222313 YDU0001995977 YDU0004812213
YDU0005342314 YDU0007874997 YDU0009592929 YDU0010114143 YDU0011654702
YDU0012501339 YDU0013165932 YDU0013511725 YDU0015398892 YDU0015704730
YDU0015872230 YDU0018893352 YDU0023798940 YDU0024943741 YDU0029028854
YDU0029629324 YDU0030990860 YDU0035349809 YDU0039337599 YDU0044439968
YDU0045091243 YDU0047892294 YDU0047936500 YDU0053252734 YDU0053582268
YDU0053889897 YDU0054324162 YDU0056871723 YDU0057362444 YDU0059776545

YDU0064762880 YDU0065739936 YDU0065892741 YDU0067017427 YDU0067271598
YDU0068714028 YDU0071592235 YDU0072054945 YDU0072905616 YDU0073446731
YDU0074216081 YDU0076511236 YDU0076967328 YDU0080623491 YDU0084102565
YDU0084791938 YDU0086558979 YDU0089151951 YDU0092317699 YDU0093793509
YDU0097648169 YDU0098326501 YDU0099477386 YDU0102547972 YDU0103415343
YDU0109904354 YDU0111769440 YDU0112592176 YDU0112640181 YDU0123041202
YDU0128184270 YDU0129091497 YDU0130065731 YDU0132070845 YDU0133676187
YDU0135408446 YDU0140499323 YDU0142668978 YDU0144341018 YDU0144626603
YDU0144626604 YDU0144626605 YDU0148702199 YDU0150532271 YDU0150589315
YDU0158173116 YDU0160578066 YDU0160769184 YDU0161082557 YDU0161752608
YDU0162000130 YDU0165816428 YDU0169902394 YDU0173108894 YDU0176742814
YDU0181639157 YDU0187125954 YDU0189863547 YDU0191480696 YDU0196742056
YDU0199652904 YDU0207836983 YDU0208767786 YDU0211097573 YDU0213031281
YDU0216695424 YDU0222788751 YDU0226307867 YDU0227127582 YDU0234300612
YDU0234983411 YDU0236027528 YDU0237370881 YDU0238181248 YDU0239303440
YDU0240350443 YDU0241445434 YDU0242373021 YDU0247161988 YDU0250379388
YDU0250446862 YDU0252175358 YDU0252721858 YDU0257649605 YDU0257793879
YDU0262071045 YDU0262420381 YDU0262424844 YDU0278262356 YDU0281453940
YDU0283089966 YDU0283670083 YDU0283970673 YDU0286312782 YDU0292722320
YDU0293591268 YDU0293828571 YDU0295441515 YDU0295451586 YDU0298383646
YDU0299160367 YDU0301056481 YDU0301956507 YDU0302633253 YDU0304807885
YDU0305577176 YDU0308807254 YDU0310771230 YDU0312745603 YDU0314082337
YDU0317471688 YDU0320124394 YDU0323716285 YDU0324183969 YDU0324550443
YDU0328993665 YDU0330250803 YDU0331939947 YDU0332451869 YDU0338189874
YDU0342312503 YDU0342530934 YDU0343788578 YDU0351589511 YDU0354521219
YDU0358584419 YDU0362615041 YDU0371045035 YDU0372786121 YDU0374591263
YDU0375115358 YDU0378706093 YDU0384303266 YDU0385237968 YDU0390670726
YDU0392936549 YDU0394930661 YDU0395245091 YDU0397651649 YDU0397955754
YDU0401256539 YDU0408173575 YDU0410561867 YDU0411449437 YDU0411490306
YDU0412693060 YDU0414648157 YDU0420807930 YDU0424959326 YDU0427692275
YDU0427932340 YDU0431355497 YDU0437042415 YDU0438510865 YDU0439121597

YDU0440190199 YDU0443993605 YDU0444599141 YDU0444891311 YDU0447375509
YDU0447602253 YDU0453755748 YDU0454193187 YDU0454714376 YDU0455939792
YDU0458592227 YDU0461049251 YDU0462692310 YDU0463010837 YDU0464687505
YDU0464807783 YDU0465797464 YDU0466459882 YDU0467332523 YDU0467901475
YDU0468102939 YDU0469690072 YDU0470548424 YDU0473619766 YDU0474403630
YDU0475404730 YDU0478065614 YDU0479585880 YDU0480309885 YDU0482080206
YDU0487785606 YDU0492103324 YDU0494834573 YDU0495019518 YDU0496144274
YDU0499378634 YDU0501930192 YDU0503742430 YDU0504645194 YDU0505961541
YDU0506254540 YDU0514125798 YDU0514945177 YDU0517164636 YDU0522020499
YDU0524712288 YDU0525361302 YDU0525952363 YDU0528002908 YDU0528538580
YDU0530924986 YDU0533056527 YDU0533591210 YDU0534613759 YDU0534933727
YDU0536600583 YDU0536815991 YDU0537023473 YDU0537806177 YDU0538266654
YDU0545977807 YDU0551446696 YDU0553542309 YDU0555771202 YDU0562769654
YDU0564064172 YDU0565394942 YDU0565423448 YDU0567840441 YDU0570016807
YDU0572225632 YDU0573801367 YDU0573886017 YDU0574278281 YDU0578677034
YDU0578983572 YDU0579243203 YDU0581858214 YDU0582900297 YDU0585146762
YDU0586894096 YDU0587062224 YDU0589008274 YDU0591163132 YDU0594803924
YDU0595222711 YDU0596167854 YDU0599439852 YDU0600634744 YDU0602750163
YDU0606595706 YDU0609307585 YDU0609903660 YDU0610697050 YDU0613795734
YDU0617101708 YDU0617727107 YDU0620373415 YDU0622271221 YDU0631752495
YDU0634589573 YDU0641077570 YDU0642034483 YDU0646366306 YDU0647273907
YDU0649159463 YDU0650105226 YDU0650249345 YDU0650438393 YDU0652496541
YDU0653358925 YDU0654026159 YDU0654590886 YDU0661006598 YDU0664631001
YDU0664820461 YDU0664834985 YDU0667291766 YDU0669849871 YDU0674053597
YDU0674229937 YDU0674551186 YDU0675498868 YDU0675540829 YDU0678185867
YDU0684229814 YDU0691192178 YDU0694055157 YDU0695677934 YDU0698800292
YDU0698821181 YDU0704235034 YDU0705545335 YDU0707097531 YDU0708081911
YDU0717139526 YDU0721010010 YDU0727898053 YDU0728688103 YDU0730528186
YDU0734425357 YDU0734865612 YDU0737242733 YDU0741139360 YDU0749812806
YDU0751582127 YDU0751628294 YDU0753522664 YDU0758123625 YDU0758647904
YDU0764084109 YDU0768930299 YDU0769581240 YDU0778356040 YDU0778940952

YDU0780406559 YDU0781195161 YDU0781277950 YDU0787349648 YDU0798189217
YDU0800760851 YDU0805764096 YDU0811681153 YDU0816608737 YDU0823515466
YDU0824509409 YDU0828734439 YDU0831473474 YDU0837183056 YDU0838540694
YDU0839400511 YDU0844151994 YDU0845965219 YDU0848276648 YDU0849844734
YDU0850335717 YDU0851512089 YDU0855956104 YDU0860348325 YDU0864080241
YDU0869377771 YDU0870360283 YDU0882098626 YDU0886438159 YDU0886886499
YDU0902369543 YDU0906205606 YDU0910172437 YDU0912559460 YDU0913173125
YDU0916680447 YDU0921517182 YDU0922196152 YDU0922777542 YDU0924526693
YDU0929748271 YDU0933400705 YDU0940271716 YDU0940735722 YDU0941955429
YDU0943754303 YDU0943845775 YDU0949859113 YDU0950904469 YDU0954509084
YDU0955864687 YDU0955925927 YDU0958525882 YDU0970563830 YDU0976032835
YDU0980517386 YDU0981813170 YDU0987821825 YDU0988048246 YDU0989652919
YDU0989910683 YDU0995589114 YDU0998363240 YDU0998793474 YDU0999167953
YDU0999595221 YDU1004474181 YDU1018192801 YDU1019325829 YDU1025873863
YDU1026162087 YDU1027293890 YDU1027866636 YDU1028639974 YDU1029011764
YDU1032476241 YDU1033010101 YDU1033166965 YDU1042713374 YDU1045480866
YDU1046946532 YDU1048144461 YDU1049621975 YDU1053341067 YDU1054979951
YDU1056713099 YDU1059427652 YDU1061153015 YDU1066618405 YDU1067499068
YDU1067591442 YDU1069605307 YDU1069689067 YDU1070761371 YDU1073292090
YDU1073799419 YDU1075401823 YDU1075513151 YDU1075602447 YDU1076746880
YDU1081155040 YDU1082606903 YDU1083764770 YDU1084061420 YDU1084407578
YDU1087372122 YDU1087625628 YDU1089143852 YDU1094596097 YDU1094792359
YDU1097676127 YDU1098238228 YDU1098613165 YDU1098676186 YDU1102586140
YDU1104643378 YDU1106314178 YDU1108148829 YDU1111607398 YDU1112702194
YDU1113209016 YDU1113658108 YDU1116013858 YDU1123212132 YDU1132720809
YDU1134241184 YDU1137749575 YDU1138427750 YDU1140722265 YDU1146481335
YDU1147653631 YDU1153137036 YDU1157000619 YDU1157225559 YDU1157616023
YDU1167267316 YDU1171197008 YDU1172918336 YDU1178670955 YDU1186694994
YDU1189108787 YDU1190396292 YDU1193489166 YDU1193816462 YDU1197182909
YDU1199864177 YDU1203734587 YDU1204778565 YDU1205811401 YDU1207960996
YDU1208343790 YDU1208828981 YDU1209781630 YDU1227785441 YDU1228624646

YDU1228911252 YDU1229663052 YDU1229688459 YDU1231033730 YDU1234149197
YDU1237112480 YDU1239770432 YDU1239874577 YDU1247064281 YDU1248852579
YDU1255693404 YDU1263198983 YDU1264799747 YDU1266004989 YDU1266015549
YDU1266029545 YDU1266620540 YDU1267709088 YDU1272005878 YDU1272653553
YDU1280636495 YDU1287356456 YDU1290498395 YDU1290989927 YDU1291267053
YDU1291817442 YDU1292008635 YDU1298824734 YDU1302230099 YDU1305086452
YDU1306656671 YDU1310342415 YDU1317778907 YDU1322016191 YDU1325312070
YDU1325693928 YDU1326786534 YDU1327470632 YDU1331081386 YDU1332665658
YDU1333024251 YDU1334168403 YDU1341233881 YDU1346712534 YDU1347656630
YDU1348116530 YDU1350419873 YDU1350453932 YDU1366841878 YDU1370822148
YDU1371447625 YDU1376419648 YDU1383646671 YDU1386001063 YDU1388846274
YDU1388949517 YDU1391198275 YDU1393274532 YDU1395390747 YDU1396045194
YDU1396481970 YDU1397162770 YDU1398012108 YDU1403145333 YDU1409616908
YDU1410526265 YDU1411877134 YDU1420555330 YDU1423219925 YDU1423480656
YDU1426401048 YDU1427416020 YDU1429018687 YDU1430048228 YDU1431300834
YDU1433870150 YDU1434401473 YDU1434651150 YDU1436119627 YDU1436212775
YDU1436718652 YDU1437865289 YDU1440368798 YDU1441812972 YDU1442661988
YDU1448081874 YDU1450978918 YDU1452734092 YDU1454605088 YDU1455185435
YDU1457112930 YDU1457461395 YDU1457559306 YDU1458035925 YDU1458697332
YDU1459723324 YDU1460920966 YDU1462987822 YDU1465585716 YDU1469606748
YDU1471226640 YDU1473040393 YDU1473494971 YDU1474151764 YDU1475716916
YDU1476838616 YDU1481304857 YDU1482201383 YDU1482261301 YDU1483910122
YDU1485579884 YDU1487544303 YDU1491902947 YDU1493344152 YDU1494051224
YDU1494484873 YDU1497574090 YDU1500830141 YDU1503656639 YDU1510970519
YDU1511018225 YDU1515921483 YDU1516281702 YDU1519799655 YDU1520296774
YDU1520880033 YDU1522045959 YDU1522657394 YDU1523118076 YDU1524067713
YDU1524451311 YDU1536408360 YDU1536706745 YDU1545032970 YDU1546193696
YDU1548061873 YDU1550789600 YDU1551141438 YDU1552370799 YDU1556434409
YDU1558265074 YDU1558755772 YDU1560204467 YDU1566807487 YDU1567803823
YDU1572730728 YDU1577687001 YDU1582964406 YDU1584830231 YDU1585548342
YDU1589894957 YDU1590727702 YDU1595048008 YDU1596647951 YDU1596983648

YDU1604922488 YDU1605586435 YDU1608374391 YDU1610201815 YDU1613509274
YDU1615248739 YDU1622808830 YDU1625107784 YDU1627519816 YDU1631593610
YDU1634422138 YDU1637367575 YDU1638514686 YDU1640468945 YDU1642858673
YDU1643871417 YDU1644128206 YDU1644468180 YDU1646309972 YDU1647612179
YDU1652770279 YDU1657129120 YDU1663778788 YDU1663829053 YDU1664510618
YDU1667474861 YDU1674127854 YDU1675389179 YDU1682873196 YDU1689758809
YDU1689787523 YDU1691892457 YDU1695114988 YDU1696730611 YDU1697532923
YDU1697856091 YDU1698781179 YDU1698831024 YDU1699086921 YDU1700123170
YDU1701135034 YDU1707228429 YDU1708289856 YDU1713864945 YDU1720238637
YDU1726209825 YDU1726780572 YDU1728485575 YDU1729797852 YDU1730405665
YDU1731592614 YDU1734015106 YDU1735364783 YDU1736452469 YDU1737236831
YDU1737670214 YDU1737982982 YDU1738794300 YDU1739497299 YDU1740855314
YDU1743768912 YDU1746297303 YDU1746833634 YDU1748602328 YDU1751835593
YDU1752378826 YDU1753511355 YDU1758879357 YDU1758978507 YDU1760727720
YDU1762462119 YDU1766128431 YDU1766227533 YDU1771628421 YDU1779470970
YDU1780078212 YDU1793443167 YDU1793715236 YDU1796066114 YDU1796075554
YDU1798007374 YDU1798309932 YDU1799588453 YDU1802365428 YDU1802620688
YDU1803059071 YDU1803870314 YDU1804919074 YDU1807952265 YDU1820484256
YDU1823478584 YDU1824770682 YDU1825636506 YDU1826778250 YDU1828831965
YDU1830830601 YDU1831318841 YDU1837520212 YDU1839401766 YDU1839553265
YDU1839604771 YDU1852374760 YDU1853009070 YDU1854608110 YDU1857163369
YDU1862724637 YDU1870990911 YDU1873137319 YDU1878818512 YDU1884908218
YDU1885274139 YDU1886670830 YDU1891448893 YDU1892874523 YDU1895759559
YDU1903803704 YDU1904348444 YDU1905167133 YDU1907222484 YDU1912505290
YDU1914584053 YDU1915587439 YDU1915794163 YDU1918982229 YDU1921640868
YDU1922832223 YDU1926398223 YDU1934541133 YDU1950620409 YDU1953159052
YDU1956444688 YDU1961713742 YDU1963425759 YDU1963906040 YDU1965420037
YDU1965439458 YDU1966488010 YDU1966654365 YDU1972167273 YDU1978221124
YDU1978625366 YDU1979921491 YDU1980877338 YDU1980935602 YDU1980985828
YDU1982458791 YDU1985007694 YDU1988753319 YDU1989497939 YDU1991313544
YDU1992065806 YDU1995809812 YDU2005771610 YDU2010621421 YDU2012172477

YDU2012926637 YDU2014953513 YDU2015396656 YDU2016974481 YDU2018030890
YDU2019838709 YDU2025209475 YDU2028258369 YDU2028487448 YDU2028832415
YDU2032502788 YDU2034231021 YDU2036360962 YDU2037301632 YDU2043994353
YDU2045447907 YDU2047227727 YDU2056986759 YDU2058414360 YDU2060021469
YDU2060068185 YDU2060512848 YDU2061476618 YDU2061880442 YDU2063938853
YDU2066673680 YDU2068721453 YDU2071078661 YDU2071144394 YDU2074377873
YDU2075451915 YDU2076742150 YDU2079473580 YDU2084403185 YDU2087296647
YDU2089822118 YDU2090469603 YDU2091430168 YDU2092354574 YDU2092607378
YDU2100188187 YDU2100206195 YDU2100492527 YDU2101920981 YDU2104848685
YDU2105440568 YDU2109668593 YDU2110880454 YDU2120888890 YDU2124900642
YDU2130040936 YDU2131553702 YDU2132564048 YDU2132719751 YDU2132851129
YDU2133487143 YDU2134138170 YDU2134735600 YDU2138197995 YDU2138701164
YDU2139094310 YDU2140910798 YDU2141877161 YDU2144586219

Bibliography

[1] Eva Arnáiz and Ove Almkvist. Neuropsychological features of mild cognitive impairment and preclinical alzheimer's disease. *Acta Neurologica Scandinavica*, 107(s179):34–41, 2003.

[2] Leontino Battistin and Annachiara Cagnin. Vascular cognitive disorder. a biological and clinical overview. *Neurochemical research*, 35(12):1933–1938, 2010.

[3] L Bocchi, G Coppini, J Nori, and G Valli. Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks. *Medical Engineering & Physics*, 26(4):303–312, 2004.

[4] Christian Borgelt. An implementation of the FP-growth algorithm. In *Proceedings of the 1st International Workshop on Open Source Data Mining: Frequent Pattern Mining Implementations*, OSDM '05, pages 1–5, 2005.

[5] Deborah A Cahn, David P Salmon, Andreas U Monsch, Nelson Butters, WC Wiederholt, Jody Corey-Bloom, and Elizabeth Barrett-Connor. Screening for dementia of the alzheimer type in the community: the utility of the clock drawing test. *Archives of Clinical Neuropsychology*, 11(6):529–539, 1996.

[6] Abréviation Mini Cog. Borson s, scanlan j, brush m, et al.(2000) the mini-cog: a cognitive ÃćÂĂÂŸvital signsÃćÂĂÂŹ measure for dementia screening in multi-lingual elderly. *Int. J. Geriatr. Psychiatry*, 15(11):1021–27, 2000.

[7] Jamie Cohen, Dana L Penney, Randall Davis, David J Libon, Rodney A Swenson, Olusola Ajilore, Anand Kumar, and Melissa Lamar. Digital clock drawing: Differentiating ÒthinkingÓ versus ÒdoingÓ in younger and older adults with depression. *Journal of the International Neuropsychological Society*, 20(09):920–928, 2014.

[8] R. Davis, D. Penney, D. Pittman, D. Libon, R. Swenson, and E. Kaplan. The digital clock drawing test (dcdt) - i: Development of a new computerized quantitative system. Presented at the 39th annual meeting of the International Neuropsychological Society Boston, MA, 2011.

[9] Randall Davis, Bruce Buchanan, and Edward Shortliffe. Production rules as a representation for a knowledge-based consultation program. *Artificial intelligence*, 8(1):15–45, 1977.

[10] Randall Davis, David J Libon, Rhoda Au, David Pitman, and Dana L Penney. Think: Inferring cognitive status from subtle behaviors. In *Twenty-Sixth IAAI Conference*, 2014.

[11] Randall Davis and Dana L Penney. Method and apparatus for measuring representational motions in a medical context, June 3 2014. US Patent 8,740,819.

[12] Marcel Dettling. Bagboosting for tumor classification with gene expression data. *Bioinformatics*, 20(18):3583–3593, 2004.

[13] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.

[14] Morris Freedman, Larry Leach, Edith Kaplan, Gordon Winocur, Kenneth I Shulman, and Dean C Delis. *Clock drawing: A neuropsychological analysis.* Oxford University Press, 1994.

[15] Brian F Gage, Amy D Waterman, William Shannon, Michael Boechler, Michael W Rich, and Martha J Radford. Validation of clinical classification schemes for predicting stroke: results from the national registry of atrial fibrillation. *Jama*, 285(22):2864–2870, 2001.

[16] L. Grande, J. Rudolph, R. Davis, D. Penney, C. Price, and R. Swenson. Clock drawing: Standing the test of time. In Libon (eds.) Ashendorf, Swenson, editor, *The Boston Process Approach to Neuropsychological Assessment.* Oxford University Press, 2013.

[17] T. Joachims. Making large-scale svm learning practical. LS8-Report 24, Universität Dortmund, LS VIII-Report, 1998.

[18] Hyungsin Kim, Young Suk Cho, and Ellen Yi-Luen Do. Computational clock drawing analysis for cognitive impairment screening. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pages 297–300. ACM, 2011.

[19] Hyungsin Kim, Young Suk Cho, and Ellen Yi-Luen Do. Using pen-based computing in technology for health. In *Human-Computer Interaction. Users and Applications*, pages 192–201. Springer, 2011.

[20] Linda Chiu Wa Lam, Helen Fung Kum Chiu, Kei On Ng, Calais Chan, Wan Fat Chan, Siu Wah Li, and Mary Wong. Clock-face drawing, reading and setting tests in the screening of dementia in chinese elderly adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 53(6):P353–P357, 1998.

[21] M. Lamar, M. Grajewski, D. Penney, R. Davis, D. Libon, and A. Kumar. The impact of vascular risk and depression on executive planning and production during graphomotor output across the lifespan. Paper at 5th Congress of the Int'l Soc. for Vascular, Cognitive and Behavioural Disorders, Lille, France, 2011.

[22] Benjamin Letham, Cynthia Rudin, Tyler H. McCormick, and David Madigan. Interpretable classifiers using rules and bayesian analysis: Building a better stroke prediction model. Department of Statistics Technical Report tr608, University of Washington, 2014.

[23] David J Libon, Rodney A Swenson, Edward J Barnoski, and Laura Prouty Sands. Clock drawing as an assessment tool for dementia. *Archives of Clinical Neuropsychology*, 8(5):405–415, 1993.

[24] K-N Lin, P-N Wang, Chien Chen, Y-H Chiu, C-C Kuo, Y-Y Chuang, and H-C Liu. The three-item clock-drawing test: a simplified screening test for alzheimerÃŢs disease. *European neurology*, 49(1):53–58, 2002.

[25] Roberto Alves Lourenço, Sergio Telles Ribeiro-Filho, Irene de Freitas Henriques Moreira, Emylucy Martins Paiva Paradela, and Aline Sobral de Miranda. The clock drawing test: performance among elderly with low educational level. *Revista Brasileira de Psiquiatria*, 30(4):309–315, 2008.

[26] Peter J Manos and Rae Wu. The ten point clock test: a quick screen and grading method for cognitive impairment in medical and surgical patients. *The International Journal of Psychiatry in Medicine*, 24(3):229–244, 1994.

[27] M Markatou, H Tian, S Biswas, and G Hripcsak. Analysis of variance of cross-validation estimators of the generalization error. *The Journal of Machine Learning Research*, 2005.

[28] Mario F Mendez, Thomas Ala, and Kara L Underwood. Development of scoring criteria for the clock drawing task in alzheimer's disease. *Journal of the American Geriatrics Society*, 1992.

[29] Ziad S Nasreddine, Natalie A Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L Cummings, and Howard Chertkow. The montreal cognitive assessment, moca: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4):695–699, 2005.

[30] Justin A Nyborn, Jayandra J Himali, Alexa S Beiser, Sherral A Devine, Yangchun Du, Edith Kaplan, Maureen K O'Connor, William E Rinn, Helen S Denison, Sudha Seshadri, et al. The framingham heart study clock drawing performance: normative data from the offspring cohort. *Experimental aging research*, 39(1):80–108, 2013.

[31] Hanchuan Peng, Fulmi Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.

[32] D. Penney, M. Lamar, D. Libon, C. Price, R. Swenson, S. Scala, J. Eppig, C. Nieves, C. Macaulay, K. Garett, R. Au, S. Devine, L. Delano-Wood, and R. Davis. The digital clock drawing test (dcdt) hooklets: A novel graphomotor measure of executive function. Abstract and poster, 6th Congress of the Int'l Soc. for Vascular, Cognitive and Behavioural Disorders, Montreal, Canada, 2013.

[33] D. Penney, D. Libon, M. Lamar, C. Price, R. Swenson, J. Eppig, C. Nieves, K. Garett, and R. Davis. The digital clock drawing test (dcdt) - i: Information contained within the "noise". 5th Congress of the 1International Society for Vascular, Cognitive and Behavioural Disorders (VAS-COG), Lille, France, 2011.

[34] D. Penney, D. Libon, M. Lamar, C. Price, R. Swenson, S. Scala, J. Eppig, C. Nieves, K. Garett, and R. Davis. The digital clock drawing test (dcdt) - iv: Clock drawing time and hand placement latencies in mild cognitive impairment and dementia. Abstract and poster, 5th Congress of the Int'l Soc. for Vascular, Cognitive and Behavioural Disorders, Lille, France, 2011.

[35] RC Petersen, B Caracciolo, C Brayne, S Gauthier, V Jelic, and L Fratiglioni. Mild cognitive impairment: a concept in evolution. *Journal of internal medicine*, 275(3):214–228, 2014.

[36] Emanuel F Petricoin and Lance A Liotta. Seldi-tof-based serum proteomic pattern diagnostics for early detection of cancer. *Current Opinion in Biotechnology*, 15(1):24–30, 2004.

[37] Brenda L Plassman, Kenneth M Langa, Gwenith G Fisher, Steven G Heeringa, David R Weir, Mary Beth Ofstedal, James R Burke, Michael D Hurd, Guy G Potter, Willard L Rodgers, et al. Prevalence of dementia in the united states: the aging, demographics, and memory study. *Neuroepidemiology*, 29(1-2):125–132, 2007.

[38] Catherine C Price, Holly Cunningham, Nicole Coronado, Alana Freedland, Stephanie Cosentino, Dana L Penney, Alfio Penisi, Dawn Bowers, Michael S Okun, and David J Libon. Clock drawing in the montreal cognitive assessment: recommendations for dementia assessment. *Dementia and geriatric cognitive disorders*, 31(3):179–187, 2011.

[39] M Prince, M Guerchet, and M Prina. Policy brief for heads of government: the global impact of dementia 2013–2050. *London, UK: Alzheimer Disease International*, 2013.

[40] Isabelle Rouleau, David P Salmon, Nelson Butters, Colleen Kennedy, and Katheryn McGuire. Quantitative and qualitative analyses of clock drawings in alzheimer's and huntington's disease. *Brain and cognition*, 18(1):70–87, 1992.

[41] Donald R Royall, Jeffrey A Cordes, and Marsha Polk. Clox: an executive clock drawing task. *Journal of Neurology, Neurosurgery & Psychiatry*, 64(5):588–594, 1998.

[42] Kenneth I Shulman, Dolores Pushkar Gold, Carole A Cohen, and Carla A Zucchero. Clock-drawing and dementia in the community: A longitudinal study. *International Journal of Geriatric Psychiatry*, 8(6):487–496, 1993.

[43] Kenneth I Shulman, Ralph Shedletsky, and Ivan L Silver. The challenge of time: Clock-drawing and cognitive function in the elderly. *International journal of geriatric psychiatry*, 1(2):135–140, 1986.

[44] Joella E Storey, Jeffrey TJ Rowland, David Basic, and David A Conforti. A comparison of five clock scoring methods using roc (receiver operating characteristic) curve analysis. *International journal of geriatric psychiatry*, 16(4):394–399, 2001.

[45] Joella E Storey, Jeffrey TJ Rowland, David Basic, and David A Conforti. Accuracy of the clock drawing test for detecting dementia in a multicultural sample of elderly australian patients. *International psychogeriatrics*, 14(03):259–271, 2002.

[46] Richard L Strub, F William Black, and Ann C Strub. *The mental status examination in neurology*. FA Davis Philadelphia, 1985.

[47] Trey Sunderland, James L Hill, Alan M Mellow, Brian A Lawlor, J Gundersheimer, PA Newhouse, and JH Grafman. Clock drawing in alzheimer's disease: a novel measure of dementia severity. *Journal of the American Geriatrics Society*, 1989.

[48] H Tuokko, T Hadjistavropoulos, JA Miller, and BL Beattie. The clock test: a sensitive measure to differentiate normal elderly from those with alzheimer disease. *Journal of the American Geriatrics Society*, 1992.

[49] Holly Tuokko, Thomas Hadjistavropoulos, Sandra Rae, and Norm O'Rourke. A comparison of alternative approaches to the scoring of clock drawing. *Archives of Clinical Neuropsychology*, 15(2):137–148, 2000.

[50] Berk Ustun and Cynthia Rudin. Methods and models for interpretable linear classification. *arXiv preprint arXiv:1405.4047*, 2014.

[51] Berk Ustun, Stefano Tracà, and Cynthia Rudin. Supersparse linear integer models for interpretable classification. *arXiv preprint arXiv:1306.6677*, 2013.

[52] Vanya MCA Van Belle, Ben Van Calster, Dirk Timmerman, Tom Bourne, Cecilia Bottomley, Lil Valentin, Patrick Neven, Sabine Van Huffel, Johan AK Suykens, and Stephen Boyd. A mathematical model for interpretable clinical decision support with applications in gynecology. *PloS one*, 7(3):e34312, 2012.

[53] Jia-Xiang Wang, Bo Zhang, Jie-Kai Yu, Jian Liu, Mei-Qin Yang, and Shu Zheng. Application of serum protein fingerprinting coupled with artificial neural network model in diagnosis of hepatocellular carcinoma. *Chinese medical journal*, 118(15):1278–1284, 2005.

[54] Yasmira I Watson, Cynthia L Arfken, and Stanley J Birge. Clock completion: an objective screening test for dementia. *Journal of the American Geriatrics Society*, 1993.

[55] Jenna Wiens, Wayne N Campbell, Ella S Franklin, John V Guttag, and Eric Horvitz. Learning data-driven patient risk stratification models for clostridium difficile. In *Open Forum Infectious Diseases*, volume 1, page ofu045. Oxford University Press, 2014.

[56] Jenna Wiens and John V Guttag. Active learning applied to patient-adaptive heartbeat classification. In *Advances in neural information processing systems*, pages 2442–2450, 2010.

[57] Gisele P Wolf-Klein, Felix A Silverstone, Arnold P Levy, Meryl S Brod, et al. Screening for alzheimer's disease by clock drawing. *Journal of the American Geriatrics Society*, 1989.

[58] Zheng Zhao, Fred Morstatter, Shashvata Sharma, Salem Alelyani, Aneeth Anand, and Huan Liu. Advancing feature selection research. *ASU feature selection repository*, 2010.

[59] Xiaobo Zhou, Kuang-Yu Liu, and Stephen TC Wong. Cancer classification and prediction using logistic regression with bayesian gene selection. *Journal of Biomedical Informatics*, 37(4):249–259, 2004.