

Detecting Task Demand via an Eye Tracking Machine Learning System

Mina Shojaeizadeh^a, Soussan Djamasbi^a, Randy C. Paffenroth^b, Andrew C. Trapp^c

a Foisie Business School, Worcester Polytechnic Institute, Worcester, MA, USA

b Department of Mathematical Sciences, Data Science Program, Worcester Polytechnic Institute, Worcester, MA, USA

c Foisie Business School, Department of Mathematical Sciences, Data Science Program, Worcester Polytechnic Institute, Worcester, MA, USA

ABSTRACT

Computerized systems play a significant role in today's fast-paced digital economy. Because task demand is a major factor that influences how computerized systems are used to make decisions, identifying task demand automatically provides an opportunity for designing advanced decision support systems that can respond to user needs at a personalized level. A first step for designing such advanced decision tools is to investigate possibilities for developing automatic task load detectors. Grounded in decision making, eye tracking, and machine learning literature, we argue that task demand can be detected automatically, reliably, and unobtrusively using eye movements only. To investigate this possibility, we developed an eye tracking task load detection system and tested its effectiveness. Our results revealed that our task load detection system reliably predicted increased task demand from users' eye movement data. These results and their implications for research and practice are discussed.

Keywords: human computer interaction, eye tracking, task demand, adaptive decision making, cognitive effort, machine learning

1 Introduction

Recent advances in specialized hardware and software provide the opportunity to capture and analyze physiological measures that can reliably recognize as well as potentially predict changes in user behavior. This opportunity in turn makes it possible to envision intelligent decision support systems that can more effectively recognize and address user needs at a personalized level [6,30]. A first step towards developing such adaptive decision support systems is to investigate intelligent tools that can reliably and unobtrusively detect user needs. One such tool is a system that can detect task demand automatically. Task demand plays a major role in how people choose to process information and as such has a significant impact on how individuals use computers to make decisions [62, 74-76]. Because computers can augment an individual's information processing capacity, it is often assumed that they can help people improve their decisions. However, studies report that this is not always the case. Studies show that people often use computers in a way to reduce their effort rather than to maximize their accuracy [74-76]. Such technology usage behavior is not due to inherent laziness or indifference; whereas accurate, rational decisions are the intention, due to limited cognitive capacity people naturally try to conserve their cognitive resources [17,62,71,74-76].

Because of the significant impact of task demand on technology usage behavior [74-76], the development of systems that can detect task demand automatically provides excellent opportunities for addressing user needs more effectively. For example, a decision tool enabled with an automatic task load detection system can provide feedback or suggestions for the user to help ease cognitive effort, or more effectively use limited cognitive resources, thereby helping the user to improve his or her decision accuracy [6,36,68,73].

We argue that an eye tracking machine learning system is an effective way to detect task demand automatically. The integration of eye tracking and machine learning technologies for developing such task load detection systems has several important advantages. Because eye movements reflect how people visually inspect stimuli, and because vision is our most dominant sense [4,18], eye tracking provides a natural method for collecting information about task demand. Eye trackers collect eye movements

continuously, thereby providing a moment-by-moment picture of behavior [30]. Modern remote eye trackers are integrated into monitors, or can be easily attached to such visual displays. Hence, they can collect eye movement behavior unobtrusively, without requiring any additional step or the need for users to wear special gear [38,20,64]. Moreover, as eye tracking technology matures, high-quality remote eye trackers become increasingly affordable [20]. This in turn, not only makes developing eye tracking task load detection systems possible, but also cost effective.

Combining machine learning with eye tracking to design task load detection systems holds promise not only for a dynamic and flexible mechanism for detecting task demand, but also one that is easily scalable. The advent of modern machine learning approaches carries the promise of discovering meaningful insights even on data sets of massive size. Because machine learning models can generalize from a given set of data [24], advanced machine learning eye tracking systems will only improve over time, as the user's gaze data set grows every time they are used.

In the following sections we establish the framework for our research by providing a brief review of relevant theory and literature. We subsequently form a hypothesis to assert that it is possible to develop an unobtrusive machine learning task detection system using only user eye movements. We then discuss the methodology that we used to design and test our proposed eye tracking task load detection system. Finally, we report the results obtained from our investigations and discuss their implication for research and practice.

2 Theoretical Background

To highlight the need for task load detection tools, we start by a brief discussion of the adaptive decision making theory which asserts that by influencing information processing behavior, task demand has a significant impact on decision outcome [62]. We use this theory as the framework for developing an eye tracking task load detection system [30]. In particular, we argue that a user's cognitive effort, attempting to

meet task demand, is likely to be reflected in his or her eye movements and this information will be distinct enough to be detected automatically by a machine learning system.

To support the suitability of eye movement for building a machine learning task load detector, we review relevant eye tracking studies that suggest eye movements can reveal information about cognitive effort. We also discuss relevant machine learning studies that have used eye movements to predict various user behaviors.

2.1 Adaptive Decision Making

Because cognitive effort (attention) is a scarce resource, human information processing behavior is strongly influenced by the demands of the task at hand [62,71]. Due to the scarce attentional capacity, people tend to intelligently choose an information processing behavior that can effectively meet the task demand while conserving valuable cognitive resources. After all, when cognitive capacity is exhausted, the ability to make good decisions is seriously compromised, and making good decisions is what people typically strive to do [61]. Supporting this point of view, literature in judgment and decision making provides ample evidence that task demand plays a significant role in how people go about solving a problem [62]. When making decisions, people often use a diverse set of information processing strategies, which differ in how long they take to execute, how much demand they put on cognitive resources, and how accurate their results would be. When cognitive demand of a decision task is reasonable, people use more effortful strategies (e.g., normative strategies), which typically yield better results. When cognitive demand is high, people use less effortful strategies to save their cognitive resources. For example, when the task requires people to consider only two alternatives, they tend to use an effortful compensatory decision strategy, which allows them to evaluate whether good values on some attributes can compensate for bad values on others. However, when the task requires people to consider more than two alternatives, they tend to resort to a less effortful non-compensatory strategy [11,60,61]. This pattern of behavior is also present in children [48]. Additionally, task demand can force people to use multiple strategies when completing the same task. For example, when

faced with a time limit, people first try to accommodate task demand by increasing their processing speed. If this approach is not sufficient to meet the time constraint, they tend to save time by filtering information. If these changes are still not sufficient to meet the time demand, people would then switch to a decision strategy that would take less time to execute, such as heuristics [61]. Naturally, tasks with higher demands will require greater cognitive effort [28].

2.2 Eye Movement Behavior and Cognitive Effort

We predominantly use our visual system to collect information from our environment, hence eye tracking provides an excellent tool for examining how people attend to and process information [20]. Not surprisingly, eye tracking is becoming increasingly popular in investigating user experience and technology usage behavior [3,51]. Gaze serves as a reliable indicator of attention, and thus it can reflect cognitive effort [10,64,67]. Grounded in eye tracking literature, we discuss four major eye movement behaviors (fixations, saccades, blinks, and pupillary responses) that are likely to reveal distinct information about cognitive effort in response to task demand. Table 1 provides a summary of the eye movement behaviors and their respective parameters that are discussed in this section.

2.2.1 Fixation

Fixation refers to a collection of relatively stable gaze points that are near in both spatial and temporal proximity. During fixation, the eyes hold steady on an object, and thus fixation reflects attention to a stimulus [38]. A number of studies have associated fixation-related metrics to cognitive effort [23,34,44]. For example, the number of fixations within an area of investigation (AOI) has been used to compare cognitive effort of millennials and baby boomers when viewing a web page [22]. How frequently people fixate on an object has also been used to assess cognitive effort in business to consumer (B2C) transactional processes, when an option must be selected prior to continuing with the transaction [37]. Additionally, the number of fixations has been shown to strongly correlate with task performance [78]. Because task performance is also correlated with effort expenditure [62], this result suggests a link between fixation

frequency and cognitive effort. Similarly, fixation duration, or the amount of time a user looks at stimuli, can be used to measure effort. To attend to a stimulus or an object, the user has to expend effort to maintain a steady gaze on the object [20]. Moreover, studies provide evidence that fixation duration increases as information processing becomes more effortful [34,57,78].

2.2.2 Saccade

Saccades refer to small, rapid eye movements when jumping from fixating on one object to another [31]. While visual information is not processed during saccadic eye movements [20], they still can provide information about viewing behavior [38,41]. For example, people tend to exhibit more saccadic eye movements when reading long pseudo-words [19]. Similarly, saccade amplitude, or the path traveled by a saccade between two consecutive fixations, tends to increase when reading longer words [19]. When interacting with an online resource, longer saccadic amplitudes can reflect whether users have become familiar with an interface. Having a better internal representation of an interface allows users to move their eyes directly to a desired location on the screen, hence producing longer saccadic amplitudes [31]. Consistent with this point of view, difficulty in locating information when browsing a webpage is likely to impact the duration of saccades. According to the theory of visual hierarchy [29], a stimulus is inspected by scanning it through a sequence of visual entry points. Each entry point acts like an anchor, which allows the user to scan for information around it. According to this perspective, longer duration of saccadic eye movements could indicate increased cognitive effort in finding a suitable entry point into a visual display [20].

2.2.3 Blink

Blinks are the involuntary act of shutting and opening the eyelids. They are known to reflect changes in attention and thus they are likely to reflect an individual's cognitive effort [64,78]. In particular, fewer blinks have been associated with increased attention [53]. For example, a study shows that surgeons had a lower number of blinks when performing surgery as compared to when they were engaged in casual

conversations [79]. In addition to the number of blinks, the duration of blinks can also indicate cognitive effort. For example, shorter blink durations were associated with increased visual workload during a traffic simulation task [2]. Similarly, comparing blink data during a hard (math problem solving) and easy task (listening to relaxing music), people exhibited shorter blink durations during the hard task [4]. Because of its observed association with cognitive effort, blink duration has been used to assess mental effort in educational games [39]. The above studies suggest that people often exhibit fewer or shorter blinks during more challenging tasks because they want to minimize missing visual information. After all, when the eyes are closed during a blink, there is no incoming visual information to process.

2.2.4 Pupillary Response

Changes in pupil size, which are controlled by the involuntary nervous system, can serve as a reliable proxy of mental effort [5,52,69]. For example, when people are asked to memorize numbers, retain them in memory, or perform multiplication, the size of their pupil seems to correlate with the difficulty of the task [7,45,63]. Similarly, variation in pupil size can also carry information about cognitive effort [14,21,30]. For example, the level of difficulty measured as the number of steps required to complete a task has been shown to impact pupil dilation variation. Increased cognitive load measured as implicit and explicit time limit also has a significant impact on pupil dilation variation. It is argued that pupil dilation variation is particularly effective in detecting the impact of complex decision tasks on users, because these tasks often involve a number of smaller subtasks. These subtasks are likely to require different types of mental activity with varying levels of difficulty. Consequently, complex decision tasks may result in variability in pupil size over the course of the task [14]. Another explanation for the suitability of pupil dilation variation in measuring cognitive load is rooted in the adaptive decision making theory which asserts people often switch their information processing strategies to conserve their limited cognitive resources. This flexibility in adjusting to the decision environment, which involves balancing one's cognitive load, is likely to be detected by the variation in pupil dilation [30].

Table 1. Eye Movement Behaviors and Parameters for Measuring Cognitive Effort

Behavior	Parameter	Source
Fixation: Relatively stable gaze points that are close in proximity and time	Fixation number	[22,23,37,78]
	Fixation duration	[20,34,44,57,78]
Saccade: Rapid eye movements between fixations	Saccade number	[19]
	Saccade duration	[20]
	Saccade amplitude	[31,19]
Blink: Involuntary act of shutting and opening the eyelids	Blink number	[53,64,78,79]
	Blink duration	[4,2,39]
Pupillary Response: Changes in pupil	Pupil dilation	[5,7,30,45,63,69]
	Pupil dilation variation	[14,21,30]

2.3 Eye Tracking and Machine Learning

As discussed in the previous section, eye tracking studies provide ample evidence that certain eye movement behaviors (i.e., fixations, saccades, blinks, and pupillary responses) have the potential to reveal information about cognitive effort. We argue that eye movement behaviors are distinct enough to serve as a suitable input for designing machine learning systems. In this section we discuss a number of relevant machine learning studies that have successfully used eye movement data to predict a variety of different behaviors. Because we use classification to design our proposed task load detection system, we focus on those studies that use supervised classification to predict categorical responses from eye movements.

Using eye movement data, a classification approach was used to predict how well people would solve a puzzle with approximately 53% accuracy [27]. In addition to predicting task performance, classification has been used to predict user intention from their eye movement data [8]. The authors developed a classification system to predict whether study participants intended to give a command to a gaze-based interface. Another study used classification from eye movement data of people collaborating on building concept maps to distinguish expert participants from novice participants [54].

Klami et al. [47] used a classification approach to predict from the eye movement data whether the retrieved images in a visual search task were relevant to the search terms used. Simola et al. [70] used classification

to predict whether a user is searching for a word, answering a question, or looking up the most interesting title in a given list from user's eye movement. Marshal [55] examined the states of relaxed and engaged users in the context of problem solving using two different statistical models. Kardan and Conti [46] classified students' performance with 71% accuracy using eye movement data. Henderson et al. [35] used classification to identify different visual activities (e.g., scene search, scene memorization, reading) using ocular events. Najar et al. [59] used eye movements to classify novice vs. advanced learners.

Steichen et al. [72] used classification to predict visualization task properties, performance on such tasks, and user cognitive abilities (visual and verbal working memory, perceptual speed) using basic eye movement features. Borji et al. [12] used classification to decode observer performance for estimating the ages of people shown in a picture from their eye movements. Finally, Krol and Krol [49] used eye movements to classify different decision making tasks.

Table 2 lists the studies that were reviewed in this section. It also provides information about eye movement features that were used in these studies. This list shows that eye movement data has been successfully used to detect a variety of behaviors with machine learning, thus providing support for the feasibility of gaze in developing a reliable task load detection system.

Table 2. Eye-tracking Machine Learning Classifiers

Eye-Movement Metrics (Features)	Source
Fixation count and fixation duration	[54]
Fixation duration, total fixation duration, fixation count, visit duration	[59]
Mean and standard deviation of fixation duration, mean and standard deviation of saccade amplitude, number of fixations per trial.	[35]
Fixation rate, number of fixations and fixation duration, saccade length, relative saccades angle and absolute saccade angle	[46]
Fixation map and histogram of scan path, fixation count, mean fixation duration, mean saccade amplitude	[12]
fixation rate, number of fixations, fixation duration, saccade amplitude, relative saccade angles, absolute saccade angles	[72]

Total and average duration of fixations, and fixation count	[47]
Fixation count, mean and standard deviation of fixation duration, mean and standard deviation of saccade amplitude and saccade direction	[70]
Fixation duration, saccade amplitude, fixation count, fixation rate	[27]
Saccade amplitude, saccade duration, saccade velocity, and saccade acceleration	[8]
Pupil size and point-of-gaze	[55]
Pupil dilation and gaze dispersion	[49]

2.4 Hypothesis

Modern remote eye tracking devices allow us to collect information about user gaze unobtrusively and seamlessly (e.g., 60, 120, or 300 samples per second) [38]. The inherently rich and vast amount of eye movement signals collected for a user have been shown to provide suitable information for developing predictive machine learning systems (Table 2). Because task demand forces decision makers to adjust their effort [62], and because eye movements have the potential to carry information about effort (Table 2), we argue that it is possible to develop a machine learning system using only eye movement data that can automatically and reliably detect task demand:

Hypothesis: Our proposed eye tracking task load detection system can reliably identify task demand.

3 Methodology

To investigate our hypothesis, we developed algorithms for designing and testing our proposed eye tracking task load detection system. In the following sections we explain our process in details.

3.1 Designing the Eye Tracking Task Load Detection System

To design our eye tracking task load detection system, we developed an algorithm to solve a classification problem. Classification refers to the process of identifying the correct category for a new piece of information based on prior observations. In this case, we were interested in developing a classifier for our system that could identify whether eye movements were collected under lower or higher level of task

demand. The design of our eye tracking task load detection system required three major steps. In the first step we developed an eye tracking feature set, or set of eye movement metrics, that based on the aforementioned literature had the potential to detect task demand. In the second step, we used adaptive decision making theory to select an algorithm for designing and testing our task load classifier. In the third and final step, we conducted an eye tracking study to capture and prepare eye movement data to implement and test our proposed task load detection system. In the following sections we explain how we completed each step.

3.1.1 Step 1: Developing a Set of Eye Movement Metrics (Feature Set)

We started this step by constructing a set of eye movement parameters that based on the literature reviewed in this paper was most likely to reveal cognitive effort (see Table 1). Machine learning feature sets are often developed using statistical properties of fundamental parameters. Hence, we expanded our feature set by including basic statistical properties, such as mean and standard deviation, for each of the parameters listed in Table 1. Recently, pupil data during the saccadic and fixation events has been shown to differ [21], thus we considered pupil data for fixations and saccades separately. In addition to average duration values for saccades, fixations, and blinks, we also considered their normalized duration metrics. Normalized metrics are obtained by dividing the total duration of each parameter by the total task completion time. Additionally, we included certain ratios for eye movement behaviors (in particular a new set of pupillometry ratios) that could provide additional insight. For example, the ratio of saccades to fixations reveal the amount of time spent searching for information, versus the amount of time spent on processing the information visually [20]. This in turn can provide insight about cognitive effort. Together, the feature set for our proposed task load detection system consisted of thirty different eye metrics. This feature set is displayed in Table 3.

Table 3. Feature Set: List of Eye Movement Metrics for the Task Load Detection System

Eye Movement	Eye Movement Metrics (Features)
--------------	---------------------------------

Fixation	Average fixation duration (millisecond) Standard deviation of fixation duration Normalized fixation number (fixation number/task completion time) Normalized fixation duration (total fixation duration/task completion time)
Saccade	Average saccade duration (millisecond) Standard deviation of saccade duration Average saccade amplitude (degree) Standard deviation of saccade amplitude Normalized saccade number (saccade number/task completion time) Normalized saccade duration (total saccade duration/task completion time)
Blink	Average blink duration (millisecond) Standard deviation of blink duration Normalized blink number (blink number/task completion time) Normalized blink duration (total blink duration/task completion time)
Pupil Dilation	Average pupil dilation (PD) during fixation (millimeter) Standard deviation of PD during fixation Average pupil dilation variation (PDV) during fixation Standard deviation of PDV during fixation Average PD during saccade (millimeter) Standard deviation of PD during saccade Average PDV during saccade Standard deviation of PDV during saccade
Eye Movement Ratios	Average (PD during saccade/PD during fixation) Standard deviation (PD during saccade/PD during fixation) Average (saccade duration/fixation duration) Standard deviation (saccade duration/fixation duration) Average (PDV during saccade/PDV during fixation) Standard deviation (PDV during saccade/PDV during fixation) Normalized saccade duration/normalized fixation duration Normalized saccade number/normalized fixation number

3.1.2 Step 2: Selecting an Algorithm

Classification algorithms are typically selected based on the complexity of the problem at hand. The purpose of our task load detection system is to identify task demand operationalized as effort expenditure captured by user eye movements during the decision-making process. According to adaptive decision making theory [62], effort expenditure resulting from the attempt to balance the conflict between maximizing accuracy and minimizing effort using various information processing strategies is highly

contingent upon task conditions. For example, when task demand is high people are likely to switch between multiple strategies to meet the task demand (e.g., they may increase their processing speed, use less information, and/or switch to a less demanding strategy such as heuristics) [61]. Such flexibility in decision behavior suggests the need for an algorithm that is suited for processing complex models. Because of its ability to identify complex boundaries in predictive models, we selected the Random Forest (RF) framework to develop our classifier.

The random forest algorithm solves a classification problem by creating several individual models, or trees, using bootstrapping [33]. Individual trees are developed by randomly selecting sub-samples from the original dataset. Each individual tree is a type of classifier that uses the divide-and-conquer methodology combined with bootstrapping. Individual trees are considered weak learners in the random forest framework. The algorithm generates a strong learner by combining the weak learners into a single overall tree that can produce more accurate results than any of the weak learners [33].

Figure 1a displays the bootstrapping algorithm that we designed for our random forest classifier. Our bootstrapping methodology causes each sample to appear exactly 200 times in the computation. Each data point is taken with equal probability, hence some of the samples may appear several times in the bootstrap set and others not at all. Consistent with prior research, we use 200 number of bootstrap replications [25]. A very large bootstrap replication is not suggested as it results in a computational burden.

Our eye tracking classifier, which is designed to identify whether eye movements are captured during lower or higher level of task demand, requires two distinct phases. In the first phase, the classifier is trained with a set of (eye movements, task condition) data. During this training phase the system has access to both the collected eye movement data as well as the task condition under which the data is collected. The second phase is the test phase, which assesses the success of the training phase. With a successful training, the

system will be able to take as input a new set of eye-movement data only (without information about task condition) and reliably detect the task condition under which the eye movement data was collected.

The dataset for our random forest classifier, which is generated by bootstrapping, is divided into two distinct “training” and “test” sets (80% and 20%, respectively) [33]. The training dataset is used in the training phase to train the classifier and the test dataset is used in the test phase to assess the performance of the trained classifier. The performance assessment in the test phase is achieved by measuring the level of error in answering questions about the task condition on the test data. The test phase in our algorithm uses resampling methods (bootstrapping) to estimate the generalization error of the classifier [9,26,65]. As shown in Figure 1b, each tree (RF_i) is trained with a bootstrapping sample (training data) and tested with the remaining data in the original set (test data). The accuracy of the classifier is then measured by comparing the output of each individual tree with the task condition of its test data. If there is a match, the error variable for that particular subtree is set to 0, or 1 otherwise. The average error value for the subtrees represents the generalized error for the random forest classifier.

1. Initialization

1.1. Set number of replications $i = 200$

2. Training and Test

2.1. At random, generate training sets out of the feature matrix dataset and use these for training the untrained classifier. Training set generation is done "with replacement".

2.2. The resulting trained classifiers are tested on the corresponding test data.

2.3. Repeat this procedure i times.

3. Classifier Accuracy

3.1. Compute the classification error at each replicate.

3.2. Calculate the bootstrapping generalized error by averaging over the errors of all i classifiers.

1a. Bootstrapping Algorithm

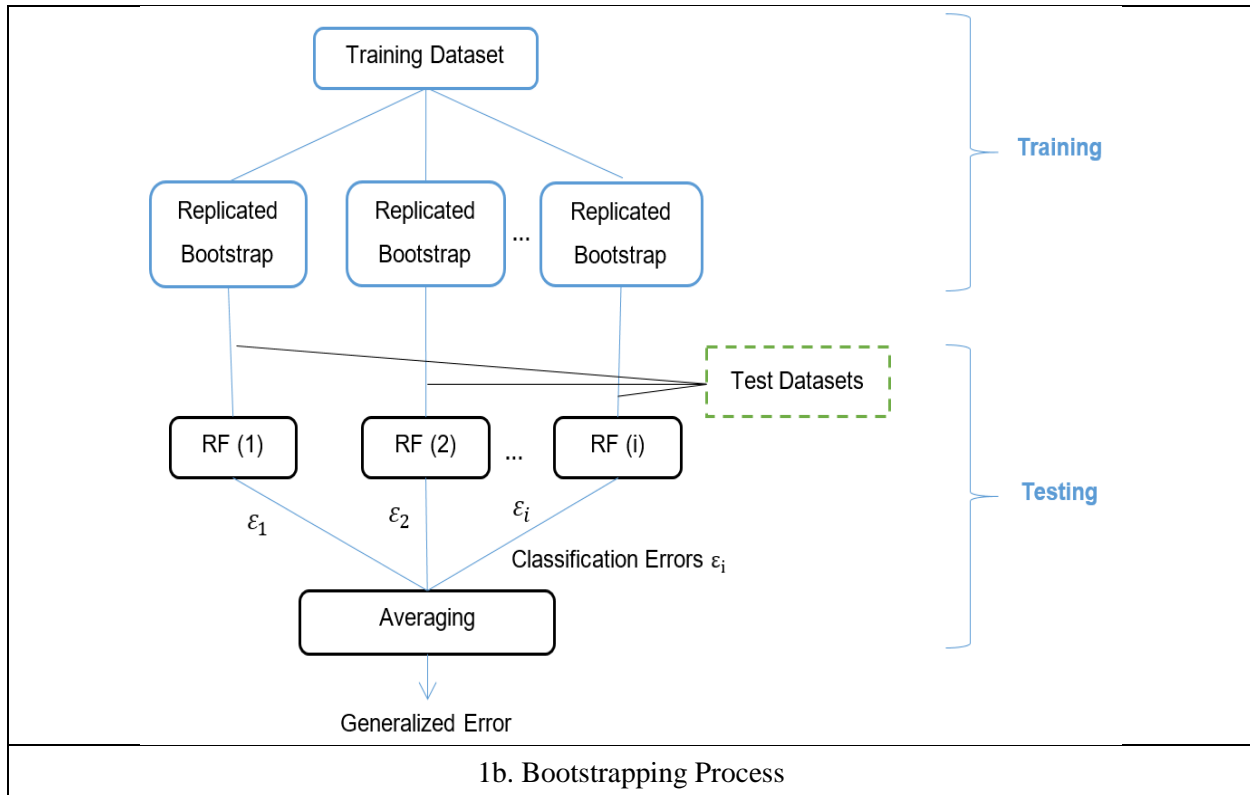


Figure 1. Bootstrapping algorithm and process

As mentioned earlier, we chose the random forest methodology to detect task load because random forest is commonly used for modeling complex behaviors (in our case classifying eye movements resulting from users' attempt to meet a higher or lower task demand). In addition to the ability to detect complex boundaries, random forests are particularly effective for eye movement datasets. Eye tracking studies typically provide rich data from a limited number of eye movement recordings. For example, as we explain in the next section, our eye tracking experiment produced rich information (30 features shown in Table 2) from 48 eye movement recordings (participants). This data is a good candidate for random forest because random forests can process the large sets of features without having to reduce the selected variables to a manageable set and because bootstrapping in random forest can address the smaller sample size that is typical in eye tracking studies. Furthermore, random forests are robust even when the data is not normally

distributed. The fast runtime of random forest is yet another attractive factor in developing a practical eye tracking task detection system [33,16].

3.2 Preparing Data Sets

To capture and prepare eye movement datasets for implementing and testing our proposed task load detection design, we conducted an eye tracking study, which required participants to complete a cognitively complex problem solving task under two different task treatments. It is well-known that a time constraint increases the use of cognitive resource by making problem-solving tasks more demanding [62]. We created the two different task treatments in our study by manipulating the time available for completing the task. In the control treatment no time limit was enforced, while in the experimental group the time available for completing the task was set to five minutes. This allowed us to create lower task demand in the control treatment and higher task demand in the experimental treatment. Participants were randomly assigned to either the control or experimental group. Participants in both groups completed the same problem-solving task, however, in the experimental group participants had to complete the task within five minutes, while in the control group they could take as long as they wished to complete the task.

The task in our study required participants to answer a set of ten mathematical questions. This set of questions were selected from a pool of problem-solving practice tests for the Graduate Record Examination (GRE), which is a standardized test required for admission to most graduate degree programs in the United States. The full set of these practice questions were retrieved from www.majortests.com.

Because we used GRE math problems for our problem-solving task, we recruited participants via email from a pool of graduate students in various technical disciplines (e.g., computer science, electrical and computer engineering, robotics engineering, etc.) in a northeastern university in the United States. Because these technical disciplines require GRE math for admission to the program, all the participants in our study had the math skillset that is typically required of graduate students in technical disciplines. Because students

are accustomed to taking timed tests, the task and setting created an appropriate and realistic environment for our participants.

We used the Tobii X300 remote eye tracker with a sampling rate of 300 Hz mounted on a 21-inch monitor at a resolution of 1920 x 1200 to collect the gaze data. To track eye movements, each participant completed a brief eye-calibration process. While seated, participants were asked to observe a moving dot on the eye-tracking monitor. This calibration process took less than one minute to complete.

Next we analyzed the captured eye movement data for quality. This process was completed by examining the quality of eye movement recordings and removing the data sets for those participants that had less than 80% gaze sample [15,50]. The gaze sample refers to percentage of the times that eyes were correctly detected by the eye tracker for each participant. For example, 100% means that one or both eyes were detected by the device throughout the recording; 50% means that one eye or both eyes were found for half of the recording duration. While screen-based eye tracking experiments typically require users to look at the screen while completing a task, some people may look away or look down (e.g., at the keyboard or mouse) to think about a problem.

We removed the data for 7 participants who did not meet the 80% or above gaze sample criteria. Thus, the final dataset for implementing our task load detection system included the eye movement recordings that were captured from a total of 48 participants (21 females and 27 males) with ages ranging between 24 and 31.

To calculate the metrics for the feature set (Table 3), we first determined basic units from the eye movement recordings. We identified fixations and saccades in the gaze stream with the widely used I-VT filter with 30°/sec saccadic velocity threshold provided in the Tobii Studio software version 3.2.3. Saccade amplitude (the distance traveled between two adjacent fixations), measured in degrees, as well as pupil dilation (size of pupil diameter) was also provided by the Tobii Studio software. Pupil Dilation Variation (PDV) or rate

of change of pupil dilation was calculated by taking the temporal derivative of pupil dilation [40,77]. Blinks were calculated as complete eye closure lasting between 100-500 milliseconds [1]. These basic units were then used to calculate the metrics in the feature set (Table 3).

4 Results

The task load detection system in our study was implemented in R version 3.4.2 on Windows 7, with Core i5 CPU and 3.30 GHz speed machine. We used R libraries such as ISLR [42], tree [66], random forest [13], e1071 [58], and caret [56].

A useful aspect of random forest is their ability to automatically establish the effectiveness of predictors in the feature set with respect to classification accuracy. Random forests can rank the importance of each metric based on its ability to predict the outcome by permuting each metric and computing the prediction accuracy of the out-of-bag portion of the data before, and after, the permutation [13]. The results of random forest variable ranking are displayed in Figure 2, highlighting the metrics ordered by *variable importance* [33,43].

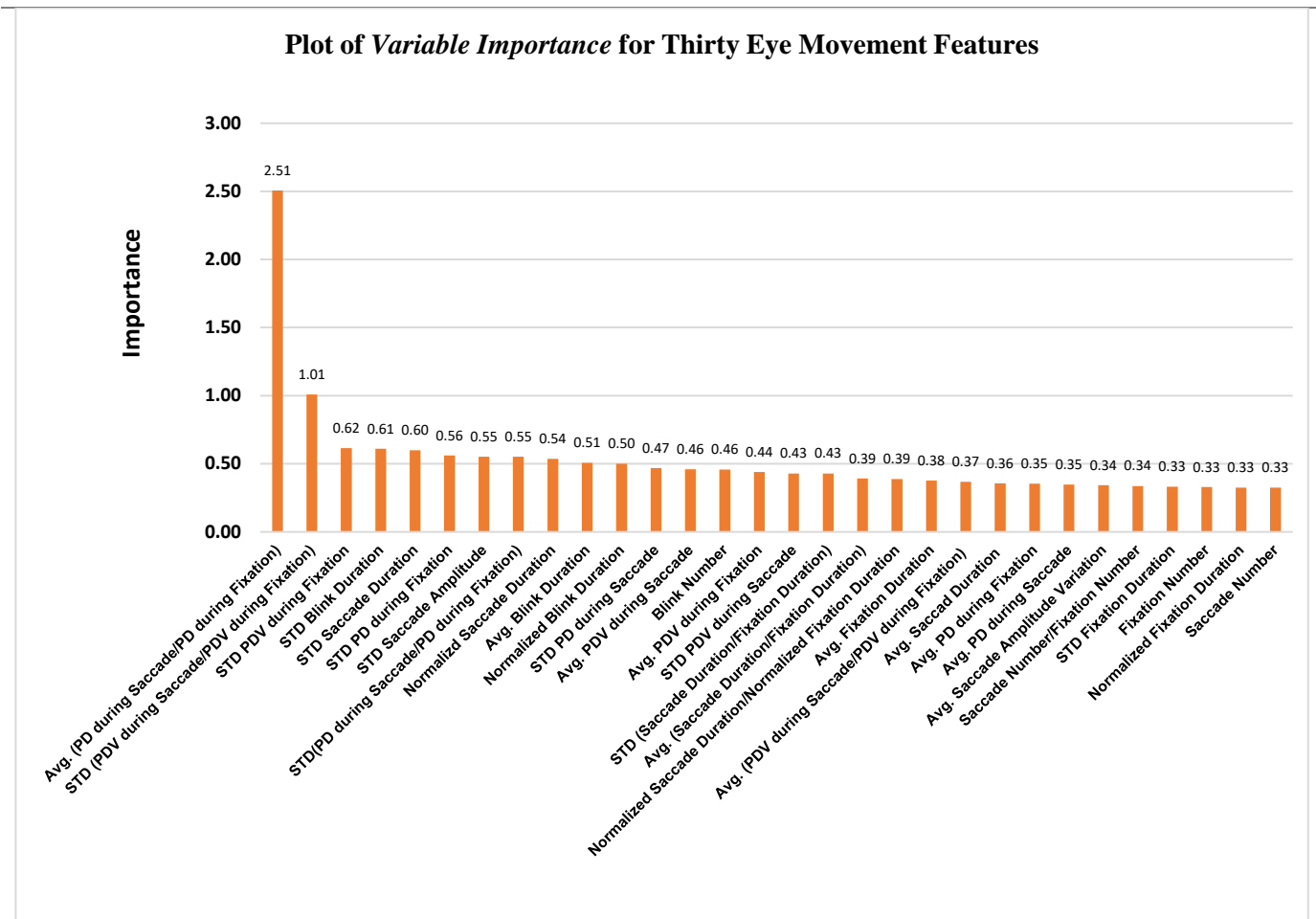


Figure 2. Variable importance plot

Next, we used the random forest variable ranking results to refine our feature set, that is, we selected those variables in our feature set that were sufficiently discriminative for our task load classifier [33]. To do this, we carried out a forward stepwise feature selection, systematically investigating the task demand detection accuracy of our random forest classifier by iteratively adding features based upon their *variable importance* [33]. This process resulted in a minimized error after adding the first ten features; additional features provided only marginal increases in the performance of detecting task demand. Accordingly, to avoid overfitting we selected only the first ten out of thirty features to develop our task load detection system [33]. These ten features are listed based on their order of importance in Table 4.

Table 4. List of features selected by variable importance

Rank	Eye Movement Metrics (Features)	Variable Importance
1.	Average (PD during saccade / PD during fixation)	2.51
2.	Standard deviation (PDV during saccade / PDV during fixation)	1.01
3.	Standard deviation of PDV during fixation	0.62
4.	Standard deviation of blink duration	0.61
5.	Standard deviation of saccade duration	0.60
6.	Standard deviation of PD during fixation	0.56
7.	Standard deviation of saccade amplitude	0.55
8.	Standard deviation (PD during saccade / PD during fixation)	0.55
9.	Normalized saccade duration	0.54
10.	Average blink duration	0.51

As apparent in Table 4, half of the top ten factors that were most effective in detecting task demand were related to pupil data: *Average saccade-to-fixation PD ratio, standard deviation of saccade-to-fixation PDV ratio, standard deviation of PDV during fixation, standard deviation of PD during fixation, standard deviation of saccade-to-fixation PD ratio*. These results support research linking pupil data and cognitive effort [7,14,21,30,45,63], as well as research advocating that valuable pupil information exists in both fixation and saccade data [21]. The ratio of pupil dilation and variation during saccades and fixations reflect the distribution of cognitive effort during information search and information processing. The distribution of effort between search and information processing, as suggested by our results, may provide valuable information about task demand.

Thirty percent of the remaining top 10 factors in our results were related to saccade parameters (*standard deviation of saccade duration, standard deviation of saccade amplitude, normalized saccade duration*),

while twenty percent were related to blink patterns (*standard deviation of blink duration, average blink duration*). These results suggest that saccade and blink eye movements had a major influence in effective classification of the eye movement data based on task demand. Hence, these results not only support the literature that indicates saccades and blinks are associated with cognitive effort, but also show that the metrics related to saccades and blinks were among most effective variables for detecting task demand.

Interestingly enough, the results did not indicate fixation parameters, such as fixation duration and number, to be major contributors to classifying task demand. This contrasts with previous research that shows a positive link between fixation duration and cognitive effort – the very nature of viewing a stimulus requires effort in keeping the gaze steady for the information to be visually processed. While fixation serves as a reliable and direct indicator of attention and thus information processing, our results indicate that more effective in classifying task demand were the saccade and blink eye movement behaviors, which take place between, and not during, fixations.

Perhaps most interesting among our results is that pupil dilation ratio values involving saccades and fixations played a major role in classifying higher/lower task demand (Table 4). In particular, the *variable importance* for *average saccade-to-fixation PD ratio* was noticeably larger than all other metrics. The importance of the average saccade-to-fixation PD ratio was more than twice as large as the standard deviation of saccade-to-fixation PDV ratio and over four times as large as the rest of the factors.

The results of variable ranking discussed above support extant literature summarized in Table 1, and also extend previous findings by showing that only pupil, saccade, and blink related data were major predictors in classifying task demand in our study. Further, average saccade-to-fixation PD ratio appears to be far more important than the rest of the feature set.

The random forest algorithm can be used to develop different sets of forests that have varying numbers of trees. To find the number of trees that correspond to a stable classifier, we constructed random forests with

the number of tree values in the range [1,100], and with 200 replications of bootstrapping. The optimal number of trees for our classifier was determined via a standard technique having to do with individual tree error rates, namely, the out of bag error rates [33]. When the error rates stabilize and reach a minimum value, the corresponding number of trees constitute the optimal number of trees. The accuracy rate of our classifier, as typical during this process, initially increased as the number of trees increased. However, once the number of trees reached approximately fifteen, the performance of the model stabilized and corresponded to an eye movement classifier with 69.6% accuracy. These results show that our proposed model can detect task demand using eye movements not only reliably but also quickly (with 15 trees).

Of course, one might wonder how such results could be improved. The stability of the results after applying fifteen trees indicates that additional computational effort will likely not improve our results beyond those already achieved for our *fixed model* and *fixed data set*. As far as the *model* is concerned, one could imagine the application of a more sophisticated or customized model giving superior results. On the other hand, overfitting is always a concern, and random forests were intentionally selected in our study for their broad applicability to complex problems. As far as the *data* is concerned, additional and more detailed measurements would likely increase performance. It is precisely our goal to pursue such improved data generation in future work.

We extended the above analysis by generating a confusion matrix and an ROC curve to investigate the performance of our classification algorithm. The confusion matrix represents the true positive, true negative, false positive and false negative of the classification task. The ROC curve shows a trade-off between (true positive rate) sensitivity and (false positive rate) specificity and is a measure of test accuracy [80]. Both the confusion matrix and the ROC curve for 15 trees are presented in Figure 3. According to this analysis, the accuracy of detecting task load is 75%, which is calculated as the sum of true positives and true negatives divided by the total number of test samples (20).

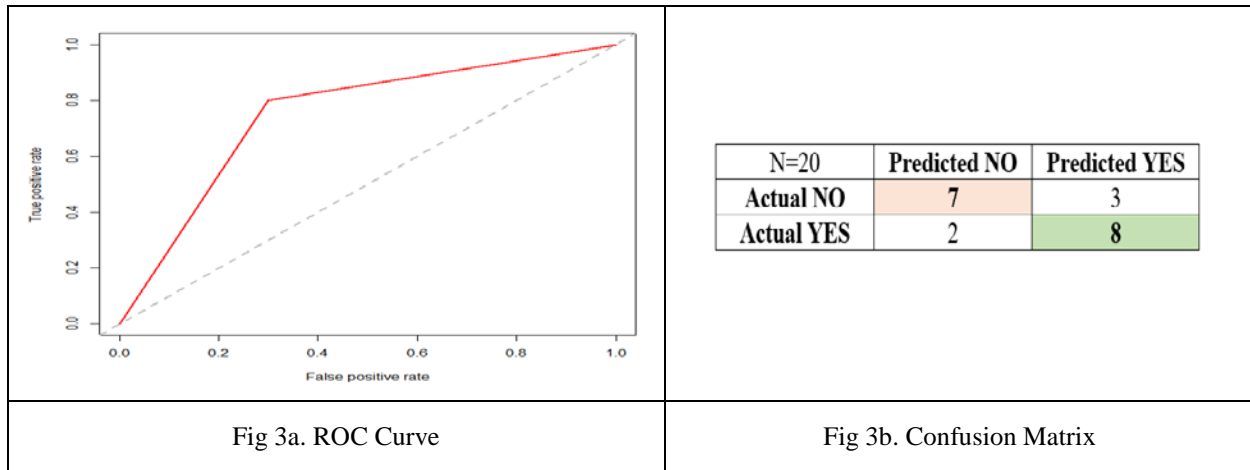


Figure 3. ROC curve and confusion matrix for 15 trees and 20 test samples

Because people tend to exhibit complex behavior such as switching between multiple strategies when making decisions, we argued that classifying task demand is likely to require an algorithm that can process complex models. Hence, we used random forests to build our classifier. The relatively high accuracy level achieved by our classifier displayed in Figure 1 suggests that using the random forest algorithmic approach in our study was indeed a good choice. To further investigate the appropriateness of random forests for developing a task load detection system we compared its performance against another set of widely used machine learning classifiers, namely linear and kernel-based Support Vector Machine (SVM) classifiers. As shown in Table 5, the linear or nonlinear SVM classifiers reached much lower accuracy levels (41% to 56%) compared to the accuracy level of our proposed random forest model (69.6%). These results show that random forest was a more suitable algorithm for classification of task demand (manipulated by time limit) based on the eye-movement data in our study.

Table 5. Support vector machine classification performance

Algorithm	Accuracy
Linear SVM	56%
Nonlinear SVM with radial basis kernel	48%
Nonlinear SVM with polynomial degree of 2	43%
Nonlinear SVM with polynomial degree of 3	41%

The results of variable importance analysis showed that half of the variables among selected features were related to pupil data (see Table 4). To further investigate the effect of each category of features based on eye-movement metrics (e.g. pupil dilation, blinks, fixation, and saccade), we created 6 different categories. Next we trained 6 different RF models with each of these different feature sets to investigate the classification performance for each category.

Table 6 presents the performance results. Interestingly, the highest accuracy (79%) was achieved from saccade-to-fixation PD and PDV ratios. The second column in Table 6 shows the features listed based on their importance order according to RF *Variable Importance* values. It is important to note that similar to when we used all the 30 features, the most effective features in the classification is average saccade-to-fixation PD ratio.

The above results together support our hypothesis by showing that our proposed task load detection system identified task demand reliably and unobtrusively. The results support our choice of algorithm for developing the task load detection system and show that eye movement data carries distinct information about task demand. Pupillary responses were more effective than other eye moment behaviors in detecting task demand in our study. In particular, saccade-to-fixation pupil dilation and pupil variation ratios, which were designed for the first time in our study, proved to be most valuable in detecting task demand (79% accuracy).

Table 6. RF classification performance using different categories of eye features

Feature Categories	Features	Accuracy
PD and PDV (Only Ratios)	1. Average saccade-to-fixation PD ratio, 2. Standard deviation of saccade-to-fixation PDV ratio, 3. Standard deviation of saccade-to-fixation PD ratio, 4. Average saccade-to-fixation PDV ratio	79%
All PD and PDV Features	1. Average saccade-to-fixation PD ratio, 2. Standard deviation of saccade-to-fixation PDV ratio, 3. Standard deviation of PDV during fixation, 4. Standard deviation (saccade-to-fixation PD ratio), 5. Average PDV during fixation, 6. Standard deviation of PD during fixation, 7. Average PDV during saccade, 8. Average saccade-to-fixation PDV ratio, 9. Standard Deviation of PD during saccade, 10. Standard deviation of PDV during saccade, 11. Average PD during saccade, 12. Average PD during fixation	70%
Blink Features	1. Standard deviation of blink duration, 2. Normalized blink duration, 3. Average blink duration, 4. Blink number	52%
Saccade Features	1. Standard deviation of saccade duration, 2. Normalized saccade duration, 3. Standard deviation of saccade amplitude, 4. Average saccade duration, 5. Average saccade amplitude, 6. Normalized saccade number	51%
Ratio of Saccade Features to Fixation Features	1. Standard deviation (saccade duration/fixation duration), 2. Average (saccade duration/fixation duration), 3. Normalized saccade duration/normalized fixation duration, 4. Normalized saccade number/normalized fixation number	43%
Fixation Features	1. Average fixation duration, 2. Normalized fixation duration, 3. Normalized fixation number, 4. Standard deviation of fixation duration	40 %

5 Discussion

Grounded in adaptive decision making theory, we argued the effort to meet task demand is likely to be reflected in eye movements. Using eye tracking literature, we argued that eye movement data is distinct enough to build a machine learning system that can automatically detect task demand. To test our assertion, we developed and tested an eye tracking task load detection system.

Our results align with our initial expectations and have important implications for designing advanced eye tracking task load detection systems. Our results suggest that combining eye tracking and machine learning

technology produces a wealth of information that is likely to help build unobtrusive detection systems that can identify changes in user behavior. This in turn will provide attractive opportunities for designing intelligent decision tools that can respond to user needs at a personalized level. The increasing availability of high quality eye trackers at affordable prices [20] makes it possible and practical to include eye tracking task load detectors into decision support systems. By recognizing the relative task demand via task load detectors, such decision support systems can respond to user needs more fully and thus provide a more effective and efficient human-technology collaboration in complex domains [30].

Decision support systems enhanced with task load detectors can be particularly effective in training novice decision makers through various feedback mechanisms that are triggered by their eye tracking sensors. For example, such advanced decision support systems can provide recommendations for the use of decision strategies that are best at optimizing accuracy at the given level of task demand recognized by their task load detector. The results of our study showing the effectiveness of eye movements to detect task demand reliably and unobtrusively provides motivation for future investigation of eye tracking task load detection systems. Our positive results suggest that eye tracking task load detectors are likely to build a productive line of research in decision support systems.

Our results showed that pupillometry measures can serve as effective eye tracking metrics for designing task load detectors. Many studies have shown that pupil data is a reliable predictor of cognitive load (e.g., see Table 1). A novel contribution of our study is that it not only supports this previous finding, but also refines it by showing that pupil data was the most prominent predictive factor in our set of thirty eye movement features (see Tables 6, and Figure 2). Another novel contribution of our study was the introduction of pupillary ratio variables in the features set for our proposed classifier. As shown by our results, the saccade-to-fixation pupil dilation ratio was far more important than other features in detecting task demand, *perhaps even more important than the absolute pupil dilation reported in previous studies*. It is well-established that visual information is processed only during fixations. Upon focusing on an object,

the eye can only see vividly and colorfully around the fixation center. To compensate for this limitation, saccades are used to rapidly collect high quality visual information. Because saccades change the center of our attention, they represent information search [20]. Because pupil dilation is linked to cognitive activity, pupil dilation during saccade suggests cognitive activity related to information search and pupil dilation during fixation indicates cognitive activity related to information processing [21]. Thus, our results showing a significant role for the ratio of pupil dilation in saccades and fixations suggest that the ratio of cognitive activity during information search and information processing can provide invaluable insight for classifying task demand.

Another key insight of our study for future eye tracking task load detection systems is that, among the top ten discriminating features selected by the machine learning model, *none* were related solely to fixation. Fixations typically convey effort in visual processing [20]. In our study, however, metrics related to saccades and blinks were more important than metrics related to fixations. In particular, saccade duration and amplitude were among the top ten factors detecting task demand. Because saccades indicate effort in locating relevant information, our results suggest how long people took to locate a fixation and how far their eyes had to travel to locate that information provided more insight about task demand than data about their fixation. Similarly, our results demonstrate that average blink duration and variation were more effective than fixation-related information in detecting task demand. Blink duration has been associated with task complexity [2,4,39]. This is substantiated in our results. Average blink duration, and variation in blink duration, are likely indicating adjustment to task load, which according to adaptive decision making theory is what people do when making complex decisions [62].

Our study also makes important contributions to the judgment and decision making literature. By showing that pupillometry plays a major role in detecting task demand, our results support a recent exploratory DSS study [30] that suggests adaptive decision making theory can serve as a suitable framework for explaining the relationship between cognitive effort and pupillometry during problem solving and decision making.

Our results extend this previous research by using the adaptive decision making theory as a framework for our study and by providing evidence for its applicability to detect task demand at a physiological level (Tables 4 and 6).

6 Limitations and Future Studies

As with any study, our results are limited by the task context, which in our study was a math problem-solving task. Future studies using different tasks are needed to verify and extend our results. Similarly, the results are limited to a fairly static population, namely graduate students in an engineering school. A more diverse population may provide a deeper understanding of user behavior. While our machine learning algorithm was suitable for smaller datasets, larger sample sizes are likely to improve the accuracy of the proposed classifier and provide additional insight. The age of participants in our study ranged from 24 to 31; future studies including participants from a wider variety of age groups will further serve to confirm and extend our results.

Another limitation is the manipulation of task demand. In our study we used a time limit, a hallmark of today's fast-paced decision environments, to manipulate task demand. Nevertheless, future studies using other relevant task characteristics are needed to extend our results. For example, people often need to justify their decisions, which can increase an individual's cognitive effort [30]. Hence, future studies can use justification to manipulate task demand.

We used 30 eye metrics to develop our proposed classifier system. Using additional single or combined eye movement metrics (e.g., ratios) may provide a more nuanced understanding of user behavior. Similarly, including other physiological measures, such as heart rate variability, in the feature set of future studies may improve the sensitivity of the proposed classifier in detecting cognitive effort.

Our results show that eye movements features had different levels of importance in detecting task demand (e.g., Table 4 and Table 6). These results can motivate future research examining the development of a theoretically-derived taxonomy of the relationship between eye movements and cognitive demand.

7 Conclusion

Because users place a high value on conserving cognitive resources [32,62,74-76], developing computerized tools to help people manage their cognitive resources can help them be more effective in decision making. A first step in designing such advanced computerized tools is to investigate possibilities for developing systems that can identify level of task demand unobtrusively and automatically.

In this study, grounded in the adaptive decision making and eye tracking literature, we argued that task demand can be detected unobtrusively and automatically via eye movement data. We developed an eye tracking machine learning task load detection system to test our assertion. Our results showed that eye movements indeed carry distinct information about task demand and that pupil data, in particular the ratio of pupil dilation during saccades and fixations, was the most important predictor factor in identifying task demand. Our results showed that our task load detector can detect task demand quickly and reliably. These results show that building such an advanced task load detection system is not only possible but also computationally practical. Hence, the results provide valuable insights as well as motivation for future studies that focus on designing advanced task load detection systems.

8 References

- [1] H. Aarts, E. Bijleveld, R. Custers, M. Dogge, M. Deelder, D. Schutter, N.E.M. Haren, Positive Priming and Intentional Binding: Eye-Blink Rate Predicts Reward Information Effects on the Sense of Agency, *Social Neuroscience* 7(1) (2012) 105–112.
- [2] U. Ahlstrom, F.J. Friedman-Berg, Using Eye Movement Activity as a Correlate of Cognitive Workload, *International Journal of Industrial Ergonomics* 36(7) (2006) 623–636.
- [3] W. Albert, T. Tullis, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*, Newnes, 2013.
- [4] M. Andrzejewska, A. Stolińska, Comparing the Difficulty of Tasks Using Eye Tracking Combined with Subjective and Behavioural Criteria, *Journal of Eye Movement Research* 9(3) (2016) 1–16.
- [5] B. Bailey, S. Iqbal, Understanding Changes, In Mental Workload During Execution of Goal-Directed Tasks and Its Application for Interruption Management, *ACM Transactions on Computer-Human Interaction (TOCHI)* 14(4) (2008) 1–28.
- [6] R. Barkhi, E. Rolland, J. Butler, W. Fan, Decision Support System induced guidance for model formulation and solution, *Decision Support Systems* 40(2) (2005) 269–281.
- [7] J. Beatty, Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources, *Psychological Bulletin* 91(2) (1982) 276–292.
- [8] R. Bednarik, H. Vrzakova, M. Hradis, What Do You Want to Do Next: A Novel Approach for Intent Prediction in Gaze-Based Interaction, in: *Proceedings of the Symposium on Eye Track. Research and Applications* (2012) 83–90, ACM.
- [9] R. Beran, Introduction to Efron (1979) Bootstrap Methods: Another Look at the Jackknife, in: S. Kotz, N. Johnson (Eds.), *Breakthroughs in Statistics*, Springer New York, 1992: pp. 565–568. doi:10.1007/978-1-4612-4380-9_40.
- [10] J.R. Bergstrom, A. Schall, *Eye Tracking in User Experience Design*, Morgan Kaufmann Publishers Inc., 2014.
- [11] S.F. Biggs, J.C. Bedard, B.G. Gaber, T.J. Linsmeier, The Effects of Task Size and Similarity on the Decision Behavior of Bank Loan Officers, *Management Science* 31(8) (1985) 970–987.
- [12] A. Borji, L. Itti, Defending Yarbus: Eye Movements Reveal Observers’ Task, *Journal of Vision* 14(3) (2014) 1-22.
- [13] L. Breiman, Random Forests, *Machine Learning* 45(1) (2001) 5–32.
- [14] R. Buettner, S. Sauer, C. Maier, A. Eckhardt, Towards Ex Ante Prediction of User Performance: A Novel NeuroIS Methodology Based on Real-Time Measurement of Mental Effort, in: *Proceedings of the 48th Hawaii International Conference on System Sciences* (2015) 533–542.
- [15] K.S. Chiew, T.S. Braver, Temporal dynamics of motivation-cognitive control interactions revealed by high-resolution pupillometry, *Frontiers in Psychology* 4 (2013) 1-15.
- [16] A. Cutler, *Random Forests, a Statistical Tool for the Science*, Banff International Research Station for Mathematical Innovation and Discovery, <https://doi.org/10.14288/1.0368938>.
- [17] F.D. Davis, Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly* 13 (1989) 319–340.

- [18] S. Dähne, N. Wilbert, L. Wiskott, Slow Feature Analysis on Retinal Waves Leads to V1 Complex Cells, *PLoS Computational Biology* 10 (2014) e1003564.
- [19] M. De Luca, M. Borrelli, A. Judica, D. Spinelli, P. Zoccolotti, Reading Words and Pseudowords: An Eye Movement Study of Developmental Dyslexia, *Brain and Language* 80(3) (2002) 617–626.
- [20] S. Djamasbi, Eye Tracking and Web Experience, *AIS Transactions on Human-Computer Interaction* 6(2) (2014) 37–54.
- [21] S. Djamasbi, M. Shojaeizadeh, A. Trapp, Does Pupillary Data Differ During Fixations and Saccades? Does it Carry Information About Task Demand? in: 13th. Annual Workshop on HCI Research in MIS, Fort Worth, Texas, USA, 13 (2015).
- [22] S. Djamasbi, M. Siegel, J. Skorinko, T. Tullis, Online Viewing and Aesthetic Preferences of Generation Y and the Baby Boom Generation: Testing User Web Site Experience Through Eye Tracking, *International Journal of Electronic Commerce*, 15(4) (2011) 121-158.
- [23] S. Djamasbi, M. Siegel, T. Tullis, Visual Hierarchy and Viewing Behavior: An Eye Tracking Study, in: J. Jacko (Ed.), *Human-Computer Interaction, Design and Development Approaches*, Lecture Notes in Computer Science, 6761 (2011) 331–340, Springer Berlin Heidelberg.
- [24] P. Domingos, A Few Useful Things to Know About Machine Learning, *Communications of the ACM* 55 (10) (2012) 78–87.
- [25] B. Efron, R.J. Tibshirani, *An Introduction to the Bootstrap*, Monographs on Statistics & Applied Probability (1994) Chapman & Hall/CRC, Boca Raton, Florida.
- [26] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, (1979) 1–26. doi:10.1214/aos/1176344552.
- [27] S. Eivazi, R. Bednarik, Predicting Problem-Solving Behavior and Performance Levels from Visual Attention Data, in: *Proceedings of 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI 2011* (2011) 9-16.
- [28] M.J. Eppler, J. Mengis, The Concept of Information Overload - A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines, *The Information Society* 20(5) (2008) 325-344.
- [29] P. Faraday, Visually Critiquing Web Pages, in: N. Correia, T. Chambel, G. Davenport (Eds.), *Multimedia '99, Eurographics* (2000) 155-166, Springer Vienna, Vienna.
- [30] D.D. Fehrenbacher, S. Djamasbi, Information Systems and Task Demand: An Exploratory Pupillometry Study of Computerized Decision Making, *Decision Support Systems* 97 (2017) 1–11.
- [31] J. H. Goldberg, X.P. Kotval, Computer interface evaluation using eye movements: methods and constructs, *International Journal of Industrial Ergonomics* 24(6) (1999) 631–645.
- [32] S. Gregor, I. Benbasat, Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice, *MIS Quarterly*. 23(4) (1999) 497–530.
- [33] T. Hastie, R. Tibshirani, J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, And Prediction*, Springer Series in Statistics, Second Edition (2016), Springer, New York, NY.
- [34] J. He, J.S. McCarley, Executive Working Memory Load Does Not Compromise Perceptual Processing During Visual Search: Evidence From Additive Factors Analysis, *Attention, Perception, & Psychophysics* 72(2) (2010) 308–316.

- [35] J. M. Henderson, S. V Shinkareva, J. Wang, S.G. Luke, J. Olejarczyk, Predicting Cognitive State from Eye Movements, *PLoS One* 8 (2013) e64937.
- [36] T.J. Hess, L.P.; Rees, T.R. Rakes, Using Autonomous Planning Agents to Provide Model based Decision-making Support, *Journal of Decision Systems* 14(3) (2005) 261–278.
- [37] M. Hogan, A. Torres, C. Barry, An Eye Tracking Pilot Study of Optional Decision Constructs in B2C Transactional Processes, in: 14th International Conference on WWW/INTERNET (ICWI) (2015).
- [38] K. Holmqvist, M. Nystrom, R. Anderson, R. Dewhurst, H. Jarodzka, J. & Van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*, OUP Oxford, 2011.
- [39] C. S. Ikehara, M.E. Crosby, P.A. Silva, Combining Augmented Cognition and Gamification, in: Schmorrow D.D., Fidopiastis C.M. (Eds.) *Foundations of Augmented Cognition (AC)*, Lecture Notes in Computer Science 8027 (2003) 676–684, Springer, Berlin, Heidelberg.
- [40] S. Iqbal, P. Adamczyk, X. Zheng, B. Bailey, Towards an Index of Opportunity: Understanding Changes in Mental Workload During Task Execution, in: *Proceedings of the SIGCHI conference on Human factors in computing systems* (2005) 311–320.
- [41] R. Jacob, K.S. Karn, Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises, in: R. Radach, H. Deubel (Eds.), *The Mind’s Eye: Cognitive and Applied Aspects of Eye Movement Research*, Oxford, England, Elsevier Science BV (2003) 573–605.
- [42] A.G. James, D. Witten, T. Hastie, R. Tibshirani, M.T. Hastie, Package “ISLR,” (2018), <https://cran.r-project.org/web/packages/ISLR/index.html>.
- [43] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, Springer New York, 2014, <https://books.google.com/books?id=at1bmAEACAAJ>.
- [44] M. A. Just, P.A. Carpenter, Eye Fixations and Cognitive Processes, *Cognitive Psychology* 8(4) (1976) 441–480.
- [45] D. Kahneman, J. Beatty, Pupil Diameter and Load on Memory, *Science* 154(3756) (1966) 1583–1585.
- [46] S. Kardan, C. Conati, Exploring Gaze Data for Determining User Learning with an Interactive Simulation, in: J. Masthoff, B. Mobasher, M.C. Desmarais, R. Nkambou (Eds.), *User Modeling, Adaptation, and Personalization (UMAP)* (2012) 126–138, Springer, Berlin, Heidelberg.
- [47] A. Klami, C. Saunders, T.E. de Campos, S. Kaski, Can Relevance of Images Be Inferred from Eye Movements? In: *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval (MIR)* (2008) 134-140.
- [48] J. Klayman, Children’s decision strategies and their adaptation to task characteristics, *Organizational Behavior and Human Decision Processes* 35(2) (1985) 179–201.
- [49] M. Król, M.E. Król, A Novel Approach to Studying Strategic Decisions With Eye-Tracking and Machine Learning, *Judgment & Decision Making* 12(6) (2017) 596–609.
- [50] J. L. Kruger, E. Hefer, G. Matthew, Measuring The Impact of Subtitles on Cognitive Load, in: *Proceedings of the 2013 Conference on Eye Tracking, South Africa (ETSA)* (2013) 62–66.
- [51] F.Y. Kuo, C.W. Hsu, R.F. Day, An Exploratory Study of Cognitive Effort Involved in Decision under Framing- An Application of the Eye-Tracking Technology, *Decision Support Systems* 48(1) (2009) 81–91.

- [52] B. Laeng, S. Sirois, G. Gredebäck, Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science* 7(1) (2012) 18–27.
- [53] H. Ledger, The Effect Cognitive Load has on Eye Blinking, *Plymouth Student Scientist* 6(1) (2013) 206–223.
- [54] Y. Liu, P.Y. Hsueh, J. Lai, M. Sangin, M.A. Nussli, P. Dillenbourg, Who Is the Expert? Analyzing Gaze Data To Predict Expertise Level In Collaborative Applications, in: *IEEE International Conference on Multimedia and Expo (ICME)* (2019) 898–901.
- [55] S. P. Marshall, Identifying Cognitive State From Eye Metrics, *Aviation, Space, and Environmental Medicine* 78(5) (2007) 165–175.
- [56] A. Max, K. Contributions, S. Weston, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, R.C. Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, T. Hunt, Package “caret” (2018), <https://cran.r-project.org/web/packages/caret/index.html>.
- [57] R. N. Meghanathan, C. van Leeuwen, A.R. Nikolaev, Fixation Duration Surpasses Pupil Size as a Measure of Memory Load in Free Viewing, *Frontiers in Human Neuroscience* 8 (2014) 1063.
- [58] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C.-C. Chang, C.-C. Lin, R Package e1071 Version 1.6-8, Gpl-2. (2017). <https://cran.rproject.org/web/packages/e1071/e1071.pdf>.
- [59] A. S. Najar, A. Mitrovic, K. Neshatian, Utilizing Eye Tracking to Improve Learning from Examples, in: C. Stephanidis, M. Antona (Eds.), *Universal Access in Human-Computer Interaction. Universal Access to Information and Knowledge, Lecture Notes in Computer Science* 8514 (2014) 410–418, Springer International Publishing, Cham.
- [60] J.W. Payne, Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis, *Organizational Behavior and Human Performance* 16 (2) (1976) 366–387.
- [61] J.W. Payne, J.R. Bettman, E.J. Johnson, Adaptive Strategy Selection in Decision Making, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14(3) (1988) 534–552.
- [62] J.W. Payne, J.R. Bettman, E.J. Johnson, The Use of Multiple Strategies in Judgment and Choice, in N. J. Castellan, Jr. (Ed.), *Individual and Group Decision Making: Current Issues* (1993) 19-39, Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- [63] T. Piquado, D. Isaacowitz, A. Wingfield, Pupillometry as a Measure of Cognitive Effort in Younger and Older Adults, *Psychophysiology* 47(3) (2010) 560–569.
- [64] A. Poole, L. Ball, Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects, in: C. Ghaoui (Ed.), *Encyclopedia of Human Computer Interaction*, Idea Group Inc. PA, USA (2005) 211-219.
- [65] R.B. Rao, G. Fung, R. Rosales, On the Dangers of Cross-Validation. An Experimental Evaluation, in: *Proceedings of the SIAM International Conference on Data Mining*, Society for Industrial and Applied Mathematics (2008) 588–596. doi:10.1137/1.9781611972788.54.
- [66] B. Ripley (2018). tree: Classification and Regression Trees. R package version 1.0-37. <https://CRAN.R-project.org/package=tree>
- [67] J.L. Rosch, J.J. Vogel-Walcutt, A Review of Eye-Tracking Applications as Tools for Training, *Cognition. Technology & Work* 15 (3) (2013) 313–327.

- [68] J. Shah, J. Wiken, B. Williams, C. Breazeal, Improved Human-Robot Team Performance using Chaski, a Human-Inspired Plan Execution System, Proceedings of the 6th International Conference on Human-Robot Interaction (2011) 29–36. ACM.
- [69] M. Shojaeizadeh, S. Djamasbi, P. Chen, J. Rochford, Task Condition and Pupillometry, in: Proceedings of 23rd Americas Conference on Information Systems, Human-Computer Interaction (SIGHCI) (2017) 19.
- [70] J. Simola, J. Salojärvi, I. Kojo, Using Hidden Markov Model to Uncover Processing States from Eye Movements in Information Search Tasks, Cognitive System Research 9(4) (2008) 237–251.
- [71] H.A. Simon, A Behavioral Model of Rational Choice, The Quarterly Journal of Economics 69(1) (1955) 99–118.
- [72] B. Steichen, C. Conati, G. Carenini, Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data, ACM Transaction on Interactive Intelligent Systems (TiiS) 4(2) (2014)
- [73] P. Todd, I. Benbasat, Evaluating the Impact of DSS, Cognitive Effort, and Incentives on Strategy Selection, Information Systems Research 10(4) (1999) 356–374.
- [74] P. Todd, I. Benbasat, The Influence of Decision Aids on Choice Strategies: An Experimental Analysis of the Role of Cognitive Effort, Organizational Behavior and Human Decision Processes 60(1) (1994) 36–74.
- [75] P. Todd, I. Benbasat, The Use of Information in Decision Making: An Experimental Investigation of the Impact of Computer-Based Decision Aids, MIS Quarterly 16(3) (1992) 373–393.
- [76] P. Todd, I. Benbasat, An Experimental Investigation of the Impact of Computer Based Decision Aids on Decision Making Strategies, Information Systems Research 2(2) (1991) 87–115.
- [77] R.L. Van den Brink, P.R. Murphy, S. Nieuwenhuis, Pupil Diameter Tracks Lapses of Attention, PLoS One 11 (2016) e0165274.
- [78] K. Van Orden, W. Limbert, S. Makeig, T.-P. Jung, Eye Activity Correlates of Workload during a Visuospatial Memory Task, Human Factors 43(1) (2001) 111–121.
- [79] K. K. Wong, W.Y. Wan, S.B. Kaye, Blinking and Operating: Cognition versus Vision, British Journal of Ophthalmology 86(4) (2002) 479–479.
- [80] M.H. Zweig, G. Campbell, Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine, Clinical Chemistry 39(4) (1993) 561–577.