Detecting Personality Traits Using Eye-Tracking Data

Shlomo Berkovsky
Data61 – CSIRO and
Macquarie University
shlomo.berkovsky@gmail.com

Ronnie Taib
Data61 – CSIRO
ronnie.taib@csiro.au

Irena Koprinska
School of Computer Science
University of Sydney
irena.koprinska@sydney.edu.au

Eileen Wang, Yucheng Zeng
School of Computer Science
University of Sydney
ewan9058,yzen7770@sydney.edu.au

Jingjie Li
Dept. of Electrical and Computer Eng.
University of Wisconsin–Madison
jingjie.li@wisc.edu

Sabina Kleitman
School of Psychology
University of Sydney
sabina.kleitman@sydney.edu.au

ABSTRACT

Personality is an established domain of research in psychology, and individual differences in various traits are linked to a variety of real-life outcomes and behaviours. Personality detection is an intricate task that typically requires humans to fill out lengthy questionnaires assessing specific personality traits. The outcomes of this, however, may be unreliable or biased if the respondents do not fully understand or are not willing to honestly answer the questions. To this end, we propose a framework for objective personality detection that leverages humans' physiological responses to external stimuli. We exemplify and evaluate the framework in a case study, where we expose subjects to affective image and video stimuli, and capture their physiological responses using a commercial-grade eye-tracking sensor. These responses are then processed and fed into a classifier capable of accurately predicting a range of personality traits. Our work yields notably high predictive accuracy, suggesting the applicability of the proposed framework for robust personality detection.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*;

KEYWORDS

Personality detection; framework; eye tracking; field study.

ACM Reference format:
Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. 2019. Detecting Personality Traits Using Eye-Tracking Data. In *Proceedings of CHI Conference on Human Factors in Computing Systems Proceedings, Glasgow, Scotland UK, May 4–9, 2019 (CHI 2019),* 12 pages.
https://doi.org/10.1145/3290605.3300451

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300451

1 INTRODUCTION

Personality is an established area of research in psychology. Generally, it refers to a set of individual patterns of behaviours, cognitions, and emotions that predict a human's interactions with their environment [46]. Although multiple personality theories exist, many of them conceptualise these patterns into *traits*, which are believed to be relatively stable and consistent predictors of human behaviour [34]. Even within the trait-based representation of personality there is no single agreed-upon model and multiple models, such as the Five Factor Model (also known as the Big-5) [35], HEXACO [4], Temperament and Character Inventory [11], and Interpersonal Circumplex [12], have been developed.

Regardless of the underlying personality model, detection of traits is a complex and error-prone task. This is traditionally carried out using validated psychological questionnaires aimed at uncovering the values of the model traits [36]. However, the fixed and long questionnaires restrict practical applications. Moreover, due to privacy considerations [5, 17] or in high-stake situations like recruitment questionnaires [3, 60], people may not be willing to genuinely answer the questions, providing rather the desired or false answers. Faking, response distortion and self-deception phenomena, and how to overcome these, are hotly debated contemporary issues [16, 61]. All the above limitations trigger an increasing interest in alternative methods for objective and valid detection of personality traits [28, 33, 41], including methods that leverage physiological signals [1, 49, 53].

In this work, we turn to the challenge of personality detection using humans' physiological responses to external stimuli. Indeed, the rich modality, increasing accuracy, and decreasing costs of modern sensing technologies pave the way for their deployment in a range of real-life applications. We propose to use such technologies for capturing physiological signals, e.g., brain activity, eye saccades, or skin conductivity, produced by the human body in response to stimuli. As many of these signals are bodily responses that cannot be consciously controlled [10], we posit that they can be considered as reliable and valid indicators of the human's reaction to the stimuli and to the emotions evoked by the stimuli, which we attribute to the personality traits. In this work we set out to study whether observable patterns in physiological responses to stimuli can serve as predictors of personality traits.

We propose a generic framework for the detection of personality traits that does not rely on self-reports and self-perceptions. The main components of the framework include: (i) an external stimulus that triggers the physiological responses; (ii) a sensing technology that captures a person's responses to the stimuli; (iii) a data processing component that extracts the features required for personality detection; and (iv) a machine learning algorithm that predicts the values of the personality traits. We initially present the framework, discuss the roles of the above four components, and focus on the dependencies between them.

Then, we proceed to a specific application of the framework, in which we use affective image and video stimuli and eye-tracking data to detect personality traits belonging to three models: the Dark Triad [40], the Reinforcement Sensitivity model, also known as BIS/BAS [9], and the HEXACO model [4]. We elaborate on the methodology applied for the data collection and analysis, and then present the evaluation results. The results demonstrate that the framework is capable of accurately predicting a range of personality traits, with the video stimuli yielding higher predictive accuracy than the image stimuli. We also present a comprehensive analysis of the factors underpinning the obtained results.

Hence, the contributions of this work are three-fold. First, we demonstrate how off-the-shelf sensing and machine learning methods can be combined into a *generic framework for detection of personality traits*. Second, we exemplify and evaluate a specific *instantiation of the framework using affective stimuli and eye-tracking data*. Third, our evaluation achieves *notably high predictive accuracy results*, some of which are associated with well-studied factors in human psychology.

The research direction of personality detection using multimodal stimuli, sensing technologies, and physiological responses, as embodied by our framework, has a promising future for building usable, tailored, and engaging interactive systems. It also has the tremendous potential to simplify and streamline various profiling and user modelling tasks, e.g., those related to recruitment and personnel appointment.

2 RELATED WORK

Personality is an organised set of characteristics, which influences the individual's behaviours, cognition, and emotions [46]. Modern personality theories conceptualise these characteristics into traits, which are believed to be relatively stable and consistent dispositions that humans possess. Within the trait-based representation of personality, multiple personality models have been proposed and studied. Influential and numerously validated models include the Big-5 Factor model [35] and its extension referred to as the HEXACO model [4], the Reinforcement Sensitivity Model (BIS/BAS) [9], the Dark Triad (D3) [40], and Interpersonal Circumplex [12].

One of the most widely-used personality detection methods entails the administration of personality inventories [36]. These are questionnaires developed and validated based on relevant personality and psychometric theory. A self-report questionnaire is typically used to measure personality traits and their facets. Most are well validated, cheap to administer and process. However, the self-report results can be distortion-prone, especially in high-stake situations [3, 16, 60]. This triggers an increasing body of research seeking for alternative or supplementary, distortion resistant methods of personality detection and user modelling. The popularity of social networks opens the opportunity to detect personality through social media activities and textual content posted by users [51]. For example, Big-5 traits were linked to social media activity [19], such that some traits were accurately detected merely through the analysis of Facebook likes [28] and linguistic features of posted tweets [41]. Beyond social networks, deep learning was applied to detect the Big-5 traits from essays [33]. These methods, however, are not without their own problems as people can use social media for impression-management.

Personality traits may also impact the autonomic nervous system and, in turn, bodily responses and physiological signals [52]. To the best of our knowledge, the first significant attempt to detect the Big-5 traits using physiological signals was in the [1, 49, 53] line of research. These works collected the EEG, GSR, EMO, and ECG signals of subjects watching video clips. The obtained prediction accuracy levels varied substantially, ranging from below-random to 90%. However, this stream of research demonstrated the feasibility of personality detection using commercial-grade sensors. Beyond the Big-5 model, the 16 Personality Factors model was predicted using facial features and a neural network [18]. Although an improved accuracy was achieved, the method was complex and required about three hours for training the model.

Eye movement parameters were extensively used to detect conscious and unconscious activities. Complex features, like gaze pattern and scan path, were found to be reliable indicators of cognitive strategies and attention [14, 42]. Pupillary response was used as an indicator of cognitive load [10, 58], whereas saccade amplitude and fixation durations were used for lie detection [32]. In the personality domain, early research established the links between eye contact, gaze aversion, and sociability [31]. With the advent of eye tracking technologies, features derived from saccades, eye fixations, and pupils were found to be associated with personality traits [44, 56]. Recently, eye movement data during an everyday task was used to predict, although with relatively low levels of accuracy, the Big-5 traits and perceptual curiosity [24].

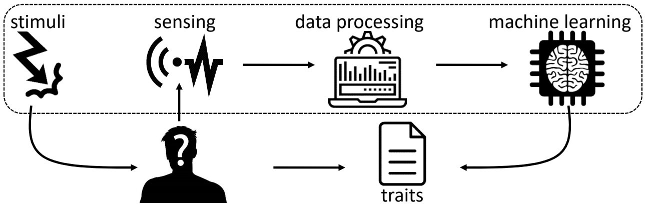

Figure 1: Framework for personality detection

This work addresses several limitations faced by prior research. We extend the psychological traits to cover the D3, BIS/BAS, and HEXACO models. We propose a framework for personality detection and exemplify it using eye-tracking data. We examine two types of affective stimuli, carefully crafted using valence-arousal and emotion metrics. Finally, we experiment with a range of machine learning methods to optimise predictive accuracy.

3 SETTING AND METHODS

Personality Detection Framework

We start by presenting our framework for detection of personality traits using physiological signals (see Figure 1, where the framework components are framed).

External stimuli to trigger physiological responses. The main idea underpinning the framework is that not consciously controllable bodily responses to stimuli can be considered as objective personality predictors. A range of stimuli can be applied: from plain text, through multimedia, to interactive tasks. We highlight the links between the nature of the applied stimuli and the detected traits. For example, when detecting the learning style, a suitable stimulus could be a series of riddles to solve. Conversely, detection of emotional stability may involve energetic songs or graphic violence as the stimuli.

Sensing technologies for physiological responses. To capture physiological responses, the framework interacts with the subject as they are exposed to the stimuli. The responses are captured by the sensing technology and fed into the framework for data processing. Revisiting the key idea of objective personality detection, we highlight the importance of technologies capturing not consciously controllable responses, e.g., skin conductivity, brain activity, or pupil size. The level of control over the captured responses the subject can exhibit drives the objectivity of the captured physiological responses and, in turn, of the detected personality traits.

Data processing and feature extraction. The captured physiological responses are raw signals, e.g., skin conductivity values or brain signals. To predict personality traits, we need to process these and extract features characterising the signals. Although the exact data processing steps depend on the sensing technology, several typical steps include: filtering, segmentation, and normalisation.The feature extraction depends even more on the selected technology. It can be divided into standard signal processing features, e.g., temporal and transform features, and sensor-specific features, e.g., phasic/tonic feature decomposition for GSR data.

Machine learning for personality detection. The last component of the framework deals with the detection of personality traits. This is performed using machine learning methods, where the learning component is trained on labelled data, e.g., from past reference subjects, and predicts labels for unknown data, i.e., traits of a new target subject, whose personality is being detected. A range of supervised machine learning methods, such as decision trees, regression models, and ensembles of classifiers, is applicable for this task [23].

This work demonstrated one instantiation of the framework. Namely, we detect personality traits using eye-tracking data reflecting autonomic nervous activity elicited by affective image and video stimuli. The following sub-sections outline the personality models and traits, discuss how the framework is applied, and finally present our experimental methodology.

Personality Models and Traits

We focus on three well-validated models: D3, BIS/BAS, and HEXACO. Table 1 briefly presents these models, the traits, and their relevant facets included in each model. Altogether, we examined 16 variables capturing the traits and their facets. In order to establish ground truth values for the 16 variables used as class labels for the machine learning component, we deploy five well-validated personality inventories. For the D3, we use (i) Levenson's Self-Report Psychopathy inventory that includes 26 items assessing psychopathy, both primary and secondary [30]; (ii) Narcissistic Personality Inventory (NPI-16), which is a short 16-item version of the full NPI-40 inventory that measures narcissism [2]; and (iii) MACH-IV, which is a trimmed 20-item inventory extracted from the full MACH inventory that measures Machiavellianism, including the tactics, morality, and views traits [43]. The BIS/BAS traits (BIS, BAS Drive, BAS Fun Seeking and BAS Reward Responsiveness) are measured using the BIS/BAS inventory containing 24 items [20], and the HEXACO traits (agreeableness, conscientiousness, extraversion, honesty, resiliency and openness) are measured using a 25-item inventory [50].

A score for the 16 variables and for each subject is discretised and used as a class label for the recorded eye activity, to create training data for the trait classifiers. The classifiers learn to predict the trait values from the eye data and are then used to determine the trait class label for a new subject.

Specific Personality Detection Application

We will describe how the generic framework presented in the previous section is instantiated for the detection of the above

Trait	Description
Dark Triad (D3)	
Primary Psychopathy	Primary psychopathy is the emotional aspect of psychopathy, characterised by a lack of empathy and deficit in processing negative feelings. It is associated with callousness, remorseless, and failure to accept responsibility [13].
Secondary Psychopathy	Secondary psychopathy is the behavioural aspect of psychopathy, characterised by antisocial acts. It is associated with instability and aggression, although it does not arise from deficit in processing negative feelings [13].
Tactics	Tactics is a component of Machiavellianism that focuses on exploitation of others. People high in tactics tend to engage in interpersonal exploitation, willingly and skilfully manipulating their peers in pursuit of personal goals [45].
Views	Views is a component of Machiavellianism that focuses on the lack of trust. People high in views hold a cynical view of the human nature, have a hyper-vigilance to being manipulated, with a view that others cannot be trusted [45].
Morality	Morality is a component of Machiavellianism that focuses on disbelief in the moral norms. People high in morality (more precisely, immorality) disregard conventional morality of the society, which would condemn their actions [45].
Narcissism	Narcissism involves excessive self-love. People high in narcissism have inflated sense of self-importance and self-admiration, with tendencies toward grandiose ideas, fantasied talents, and defensiveness to criticism [37].
Behavioural Inhibition System and Behavioural Activation System (BIS/BAS)	
BIS	BIS measures the motivation to avoid aversive outcomes. BIS is responsible for the experience of negative feelings like fear, frustration, and sadness in anticipation for punishment. People high in BIS are more prone to anxiety [9].
BAS Drive	BAS Drive measures the motivation to persistently pursue the desired goals. People high in BAS Drive are more eager to engage in goal-directed efforts and to pursue their goals with perseverance [9].
BAS Fun Seeking	BAS Fun Seeking measures the motivation to find novel rewards spontaneously. People high in BAS Fun Seeking have a stronger desire for new rewards and a willingness to approach rewarding events on the spur of the moment [9].
BAS Reward Responsiveness	BAS Reward Responsiveness measures the sensitivity to pleasant reinforcers in the environment. People high in BAS Responsiveness are sensitive to rewards and positive stimuli, and positively respond to the anticipation of reward [9].
HEXACO Personality Traits	
Agreeableness	Agreeableness concerns how people interact and maintain relationships with others. People high in agreeableness build warm relationships, are empathetic, altruistic, good-tempered, and less prone to conflicts [47].
Conscientious.	Conscientiousness relates to the people's will to achieve their goals. People high in conscientiousness are more diligent, dutiful, organised, self-disciplined, and strive for achievements [47].
Extraversion	Extraversion relates to the sociability and assertiveness of people. People high in extraversion are sociable, gregarious, and seek excitement in interpersonal interactions with others [47].
Honesty	Honesty is associated with humility and sincerity of people. People high in honesty are generally loyal, truthful and direct, less hypocritical, less manipulative, and less deceitful [47].
Resiliency	Resiliency (or, the inverse trait, Neuroticism) concerns the emotional stability of people. People high in resiliency are better at emotional control, less impulsive, and less prone to anxiety and depression [47].
Openness	Openness is associated with people's acceptance of experiences and their creativity. People high in openness are more creative, curious, and have a stronger desire for novel experiences and intellectual exploration [47].

Table 1: Summary of the studied personality models and traits

16 traits using affective image and video stimuli and eye-tracking data. We will discuss the components individually.

Affective images and videos. Since some of the predicted traits are affect-related, in this case study we used affective stimuli expected to evoke subjects' emotional responses. In particular, we opted for still images and short video stimuli, because we wanted to develop a system able to provide a fast psychological profiling.

Images. We used a subset of images from the International Affective Picture System (IAPS) dataset [29]. This is a well-studied dataset, where each image is assigned scores corresponding to different emotions. We focused on the arousal and valence scores and clustered images into five categories: high arousal and high valence (HAHV, strongly positive emotions), low arousal and high valence (LAHV, mildly positive emotions), low arousal and low valence (LALV, mildly negative emotions), high arousal and low valence (HALV, strongly negative emotions), and neutral images (neutral emotions, no specific arousal applies). A set of 50 IAPS images – ten for each category – was selected as the image stimuli.

The images from each category were shown for 8 seconds each, in blocks of five and in the same order, to all the subjects. Each block was preceded by a cool-down period of 15 seconds, during which a black cross on a white background was shown, allowing recovery from previous stimuli. The subjects could not pause the presentation or skip images, so the overall duration of the image stimuli was 9.2 minutes.

Videos. We used videos from the English version of the FilmStim dataset [48]. Video stimuli representing seven emotion types – fear, tenderness, anger, neutral, sadness, amusement, and disgust – were selected based on their pre-annotated arousal-valence scores. These video stimuli were extracted from the following movies: “Seven”, “Life is Beautiful”, “American History X”, “Blue”, “Dangerous Mind”, “A Fish Called Wanda”, and “Trainspotting”. The duration of the videos was between 25 and 132 seconds, to ensure that the emotional peak is reached, while avoiding subject fatigue.

The videos were shown in the same order to all the subjects, and broken down into two blocks, to minimise carry-over effects. The first block included videos evoking fear, tenderness, and anger, while the second included videos evoking neutral emotions, sadness, amusement, and disgust. A cool-down period of 30 seconds was allocated after each video, during which a black cross on a white background was displayed. The overall duration of the video stimuli was 14 minutes.

For the analysis purposes, we can aggregate the stimuli in various ways. At the fine-grained level, we have ten image blocks (two for the HAHV, LAHV, LALV, HALV, and neutral images), and seven video blocks (for the fear, tenderness, anger, neutral, sadness, amusement, and disgust videos), obtained by segmenting and filtering out the cool-down periods. We can aggregate the stimuli into a single image block and a single video block. Finally, aggregating all the blocks yields the complete subject’s response to all the stimuli.

Eye-tracking glasses. In this instantiation of the framework, we focus exclusively on eye-tracking data. This data provides valuable information about human autonomic nervous system, which is acquired in a non-intrusive and continuous manner.

We deployed the SMI Eye-Tracking Glasses (ETG) as the underlying sensing technology. SMI ETG are light-weight wearable glasses that can capture natural eye and gaze behaviour through two infrared cameras focusing on each eye. Eye data is estimated in real-time and transmitted to a server storing the data and producing various metrics. Eventually, the sensor captures and provides pupil dilation (along X and Y axes), eye saccades and fixations, blinks, and relative gaze direction data. A relatively wide range of field-of-view angles is captured: 60° horizontally and 46° vertically.

Eye-tracking data and features. Features extracted from the captured eye-tracking data and related to eye activity can be categorised into three groups: eye blink measures, eye movement (saccades and fixations) measures, and pupillary response measures. Specifically, we extracted ten unique features for each temporal block:

– *Blink Rate (BR)* - average number of blinks per second; blink count divided by block duration.
– *Saccade Rate (SR)* - average number of saccades per second; saccade count divided by block duration.
– *Saccade Amplitude (SA)* - average angular distance of the saccades (in °), over all saccades in the block.
– *Average Saccade Velocity (ASV)* - average angular velocity of the saccades (° per second), over all saccades in the block.
– P*eak Saccade Velocity (PSV)* - average of the peak angular velocities (° per second) of each saccade in the block.
– *Fixation Rate (FR)* - average number of fixations; fixation count divided by block duration.
– *Fixation Duration (FD)* - average duration of fixations; cumulative duration of fixations divided by fixation count.
– *Saccade-Fixation Ratio (SFR)* - ratio between the duration of saccades (search) and fixations (processing) in the block.
– *Horizontal pupil size (PX)* - average horizontal diameter (in pixels) of the pupils over the duration of the block.
– *Vertical pupil size (PY)* - average vertical diameter (in pixels) of the pupils over the duration of the block.

These features are in line with the features extracted in previous works using eye-tracking data [10, 38, 59].

The feature extraction process started by normalising the raw feature values in each temporal block with respect to the baseline observed in the 15- or 30-second cool-down period immediately preceding the block. Then all features were extracted for each of the 17 blocks. Finally, the ten image blocks were collapsed into five blocks by averaging the extracted feature values, as they consisted of two repeats of the five arousal-valence categories: HAHV, LAHV, LALV, HALV and neutral. Thus, the overall number of extracted features from the image and video blocks is (5+7)*10 = 120.

To prevent overfitting, we conducted feature selection using the Correlation-based Feature Selection (CFS) algorithm [22]. Its main idea is that a good feature subset should contain features that are highly correlated with the class label, i.e., very informative, but weakly correlated with other features, i.e., not redundant. CFS defines a heuristic measure based on these criteria and uses a search algorithm to find the feature subset that maximizes this measure. CFS was applied to each trait separately, yielding feature sets ranging from 3 to 10 features, which is a considerable reduction from the original 120 features. The smallest feature set (3) was selected for BAS Reward Responsiveness and the largest (10) – for Originality.

Classification for trait predictions. The raw trait values obtained through the aforementioned personality inventories were discretised into three classes – low, medium and high – for each trait, using equal-frequency binning. A separate classifier was trained for the predictions of each trait. One training data point corresponds to one subject and includes the values of the selected features and the trait label assigned based on the discretised scores for the given trait. Due to the equal-frequency binning, the classification problem was

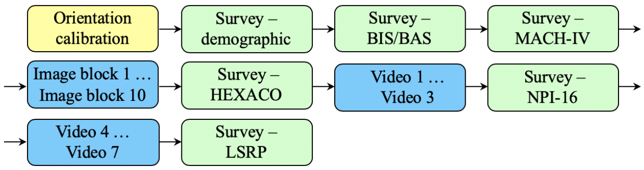

Figure 2: Experimental workflow: stimuli and inventories

class-balanced, as the number of training data points in every class, and for every trait, was identical.

To deploy the classifiers, we used standard implementations of Weka, an open-source data mining toolbox [23]. Specifically, seven classifiers were deployed: AdaBoost (AB), Decision Tree (DT), Logistic Regression (LR), Naive Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and k-Nearest Neighbour (kNN). To evaluate the performance of the classifiers, we applied the leave-one-out methodology: data of all but one subjects was used to train the classifiers, while the data for the last (target) subject was used for testing. At each run, the goal of the classifier was to predict the trait class label for the target subject and the accuracy of the prediction was assessed. The process was repeated using another subject as a target until all subjects were tested. Finally, the average accuracy over all runs was calculated and reported.

Data Collection and Evaluation Setting

In total, 21 subjects were recruited for this experiment, under appropriate ethics clearance. The subjects were either students or staff of a research organisation. As the subjects were not recruited on psychological grounds, they were unlikely to exhibit particularly low/high trait values, especially with for the Dark Triad. All but one subjects reported good or native English proficiency. 18 subjects were aged 18 to 30 whereas the other 3 were older than 30. The data was collected in a controlled laboratory setting under fixed illumination and room temperature conditions.

The workflow of the conducted data collection, which was performed individually for every subject, is shown in Figure 2. The whole procedure took, on average, 55 minutes. At the start, the subjects were given an overview of the experiment and their consent was obtained. The eye tracking glasses were then mounted and calibrated to achieve the best signal quality. The subjects were instructed to sit down and relax, to acquire a baseline signal and minimise artefacts related to movement.

The ground truth data (target class label) for the traits was obtained by administering the personality inventories for the 16 traits listed in Table 1. Note that the inventories were interleaved with the image and video stimuli, in order to avoid fatigue and provide additional cool-down time that further diminished carry-over effects. As shown in Figure 2, there were

Trait	Range	Mean	SD	Low	Medium	High
Prim. Psychopathy	[16,64]	30.05	5.21	[21,26]	[28,32]	[33,40]
Second. Psychopathy	[10,40]	20.68	2.34	[17,19]	[20,21]	[22,25]
Tactics	[9,45]	23.91	3.96	[19,20]	[21,25]	[26,32]
Views	[9,45]	23.68	4.06	[17,21]	[22,24]	[25,32]
Morality	[2,10]	5.18	1.53	[2,4]	[5,5]	[6,8]
Narcissism	[0,100]	29.47	22.14	[0,13]	[18,32]	[37,82]
BIS	[7,28]	20.23	3.04	[15,18]	[19,20]	[21,26]
BAS Drive	[4,16]	11.00	1.67	[9,9]	[10,11]	[12,14]
BAS Fun Seeking	[4,16]	11.59	1.94	[8,10]	[11,12]	[13,15]
BAS Reward Resp.	[5,20]	15.64	1.89	[13,14]	[15,15]	[16,20]
Agreeableness	[4,20]	13.95	2.17	[10,13]	[14,14]	[15,18]
Conscientiousness	[4,20]	13.14	2.80	[9,11]	[12,14]	[15,20]
Extraversion	[4,20]	14.14	2.42	[11,12]	[12,14]	[15,19]
Honesty	[5,25]	17.91	2.91	[14,15]	[16,19]	[20,24]
Resiliency	[4,20]	13.64	2.63	[9,12]	[13,15]	[16,19]
Openness	[4,20]	13.82	2.86	[9,12]	[13,14]	[16,19]

Table 2: Descriptive statistics of personality scores (the Range column shows theoretical trait ranges, while Low, Medium, and High reflect our study observations)

six inventories (marked in green). The average cumulative completion time of these inventories was 11 minutes.

The subjects were exposed to the image and video stimuli described in the previous sections. At all times, the subject's physiological responses to the stimuli were captured by the SMI ETG. We processed the captured data, populated the features for the temporal blocks, and fed them into the classifiers that were trained to predict the personality traits' class labels, building a separate classifier for each trait. We used the classification *accuracy* metric to evaluate the performance of the classifiers. This is the ratio (shown in %) between the number of correctly predicted traits class labels (low, medium, or high) and the total number of predictions made.

4 RESULTS

The following questions guide our analysis:

– Q1: What machine learning methods should be used for detecting personality traits?
– Q2: What traits can be detected with high/low accuracy?
– Q3: What stimuli – images, videos, or both – are most predictive for trait detection?
– Q4: What features are predictive of each personality model?

Descriptive Statistics of Personality Scores

We start by showing in Table 2 the descriptive statistics of the raw scores obtained for the 16 personality traits. The table shows the overall range of values in the inventory, the mean and standard deviation obtained for the 21 subjects, and the brackets of the low, medium, and high classes within each trait. As can be seen, a few scores are placed at the extremities of the trait ranges, although the equal-frequency binning ensures uniform distribution of subjects across the classes.

Classifier	AB	DT	LR	NB	RF	SVM	kNN
Accuracy	53.87	35.42	78.87	**85.71**	61.01	75.89	72.62

Table 3: Accuracy (in %) of the seven classifiers

Classifier Performance

Q1 deals with the performance of various machine learning methods applied for trait predictions. We assess the classifiers through their mean accuracy computed across all the traits and all the subjects. Table 3 shows the mean accuracy of the seven classifiers, using both the image and video stimuli. The best performing method is highlighted in bold. We observe that NB, the most accurate classifier, achieves 85.71% accuracy and outperforms all other classifiers by 8.7% or more. LR, the second best classifier, achieve promising accuracy of 78.87%. Then, the SVM and kNN classifiers achieve accuracy of 75.89% and 72.62%, respectively. The remaining classifiers are substantially lower – between 35.42% and 61.01%.

The results demonstrate that the NB classifier performs well in combination with CFS, a correlation-based feature selection method. NB's performance is adversely affected by highly correlated input features, since it assumes that the features are independent from each other within the class [22]. By using CFS, we select a subset of less correlated features, which aligns with NB's underlying assumption and contributes to its predictive performance. Likewise, LR achieves high classification accuracy, when preceded by the CFS feature selection of non-correlated input features [25, 57].

The DT classifier demonstrates the lowest performance. A closer examination shows that the generated trees are shallow, testing 4-5 features only, which is insufficient for the target classification task. As could be expected, the tree ensembles, implemented by the AB and RF classifiers, improve the performance of the single DT. However, their performance is still not competitive with the NB and LR classifiers. The similarity- and separation-based kNN and SVM classifiers achieve better classification accuracy than the tree-based methods, but are still inferior to NB and LR; presumably, due to the relatively small size of the training data.

Hence, revisiting Q1 we conclude that *Naive Bayes is the most appropriate machine learning method* for the personality trait detection task. In the following subsections we will primarily focus on the results achieved by the NB classifier.

Individual Trait Detection

Q2 deals with the detection of individual personality traits. Table 4 shows the accuracy of the NB classifier when predicting each of the 16 traits. For benchmarking purposes, we also show the highest accuracy achieved for the trait by any other classifier (the best-other classifier is named in brackets). The best performing method for each trait is highlighted in bold.

Trait	Best-Other	NB
Primary Psychopathy	76.19 (kNN)	76.19
Secondary Psychopathy	80.95 (SVM)	80.95
Tactics	90.48 (SVM)	90.48
Views	90.48 (SVM)	90.48
Morality	80.95 (SVM)	**90.48**
Narcissism	76.19 (LR)	**80.95**
Mean D3		*84.92*
BIS	90.48 (LR)	90.48
BAS Drive	**85.71** (SVM)	80.95
BAS Fun Seeking	95.24 (SVM)	95.24
BAS Reward Responsiv.	**90.48** (SVM)	85.71
Mean BIS/BAS		*88.10*
Agreeableness	85.71 (LR)	**90.48**
Conscientiousness	71.43 (LR)	**80.95**
Extraversion	71.43 (LR)	**80.95**
Honesty	76.19 (LR)	**80.95**
Resiliency	80.95 (AB)	**90.48**
Openness	85.71 (LR)	85.71
Mean HEXACO		*84.92*
Mean Overall		*85.71*

Table 4: Accuracy of NB and best-other classifiers

The average accuracy of NB for each psychological model and across all the 16 traits/facets is also shown.

Comparing NB with the other best-performing classifier, we observe several trends. NB yields the highest accuracy for seven traits, for two traits SVM outperforms NB, and in seven cases NB is as good as the other best-performing classifier. We also compare the performance of NB across the models. For D3, NB beats the best-other classifier for two traits (Morality and Narcissism) and SVM is the dominant best-other classifier for four traits, including three on par with NB. For BIS/BAS, SVM is again the strongest best-other classifier, beating NB for two traits (BAS Drive and BAS Reward Responsiveness). For HEXACO, NB is clearly the best performer; it beats best-other LR classifier for all traits except for Openness. These results re-affirm that NB is the most appropriate classification method for personality predictions. Considering the second-best classifier, we refine our previous finding and conclude that SVM can be applied for the D3 and BIS/BAS traits, while LR – for HEXACO.

Focussing on NB predictions only, we observe reasonably high classification accuracy. Namely, all the traits are predicted with accuracy greater than 76%, which is more than twice higher than the accuracy of a random guess in a 3-class classification task. Notably, six traits – Tactics, Views, Morality, BIS, BAS Fun Seeking, Agreeableness, and Resiliency – are predicted with accuracy greater than 90%, and eight others are predicted with accuracy greater than 80%. Overall, we observe a mean accuracy close to 88% for the BIS/BAS model, and 85% for the D3 and HEXACO models.

To explain these results, we resort to the nature of the predicted traits. Psychology research classified personality traits into those driven by affect and, on the other hand, by cognitions or behaviours, i.e., non-affective. Specifically, [54] and [26] associated Machiavellianism with the affective rather than cognitive assessment. As the deployed stimuli were validated affective images and videos, they presumably evoked emotional responses and, as a result, we observe more accurate detection of the Tactics, Views, and Morality traits, which all achieve classification accuracy greater than 90%.

Similarly, [27] analysed the links between the BIS/BAS traits and affect, and found BAS Fun Seeking and BIS to be correlated with positive and negative affect, respectively. We observe that these BIS/BAS traits are predicted with accuracy levels greater than 90%, which, given the affective stimuli, is consistent with [27]. Considering the HEXACO traits, [62] identified Neuroticism, the Big-5's counter-part of Resiliency, to be the only trait associated with affect. Inspecting the results in Table 4, we find Resiliency and Agreeableness being predicted with accuracy greater than 90%, while the predictions of other HEXACO traits are less accurate.

Hence, we summarise Q2 and conclude that the affective nature of the stimuli allows us to generate *more accurate predictions for traits associated with affect*. Other personality traits, associated with either behaviours or cognitions, are generally predicted with a lower degree of accuracy.

Image vs. Video Stimuli

Next, we turn to Q3 and assess whether image or video stimuli are more predictive of the personality traits. For this, we separate the signals captured in response to the image stimuli from those captured in response to the video stimuli, and use them individually for feature extraction, classifier training, and trait predictions. In the first two columns of Table 5, we summarise the mean accuracy obtained for each trait by the NB classifier using either the image or video stimuli. The best performing stimuli for every trait is highlighted in bold.

We observe that the video stimuli achieved a higher overall mean accuracy than images, 76.19% vs. 73.81%. This observation is valid for the mean trait scores obtained for the D3 (76.98% vs. 74.60%) and BIS/BAS (76.19% vs. 70.24%) psychological models, whereas for the HEXACO traits images and videos exhibit the same mean predictive accuracy. The superiority of the video stimuli over the images can be explained by their stronger affective nature, which presumably evokes stronger emotional and physiological responses that allow for an easier detection of the traits [8].

Analysing the traits individually, we observe that the video stimuli outperform the image stimuli for nine traits, images outperform the videos for five traits, and for two traits they achieve the same accuracy. The dominance of the video stimuli is particularly pronounced in the BIS/BAS model where it

Trait	Image	Video	Both
Primary Psychopathy	**66.67**	61.90	**76.19**
Secondary Psychopathy	80.95	80.95	**80.95**
Tactics	**80.95**	66.67	**90.48**
Views	71.43	**90.48**	**90.48**
Morality	80.85	**85.71**	**90.48**
Narcissism	66.67	**76.19**	**80.95**
Mean D3	*74.60*	***76.98***	***84.92***
BIS	71.43	**76.19**	**90.48**
BAS Drive	66.67	**71.43**	**80.95**
BAS Fun Seeking	**85.71**	71.43	**95.42**
BAS Reward Responsiv.	57.14	**85.71**	**85.71**
Mean BIS/BAS	*70.24*	***76.19***	***88.10***
Agreeableness	**85.71**	76.19	**90.48**
Conscientiousness	61.90	**76.19**	**80.95**
Extraversion	**80.95**	61.90	**80.95**
Honesty	71.43	**80.95**	**80.95**
Resiliency	76.19	**80.95**	**90.48**
Openness	76.19	76.19	**85.71**
Mean HEXACO	***75.40***	***75.40***	***84.92***
Mean Overall	*73.81*	***76.19***	***85.71***

Table 5: Accuracy of the image and video stimuli

outperforms the images for three out of four traits. Notably, the video stimuli outperform the images by more than 10% for five traits: Views, Narcissism, BAS Reward Responsiveness, Conscientiousness, and Honesty. However, the images outperform the videos by more than 10% for four traits: Tactics, BAS Fun Seeking, Agreeableness, and Extraversion.

We also consider the combination of the image and video stimuli, shown in the right column of Table 5. We highlight in bold the traits, where the combined stimuli yielded accuracy on par with or superior to the best-performing individual type of stimuli. We observe that when both stimuli are used, the mean accuracy increases to 85.71%, which is 12.50% higher than for videos only and 16.12% higher than for images only.

Analysing the individual traits, we observe that the performance of the combined stimuli matches or outperforms the image and video stimuli individually for all the traits. The combined image and video stimuli outperform the best-performing individual stimuli by more than 10% for seven traits: Secondary Psychopathy, Tactics, BIS, BAS Drive, BAS Fun Seeking, Resiliency, and Openness. The mean improvement of the combined stimuli for D3, BIS/BAS, and HEXACO is 10.31%, 15.63%, and 12.62%, respectively.

With respect to Q3 formulated at the outset of this section, we conclude that *the video stimuli are more predictive than the image stimuli* in the context of the trait detection task. It is also important to highlight that *the combined image and video stimuli yield more accurate predictions* than either the image or video stimuli considered individually.

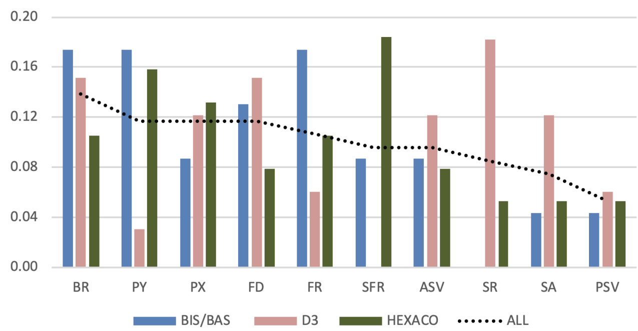

Figure 3: Relative importance of features for each model

Predictive Stimuli and Features

Finally, we turn to Q4 that deals with the predictive value of the extracted features. For this, we calculate the normalised selection frequency for each of the ten extracted features, for the three personality models (D3, BIS/BAS, and HEXACO) separately. The rationale is that the more frequently selected features have more important predictive value; thus, the higher the value, the more important the feature is. Specifically, we counted how many times a feature was selected in a feature subset for a trait from each personality model and normalised this value per model. The results are shown in Figure 3.

Considering the detection of D3's traits, we note the dominance of the saccade rate (SR), which is supported by previous research that discovered links between reduced saccade movements and several facets of psychopathy [6]. Fixation duration (FD) is the second most used feature, which may be linked to the reduced saccade movements. Additionally, we highlight the importance of blink rate (BR). This supports the work of [39] which found that those with psychopathic traits tended to display unusual blink responses. The least selected feature is vertical pupil size (PY), in line with the findings of [7], which showed no relations between the subjects' psychopathy scores and pupil diameter changes in response to affective stimuli. However, the trend is not supported by PX data. No links to other D3 components were found in prior literature.

Looking into the BIS/BAS predictions, we highlight the importance of the blink rate (BR) feature. This aligns with the findings of [21] that found significant correlations between the BAS scores and eye blink responses. In addition, we found that fixations (FR) are predictive of BIS/BAS traits, which is supported by previous research that linked the number and duration of fixations to BAS Drive and BAS Fun Seeking [44]. Vertical pupil dilation (PY) is also an important feature, as explained by the strong association between pupillary reactivity and both fear and anxiety [55], components of BIS.

Predictions of the HEXACO traits are dominated by the Saccade-Fixation Ratio (SFR), in line with [44], which associated fixations with Extraversion, Agreeableness, and Neuroticism (inverse to Resiliency), although the individual FR and SR features are not selected often. Pupil size contributes to the second most selected feature (both PX and PY), which is aligned with early works that studied traits like Extraversion and Neuroticism (in this case, Resiliency) [15].

Revisiting Q4 we conclude that *different features best informed the predictions of different personality models*. Namely, saccade rate was most predictive of D3, blink rate and pupil size – of BIS/BAS, and saccade-fixation rate – of HEXACO traits. Overall, blinks and pupil size were found to be the most predictive eye-tracking data features.

5 DISCUSSION

In this work we developed a framework for predicting human personality traits using physiological responses to external stimuli. We found that the Naive Bayes algorithm, in conjunction with feature selection, substantially outperformed other machine learning algorithms. Seven traits were predicted with accuracies greater than 90%. Comparing the image and video stimuli, we found that the latter performed better, while their combination improved the predictive accuracy.

It is important to analytically compare our results to the closest line of work on personality detection using physiological signals [1, 49, 53]. Our findings largely align with the main observations made in those papers; however, it is important to highlight the notable advantages of our work:

– *Personality traits.* Previous works focused on the predictions of the Big-5 traits only. In our work, not only we use the more recent HEXACO model which introduced an additional trait of Honesty to the Big-5, but we also complement this with traits from the D3 and BIS/BAS models. Altogether, our method is capable of predicting more than three times the number of traits predicted in [1, 49, 53]. These offer a more encompassing perspective on human personality and may be useful in practical scenarios, like hiring decisions, particularly for the jobs that would require screening out people scoring high or low on a particular trait, or interface customisation.
– *Classification accuracy.* Compared to previous research, our method achieves substantially higher classification accuracy. Specifically, [1, 49, 53] conducted a 2-class classification, whereas our work addresses a 3-class classification (random guess baselines of 50% and 33%, respectively). The F1-scores reported in [1, 49, 53] generally hover between the 0.5 and 0.8 marks, while our results achieve accuracy levels as high as 95%, with the mean accuracy of 85.71% across the 16 traits being predicted. Hence, previous work achieves a 30-40% improvement, while we achieve more than 250% improvement over the baseline.

– *Duration of stimuli.* Previous works [1, 49, 53] required the subjects to be exposed to the stimuli for substantially longer periods of time. For example, [53] and [49] used 36 videos, on average 80 seconds long, which brought the duration of the video stimuli to 48 minutes. In the more recent work, they used four longer videos that summed up to 85 minutes. In our case, the image and video stimuli required 9 and 14 minutes, respectively, which is much shorter than the above times. Also, reasonably high levels of classification accuracy were achieved with one type of stimuli only, which would require even less time.
– *Deployed sensors.* In our work we deployed only the SMI ETG eye-tracking sensor, while previous works [1, 49, 53] used a substantially larger range of sensing technologies. Specifically, in all three papers, the authors used the EEG (brain signals), ECG (heart rate), and GSR (skin conductivity) sensors, as well as the EMO face feature tracker. These sensing technologies are more complex, often more expensive, and are less available than the SMI ETG eye-tracker that was deployed in this work.

Limitations

While our results are promising, there are several limitations that require further attention. The first one refers to the reasonably small sample size. Although the leave-one-out validation with 21 subjects produced solid results, more subjects should be recruited to better understand the results, validate our findings, and replicate them for other sensors and traits.

The second limitation refers to the subject recruitment, not based on any psychological or clinical criteria. Thus, we were unlikely to have subjects on the extreme ends of the scales for some traits, especially the D3 traits, also evident from the range of personality scores in Table 2. Despite the equal-frequency binning of subjects, our results likely reflect general population of normative subjects, and a targeted recruitment is required for validation with extreme trait values.

The third limitation is the use of the equal-frequency binning of the subjects, not based on norms or psychological theories. Given the second limitation, the replication on a larger sample, using norms, whenever they are available, is needed to determine generalisability of the findings. Future research should also employ a full spread of scores on psychological traits of interest in the MLA to increase authenticity of results within predictive models.

The last limitation refers to the “off-the-shelf” nature of the stimuli, sensors, feature extraction, and classifiers. While this can be interpreted as a limitation, e.g., with regards to the signal quality and predictive accuracy, it is also a door-opener for future improvements and a strength. Although we managed to accurately detect personality with these off-the-shelf components, accuracy can be improved by tailoring the components of the framework to the trait prediction task.

6 CONCLUSIONS AND FUTURE WORK

This paper presents our work on objective detection of personality traits using physiological responses to external stimuli. Specifically, we propose a framework, which combines external stimuli that trigger physiological responses and sensing technologies that capture these responses, with machine learning methods for detecting personality traits. We evaluate a specific instantiation of the proposed framework, which uses affective image and video stimuli and eye-tracking data. Our work demonstrates that personality traits can be accurately detected, suggesting possible use in practical applications to supplement the traditional forms of assessment or to provide a possible alternative for tailored human-computer interaction.

Revisiting the research questions, we established that: (i) Naive Bayes was the most accurate classification method, (ii) traits associated with affect were predicted more accurately than traits associated with behaviour and cognition, (iii) video stimuli were more predictive than image stimuli, although best predictions were obtained by combining the two, and (iv) predictive features differed across the models, consistently with previous psychology research. Our findings enhanced prior research by considering a broader range of traits and models, and improving the classification accuracy, while deploying only one non-invasive commercial sensing technology and greatly reducing data acquisition times.

Future research should address the identified limitations including experimenting with a larger cohort of participants and in different scenarios, such as gaming, driving, and more. For psychology practitioners and clinicians, it will be important to validate our method with populations having an established pathology. In addition to eye-tracking signals, other physiological signals such as EEG and GSR, as well as their combinations, should be investigated, which may further improve the results. The individual components of our method (stimuli, sensors, feature extraction, classifiers) may also be refined and tuned in the future, to support the generalisation of our findings. While we managed to establish high levels of accuracy using off-the-shelf components, the performance can be further improved by tailoring the components of the generic framework to the specific trait prediction task.

We should also highlight an important area calling for future research. This refers to a thorough investigation of how each scenario and type of stimulus influences personality detection, as certain scenarios and stimuli can be linked to certain traits stronger than to others. Finally, we would like to study physiological responses beyond personality detection tasks, e.g., for gauging the effect of social media on users. These exciting veins of future work may evolve into cross-disciplinary activities pulling capabilities from human-computer interaction, psychology, sensing technologies, signal processing, machine learning, and physiology.

REFERENCES

[1] Mojtaba K. Abadi, Juan A. Miranda-Correa, Julia Wache, Heng Yang, Ioannis Patras, and Nicu Sebe. 2015. Inference of personality traits and affect schedule by analysis of spontaneous reactions to affective videos. In *International Conference on Automatic Face and Gesture Recognition (FG)*. 1–8.

[2] Daniel R. Ames, Paul Rose, and Cameron P. Anderson. 2006. The NPI-16 as a short measure of narcissism. *Journal of Research in Personality* 40, 4 (2006), 440–450.

[3] Jeromy Anglim, Stefan Bozic, Jonathon Little, and Filip Lievens. 2018. Response distortion on personality tests in applicants: comparing high-stakes to low-stakes medical settings. *Advances in Health Sciences Education* 23 (10 2018), 311–321.

[4] Michael C. Ashton, Kibeom Lee, Marco Perugini, Piotr Szarota, Reinout E. De Vries, Lisa Di Blas, Kathleen Boies, and Boele De Raad. 2004. A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *Journal of personality and social psychology* 86, 2 (2004), 356.

[5] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. 2012. The impact of data obfuscation on the accuracy of collaborative filtering. *Expert Syst. Appl.* 39, 5 (2012), 5033–5042.

[6] Sabrina Boll and Matthias Gamer. 2016. Psychopathic traits affect the visual exploration of facial expressions. *Biological Psychology* 117 (2016), 194–201.

[7] Daniel T. Burley, Nicola S. Gray, and Robert J. Snowden. 2017. As far as the eye can see: relationship between psychopathic traits and pupil response to affective stimuli. *PloS One* 12, 1 (2017), 1–22.

[8] Suzanne R. Byrnes. 1996. The effect of audio, video, and paired audio-video stimuli on the experience of stress. *Journal of Music Therapy* 33, 4 (1996), 248–260.

[9] Charles S. Carver and Teri L. White. 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. *Journal of Personality and Social Psychology* 67, 2 (1994), 319.

[10] Fang Chen, Natalie Ruiz, Eric Choi, Julien Epps, M. Asif Khawaja, Ronnie Taib, Bo Yin, and Yang Wang. 2013. Multimodal Behavior and Interaction As Indicators of Cognitive Load. *ACM Transactions on Interactive Intelligent Systems* 2, 4 (2013), 22:1–22:36.

[11] C. Robert Cloninger, Dragan M. Svrakic, and Thomas R. Przybeck. 1998. A psychobiological model of temperament and character. *The development of psychiatry and its complexity* 50, 12 (1998), 1–16.

[12] Hope R. Conte and Robert Plutchik. 1981. A circumplex model for interpersonal personality traits. *Journal of personality and social psychology* 40, 4 (1981), 701.

[13] Ariel L. Del Gaizo and Diana M. Falkenbach. 2008. Primary and secondary psychopathic-traits and their relationship to perception and experience of emotion. *Personality and Individual Differences* 45, 3 (2008), 206–212.

[14] Susan T. Dumais, Georg Buscher, and Edward Cutrell. 2010. Individual differences in gaze patterns for web search. In *Proceedings of the Symposium on Information Interaction in Context*. 185–194.

[15] Hans J. Eysenck. 1983. Psychophysiology and personality: Extraversion, neuroticism and psychoticism. In *Individual Differences and Psychopathology*. 13–30.

[16] Gerry Fahey. 2018. Faking Good and Personality Assessments of Job applicants: A Review of the Literature. *DBS Business Review* 2 (2018), 45–68.

[17] Arik Friedman, Bart P. Knijnenburg, Kris Vanhecke, Luc Martens, and Shlomo Berkovsky. 2015. Privacy Aspects of Recommender Systems. In *Recommender Systems Handbook*. 649–688.

[18] Mihai Gavrilescu and Nicolae Vizireanu. 2017. Predicting the Sixteen Personality Factors (16PF) of an individual by analyzing facial features. *EURASIP Journal on Image and Video Processing* 2017, 1 (2017), 59.

[19] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. 2011. Predicting personality from Twitter. In *Proceedings of the International Conference on Social Computing (SocialCom)*. 149–156.

[20] Jeffrey A. Gray. 1990. Brain systems that mediate both emotion and cognition. *Cognition & Emotion* 4, 3 (1990), 269–288.

[21] Daniel F. Gros. 2011. Startle inhibition to positive-activated compared to neutral stimuli: Variations in self-reported behavioral approach. *Journal of Psychopathology and Behavioral Assessment* 33, 3 (2011), 308–314.

[22] Mark A. Hall. 2000. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*. 359–366.

[23] Mark A. Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explorations* 11, 1 (2009), 10–18.

[24] Sabrina Hoppe, Tobias Loetscher, Stephanie A. Morey, and Andreas Bulling. 2018. Eye Movements During Everyday Behavior Predict Personality Traits. *Frontiers in Human Neuroscience* 12 (2018), 105.

[25] David W. Hosmer Jr, Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied logistic regression*. Vol. 398. John Wiley & Sons.

[26] Peter K. Jonason and Laura Krause. 2013. The emotional deficits associated with the Dark Triad traits: Cognitive empathy, affective empathy and alexithymia. *Personality and Individual Differences* 55, 5 (2013), 532 – 537.

[27] Anthony F. Jorm, Helen Christensen, Alexander S. Henderson, Patricia A. Jacomb, Ailsa E. Korten, and Bryan Rodgers. 1998. Using the BIS/BAS scales to measure behavioural inhibition and behavioural activation: Factor structure, validity and norms in a large community sample. *Personality and Individual Differences* 26, 1 (1998), 49 – 58.

[28] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5802–5805.

[29] Peter J. Lang, Margaret M. Bradley, and Bruce N. Cuthbert. 2008. International affective picture system (IAPS): Affective ratings of pictures and instruction manual, Technical Report. (2008).

[30] Michael R. Levenson, Kent A. Kiehl, and Cory M. Fitzpatrick. 1995. Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology* 68, 1 (1995), 151.

[31] William L. Libby and Donna Yaklevich. 1973. Personality determinants of eye contact and direction of gaze aversion. *Journal of Personality and Social Psychology* 27, 2 (1973), 197.

[32] Kai K. Lim, Max Friedrich, Jenni Radun, and Kristiina Jokinen. 2013. Lying through the eyes: detecting lies through eye movements. In *Proceedings of the Workshop on Eye gaze in Intelligent Human Machine Interaction: Gaze in Multimodal Interaction*. 51–56.

[33] Navonil Majumder, Soujanya Poria, Alexander Gelbukh, and Erik Cambria. 2017. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems* 32, 2 (2017), 74–79.

[34] Gerald Matthews, Ian J Deary, and Martha C Whiteman. 2003. *Personality traits*. Cambridge University Press.

[35] Robert R. McCrae and Paul T. Costa Jr. 1999. A five-factor theory of personality. *Handbook of personality: Theory and Research* 2 (1999), 139–153.

[36] Leslie C. Morey. 2015. *Personality Assessment Inventory (PAI)*. Wiley Online Library.

[37] Carolyn C. Morf and Frederick Rhodewalt. 2001. Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry* 12, 4 (2001), 177–196.

[38] Tim Morris, Paul Blenkhorn, and Farhan Zaidi. 2002. Blink detection for real-time eye tracking. *Journal of Network and Computer Applications* 25, 2 (2002), 129–143.

[39] Christopher J. Patrick, Margaret M. Bradley, and Peter J. Lang. 1993. Emotion in the Criminal Psychopath: Startle Reflex Modulation. *Journal of Abnormal Psychology* 102, 1 (1993), 82–92.

[40] Delroy L. Paulhus and Kevin M. Williams. 2002. The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality* 36, 6 (2002), 556–563.

[41] Lin Qiu, Han Lin, Jonathan Ramsay, and Fang Yang. 2012. You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality* 46, 6 (2012), 710–718.

[42] George E. Raptis, Christos A. Fidas, and Nikolaos M. Avouris. 2017. On implicit elicitation of cognitive strategies using gaze transition entropies in pattern recognition tasks. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1993–2000.

[43] John F. Rauthmann. 2013. Investigating the MACH–IV with item response theory and proposing the trimmed MACH. *Journal of Personality Assessment* 95, 4 (2013), 388–397.

[44] John F. Rauthmann, Christian T. Seubert, Pierre Sachse, and Marco R. Furtner. 2012. Eyes as windows to the soul: Gazing behavior is related to personality. *Journal of Research in Personality* 46, 2 (2012), 147–156.

[45] John F. Rauthmann and Theresa Will. 2011. Proposing a multidimensional Machiavellianism conceptualization. *Social Behavior and Personality: An International Journal* 39, 3 (2011), 391–403.

[46] Richard M. Ryckman. 2012. *Theories of Personality*. Cengage Learning.

[47] Gerard Saucier. 2009. Recurrent personality dimensions in inclusive lexical studies: Indications for a Big Six structure. *Journal of Personality* 77, 5 (2009), 1577–1614.

[48] Alexandre Schaefer, Frédéric Nils, Xavier Sanchez, and Pierre Philippot. 2010. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion* 24, 7 (2010), 1153–1172.

[49] Ramanathan Subramanian, Julia Wache, Mojtaba Abadi, Radu Vieriu, Stefan Winkler, and Nicu Sebe. 2016. ASCERTAIN: Emotion and Personality Recognition using Commercial Sensors. *IEEE Transactions on Affective Computing* 9, 2 (2016), 147–160.

[50] Isabel Thielmann, Benjamin E. Hilbig, Ingo Zettler, and Morten Moshagen. 2016. On measuring the sixth basic personality dimension: a comparison between HEXACO honesty-humility and Big Six honesty-propriety. *Assessment* 24, 8 (2016), 1024–1036.

[51] Amit Tiroshi, Shlomo Berkovsky, Mohamed Ali Kaafar, Terence Chen, and Tsvi Kuflik. 2013. Cross Social Networks Interests Predictions Based Ongraph Features. In *Proceedings of the ACM Conference on Recommender Systems*. 319–322.

[52] Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A survey of personality computing. *IEEE Transactions on Affective Computing* 5, 3 (2014), 273–291.

[53] Julia Wache, Ramanathan Subramanian, Mojtaba Khomami Abadi, Radu-Laurentiu Vieriu, Nicu Sebe, and Stefan Winkler. 2015. Implicit User-centric Personality recognition based on physiological responses to emotional videos. In *Proceedings of the International Conference on Multimodal Interaction*. 239–246.

[54] Michael Wai and Niko Tiliopoulos. 2012. The affective and cognitive empathic nature of the dark triad of personality. *Personality and Individual Differences* 52, 7 (2012), 794 – 799.

[55] Tara L. White and Richard A. Depue. 1999. Differential association of traits of fear and anxiety with norepinephrine- and dark-induced pupil reactivity. *Journal Of Personality And Social Psychology* 77, 4 (1999), 863–877.

[56] Anne-Kathrin Wilbers, Alina Vennekoetter, Moritz Koster, Kai-Christoph Hamborg, and Kai Kaspar. 2015. Personality traits and eye movements: an eye-tracking and pupillometry study. In *Proceedings of the European Conference on Eye Movements*. 269.

[57] Ian H. Witten, Eibe Frank, and Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann Publishers.

[58] Jie Xu, Yang Wang, Fang Chen, and Eric Choi. 2011. Pupillary response based cognitive workload measurement under luminance changes. In *IFIP Conference on Human-Computer Interaction*. 178–185.

[59] Zhefan Ye, Yin Li, Alireza Fathi, Yi Han, Agata Rozga, Gregory D. Abowd, and James M. Rehg. 2012. Detecting eye contact using wearable eye-tracking glasses. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 699–704.

[60] Michael J. Zickar and Fritz Drasgow. 1996. Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement* 20, 1 (1996), 71–87.

[61] Matthias Ziegler, Carolyn MacCann, and Richard Roberts. 2011. *New perspectives on faking in personality assessment*. Oxford University Press.

[62] Lisa M. Pytlik Zillig, Scott H. Hemenover, and Richard A. Dienstbier. 2002. What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin* 28, 6 (2002), 847–858.