Article

Associations between the Avatar Characteristics and Psychometric Test Results of VK Social Media Users

Valeriia Stoliarova *,†, Fedor Bushmelev † and Maxim Abramov †

St. Petersburg Federal Research Center of the Russian Academy of Sciences, 39, The 14th Line V.O,
St. Petersburg 199178, Russia; fvb@dscs.pro (F.B.); mva@dscs.pro (M.A.)
* Correspondence: vfs@dscs.pro; Tel.: +7-812-328-33-37
† These authors contributed equally to this work.

Abstract: Online social media has an increasing influence on people's lives, providing tools for communication and self–representation. People's digital traces are gaining attention as a reflection of their personality traits, enhancing the personality computing tasks in various areas. This study aims at the identification of statistical associations between psychometric scores from three questionnaires—the Big Five Inventory, Plutchik's Lifestyle Index and the Eysenck Personality Questionnaire—and a set of graphical features of avatar images from the VK online social media that include the pixel characteristics from the HSV and RGB color models and the number of persons and faces depicted in an avatar. The problem is considered from the statistical point of view. The dependency between psychometric scores and the number of faces/persons is assessed with the Kruskal–Wallis test with Dunn test pairwise comparisons. The color-pixel characteristics that are associated with the psychometric scores are selected with several fits of the regularized regression with L_2 and MCP penalties. The data for the study were collected via a specially developed application for the online social media platform VK. The results of the analysis support existing research on how colors express personality and discover certain color-pixel image characteristics that could be used for personality computing models.

Keywords: online social network; digital traces; Big Five Inventory; Plutchik's Lifestyle Index; Eysenck Personality Questionnaire; computer vision; elastic net

MSC: 62P15

1. Introduction

Colors and images accompany a person's everyday life, reflecting his, her or their personality. Its characteristics are often perceived subconsciously during social interaction, affecting the opinions and actions of a person, and online social media (OSM) is not an exception. Various decision support tools that analyze information from user's profile are demanded nowadays, especially in the areas of career counseling and marketing. A user's profile in turn contains a lot of user-generated information—textual and graphical, self-reported (like posts or biographies) and technical (like dates of posts or number of likes)—forming person's digital trace. Such graphical digital traces have been studied in several contexts of real-world behaviors and are the focus of this study. For example, the OSM profile avatar influences the decision to hire [1], and the use of a high-quality image with a face depicted increases the involvement of users in online activities [2,3]. Such a positive response to published content is influenced not only by the quality of the media content [4] but also by the color characteristics of the image [2]. At the same time, different colors tend to cause different emotional responses among users of online social networks [5]. The relationship between graphical digital traces and a user's personality traits is the core for such associations.

Thus, one of the problems in the area of digital trace research is establishing this relationship [1,6]. Some associations were revealed for photo or video material published on Instagram [7] or TikTok [8], where graphical information is the main source for communication. A set of posts for a given period of time usually serves as a base for such research. The avatars, in turn, allow for concise, intuitive self-representation. The relations with the personality traits were found not only for the OSM photo avatars [1,9] but for people's avatars in other virtual activities [10,11]. In the area of psychological research, it was shown that avatars of individuals who are prone to depression often depict a single person and are biased toward the gray spectrum [12]. Similar relationships between the graphic content of a user's online social network profile and his, her or their psychological peculiarities can be traced for other conditions as well [12–14].

As noted earlier, such quantified relationships between an avatar's graphical characteristics and personality traits can be used in automated decision support systems [15] in various fields, such as marketing and psychological and career counseling. For example, there exist automated systems for determining depression [12,16]. Moreover, these associations can be viewed as a part of the user profiling process that plays an important role in assessing an estimation of the users' protection from social engineering attacks [17–19].

This study contributes to the identification of the avatar's graphical characteristics that may reflect various personality traits. The results of three psychometric tests were investigated: the Big Five Inventory, Plutchik's Lifestyle Index and the Eysenck Personality Questionnaire. The main research question is the following: *is there significant statistical association between the semantic and color-pixel characteristics of an OSM user avatar image and his, her or their psychometric scores from three tests: the Big Five Inventory, Plutchik's Lifestyle Index and the Eysenck Personality Score? What personality dimensions are mirrored in a user's profile avatar in terms of statistical associations?* The answers to these questions lay the foundation for personality computing models from a user's online profile data.

We note that the Big Five Inventory is often used for the analysis of digital traces [20,21], and therefore the primary question is intended to add some evidence to existing research on this topic from a novel dataset. Two latter psychometric scores rarely arise among the scientific literature devoted to studying person's representation in online social networks, and the current study also contributes to the novel discussion of how the personality traits measured by those tests are reflected in a person's digital traces [22]. Therefore, the secondary research question is the following: are the semantic and color-pixel characteristics of an OSM user's avatar image related to the scores reflecting his, her or their emotional management, as measured by Plutchik's Lifestyle Index and the Eysenck Personality Questionnaire?

1.1. Analysis of Relevant Studies

The Big Five Inventory serves as a background for plenty of research on the relationship between graphical digital traces and user's personality traits. In this section, we list some of the obtained results in this direction regarding the color-pixel and semantic characteristics of an image.

Extraversion is associated with the brightness and colorfulness of the images in a user's profiles [7,9] along with the multiple faces depicted [9]. The authors of [23] also found the positive correlation of extraversion score with the color diversity of the images [24]. *Agreeableness* is correlated with colorfulness [9]. Agreeable people prefer avatars with faces. Moreover, images posted by agreeable people are likely to have medium contrast [24]. *Conscientiousness* is associated with avatars with unique, clearly visible faces [9]. *Neuroticism* is anticorrelated with colorfulness and correlated with the absence of faces in an image [9]. A negative association is also found with color harmony [23]. It is also correlated with the value characteristic of an image [24]. *Openness to experience* is associated with photos high in contrast, sharpness and saturation and low in blur [9]. This characteristic is also negatively associated with color diversity and color harmony [23]. Correlations were revealed with the value and saturation characteristics of images in [24]. This score is also correlated with the number of persons and number of faces depicted [24].

While the Big Five Inventory is frequent in graphical digital trace studies, we found no papers that explore the reflections of the emotion–related scores in the color-pixel and semantic contents of an image. Most studies concerning this question are dedicated to the extraction of emotions from posts and text. In [9], the color emotion characteristics of an image were used along with other image characteristics. Thus, the problem of identification of relationships between various personality traits depicted in the results of psychological tests (especially the person's emotion management traits) and the graphical parameters of the user's avatar arises. The avatar photos of users of the largest OSM in Russian-speaking countries—VK—are the focus of this study. We note that there is no similar research for this segment of users.

1.2. Outline of the Study

This study aims at the identification of statistical associations between the psychometric scores from three questionnaires and a set of graphical features of the avatar images of the OSM VK users. The study design does not suppose an in-depth psychological interpretation, but the statistical relationship between instances of interest is considered. Nonetheless, the results are useful for future research planning, as they allow one to narrow down the range of features to consider when building models. This point may be crucial when the available data are limited due to cost or time constraints.

This study concentrates on two levels of image contents: the pixel level and the semantic level. The former contains the color-pixel characteristics from the RGB and HSV models, while the latter involves analysis of the semantic content of an image and includes the number of persons and number of faces depicted. Two research hypotheses were formulated correspondingly:

H1. *The psychometric scores from the Big Five Inventory, Plutchik's Life Style Index and the Eysenck Personality Questionnaire vary between groups of users with different numbers of faces or persons depicted in the avatar of an OSM VK profile.*

H2. *The psychometric scores from the Big Five Inventory, Plutchik's Life Style Index and the Eysenck Personality Questionnaire can be predicted from the set of color-pixel features of a user's avatar image, and the best explaining subsets of features for prediction vary between scores.*

2. Materials and Methods

This study used the data collected with a specially developed application in the VK social media platform. VK is an ecosystem of products and services with a wide range of functions for users, content creators and advertisers, similar to other popular online social media. It consists of pages, which are a kind of a blog, portfolio or personal brand, where users can publish various content. VK is the largest OSM in Russia, with a significant portion of users in the Commonwealth of Independent States (CIS countries), some Eastern European countries and the United States. Currently, more than 100 million unique users from all over the world visit VK every month, and there are around 50 million unique users daily.

The application used for data collection was developed for scientific purposes. It includes a variety of psychometric tests and is accessible for all VK users. This app does not involve any psychological counseling. Only the test results and their descriptions were available to users. The tests could be passed only after viewing and accepting the informed consent form, which stated the purposes of data collection and how they would be used (namely, only for scientific purposes after proper depersonalization). After a person finished one of the tests, the app gathered open access information from the user's profiles, like avatar images, biographies and lists of subscriptions.

The study was totally survey based, and no recruiting was performed. The sample was formed by convenience from users that finished all three questionnaires from the 26 July 2022, to the 8 August 2022, and had an avatar image in their profiles. Our goal was

to discover how the personality traits were expressed in the profile's avatar image, and therefore two objects were retrieved using the app: the results of the psychometric tests and the user avatars. Its semantic and color-pixel features were extracted with the use of existing models as discussed in further subsections. The data collection and processing workflow is visualized in Appendix C.

2.1. Psychometric Questionnaires

2.1.1. The Big Five Inventory

The Big Five Inventory (BFI) reflects a dispositional personality model that characterizes a person's adaptation to a social environment. This model is biologically inspired and reflects the personality in terms of five independent factors: "extraversion" (referred to as BF1 in this paper), "agreeableness" (BF2), "conscientiousness" (BF3), "neuroticism" (BF4) and "openness to experience" (BF5). The current study uses the adaptation of the BFI for CIS countries in the form of a five-factor personality questionnaire (5PFQ) [25]. It consists of 75 statements with a Likert-scaled answer, with each level being associated with a certain score. The cumulative score was obtained for questions relevant to the mentioned factors, ranging from 15 to 75 points, with extreme values indicating the tendency for a particular disposition.

2.1.2. Plutchik's Lifestyle Index

Plutchik's Lifestyle Index is rarely considered in the context of digital footprint analysis, although it arises in the area of human risk analysis. This index reflects a person's abilities for emotion management. It was developed in 1979 on the basis of the psychoevolutionary theory of R. Plutchik and the structural theory of personality of G. Kellerman [26]. The results of this test reflect the following psychological defense mechanisms:"denial" (referred to in the following as PD1), "regression" (PD2), "compensation" (PD3), "rationalization" (PD4), "hypercompensation" (PD5), "displacement" (PD6), "projection" (PD7) and "substitution" (PD8). In this, study its Russian language adaptation was used [27], which consists of 97 "yes or no" statements. The score for a certain defense mechanism is the number of positively answered questions that correspond to this factor. For convenience, these counts were transformed into percentages.

2.1.3. The Eysenck Personality Questionnaire

The Eysenck Personality Questionnaire (EPQ), also known as psychoticism, extraversion and neuroticism (PEN) or the test for temperament [28], was also inspired by a biologically based theory of personality. The questionnaire consists of 101 "yes or no" statements, which belong to four scales reflecting the following personality traits: "extraversion or introversion" (referred to in the following as T1), "neuroticism or stability" (T2), "psychoticism or socialization" (T3) and "lie or social desirability" (T4). In this study, the percentage-based scores of these factors were used also.

2.2. Graphical Characteristics of the Avatar Image

Graphical digital traces of an OSM user consist of various content: posts and reposts with pictures or videos, emoticons, photos and the user's avatars. The latter source is the focus of this study. In order to check the hypotheses of the study, a complex set of characteristics traditional for the field of computer vision was retrieved from the avatar images [7,29,30]. This included the quantitative and stylometric color characteristics and counts of persons and faces depicted. An example of the extracted data is shown in Figure A1.

The first group of attributes contains the color–pixel characteristic of an image from the RGB and HSV color models and the stylometric color characteristic proposed in [31], colorfulness.

The features of the red–green–blue (RGB) color model were extracted using the OpenCV library for Python [32]. We calculated the total number of pixels that belonged

to three equal regions of R, G or B channels histogram, indicating "high", "medium" and "low" values, and then normalized this number to the total pixel count. An example of the Red channel histogram divided inti three equal regions is presented on Figure A2. Thus, nine characteristics were defined: "r_low", "r_mid" and "r_high" for the red spectrum, "g_low", "g_mid" and "g_high" for the green spectrum and "b_low", "b_mid" and "b_high" for the blue spectrum.

The hue–saturation–value (HSV) model is an interpretable color model designed in 1978 by Alvy Ray Smith to correspond with human vision perception of the color attributes [33]. The HSV color model is often used by visual arts specialists such as artists, photographers and designers. The features of the HSV model for avatar images were extracted with the OpenCV library as well. Values for further analysis were calculated in a way similar to the calculation of the RGB image features by dividing the histogram of hue, saturation and value components of all pixels into a certain number of intervals. For the saturation component, three equal ranges were considered: pale shades ("s_low"), medium or solid colors ("s_mid") and saturated shades ("s_high"). For the value component, which expresses the brightness of a pixel, 5 ranges were defined with the proportion 2/5/7/5/2: black shades ("v_black"), dark shades ("v_shadow"), medium tones ("v_expose"), bright shades ("v_highlight") and accents ("v_whites"), respectively. The hue component reflects the color tone, and thus the grouping was defined by the main colors: red ("h_red"), blue ("h_blue"), green ("h_green"), yellow ("h_yellow"), cyan ("h_cyan") and magenta ("h_magenta"). Thus, 23 quantitative color characteristics were obtained, ranging from 0 to 1.

The colorfulness [31] characteristic is a combination of the average value and the standard deviation of the difference between the RGB color channels for each pixel. There exist several approaches for measuring this characteristic. The colorfulness reflects the general deviance of the image characteristics from the grayscale. It was computed using the OpenCV package for Python after calculating the RGB features of an image. To obtain the mean and standard deviation over the entire image, the Euclidean norm was calculated for two pairs of channel values, where the former was defined as the difference between the red and green channels and the latter was the difference between the half-sum of the red-green channel and the blue channel.

The second aspect of the image contents is its semantics. The task of pattern recognition is quite extensive, as each image could contain dozens or even hundreds of objects. Given the specifics of the data (we considered only the avatars of a user's OSM profile), we assumed that a significant part of the images would contain an image of a person or several persons. Therefore, two semantic characteristics of the avatar image were estimated with the use of existing software: the number of persons (Yolov4 neural network [34]) and the number of faces (MTCNN [35]) depicted. The PyTorch model YOLOv4-large gives the coordinates of the area with people and the corresponding confidence score. In this study, we marked areas where the score was above 50% as containing a person and used the number of such areas in the statistical analysis as the variable *person_count*. The number of faces (variable *face_count* in further analysis) of the image was obtained in similar way with the use of the cascade model of convolutional neural networks (MTCNN).

We intended to use two variables with similar semantics (number of persons and number of faces in the avatar image) to lower the bias induced by imprecision of the feature extraction techniques.

2.3. Statistical Methods

Two statistical procedures were used to check proposed Hypotheses H1 and H2: the Kruskall–Wallis test and regularized regression.

To quantify the dependency between the measured personality traits and the number of faces or persons, the nonparametric approach was chosen. This approach allows for verification of stochastic dominance of distributions between groups of interest rather than checking the equivalence of the means. This is important for justification of the hypothesis

about psychometric traits that are measured by the sum of the scores. Such variables are intrinsically discrete with a large number of states. Medians may be more appropriate for comparing them. In order to identify how the scores from the three tests of interest varied between groups of users without faces or persons, users with exactly one face or person and users with several faces or persons depicted in their avatars, the following statistical procedure was applied:

1. For all psychometric scores, we checked if the variances between the mentioned groups were homogeneous with the Brown–Forsythe test [36]. This test relies on the F–statistic for a median-centered response variable. If the variances can be considered equal, then the underlying distributions can be considered identically scaled, and the Kruskal–Wallis test indicates differences the true difference in the medians.
2. We performed the Kruskal–Wallis test to compare the psychometric scores between groups with various numbers of faces or persons depicted in the avatar images.
3. We performed the post–hoc Dunn test of multiple comparisons.

The second research hypothesis concerns revealing the associations between the color characteristic of the avatar image and the user's psychometric scores. Linear regression is a first-choice technique for this task, however it could not be applied directly since color-pixel characteristics cannot be assumed to be independent. Moreover, there are a lot of features that may be used for prediction. In this paper, we considered 24 dependent characteristics collected from 548 users, and more features could be extracted. Therefore, we needed to identify the best subset of features that may be used for prediction of the variables of interest. In this case, the regression coefficients and their standard deviations estimated with the ordinary maximum likelihood may be unreliable and have very large values. In such cases, penalized likelihood functions may be used for fitting the regression model [37]. There exist various penalties, each of them having certain properties. Adding the ℓ_1-norm restriction ($\sum_{i=1}^{p} |\beta_i| \leq t$, where p is the number of dependent variables) on the regression coefficients, led to least absolute shrinkage and selection operator (LASSO) regression and allowed setting some coefficients to strictly zero, thereby solving the problem of choosing the best subset of factors. The L_2-norm restriction ($\sum_{i=1}^{p} \beta_i^2 \leq t^2$) was the base for the ridge regression that shrank all coefficients to zero and worked well for the multicollinearity problem. The restrictions may be combined in the elastic net regression model.

The ℓ_1- and L_2-norm restrictions are concave and easy to use for computational purposes, though they may produce biased coefficients while fitting. There exist approaches to overcome this problem, based on the nonconcave restrictions on the regression coefficients, like the minimax concave penalty (MCP) and smoothly clipped absolute deviation (SCAD) [38]. These regularizations allowed us to build estimates with good statistical properties, though the calculations may be more complex and computationally intensive. It was shown that the estimates of the regression coefficients obtained with the MCP or SCAD regularized regressions had "the oracul property" (i.e., they were asymptotically consistent) [38]. The elastic net idea can be applied with the nonconcave restrictions on the regression coefficients. In this case, the additional tuning parameter is added to the model to balance the introduced restrictions. In order to make feature selection with more stable coefficients, we used both the MCP and L2 penalties. In this case, the target function is

$$Q_{\lambda,\gamma,\alpha}(\beta) = \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i\beta)^2 + \alpha \sum_{j=1}^{p} p_{\lambda\gamma}(|\beta_j|) + (1 - \alpha) \sum_{j=1}^{p} \beta_j^2, \tag{1}$$

where n is the sample size, y_i is the dependent variable value for individual i (psychometric score), $x_i = (x_{i1}, \ldots, x_{ip})$ are the independent factor values (graphical features), $\beta = (\beta_1, \ldots, \beta_p)$ are the regression coefficients and $p_{\lambda\gamma}$ represents functions of the MCP form with the parameters λ and γ. As mentioned earlier, α is a tuning parameter that indicates the balance between two restrictions. If it gets closer to zero, then more coefficients become nonzero (i.e., the regression behaves more like ridge regression). In this study, we explored the sets of nonzero regression coefficients for various values of α.

2.4. Limitations

Among the limitations of this study, first of all, we mention the data collection method: the convenience sampling. Data were collected among those individuals who decided to use the application that was freely available on the VK platform. Therefore, the statistical conclusions have limited possibilities of generalization to the population level. Nonetheless, the results of this study are interesting for the field of graphical digital trace modeling as they reveal trends in dependencies between personality traits and image characteristics for a certain group of active OSM users.

Another limitation concerns the main statistical method used. The regularized regression requires standardization of the initial data to the interval 0–1 and therefore does not allow a straightforward interpretation of the obtained coefficients. Although the RGB and HSV characteristics of an avatar image were normalized to a $[0, 1]$ interval during the feature extraction step, the colorfulness characteristic was not. Thus, although the trends of relationships between the graphical characteristics of the image and the results of the psychometric tests have been established, additional statistical modeling is required.

This study relies on the classical computer vision features, like the frequencies of RGB channels, which did not have immediate interpretations because they were considered separately. The definition of the three components was data-driven, based on the histogram of each component. The low, medium and high groups of the saturation component were defined from data too. Though the value component's discretization is common in computer vision practice, additional research may be needed in order to validate the 2/5/7/5/2 ratio used in this study.

3. Data Analysis

With the use of the application, the data on 548 social media users, who completed the proposed psychological questionnaires, were gathered. After test completion, the profile avatar image was extracted only for open profiles. In most cases (463 observations), the user's avatar was the same for the moments of completion of all three tests. In other cases, the actual avatar for the moment of Plutchik's Lifestyle Index completion was used for the analysis. R software, v. 4.3.1, [39] was used for the statistical modeling.

We note that the study design does not suppose an interaction between researchers and participants, and belongs to the class of descriptive survey-based studies. All information about the participants comes from their publicly available VK profiles, and due to the study design, no means of identification of the true age or sex were intended. The demographics of the profiles are described below. There were 363 female profiles and 165 male profiles, where 306 profiles were missing information on the user's age. The mean age mentioned in the profiles from the sample was 27.7 years, and the median was 22. Eight users mentioned that their age was above 95, and if we consider those observations as outliers, the mean age of the sample will be 24.9 years, with a similar age distribution between the male and female profiles. This information should be used with caution as users can put incorrect information in their profiles.

In the first step of data analysis, Hypothesis H1 was investigated To test this hypothesis, the counts of faces or persons in the avatar images were transformed to a factor with levels: no faces (persons), exactly one face (person) and several faces (persons). These levels were denoted as "0", "1" and "2" respectively. It is interesting to note that the number of persons and number of faces depicted on the avatar image differ in some cases (see Table 1). This fact may be associated with the peculiarities of the methods that were used for feature extraction.

Table 1. Contingency table for the number of faces and number of persons depicted in the avatars in the collected dataset.

		Number of Faces		
		0	**1**	**2**
	0	188	161	27
Number of persons	1	20	110	15
	2	1	12	14

First, the homogeneity of variances was examined with the Brown–Forsythe test. The results are presented in Table 2. This indicated that nearly all cases were homoscedastic except for BF1 and T1 (marked in bold). In those cases, the Kruskal–Wallis analysis would be valid, though it indicated some shift in distributions, which was not only in the medians.

Table 2. Results of the Brown–Forsythe test for comparison of variances and the Kruskal–Wallis test for stochastic dominance between groups with different numbers of faces or persons depicted in the avatar for psychometric scores in interest.

Psychometric Score	Test	Person Count	Face Count
BF1	BF	**5.14 ($p < 0.05$)**	**5.82 ($p < 0.05$)**
	KW	**7.73 ($p < 0.05$)**	**10.63 ($p < 0.05$)**
BF2	BF	0.69 ($p = 0.51$)	0.11 ($p = 0.89$)
	KW	1.40 ($p = 0.50$)	0.41 ($p = 0.81$)
BF3	BF	2.58 ($p = 0.09$)	**3.60 ($p < 0.05$)**
	KW	**7.45 ($p < 0.05$)**	**7.48 ($p < 0.05$)**
BF4	BF	2.76 ($p = 0.07$)	1.83 ($p = 0.16$)
	KW	4.37 ($p = 0.11$)	3.70 ($p = 0.16$)
BF5	BF	1.31 ($p = 0.27$)	2.83 ($p = 0.06$)
	KW	2.20 ($p = 0.33$)	5.49 ($p = 0.06$)
PD1	BF	2.13 ($p = 0.12$)	3.12 ($p = 0.05$)
	KW	4.81 ($p = 0.09$)	6.22 ($p = 0.04$)
PD2	BF	0.77 ($p = 0.47$)	1.03 ($p = 0.36$)
	KW	1.35 ($p = 0.51$)	1.78 ($p = 0.41$)
PD3	BF	0.14 ($p = 0.87$)	0.15 ($p = 0.86$)
	KW	0.45 ($p = 0.78$)	0.35 ($p = 0.84$)
PD4	BF	1.10 ($p = 0.34$)	0.13 ($p = 0.87$)
	KW	2.21 ($p = 0.33$)	0.18 ($p = 0.91$)
PD5	BF	0.31 ($p = 0.74$)	0.64 ($p = 0.53$)
	KW	0.26 ($p = 0.88$)	1.51 ($p = 0.47$)
PD6	BF	2.83 ($p = 0.07$)	2.87 ($p = 0.06$)
	KW	4.78 ($p = 0.09$)	5.67 ($p = 0.06$)
PD7	BF	0.56 ($p = 0.57$)	0.33 ($p = 0.72$)
	KW	1.21 ($p = 0.54$)	0.38 ($p = 0.83$)
PD8	BF	0.10 ($p = 0.91$)	0.52 ($p = 0.59$)
	KW	0.53 ($p = 0.78$)	1.19 ($p = 0.55$)
T1	BF	2.25 ($p = 0.11$)	**7.57 ($p < 0.05$)**
	KW	5.45 ($p = 0.07$)	**13.66 ($p < 0.05$)**
T2	BF	0.02 ($p = 0.98$)	0.14 ($p = 0.87$)
	KW	0.13 ($p = 0.94$)	0.09 ($p = 0.96$)
T3	BF	1.50 ($p = 0.23$)	0.82 ($p = 0.44$)
	KW	3.60 ($p = 0.17$)	1.85 ($p = 0.40$)

The Kruskal–Wallis test results are also presented in Table 2. There was a significant effect from *the number of persons* on the BF1 score (KW test statistic = 7.73, p = 0.02 and $\epsilon^2 = 0.0141$) and BF3 score (KW test statistic = 7.45, p = 0.02 and $\epsilon^2 = 0.0136$) at the 0.05 significance level. The Bayesian factor indicates substantial evidence only for dependency on the number of persons and BF1 (Bayesian factor = 3.27). There was a significant effect from *the number of faces* on the BF1 score (KW test statistic = 10.63, $p < 0.01$ and $\epsilon^2 = 0.0194$) and BF3 score (KW test statistic = 7.48, p = 0.02 and $\epsilon^2 = 0.0137$) at the 0.05 significance level and the T1 score (KW test statistic = 13.66, $p < 0.01$ and $\epsilon^2 = 0.025$). The Bayesian factor indicates substantial evidence only for dependency on the number of faces and the extraversion scores for BF1 (Bayesian factor = 3.51) and T1 (Bayesian factor = 14.01). No significant effect of the semantic content of the avatar image on the psychological defense scores was identified.

As for the post-hoc pairwise comparisons, the Dunn test indicated stochastic dominance of scores BF1 and BF3 between the group with exactly one person and group without persons in their avatar images. Similarly, the distributions of BF1, BF3 and T1 for the group with exactly one face stochastically dominated the distribution of those scores for the group without any faces.

From this analysis, the following patterns could be revealed (refer to Figures A3–A5):

- The users with exactly one face (person) in their avatars tend to have higher BF1 ("extraversion"), T1 ("extraversion") and BF3 ("conscientiousness") scores.
- The difference of the BF3 score ("conscientiousness") between the group without faces (persons) and the group with exactly one face (person) in their avatars needs further investigation, as the Bayesian factor for this dependency approaches one.
- The absence of a face or person in the avatar image indicates lower scores for the extraversion of the user.

Thus, we conclude that the number of faces and the number of persons in avatars can serve as a predictor of the extraversion scores for the Big Five Inventory and the Eysenck Personality Questionnaire. No statistical associations were revealed for dependency between the psychological defense styles and image semantic content.

The second step involved analysis of the dependencies between the color-pixel characteristics of the avatars and the psychometric scores in terms of linear regression. Figure A6 represents the correlation structure of the RGB and HSV components. As we supposed earlier, many features were dependent, making it difficult to apply the classical regression techniques directly. Therefore, we used a regularized regression technique for identification of the best subset of features. We relied on the model with two restrictions: the L_2-norm restriction in order to manage the multicollinearity and the MCP restriction in order to make feature selection. As this method is stochastic, the regression model was fitted 50 times for every test score, and the frequency of the coefficient being nonzero was calculated. We also used several values for the tuning parameter α (0.05, 0.01, 0.005, 0.003). The closer the value of α came to zero, the more weight the L_2 restriction gained, and therefore more variable coefficients were nonzero. For λ, we used the default sequence of equidistant on log-scale values. The statistical modeling was performed with the use of the ncvreg package [38].

Along with analysis of the frequencies of appearance of nonzero coefficients in 50 fits of a penalized regression, the stability of a sign of those coefficients was assessed. We calculated the number of times each coefficient had positive or negative values in those fits. It turned out that the coefficient sign was stable, and the results can be found in Table 3.

Table 3. Variables that may serve as predictors in linear models for various psychometric scores with their signs.

	RGB Model	HSV Model: Tones	HSV Model: Brightness	HSV Model: Saturation
BF1: "extraversion"	Red (r_low, negative and r_mid, positive)	Red (h_red, positive); Green (h_Green, negative) and Blue (h_blue, negative)		
BF2: "agreeableness"	Red (r_mid, positive) and Green (g_low, negative and g_mid, positive)		Middle (v_expose, positive)	
BF3: "conscientiousness"	Red (r_mid, positive); Green (g_low, negative and g_mid, positive) and Blue (b_high, positive)	Blue (h_blue, positive) and Cyan (h_cyan, positive)	Middle and low (v_expose, positive and v_shadow, negative)	
BF4: "neuroticism"				Low and high (s_low, positive and s_high, negative)
PD4: "rationalization" psychological defense style	Red (r_mid, positive; r_high, negative)		Black, low and high (v_black, positive; v_shadow, negative and v_highlight, negative)	High (s_high, positive)
T1: "extraversion"		Yellow (h_yellow, positive) and Blue (h_blue, negative)	Black (v_black, negative)	
T2: "neuroticism"	Red (r_mid, negative and r_high, positive)	Red (h_red, positive) and Yellow (h_yellow, negative)	Middle (v_expose, negative)	

3.1. Color-Pixel Determinants of the Big Five Inventory Scores

Figure A7 along with Table A1 represents the frequencies of nonzero variable coefficient appearance in the 50 regularized regression fits for 4 values of the tuning parameter α. The analysis of such tables lays the foundation for discovering the tendencies of dependencies between the variables of interest; if the variable coefficient turned out to be nonzero in a significant number of cases, then it could serve as a predictor in a regression model.

In the following subsections, the results are presented for each of the three psychometric tests.

Therefore, the following graphical characteristics may serve as predictors for BF–scores:

- For the prediction of the BF1 score ("extraversion"), the red spectrum characteristics (r_low, r_mid, h_red) as well as the expression of green and blue color tones (h_green and h_blue, respectively) are relevant;
- For the prediction of the BF2 score ("agreeableness"), the expression of red and green colors (g_low, r_mid and g_mid) and middle-ranged pixel brightness (v_expose) may serve as predictors;
- The BF3 score, which refers to "conscientiousness", can be predicted from expression of the blue, red and green color tones (g_low, r_mid, g_mid, b_high, h_blue and h_cyan) and highly shadowed areas (v_shadow and v_expose);
- The BF4 score, "neuroticism", is associated with expression of the contrast color (gray (s_low) and highly saturated (s_high));
- For the BF5 score, there were no color-pixel features among those considered in this paper that could be used for prediction.

3.2. Color-Pixel Determinants of Plutchik's Life Style Index

The scores of this psychometric test were weakly associated with various graphical parameters, as only two psychological defense mechanisms may be related to the color-pixel characteristics of an image in a statistical sense. Table A1 and Figure A8 represent the

frequencies of a nonzero variable coefficient's appearance in the 50 regularized regression fits for 4 values of the tuning parameter for the PD4 and PD6 scores, which stand for expression of the "rationalization" and "substitution" psychological defense mechanisms.

The following tendencies of the dependencies may be indicated:

- The PD4 score, which refers to the "rationalization" psychological defense mechanism, was associated with expression of the red spectrum characteristics (r_mid and r_high), contrast brightness values (v_black, v_shadow and v_highlight) as well as high saturation (s_high) levels and the overall colorfulness of an image (colorfulness);
- The PD8 score, which refers to a "substitution" psychological defense mechanism, was also associated with the image's colorfulness and color saturation (s_low, s_mid and s_high), as well as with expression of the red spectrum characteristics (r_mid), yellow (h_yellow) and cyan (h_cyan) colors and brightness of the pixels (v_expose).

3.3. Color-Pixel Determinants of the Eysenck Personality Questionnaire

Table A1 and Figure A9 represent the frequencies of nonzero variable coefficient appearance in the 50 penalized regression fits for 4 values of the tuning parameter of the scores of the Eysenck Personality Questionnaire (test for temperament). We note that only one factor was relevant to scores T3 and T4, indicating that linear regression may not be a good choice for modeling this dependency.

The following graphical characteristics may serve as a predictors for the BF scores:

- The T1 score, or "extraversion", was associated with expression of areas of a black color (v_black) and areas of yellow and blue color tones (h_yellow and h_blue, respectively);
- The T2 score, or "neuroticism", was associated with expression of the red spectrum (characteristics r_mid, r_high and h_red), yellow color tone (h_yellow) and mid–ranged pixel brightness (v_expose).

4. Discussion and Results

This study aimed for a comprehensive analysis of the relationship between the graphical features of user's OSM avatar images and the psychometric scores of the Big Five Inventory, Plutchik's Lifestyle Index and the Eysenck Personality Questionnaire. Two levels of the image graphical features were considered: the color-pixel characteristics and number of faces or persons depicted. The former set of characteristics consists of the relative frequencies of pixels with certain RGB and HSV components and the image's colorfulness. The latter set represents the semantic level of the avatar image. As we considered the avatars, it was natural to suppose that it contained faces or persons. Those two levels of the image were considered separately, and two research hypothesis were formulated.

The first Hypothesis H1 was addressed with the one-way nonparametric Kruskall–Wallis test and post hoc pairwise Dunn test comparisons. The result of the statistical testing partially confirmed Hypothesis H1. The psychometric BF1 ("extraversion"), BF3 ("conscientiousness") and T1 ("extraversion") scores were lower for the users that chose an avatar without a face or person than for the users that chose an avatar with exactly one face or person. No statistically significant differences were found for the other scores. We note that the all psychological defense style scores did not differ between groups with different numbers of faces or persons. As was shown in the second step of the data analysis, this dimension of personality had weak representation in all considered users' avatar image characteristics.

The second stage of the data analysis was devoted to the analysis of the linear relationships between the color-pixel avatar features and psychometric scores (Hypothesis H2). We considered 23 relative counts of pixels with certain HSV and RGB color characteristics and the overall colorfulness of the images. Those features were highly correlated and excessive for the collected dataset, and therefore the penalized regression with the $L2$ and MCP penalties was used in order to identify the most explanatory subset. As this method is not deterministic, the regression was fitted 50 times for every psychometric score of interest, aiming at the identification of the subset of the most explanatory variables. If a variable

coefficient differed from zero for a significant number of times, then it could be considered a predictor of the psychometric score. The stability of the coefficient sign was assessed too. Table 3 reflects the best subsets for the psychometric scores except for the dependency between the PD4 and PD8 scores and the colorfulness characteristic of an avatar image.

The results support the correlation analysis from [24] for the Big Five Questionnaire; "extraversion" is associated with various colors in the HSV model, and "agreeableness" with is associated with the middle brightness of an image. If we consider the best subsets for "extraversion" expressed by the BF1 and T1 scores, we notice that these scores were positively associated with the red and yellow colors and negatively associated with the number of black areas and green and blue colors. This observation indirectly supports existing research for graphical user-generated content from [23] and adds more specifics to the expression of particular colors. As for the BF2 score, "agreeableness", it was positively associated with the expression of a mild green color and areas with medium values, which relates to the results from [24]. It is interesting to note that the BF3 score, conscientiousness, had significant associations with the areas of green and blue colors, as opposed to the expression of the BF1 score. We found no discussion of similar results in the area of personality computing. "Neuroticism", reflected by the BF4 and T2 scores, was in general associated with the expression of red and yellow colors along with large low-saturation and small high-saturation areas. This adds new knowledge to the analysis from [9,24], which indicated a negative association of BF4 with color diversity and color harmony. In general, this study supports existing research on how personality is expressed in the colors that a person chooses and adds some more specifics to the particular features of an avatar image. This study also adds novel discussion on expression of the psychological defense mechanism in avatar images.

We also note that the selected subsets of explanatory variables should be analyzed comprehensively and with regard to the specific situation. For example, expression of some colors could be associated with the peculiarities of the device color rendering.

5. Conclusions

Analysis of the digital traces of online social media users refers to the area of personality computing, a field of research at the intersection of artificial intelligence and psychology, where computational methods are used in order to extract personality traits from the data from various sources, including heterogeneous information from social media. As there is an enormous number of possible features that can be extracted from a user's profile, the problem of identification of ones that are relevant to a particular problem arises. This problem is crucial in the context of limited available data.

This study contributes to the selection of an avatar's graphical characteristics that are relevant to various personality traits. Among the psychometric tests, three were chosen for the purposes of this study: the Big Five Inventory, Plutchik's Lifestyle Index and the Eysenck Personality Questionnaire. From the avatar image, the following features were extracted: color-pixel characteristics from two color models, RGB and HSV, the image colorfulness and the number of persons and faces depicted. Two research hypotheses were formulated, where the first stated that the personality traits of users of an online social media platform (measured by the Big Five Inventory, Plutchik's Life Style Index and the Eysenck Personality Questionnaire) vary depending on the semantic content of their avatars, namely the number of depicted persons and faces, and the second stated that the graphical color-pixel characteristics of user avatars can serve as predictors in the linear regression model for the scores of the psychometric tests under consideration. The data of 548 observations were collected with the specially developed application, which was posted on the online social media platform VKontakte and freely available for every user. The data for the analysis consisted of psychometric test scores and avatar characteristics. The first hypothesis was investigated with the nonparametric Kruskall–Wallis test. The second question was explored with statistical modeling procedure: he

regularized regression was fitted several times, and the frequency of the variable coefficient being nonzero was determined.

The first hypothesis was confirmed only for an extraversion score (BF1 from the Big Five Inventory and T1 from the Eysenk Personality Questionnaire). The second hypothesis was confirmed partially, as not all psychometric scores were linearly associated with the color-pixel characteristics.

The proposed approach, as well as the models obtained during the research, lay the foundation for further work on conducting relevant experiments and clarifying existing and developing new approaches, models and algorithms. The obtained results contribute to the construction of more accurate ideas about the expression of a user's psychological traits and are useful to specialists in such areas as human resources, computer security, marketing, lending and social sciences, among others.

Appendix A

	Test scores			Pixel characteristics			Image content
PD1	BF1	T1		r_low	r_mid	r_high	person_count
1	3	3		0.527	0.264	0.209	1
PD2	BF2	T2		g_low	g_mid	g_high	face_count
0	3	1		0.552	0.269	0.179	1
PD3	BF3	T3		b_low	b_mid	b_high	
1	3	2		0.506	0.326	0.168	
PD4	BF4	T4		s_low	s_mid	s_high	
1	2	1		0.646	0.328	0.026	
PD5	BF5			v_black	v_shadow	v_expose	
1	3			0.126	0.362	0.283	
PD6				v_highlight	v_whites	colourness	
0				0.212	0.022	830.3	
PD7				h_red	h_green	h_blue	
1				0.221	0.023	0.517	
PD8				h_yellow	h_cyan	h_magenta	
1				0.082	0.091	0.066	

Figure A1. The psychometric scores and avatar characteristics (color-pixel and semantic) were obtained via the specially developed software.
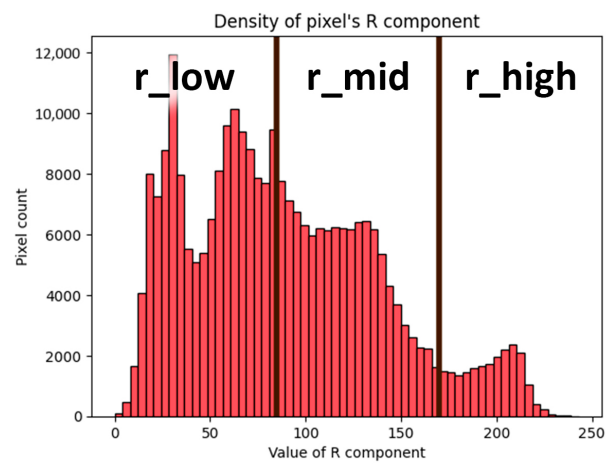
Figure A2. Histogram of an R component of the RGB characteristics of the sample avatar image with discretization on three equal-length components.
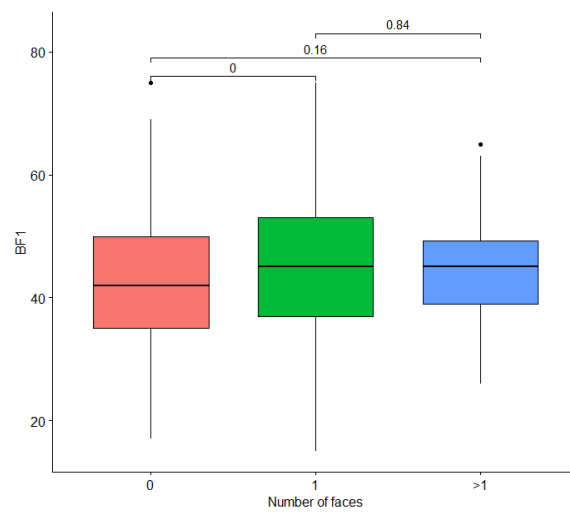
(a)
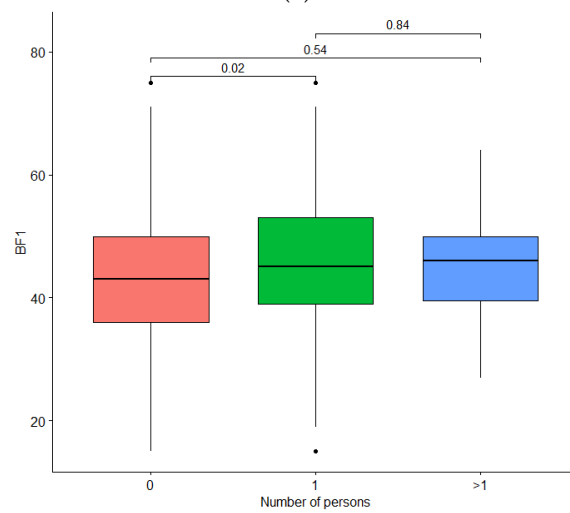
(b)

Figure A3. Post hoc analysis of the group difference for the BF1 scores with the Dunn test *p* values. (**a**) Comparison of BF1 scores between groups with different numbers of faces in their avatar images. (**b**) Comparison of BF1 scores between groups with different numbers of persons in their avatar images.
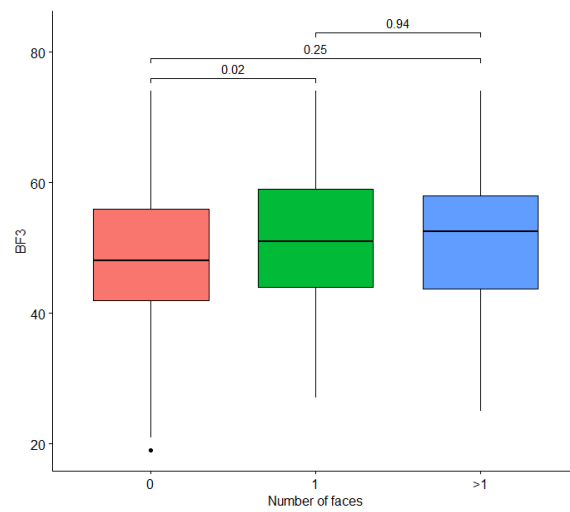
(a)
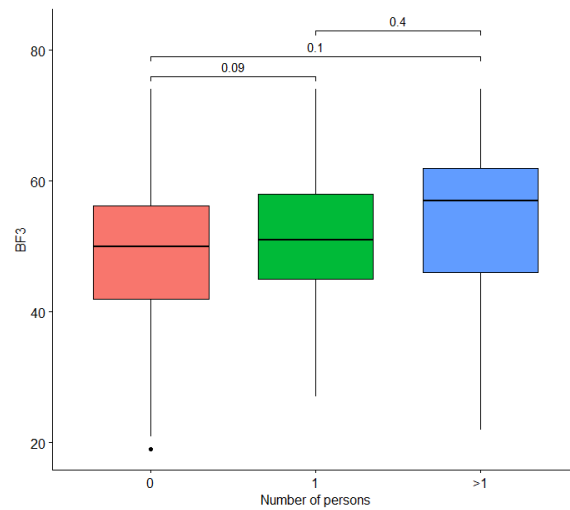
(b)

Figure A4. Post hoc analysis of the group difference for the BF3 scores with the Dunn test *p* values. (**a**) Comparison of BF3 scores between groups with different numbers of faces in their avatar images. (**b**) Comparison of BF3 scores between groups with different numbers of persons in their avatar images.
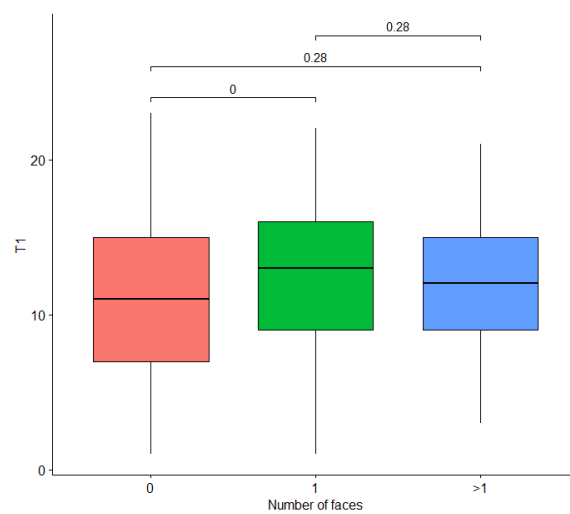
Figure A5. Post hoc analysis of the difference in the T1 scores between groups with different numbers of faces with the Dunn test *p* values.
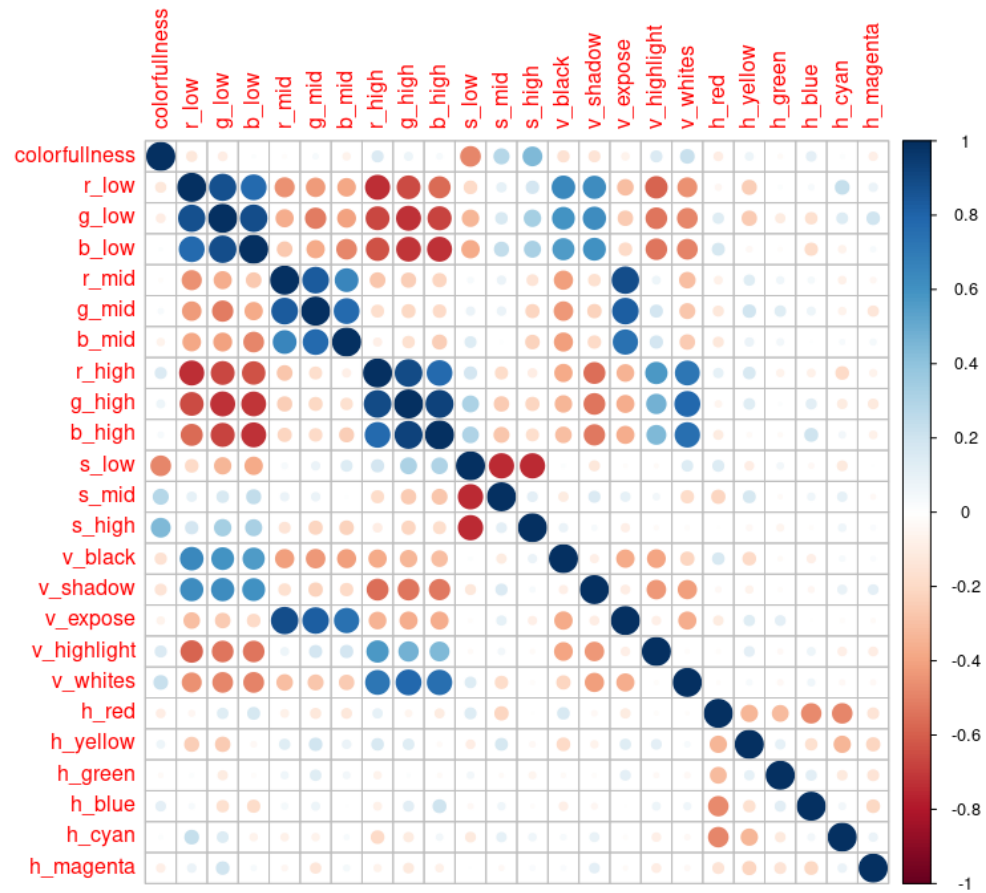
Figure A6. The correlation structure between the color-pixel characteristics of the avatar images.
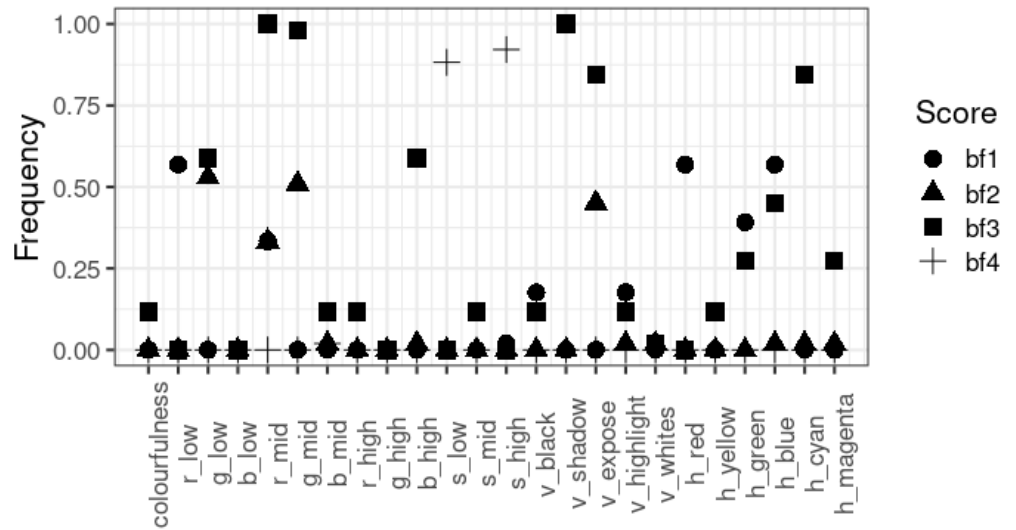
Figure A7. Frequencies of nonzero variable coefficient appearance in the 50 regularized regression fits with MCP and L2 restrictions for the Big Five Inventory scores for tuning parameter value $\alpha = 0.005$.
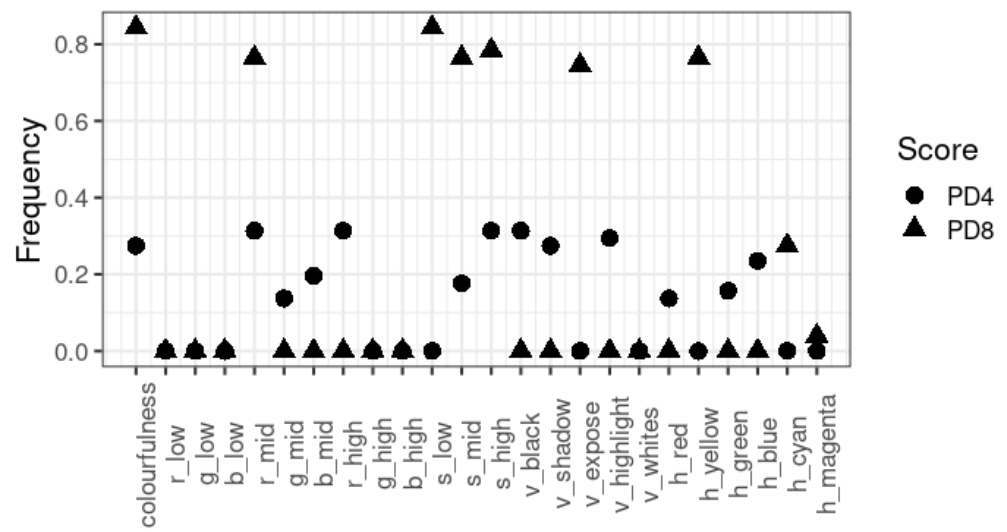
Figure A8. Frequencies of nonzero variable coefficient appearance in the 50 regularized regression fits with MCP and L2 restrictions for Plutchik's Lifestyle Index scores for tuning parameter value $\alpha = 0.005$.
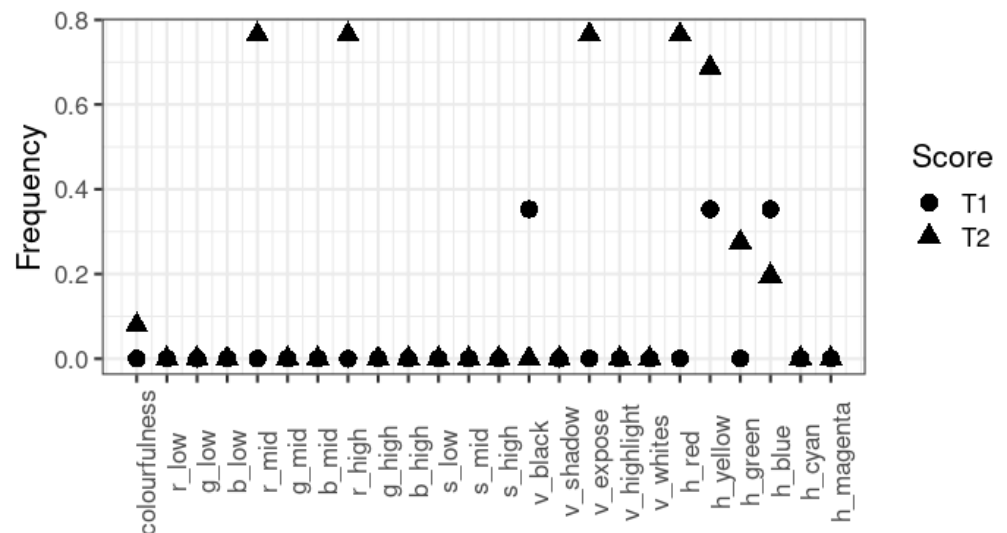
Figure A9. Frequencies of nonzero variable coefficient appearance in the 50 regularized regression fits with MCP and L2 restrictions for the Eysenck Personality Questionnaire scores for tuning parameter value $\alpha = 0.005$.

Appendix B

Table A1. Frequencies of nonzero variable coefficient appearance in the 50 regularized regression fits with MCP and L2 restrictions for the Big Five Inventory scores.

Color-Pixel Characteristic	Tuning Parameter	BF1	BF2	BF3	BF4	PD4	PD8	T1	T2	T3
Colorfulness	0.05	0	0	0	0	0.06	0.74	0	0.02	0
	0.01	0	0	0	0	0.35	0.86	0	0.08	0
	0.005	0	0	0.14	0	0.24	0.92	0	0.25	0.02
	0.003	0	0	0.40	0	0.29	0.92	0	0.39	0

Table A1. *Cont.*

Color-Pixel Characteristic	Tuning Parameter	BF1	BF2	BF3	BF4	PD4	PD8	T1	T2	T3
r_low	0.05	0.06	0	0	0	0	0	0	0	0
	0.01	0.35	0	0	0	0	0	0	0	0
	0.005	0.55	0	0	0	0	0	0	0	0
	0.003	0.61	0.22	0	0	0	0	0	0	0
g_low	0.05	0	0.75	0	0	0	0	0	0	0
	0.01	0	0.67	0	0	0	0	0	0	0
	0.005	0	0.61	0.57	0	0	0	0	0	0
	0.003	0	0.47	0.96	0	0	0	0	0	0
r_mid	0.05	0	0	0.98	0	0.80	0.10	0	0.37	0
	0.01	0.16	0	1	0	0.49	0.65	0	0.75	0
	0.005	0.39	0.45	1	0	0.31	0.88	0	0.96	0
	0.003	0.49	0.39	0.98	0	0.35	0.88	0	0.94	0
g_mid	0.05	0	0.17	0	0	0	0	0	0	0
	0.01	0	0.65	0	0	0	0	0	0	0
	0.005	0	0.61	0.96	0	0.06	0	0	0.92	0
	0.003	0	0.49	0.96	0	0.16	0	0	0.94	0
b_mid	0.05	0	0.62	0.21	0.02	0.04	0	0	0	0
	0.01	0	0.14	0	0.02	0.27	0	0	0	0
	0.005	0	0.04	0.14	0.02	0.10	0	0	0	0
	0.003	0	0.02	0.39	0.08	0.18	0	0	0	0
r_high	0.05	0	0	0.1	0	0	0	0	0.40	0
	0.01	0	0	0	0	0.22	0	0	0.80	0
	0.005	0	0	0.14	0	0.31	0	0.04	0.96	0
	0.003	0	0	0.39	0	0.35	0	0.12	0.94	0
b_high	0.05	0	0	0	0	0	0	0	0	0
	0.01	0	0.1	0.04	0	0	0	0	0	0
	0.005	0	0.04	0.51	0	0	0	0	0	0
	0.003	0	0.05	0.80	0	0	0	0	0	0
s_low	0.05	0	0	0	0	0	0.71	0	0	0
	0.01	0	0	0	0.78	0	0.86	0	0	0
	0.005	0	0	0	0.96	0	0.92	0	0	0
	0.003	0	0	0.39	0.90	0.02	0.92	0	0	0
s_mid	0.05	0	0	0	0	0	0	0	0	0
	0.01	0	0	0	0	0.02	0	0	0	0
	0.005	0	0	0.14	0	0.08	0.88	0	0	0
	0.003	0	0	0.43	0	0.16	0.90	0	0	0
s_high	0.05	0	0	0	0.94	0.80	0	0	0.02	0.57
	0.01	0	0	0	0.86	0.49	0.75	0	0.06	0.55
	0.005	0	0	0	0.96	0.31	0.90	0	0.12	0.69
	0.003	0.02	0	0	0.90	0.35	0.92	0	0.16	0.69
v_black	0.05	0	0	0	0	0.80	0	0.22	0	0
	0.01	0	0	0	0	0.49	0	0.37	0	0
	0.005	0.22	0	0.14	0	0.31	0	0.25	0	0
	0.003	0.37	0	0.39	0	0.35	0	0.27	0	0
v_shadow	0.05	0	0	0.98	0	0	0	0	0	0
	0.01	0	0	1	0	0.29	0	0	0	0
	0.005	0	0	1	0	0.24	0	0	0	0
	0.003	0	0	0.98	0	0.29	0	0	0	0
v_expose	0.05	0	0.63	0	0	0	0	0	0.02	0
	0.01	0	0.57	0	0	0	0.55	0	0.80	0
	0.005	0	0.57	0.73	0	0	0.84	0	0.96	0
	0.003	0	0.49	0.92	0	0	0.88	0	0.94	0
v_highlight	0.05	0	0.47	0	0	0	0	0	0	0
	0.01	0	0.14	0	0	0.35	0	0	0	0
	0.005	0.22	0.04	0.12	0	0.31	0	0	0	0
	0.003	0.37	0	0.39	0	0.35	0	0	0	0

Table A1. *Cont.*

| Color-Pixel Characteristic	Tuning Parameter	BF1	BF2	BF3	BF4	PD4	PD8	T1	T2	T3
v_whites	0.05	0	0	0	0	0	0	0	0	0
	0.01	0	0	0	0	0	0	0	0	0
	0.005	0	0.04	0.08	0	0	0	0	0	0
	0.003	0	0.02	0.31	0	0	0	0	0	0
h_red	0.05	0.06	0	0	0	0.02	0	0	0.39	0
	0.01	0.35	0	0	0	0.02	0	0	0.80	0
	0.005	0.55	0	0	0	0.06	0	0	0.96	0
	0.003	0.61	0	0	0	0.16	0	0	0.94	0
h_yellow	0.05	0	0	0	0	0	0.31	0.08	0.06	0
	0.01	0	0	0	0	0	0.76	0.33	0.67	0
	0.005	0	0	0.14	0	0	0.88	0.25	0.96	0
	0.003	0	0.02	0.39	0	0	0.92	0.25	0.94	0
h_green	0.05	0	0	0.22	0	0	0	0	0.06	0
	0.01	0.16	0	0.04	0	0.14	0	0	0.24	0
	0.005	0.45	0	0.24	0	0.06	0	0	0.71	0
	0.003	0.49	0	0.49	0	0.16	0	0	0.63	0
h_blue	0.05	0.06	0	0.22	0	0.04	0	0.22	0	0
	0.01	0.35	0.14	0.06	0	0.29	0	0.37	0.20	0
	0.005	0.55	0.04	0.41	0	0.18	0	0.25	0.90	0
	0.003	0.61	0.02	0.61	0	0.22	0.24	0.27	0.94	0
h_cyan	0.05	0	0.02	0.63	0	0	0	0	0	0
	0.01	0	0.14	0.59	0	0	0.02	0	0	0
	0.005	0	0.04	0.73	0	0	0.22	0	0	0
	0.003	0	0.02	0.88	0	0	0.57	0	0	0
h_magenta	0.05	0	0	0	0	0	0	0	0	0
	0.01	0	0	0	0.04	0	0	0	0	0
	0.005	0	0.04	0	0.24	0	0	0	0	0
	0.003	0	0.02	0	0.49	0	0.49	0	0	0

Appendix C
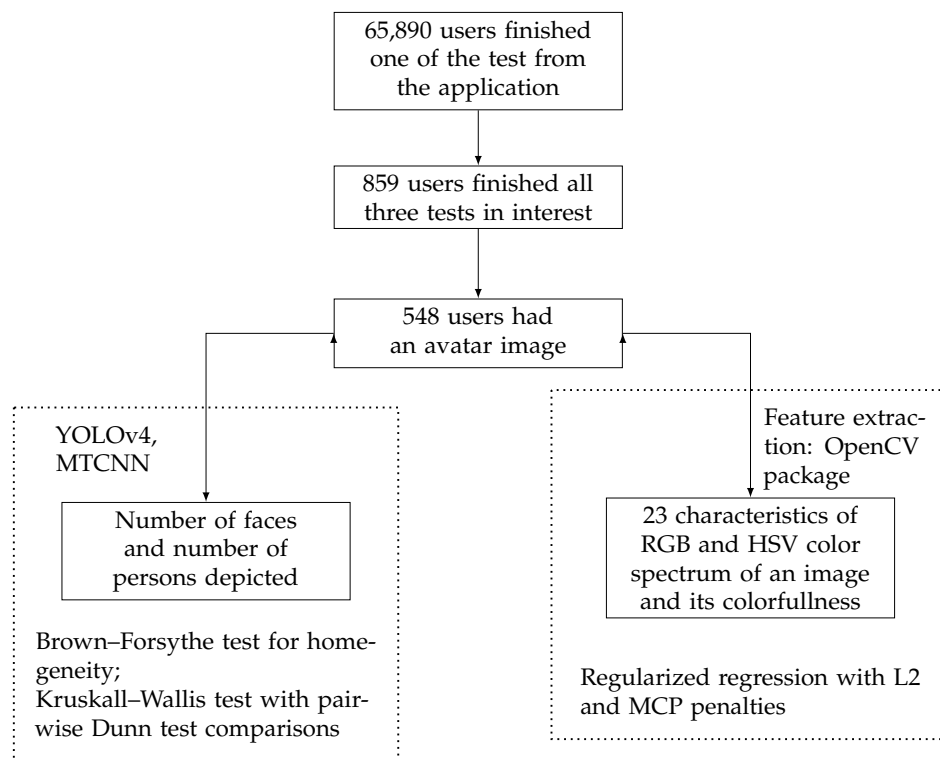
Figure A10. Data collection and processing flowchart, the YOLOv4 model is described in [34], MTCNN in [35] and OpenCV package in [32].

References

1. Ko, S.; Li, E.Y.; Wu, Y. Using artificial intelligence to study the impact of jobseekers' Facebook profile pictures on recruiters' interview decision. In Proceedings of the International Conference on Electronic Business, Bangkok, Thailand, 13–17 October 2022; Volume 22, pp. 585–592.
2. Li, Y.; Xie, Y. Is a picture worth a thousand words? An empirical study of image content and social media engagement. *J. Mark. Res.* **2020**, *57*, 1–19. [CrossRef]
3. Rietveld, R.; Van Dolen, W.; Mazloom, M.; Worring, M. What you feel, is what you like influence of message appeals on customer engagement on Instagram. *J. Interact. Mark.* **2020**, *49*, 20–53. [CrossRef]
4. Schreiner, M.; Fischer, T.; Riedl, R. Impact of content characteristics and emotion on behavioral engagement in social media: Literature review and research agenda. *Electron. Commer. Res.* **2021**, *21*, 329–345. [CrossRef]
5. Yu, C.; Xie, S.Y.; Wen, J. Coloring the destination: The role of color psychology on Instagram. *Tour. Manag.* **2020**, *80*, 104110. [CrossRef]
6. Dudău, D.P.; Sava, F.A.; Rusu, A.; Cervicescu, V. Detecting Individuals High in Neuroticism based on the Color Features of the Facebook Profile Picture. In Proceedings of the 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 1–4 September 2020; pp. 286–293.
7. Kim, Y.; Kim, J.H. Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study. *Inf. Process. Manag.* **2018**, *54*, 1101–1114. [CrossRef]
8. Meng, K.S.; Leung, L. Factors influencing TikTok engagement behaviors in China: An examination of gratifications sought, narcissism, and the Big Five personality traits. *Telecommun. Policy* **2021**, *45*, 102172. [CrossRef]
9. Liu, L.; Preotiuc-Pietro, D.; Samani, Z.R.; Moghaddam, M.E.; Ungar, L. Analyzing personality through social media profile picture choice. In Proceedings of the International AAAI Conference on Web and Social Media, Cologne, Germany, 17–20 May 2016; Volume 10, pp. 211–220.
10. Mancini, T.; Sibilla, F. Offline personality and avatar customisation. Discrepancy profiles and avatar identification in a sample of MMORPG players. *Comput. Hum. Behav.* **2017**, *69*, 275–283. [CrossRef]
11. Raveendhran, R.; Fast, N.J.; Carnevale, P.J. Virtual (freedom from) reality: Evaluation apprehension and leaders' preference for communicating through avatars. *Comput. Hum. Behav.* **2020**, *111*, 106415. [CrossRef]
12. Ramírez-Cifuentes, D.; Freire, A.; Baeza-Yates, R.; Sanz Lamora, N.; Álvarez, A.; González-Rodríguez, A.; Lozano Rochel, M.; Llobet Vives, R.; Velazquez, D.A.; Gonfaus, J.M.; et al. Characterization of anorexia nervosa on social media: Textual, visual, relational, behavioral, and demographic analysis. *J. Med. Internet Res.* **2021**, *23*, e25925. [CrossRef]
13. Scherr, S.; Arendt, F.; Frissen, T.; Oramas M, J. Detecting intentional self-harm on Instagram: Development, testing, and validation of an automatic image-recognition algorithm to discover cutting-related posts. *Soc. Sci. Comput. Rev.* **2020**, *38*, 673–685. [CrossRef]
14. Chancellor, S.; De Choudhury, M. Methods in predictive techniques for mental health status on social media: A critical review. *NPJ Digit. Med.* **2020**, *3*, 43. [CrossRef] [PubMed]
15. Smirnov, A.; Levashova, T. Context-Aware Approach to Intelligent Decision Support Based on User Digital Traces. *Inform. Autom.* **2021**, *19*, 915–941. [CrossRef]
16. Guntuku, S.C.; Preotiuc-Pietro, D.; Eichstaedt, J.C.; Ungar, L.H. What twitter profile and posted images reveal about depression and anxiety. In Proceedings of the International AAAI Conference on Web and Social Media, Münich, Germany, 11–14 June 2019; Volume 13, pp. 236–246.
17. Abramov, M.; Tulupyeva, T.; Tulupyev, A. *Social Engineering Attacks: Social Media and Users Vulnerability Assessment*; SUAI: St. Petersburg, Russia, 2018; p. 266. (In Russian)
18. Khlobystova, A.; Abramov, M. Time–Based Model of the Success of a Malefactor's Multistep Social Engineering Attack on a User. In *Proceedings of the Fifth International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'21)*; Springer International Publishing: Cham, Switzerland, 2022; pp. 216–223.
19. Korepanova, A.; Abramov, M.; Tulupyev, A. Social media user identity linkage by graphic content comparison *Sci. Tech. J. Inf. Technol. Mech. Opt.* **2021**, *21*, 942–950.
20. Huang, C. Social network site use and Big Five personality traits: A meta-analysis. *Comput. Hum. Behav.* **2019**, *97*, 280–290. [CrossRef]
21. Shchebetenko, S. Do personality characteristics explain the associations between self-esteem and online social networking behaviour? *Comput. Hum. Behav.* **2019**, *91*, 17–23. [CrossRef]
22. Alqahtani, G.; Alothaim, A. Predicting emotions in online social networks: Challenges and opportunities. *Multimed. Tools Appl.* **2022**, *81*, 9567–9605. [CrossRef]
23. Kim, J.H.; Kim, Y. Instagram user characteristics and the color of their photos: Colorfulness, color diversity, and color harmony. *Inf. Process. Manag.* **2019**, *56*, 1494–1505. [CrossRef]
24. Ferwerda, B.; Schedl, M.; Tkalcic, M. Using Instagram picture features to predict users' personality. In *MultiMedia Modeling, Proceedings of the 22nd International Conference, MMM 2016, Miami, FL, USA, 4–6 January 2016*; Proceedings, Part I 22; Springer International Publishing: Cham, Switzerland, 2016; pp. 850–861.
25. Chromov, A. *Five–Factor Personality Questionnaire*; KSU: Kurgan, Russia, 2000. (In Russian)
26. Plutchik, R.; Kellerman, H.; Conte, H.R. A structural theory of ego defenses and emotions. In *Emotions in Personality and Psychopathology*; Springer: Boston, MA, USA, 1979; pp. 227–257.

27. Wasserman, L.I.; Yeryshev, O.F.; Klubova, E.B.; Petrova, N.N. *Psychological Diagnostics of the Lifestyle Index*; SPbNIPNI im. VM Bekhtereva: St. Petersburg, Russia, 2005.
28. Eysenck, H.J. *A Model for Personality*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
29. Khorrami, M.; Khorrami, M.; Farhangi, F. Evaluation of tree-based ensemble algorithms for predicting the big five personality traits based on social media photos: Evidence from an Iranian sample. *Personal. Individ. Differ.* **2022**, *188*, 111479. [CrossRef]
30. Biswas, K.; Shivakumara, P.; Pal, U.; Chakraborti, T.; Lu, T.; Ayub, M.N.B. Fuzzy and genetic algorithm based approach for classification of personality traits oriented social media images. *Knowl.–Based Syst.* **2022**, *241*, 108024. [CrossRef]
31. Hasler, D.; Suesstrunk, S.E. Measuring colorfulness in natural images. In Proceedings of the Human Vision and Electronic Imaging VIII, Santa Clara, CA, USA, 20 January 2003; Volume 5007, pp. 87–95.
32. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools*. **2000**, *25*, 120–123.
33. Smith, A.R. Color gamut transform pairs. *ACM Siggraph Comput. Graph.* **1978**, *12*, 12–19. [CrossRef]
34. Bochkovskiy, A.; Wang, C.; Liao, H.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
35. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
36. Brown, M.B.; Forsythe, A.B. Robust tests for the equality of variances. *J. Am. Stat. Assoc.* **1974**, *69*, 364–367. [CrossRef]
37. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; CRC Press: Boca Raton, FL, USA, 2015.
38. Breheny, P.; Huang, J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Ann. Appl. Stat.* **2011**, *5*, 232. [CrossRef]
39. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.