Personality and Stress: Analyzing Social Media Usage's Stress through the Big Five Model

Minh Tran (1100203), Yangfan Chen (5194652)

1. Introduction

With the development of the internet and the world wide web leading to the rise of social media platforms, the way people communicate and interact has fundamentally changed. As of April 2024, there are hundreds of millions to billions of active users for the most popular social networks each month. Facebook, for example, has roughly 3 billion active users, and YouTube has 2.5 billions, with its contemporary rival, TikTok, having 1.5 billions [1].

Given how prevalent social media is in our lives, its influence extends well beyond casual communication. These platforms went from simply avenues for personal expression to commercial spaces, news sources, and discourse forums all in one. While convenient and entertaining, and some might say the services they provide are indispensable, there are rising concerns of how these platforms affect a person's mental health. Thus it is not surprising that there is a wealth of studies on how the human psyche has been impacted by the exposure to social media.

Some studies have found that social network usage leads to an overall negative experience and lasting effects on person's mental health. A study of 1787 U.S. young adults finds that time spent on social media is positively correlated with depression [2]. Another study with a sample of 467 Scottish teenagers finds that increased usage of social networks not only increases the likelihood of depression, but also anxiety and lower sense of self-esteem, along with a poorer quality of sleep [3]. On the other hand, some researchers have called attention to whether social media is a reliable indicator of mental health issues, arguing that the focus should be on how social media is used rather than time spent [4]. Indeed, there seem to be two camps of thoughts when it comes the topic: either social media usage is detrimental to mental health and there is significant evidence to back up [5], or that the evidence for claiming so is statically weak [6].

Regardless, we believe that there is enough literature that warrants a further look into the relationship between mental health and social media. To be more specific, we want to explore whether personality plays a factor regarding stress when using social media services. As mentioned by prior studies, heavy social media usage is positively correlated with poor mental health, so there might exist a connection between different personalities and an individual being susceptible to harm, or them using social media as part of their coping method. By the time of this writing, we find a severe lack of research that connect these aspects together. We thus propose our research question as follow:

How do varying degrees of the Big Five personality traits impact stress indicators including heart rate variability (HRV) and electrodermal activity (EDA) during social media usage?

2. Related work

How personality is strictly measured and defined has been an ongoing process. Personality, according to the American Psychology Association (APA), is "the enduring configuration of characteristics and behavior that comprises an individual's unique adjustment to life, including major traits, interests, drives, values, self-concept, abilities, and emotional patterns" [7]. To study personalities scientifically, various models have been developed, such as the OCEAN model, otherwise known as the Big Five [8], and the HEXACO model [9]. The former identifies five basic dimensions in describing personality factors: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. Each factor exists on a scale; for example, people scoring high in extraversion are typically more assertive and extroverted compared to those with a lower score, who are more reserved and introverted. The HEXACO framework builds upon the Big Five by adding another dimension, namely honesty-humility, which can be seen as an indicator of a person's integrity and ethics. While both frameworks received criticisms regarding their methodology validity and lack of theoretical foundation [10], [11], they are still widely used within the field of Psychology given its statistical robustness and consistency, especially the Big Five [12], [13].

Stress, again defined by the APA, as "the physiological or psychological response to internal or external stressors" [14]. The experience is often linked to flight-or-fight responses and coping mechanisms, where the human body reacts in certain ways when encountering a "stressor" - anything that challenges the human's homeostasis [15]. A popular framework that is used to study stress is the Transactional Stress Theory, also known as Transactional Model of Stress and Coping [16]. The theory suggests a dynamic multi-step approach to how an individual cognitively processes stressors and in turn how they cope.

There are different variations of stresses and their responses depending on the causes, and thus their possible indicators which guide how stress can be measured [17], [18]. One major way of measuring stress is using self-reports to record how individuals interact with stressors such as the Life Events and Difficulties Schedule (LEDS), Trier Inventory for Chronic Stress (TICS), and Perceived Stress Scale (PSS). Another way of measurement is to use biomarkers, such as heart rate/pulse and Electrodermal activity (EDA). While neither approaches perfectly capture stress, self-reports still provide a good subjective experience of stress while biomarkers offer an objective biological measurement [17].

There are a multitude of research papers that link stress to social media usage. One article finds that having a high level of stress leads to an increased risk of being addicted to social media through regression analysis of various biopsychosocial factors [19]. Another study done with 402 adolescents finds that stress is associated with the fear of missing out and the need of belonging on Facebook [20]. In fact, a model is developed for studying the five social media related stressors - invasion, pattern, complexity, uncertainty, and disclosure; under the model's lens, social media is strongly observed to be a source of stress for the users [21]. But, it should be noted that social media not only serves as a source of stress, but also as a coping mechanism for dealing with stress, although the effectiveness is uncertain [22].

Personalities play a significant role in how individuals experience stress. Using the Big Five model, scientists in the 90s have identified how a person judges a situation to be stressful and their coping mechanism can be influenced by different personality traits [13], [23], [24]. Recent research also supports this claim empirically [25]. Neuroticism is especially noticeable as studies have consistently find a link to increased to negative stressful responses [26]–[28].

3. Methodology

3.1 Data

We use K-EmoPhone [29], a large dataset collected from 77 Korean participants from age 17 to 38 (mean = 21.9; SD = 3.9). It records participant's personal activities and statuses across 7 days in the real-world context, which includes: (1) physiology and mobility data collected via MS Band 2 worn by participants, (2) contextual and interaction data from participants' smartphones using Android OS, and (3) self-reported affect states from participants. This extensive dataset can be utilized for various research topics in affective computing, attention management etc. [29]. It provides all the data we need to address our research question.

In this research, we utilize the participant information from the dataset provided as a CSV file. It contains participants' age, genders, times of participation, scores of the Big Five personality traits and individuals' mental health scores after finishing the experiment. In our study, we focus on participant's personality traits. The K-EmoPhone dataset employs a shortened version of the Big Five Inventory (BFI) questionnaire named the K-BFI-15[30]. This questionnaire, written in Korean, contains 15 five-point Likert-scale items. Each personality trait is measured by the sum of three items, therefore ranging from 3 to 15.

We also utilize inter-beat interval (RRI) data and skin resistance data from the dataset to extract features that represent Heart Rate Variability (HRV) and Electrodermal Activity (EDA). The RRI.csv files have two columns: the time between the last two consecutive heartbeats in milliseconds (ms) and its corresponding timestamp. The skin resistance data file (EDA.csv) contains the sampling timestamp and skin resistance value measured in kilo-ohms (kΩ) sampled at 5Hz.

We extract social media application usage time from the AppUsageEvent.csv files in this dataset. Each of these files records one participant's application usage events by sampling when particular interaction events occur, with columns of application name, package name, system application marks, types of interaction, and category retrieved from Google Play. The types of interaction mainly include MOVE_TO_BACKGROUND, MOVE_TO_FOREGROUND, and USER_INTERACTION events, as defined by the corresponding Android API[31]. The ScreenEvent.csv files contain the timestamps of screen on/off and phone unlock events. We use this data to filter out application events outside the screen-on time frame. ActivityEvent.csv files contain the inferred confidence level of a user in specific activity statuses using an Android API[32]. The status includes: still, on foot (both walking and running), walking, running, in vehicle, on bicycle, tilting and unknown[32]. We use this data to exclude social media usage accompanied by physical activity which will be explained in the next subsection.

3.2 Preprocessing and feature extraction

3.2.1 Preprocessing of social media usage timeframes

In this study, the independent variables are the Big Five personality traits: extroversion, agreeableness, conscientiousness, neuroticism, and openness to experience. The personality trait scores are provided directly by the K-EmoPhone dataset without additional processing.

Given the complicated nature of the dataset, we prioritized in filtering out unnecessary features to get the most accurate data points as possible. To identify when the participants were actively using their phone for social applications, we made use of the following data files as mentioned previously: ScreenEvent.csv, AppUsageEvent.csv, and ActivityEvent.csv.

Most if not all applications used by the participants were logged to AppUsageEvent.csv, so we could pin point the specific applications they used. While the file provides the categories of apps from Google Play - we were only interested in SOCIAL apps and COMMUNICATION apps - we still decided to specify which applications for the filter as we found some did not fit our social application requirement. For example, one application we found is associated with banking, while another app is made for shopping. Specifically, for SOCIAL apps we exclude CashSlide, Samsung Dream Class, Zenly, Karrot, Bitterness, HelloBot and Fandom. For Communication apps, we only include Kakao, Messenger, and WhatsApp.

We also had to consider whether an application was being used actively by the participants, not it being used by the phone system as part of background process. We processed this by making use of the ScreenEvent.csv file in combination with the previous file. In general, events in ScreenEvent.csv indicates when the phone is turned on, off, or is unlocked, and in AppUsageEvent.csv, each event can be identified as a Foreground process or a Background process. To be specific, by combining and filtering these two data files using the timeframe of which the phone was unlocked then turned off, and the timeframe that an application was being used in the Foreground, we can identify the social media usage time-frames of each participant. Additionally, we merged social media usage time-frames that are within 30 seconds of each other into a single, larger time-frame to have more available data.

There was a possibility that participants were physically active while using social media, which would not only introduce a covariate affecting EDA and HRV but also influence the reliability of the collected sensor data. For HRV, the RRI data essentially consisted of pulse-to-pulse intervals collected by the MS 2 Band worn on the hand. Thus, hand movement could influence blood flow and cause noise in the data. For the EDA data (skin resistance), hand movement could cause poor contact between the sensor and the skin, also introducing noise. Therefore, we introduced a procedure to mitigate these problems. ActivityEvent.csv was used as its events indicate how confident the device is in determining whether the participants were being active or not. Similar the logic behind how ScreenEvent.csv was utilized, we considered the timeframe of which an application was being actively used and matched that to the timeframe in the ActivityEvent.csv. We calculated the average confidence values during the timeframe and filtered out those with a sum of confidence on foot (walking or running) and on bicycle greater than 0.25. The value of 0.25 was selected to reduce the influence of physical activity while retaining enough timeframes. Afterwards, we filtered out timeframes that have less than two minutes duration. The reason for this will be addressed in the next subsection.

By the end, based on the above filtering methods, we had the usage timeframes of specific social and communication applications that were actively used by the participants when they were likely to be still.

3.2.2 Preprocessing and feature extraction of HRV

The HRV metric in our study is the Coefficient of Variance of Standard Deviation (CVSD), which is essentially the Root Mean Square of Successive Differences (RMSSD) divided by the mean inter-beat interval (RRI). RMSSD, based on beat-to-beat differences, is indicative of short-term heart rate variations [33]. It is influenced by the activity of the parasympathetic nervous system (PNS), which predominates in quiet and relaxing states while the sympathetic nervous system (SNS) is related to elevated activity and stressful states [34]. Therefore, RMSSD can effectively reflect participants' stress levels: a higher RMSSD value suggests a lower level of stress and vice versa. Previous research suggests that RMSSD is relatively free of respiratory influences [35] and is statistically robust [33]. However, there is an inverse relationship between the inter-beat interval (RRI) and heart rate: higher RMSSD tends to be associated with lower heart rates and vice versa, which makes RMSSD comparisons among populations less reliable [36]. Consequently, we opted to use CVSD (RMSSD divided by the mean RRI) as our HRV metric. Research indicates that standardizing RMSSD by heart rate allows for more meaningful comparisons across populations [37].

In our study, for each participant, we processed HRV data by initially extracting RRI signal segments that correspond to the period when participant was using social media. This is based on the previously mentioned time processing procedure. According to existing literature, RMSSD values derive from one minute of RRI data are generally deem to be adequate for estimating those from a five-minute and longer period of RRI data. However, some studies have reported distortions of RMSSD data from one-minute time windows. To ensure a relatively high reliability, we chose a two-minute minimum threshold of social media usage, as it is reported to have more stable RMSSD estimations for those from longer duration. We also exclude segments that do not have corresponding RRI data, which is probably because they were not wearing the MS Band 2 during this social media usage. Afterwards, the CVSD of the RRI segement will be calculated using NeuroKit 2 [38], which is a Python library for processing physiological signals. The result for each segment, together with the beginning and end timestamps, will be written in the CSV file for each participant. These CSV files will be stored for HRV data analysis which will be discussed in the next section.

3.2.3 Preprocessing and feature extraction of EDA

The EDA metric in our study is the frequency of Skin Conductance Response (SCR), which can be reflective of arousal and stress level [39]. To extract SCR, we applied the SparsEDA algorithm which is reported to be fast and efficient [40] and support automated extraction of SCRs from both large and small EDA segments [40], which suits data originally collected from wearable devices in this research.

Specifically in our study, We processed the EDA data using a procedure similar to that for HRV. First, skin resistance signal segments were extracted based on the social media usage time points. Second, skin conductance was calculated from skin resistance in microsiemen (μS). To be consistence with the data processing procedure of HRV data, we used the same two-minute minimum threshold of social media usage. The SCR extraction is achieved by also using the NeuroKit 2 library. Afterwards, the frequency of SCR was calculated for each segment. We applied a specific procedure for EDA feature extraction to exclude outliers of SCR frequency, because our investigation showed that the majority of outliers are caused by poor contact between the sensor and skin, resulting in highly fluctuated skin resistance or a non-changing maximum value of skin resistance. These issues lead to abnormally high or low SCR frequency. We employed the standard Interquartile Range (IQR) procedure to remove outliers. The results are written to a CSV file for each participant, along with the corresponding start and end timestamps. These CSV files were stored for subsequent EDA data analysis.

The pipeline of the complete preprocessing and feature extraction procedure is shown at Figure 1.

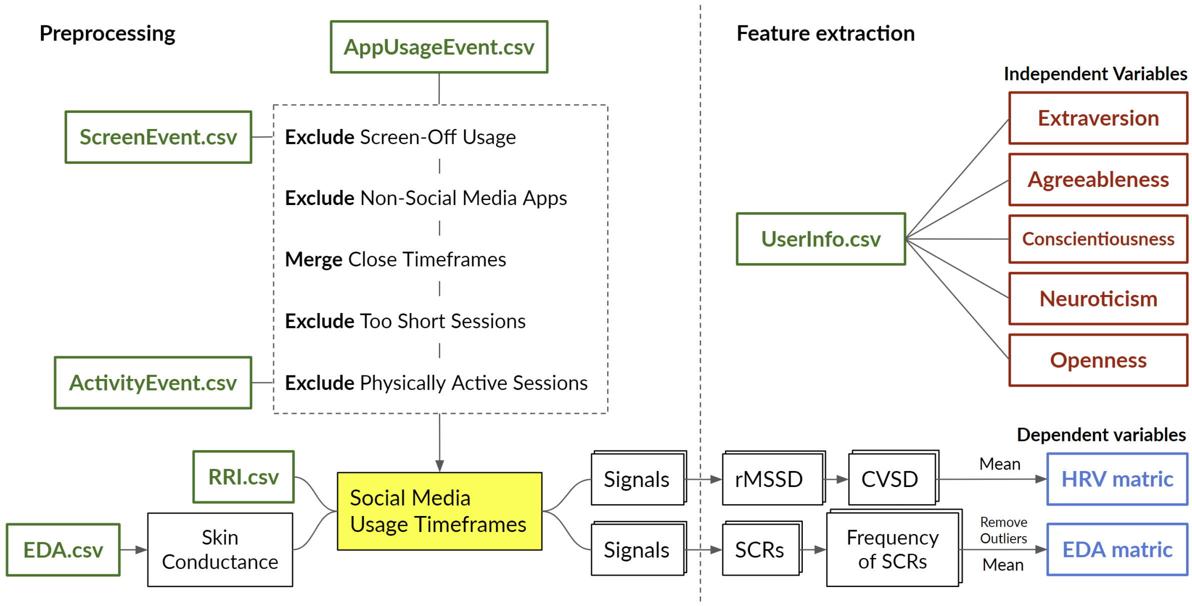

Figure 1: Pipeline of the preprocessing and feature extraction procedure

3.3 Analysis

The described processing procedure produced the HRV and EDA results for each participant. Before our analysis, we exclude participants who have fewer than five records of HRV or EDA results. This exclusion threshold is selected to filter out participants with insufficient data while maintaining a reasonable number of participants.

We conducted a multivariate multiple regression using the five personality traits as independent variables, with HRV and EDA as the dependent variables. Given that all variables are continuous, and there are multiple independent and dependent variables, it made sense that multivariate regression was used. It should be noted that the method was picked assuming the dependent variables are independent from each other. We could not find any research that confirms if HRV and EDA are related or unrelated to one another. It could be argued for both cases; while these aspects can serve as indicators for stress as mentioned before, belonging within the same human body as biological functions, they also measure different physiological responses with HRV primarily associated with autonomic nervous system regulation and cardiovascular function, and EDA measuring the skin's electrical activity. Because these two metrics capture different facets of physiological response, it's plausible that they may not exhibit a direct relationship.

Before the multivariate multiple regression is done, we performed a correlation test for each personality trait to another to check for co-linearity; for independent traits that are too highly correlated to one another, such traits would then be removed from the regression analysis to ensure its integrity.

A multivariate multiple regression model was done by performing separate multiple regression on each dependent variable. Afterwards, multivariate analysis was performed to determine which independent variable significantly influence all the dependent variables. Then based on the result, new regression models would be developed if exclusion of existing independent variables were supported.

After all the analysis was done, we performed statistical tests to validate whether any assumption was violated. Multivariate multiple regression has the same assumptions as multiple regression, which are linearity, independence of errors (normality of residuals), and homoscedasticity.

4. Results

We ran the correlation test at the beginning and found that none personality trait pairs have a correlation magnitude of 0.7 or above. Thus for the regressions, we ran with all five personality traits.

Based on table 1 and 2, the regression results for HRV show that the model explains 16.4% of the variance in CVSD's average, with conscientiousness having a significant negative impact (coef = -0.0042, p = 0.010). For the regression for EDA, it explains 15.4% of the variance in SCR count per minute, with openness having a significant negative effect (coef = -0.0546, p = 0.003). Other personality traits in both models do not show significant effects. Yet we found only the regression model for HRV is shown to be significant, while it is not the case for EDA for the alpha value of 0.05.

The Shapiro-Wilk test suggests that the residuals of both models are normally distributed, with p-values of 0.9605 and 0.9774 for the HRV and EDA models, respectively. The White test suggests homoscedasticity in both models, with p-values of 0.7814 and 0.2188 for the HRV and EDA models, respectively.

Table 1: Multiple Regression Results - HRV - All personality traits

Dep. Variable:	CVSD_mean	**R-squared:**	0.164
Model:	OLS	**Adj. R-squared:**	0.095
Method:	Least Squares	**F-statistic:**	2.391
No. Observations:	67	**Prob (F-statistic):**	0.0480
Df Residuals:	61	**Log-Likelihood:**	146.06
Df Model:	5	**AIC:**	-280.1
Covariance Type:	nonrobust	**BIC:**	-266.9

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.2000	0.030	6.733	0.000	0.141	0.259
openness	-0.0002	0.001	-0.154	0.878	-0.003	0.002
conscientiousness	-0.0042	0.002	-2.675	0.010	-0.007	-0.001
neuroticism	0.0018	0.001	1.247	0.217	-0.001	0.005
extraversion	-0.0011	0.001	-0.934	0.354	-0.004	0.001
agreeableness	0.0017	0.001	1.169	0.247	-0.001	0.005

Table 2: Multiple Regression Results - EDA - All personality traits

Dep. Variable:	scr_count_per_minute	**R-squared:**	0.154
Model:	OLS	**Adj. R-squared:**	0.085
Method:	Least Squares	**F-statistic:**	2.229
No. Observations:	67	**Prob (F-statistic):**	0.0627
Df Residuals:	61	**Log-Likelihood:**	-32.013
Df Model:	5	**AIC:**	76.03
Covariance Type:	nonrobust	**BIC:**	89.25

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	2.8302	0.424	6.681	0.000	1.983	3.677
openness	-0.0546	0.018	-3.120	0.003	-0.090	-0.020
conscientiousness	-0.0099	0.022	-0.442	0.660	-0.055	0.035
neuroticism	-0.0019	0.020	-0.095	0.925	-0.042	0.039
extraversion	0.0184	0.017	1.064	0.292	-0.016	0.053
agreeableness	-0.0057	0.021	-0.266	0.791	-0.048	0.037

Then we ran the multivariate analysis, which the results show that openness and conscientiousness significantly affect the dependent variables across all test statistics, thus we decided to model the regression using only these traits as independent variables. We found when taking into account only openness and conscientiousness, the EDA regression model is statistically significant with a similar coefficient. At the same time, what is found for HRV with regards to 5 independent variables still hold for openness and conscientiousness. The effect size does drop by a small margin from what is reported previously to 13.6% and 11.4% for EDA and HRV respectively.

The Shapiro-Wilk test suggests that the residuals of both models are normally distributed, with p-values of 0.8606 and 0.8619 for the HRV and EDA models, respectively. The White test suggests homoscedasticity in both models, with p-values of 0.4434 and 0.2087 for the HRV and EDA models, respectively.

Table 3: Multivariate linear model

Intercept	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.4081	2.0000	60.0000	43.5086	0.0000
Pillai's trace	0.5919	2.0000	60.0000	43.5086	0.0000
Hotelling-Lawley trace	1.4503	2.0000	60.0000	43.5086	0.0000
Roy's greatest root	1.4503	2.0000	60.0000	43.5086	0.0000

Openness	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.8622	2.0000	60.0000	4.7931	0.0117
Pillai's trace	0.1378	2.0000	60.0000	4.7931	0.0117
Hotelling-Lawley trace	0.1598	2.0000	60.0000	4.7931	0.0117
Roy's greatest root	0.1598	2.0000	60.0000	4.7931	0.0117

Conscientiousness	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.8930	2.0000	60.0000	3.5964	0.0335
Pillai's trace	0.1070	2.0000	60.0000	3.5964	0.0335
Hotelling-Lawley trace	0.1199	2.0000	60.0000	3.5964	0.0335
Roy's greatest root	0.1199	2.0000	60.0000	3.5964	0.0335

Neuroticism	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.9749	2.0000	60.0000	0.7709	0.4671
Pillai's trace	0.0251	2.0000	60.0000	0.7709	0.4671
Hotelling-Lawley trace	0.0257	2.0000	60.0000	0.7709	0.4671
Roy's greatest root	0.0257	2.0000	60.0000	0.7709	0.4671

Extraversion	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.9677	2.0000	60.0000	1.0025	0.3730
Pillai's trace	0.0323	2.0000	60.0000	1.0025	0.3730
Hotelling-Lawley trace	0.0334	2.0000	60.0000	1.0025	0.3730
Roy's greatest root	0.0334	2.0000	60.0000	1.0025	0.3730

Agreeableness	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.9768	2.0000	60.0000	0.7123	0.4946
Pillai's trace	0.0232	2.0000	60.0000	0.7123	0.4946
Hotelling-Lawley trace	0.0237	2.0000	60.0000	0.7123	0.4946
Roy's greatest root	0.0237	2.0000	60.0000	0.7123	0.4946

Table 4: OLS Regression Results - HRV - Openness and Conscientiousness only

Dep. Variable:	CVSD_mean	**R-squared:**	0.114
Model:	OLS	**Adj. R-squared:**	0.086
Method:	Least Squares	**F-statistic:**	4.102
No. Observations:	67	**Prob (F-statistic):**	0.0211
Df Residuals:	64	**Log-Likelihood:**	144.10
Df Model:	2	**AIC:**	-282.2
Covariance Type:	nonrobust	**BIC:**	-275.6

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	0.2256	0.019	11.679	0.000	0.187	0.264
openness	-0.0006	0.001	-0.530	0.598	-0.003	0.002
conscientiousness	-0.0041	0.001	-2.758	0.008	-0.007	-0.001

Table 5: OLS Regression Results - EDA - Openness and Conscientiousness only

Dep. Variable:	scr_count_per_minute	**R-squared:**	0.136
Model:	OLS	**Adj. R-squared:**	0.109
Method:	Least Squares	**F-statistic:**	5.055
No. Observations:	67	**Prob (F-statistic):**	0.00915
Df Residuals:	64	**Log-Likelihood:**	-32.721
Df Model:	2	**AIC:**	71.44
Covariance Type:	nonrobust	**BIC:**	78.06

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	2.8786	0.270	10.644	0.000	2.338	3.419
openness	-0.0512	0.016	-3.105	0.003	-0.084	-0.018
conscientiousness	-0.0086	0.021	-0.412	0.682	-0.051	0.033

Given that the coefficients for both personality traits remain similar across both regression models, we decided to pick the latter models to graph scatter-plots to visualize the best fit lines generated using the regression models. We also attempted visualizing the 95% confidence interval of the best fit line's coefficient by using the calculated standard error from the regression model.

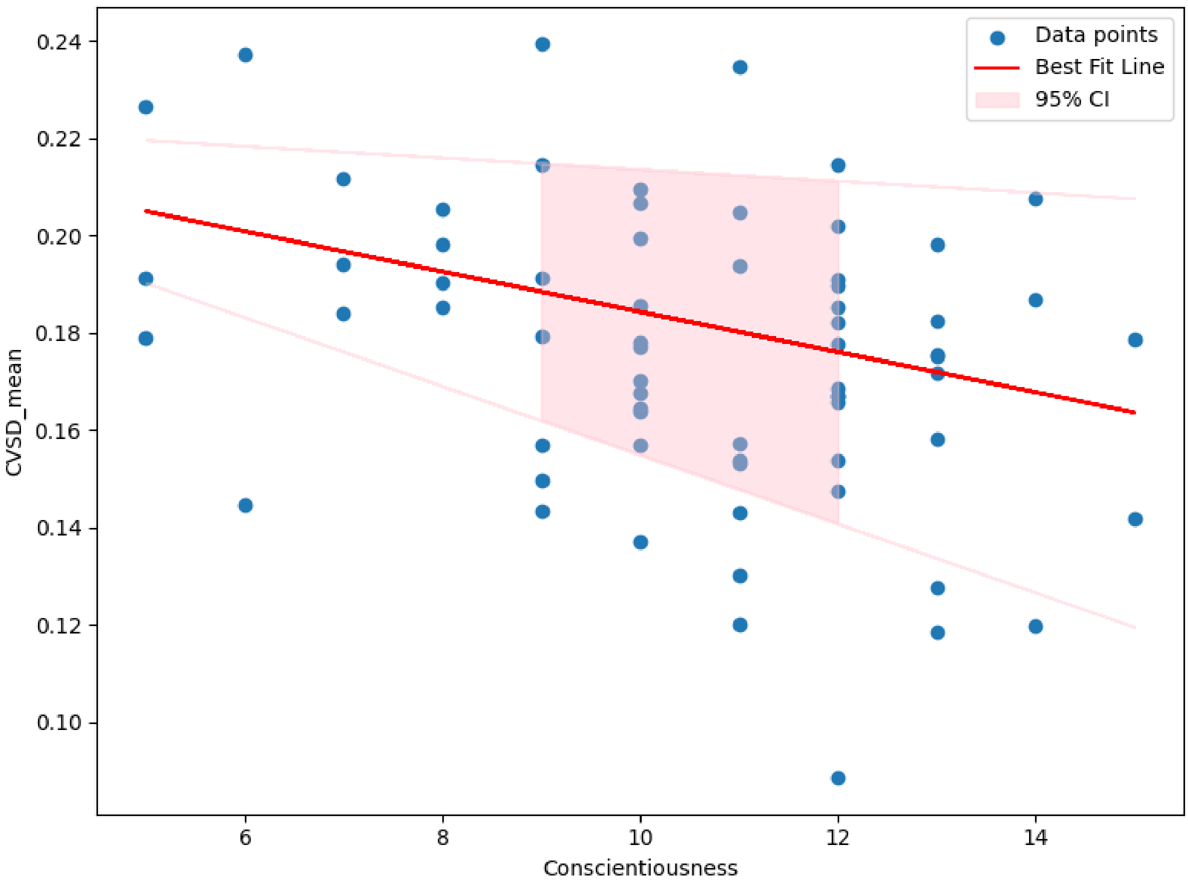

Figure 2: Scatter Plot with Best Fit Line for HRV - Conscientiousness

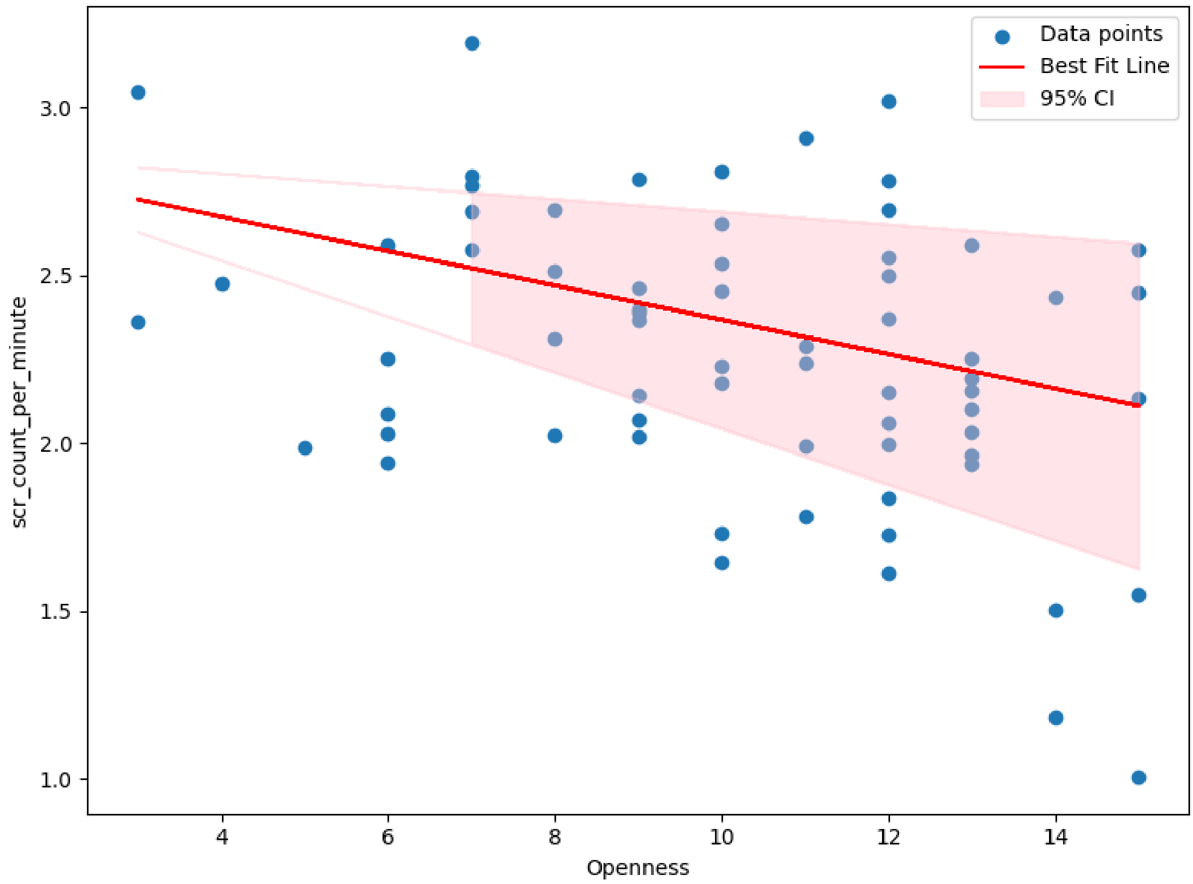

Figure 3: Scatter Plot with Best Fit Line for EDA - Openess

5. Discussion and conclusion

5.1 Results Analysis

Looking at the scatter-plots, we see both best fit lines have a downward trend as the personality trait score gets lower, and that the magnitude of the trend itself looks visually noticeable. The standard error regions for both plots are large, but even in the most extreme case, the best fit lines should still follow an inverse relation between the two variable. It is noted that some regions are uncolored, presumably because in such cases the confidence interval is very narrow - the standard error is close the coefficient, which leads to little or no color fill.

According to existing literature, a person with high stress is likely to have low HRV and high EDA [17]. Our results suggest that participants who scored highly on conscientiousness and low on openness likely experience low HRV and high EDA respectively when using social media, which in turn implies that they experience the most stress at that particular moment. The suggestion is only true if we consider the regression model to take only openness and conscientiousness into account; if 5 personality traits are used as predictors, then only conscientiousness is significantly affecting the model with a 0.05 alpha value. Another aspect to note is that the effect size for all regression models fall within 0.11 and 0.16, meaning the impacts on HRV/EDA by personalities are modest, but not insignificant.

While we can infer the previous statement from the results, we cannot meaningfully conclude whether they experience stress differently due to a variation in personality traits, or that it is due to the social media usage without further study. Another way to interpret these results is that social media is being used to cope with stress. Some papers find that conscientiousness and openness are positively correlated with finding and utilizing adaptive coping methods, which can suggest that people scoring high/low in these traits may use social media for various activities to help them manage stress more effectively [23], [24]. Again, additional work needs to be done to confirm.

5.2 Limitation

Although the K-EmoPhone dataset provides a diverse range of data for various affective computing studies, the data we used in this research contains considerable noise and artifacts due to the accuracy of sensors and the wearing position of the MS Band 2. For the RRI data that we used for HRV feature extraction, the raw data is in essence Pulse-Pulse intervals instead of inter-beat intervals. Therefore, we are actually measuring Pulse Rate Variability (PRV) instead of HRV. Previous research suggests that short-term variability is often overestimated by PRV, possibly because of respiration and cardiovascular activities [41].

Also, physical activity and some stressors seem to enlarge the difference between PRV and HRV results, often to an unacceptable extent, partly because when the sensor position is positioned around the wrist, the movement of the arm can significantly influence blood flow [41]. Specifically, in our research, our feature extraction process results in generally higher average RMSSD values. This issue can be mitigated by using electrocardiogram (ECG) sensor instead of photoplethysmography (PPG) sensor. Furthermore, for both the RRI and EDA data, the fit between the sensor and the skin when wearing the MS Band 2 can significantly influence the reliability of the collected data, especially for EDA. In our study, the raw EDA data contains a considerable proportion of skin resistance values at 340,330 kΩ, which is the maximum value that the sensor can provide. This suggests that the sensor is not in contact with skin. We have added procedures to remove outliers caused by this issue. However, the general reliability of the data is still affected.

One other limitation of this study is that we only measure HRV and EDA during social media usage, without establishing a baseline for each participant. This absence of baseline measurements prevents within-subject comparisons of the effects of social media use, thereby reducing the reliability of our conclusions. Another is the fact that the K-Emophone dataset contains information of presumably most, if not all Korean national, meaning that the cultural aspect is worth considering as well regarding stress behaviors [42].

5.3 Conclusion

Our study aimed to explore the relationship between personality traits and stress indicators, specifically HRV and EDA, during social media usage. By processing the K-Emophone dataset to remove unwanted features, and analysing the processed data with multivariate multiple regression, our findings suggest that high conscientiousness is associated with low HRV and high openness is linked to high EDA, indicating higher stress levels for these personality traits combination when using social media. However, these correlations are only significant when focusing on openness and conscientiousness alone, at least when referring to EDA. The results align with existing literature that suggests personality traits and stress are connected. Given the lack of research done to find the relationship between personality traits and stress in the context of social media usage, our research can serve as a foundation for future studies. Further research is needed to clarify whether these stress responses are directly caused by social media usage or influenced by personality traits directly, whether as stressors or as coping mechanism.

5.4 Future work

Future work that addresses the aforementioned limitations can enhance the validity of the findings. First, to address the issue of noise and artifacts of HRV and EDA data, future studies could: (1) Use ECG sensor instead of PPG sensor for HRV measurements to reduce noise and artifacts; (2) Design a more controlled environment that ensures the EDA sensor maintains good contact with the skin for reliable data collection; (3) Design a more controlled environment to avoid major physical activities. Second, further work can be done to apply a within-subject approach, further investigating the changes before and after using social media; with this approach, perhaps we can explore if social media can act as a stressor or a coping method against stress. Future researchers can also investigate the correlation between HRV, EDA and the self-reported stress level, which is also provided by the K-EmoPhone dataset, to determine whether a person's own perceived sense of stress can be correctly predicted by the biosignals.

References

[1] S. J. Dixon, *Biggest social media platforms 2024*, Apr. 2024. [Online]. Available: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

[2] L. y. Lin, J. E. Sidani, A. Shensa *et al.*, 'Association between social media use and depression among u.s. young adults,' *Depression and Anxiety*, vol. 33, no. 4, pp. 323–331, Jan. 2016. DOI: 10.1002/da.22466.

[3] H. C. Woods and H. Scott, 'sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem,' *Journal of Adolescence*, vol. 51, no. 1, pp. 41–49, Jun. 2016. DOI: 10.1016/j.adolescence.2016.05.008.

[4] C. Berryman, C. J. Ferguson and C. Negy, 'Social media use and mental health among young adults,' *Psychiatric Quarterly*, vol. 89, no. 2, pp. 307–314, Nov. 2017. DOI: 10.1007/s11126-017-9535-6.

[5] P. M. Valkenburg, A. Meier and I. Beyens, 'Social media use and its impact on adolescent mental health: An umbrella review of the evidence,' *Current Opinion in Psychology*, vol. 44, pp. 58–68, Apr. 2022. DOI: 10.1016/j.copsyc.2021.08.017.

[6] A. Orben and A. K. Przybylski, 'The association between adolescent well-being and digital technology use,' *Nature Human Behaviour*, vol. 3, no. 2, pp. 173–182, Jan. 2019. DOI: 10.1038/s41562-018-0506-1.

[7] Apr. 2018. [Online]. Available: https://dictionary.apa.org/personality.

[8] R. R. McCrae and O. P. John, 'An introduction to the five-factor model and its applications,' *Journal of Personality*, vol. 60, no. 2, pp. 175–215, Jun. 1992. DOI: 10.1111/j.1467-6494.1992.tb00970.x.

[9] M. C. Ashton, K. Lee, M. Perugini *et al.*, 'A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages.,' *Journal of Personality and Social Psychology*, vol. 86, no. 2, pp. 356–366, 2004. DOI: 10.1037/0022-3514.86.2.356.

[10] H. Eysenck, 'Four ways five factors are not basic,' *Personality and Individual Differences*, vol. 13, no. 6, pp. 667–673, Jun. 1992. DOI: 10.1016/0191-8869(92)90237-j.

[11] J. Block, 'A contrarian view of the five-factor approach to personality description.,' *Psychological Bulletin*, vol. 117, no. 2, pp. 187–215, 1995. DOI: 10.1037/0033-2909.117.2.187.

[12] S. E. Babcock and C. A. Wilson, 'Big five model of personality,' *The Wiley Encyclopedia of Personality and Individual Differences*, pp. 55–60, Sep. 2020. DOI: 10.1002/9781119547174.ch186.

[13] M. Vollrath, 'Personality and stress,' *Scandinavian Journal of Psychology*, vol. 42, no. 4, pp. 335–347, Sep. 2001. DOI: 10.1111/1467-9450.00245.

[14] Apr. 2018. [Online]. Available: https://dictionary.apa.org/stress.

[15] H. Selye, *The stress of life*. McGraw-Hill, 1978.

[16] R. S. Lazarus and S. Folkman, *Stress, appraisal, and coping*. Springer, 1985.

[17] A. Dorsey, E. M. Scherer, R. Eckhoff and R. Furberg, ‘Measurement of human stress: A multidimensional approach,’ *Research Triangle Park (NC): RTI Press*, Jun. 2022. DOI: 10.3768/rtipress.2022.op.0073.2206.

[18] A. D. Crosswell and K. G. Lockwood, ‘Best practices for stress measurement: How to measure psychological stress in health research,’ *Health Psychology Open*, vol. 7, no. 2, p. 205 510 292 093 307, Jul. 2020. DOI: 10.1177/2055102920933072.

[19] S. L. Dailey, K. Howard, S. M. Roming, N. Ceballos and T. Grimes, ‘A biopsychosocial approach to understanding social media addiction,’ *Human Behavior and Emerging Technologies*, vol. 2, no. 2, pp. 158–167, Feb. 2020. DOI: 10.1002/hbe2.182.

[20] I. Beyens, E. Frison and S. Eggermont, ‘“i don’t want to miss a thing”: Adolescents’ fear of missing out and its relationship to adolescents’ social needs, facebook use, and facebook related stress,’ *Computers in Human Behavior*, vol. 64, pp. 1–8, Nov. 2016. DOI: 10.1016/j.chb.2016.05.083.

[21] C. Maier, S. Laumer, A. Eckhardt and T. Weitzel, ‘Online social networks as a source and symbol of stress: An empirical analysis,’ *International Conference on Information Systems, ICIS 2012*, vol. 2, pp. 1399–1417, Jan. 2012.

[22] L. N. Wolfers and S. Utz, ‘Social media use, stress, and coping,’ *Current Opinion in Psychology*, vol. 45, p. 101 305, Jun. 2022. DOI: 10.1016/j.copsyc.2022.101305.

[23] J. A. Penley and J. Tomaka, ‘Associations among the big five, emotional responses, and coping with acute stress,’ *Personality and Individual Differences*, vol. 32, no. 7, pp. 1215–1228, May 2002. DOI: 10.1016/s0191-8869(01)00087-3.

[24] S. Grant and J. Langan-Fox, ‘Occupational stress, coping and strain: The combined/interactive effect of the big five traits,’ *Personality and Individual Differences*, vol. 41, no. 4, pp. 719–732, Sep. 2006. DOI: 10.1016/j.paid.2006.03.008.

[25] J. Luo, B. Zhang, M. Cao and B. W. Roberts, ‘The stressful personality: A meta-analytical review of the relation between personality and stress,’ *Personality and Social Psychology Review*, vol. 27, no. 2, pp. 128–194, Jul. 2022. DOI: 10.1177/10888683221104002.

[26] T. R. Schneider, ‘The role of neuroticism on psychological and physiological stress responses,’ *Journal of Experimental Social Psychology*, vol. 40, no. 6, pp. 795–804, Nov. 2004. DOI: 10.1016/j.jesp.2004.04.005.

[27] K. C. Gunthert, L. H. Cohen and S. Armeli, ‘The role of neuroticism in daily stress and coping.,’ *Journal of Personality and Social Psychology*, vol. 77, no. 5, pp. 1087–1100, 1999. DOI: 10.1037/0022-3514.77.5.1087.

[28] T. R. Schneider, T. A. Rench, J. B. Lyons and R. R. Riffle, ‘The influence of neuroticism, extraversion and openness on stress responses,’ *Stress and Health*, vol. 28, no. 2, pp. 102–110, Jun. 2011. DOI: 10.1002/smi.1409.

[29] S. Kang, W. Choi, C. Y. Park *et al.*, ‘K-emophone: A mobile and wearable dataset with in-situ emotion, stress, and attention labels,’ *Scientific Data*, vol. 10, no. 1, Jun. 2023. DOI: 10.1038/s41597-023-02248-2.

[30] , and, ‘ bfi (big five inventory),’, vol. 32, no. 1, pp. 47–65, 2011.

[31] [Online]. Available: https://developer.android.com/reference/android/app/usage/UsageEvents.Event#MOVE_TO_BACKGROUND.

[32] [Online]. Available: https://developer.android.com/develop/sensors-and-location/location/transitions.

[33] T. Pham, Z. J. Lau, S. H. Chen and D. Makowski, 'Heart rate variability in psychology: A review of hrv indices and an analysis tutorial,' *Sensors*, vol. 21, no. 12, p. 3998, Jun. 2021. DOI: 10.3390/s21123998.

[34] O. Mayo, M. Lavidor and I. Gordon, 'Interpersonal autonomic nervous system synchrony and its association to relationship and performance – a systematic review and meta-analysis,' *Physiology Behavior*, vol. 235, p. 113 391, 2021, ISSN: 0031-9384. DOI: https://doi.org/10.1016/j.physbeh.2021.113391. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031938421000834.

[35] S. Laborde, E. Mosley and J. F. Thayer, 'Heart rate variability and cardiac vagal tone in psychophysiological research – recommendations for experiment planning, data analysis, and data reporting,' *Frontiers in Psychology*, vol. 08, Feb. 2017. DOI: 10.3389/fpsyg.2017.00213.

[36] J. Sacha and W. Pluta, 'Alterations of an average heart rate change heart rate variability due to mathematical reasons,' *International Journal of Cardiology*, vol. 128, no. 3, pp. 444–447, 2008, ISSN: 0167-5273. DOI: https://doi.org/10.1016/j.ijcard.2007.06.047. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167527307012223.

[37] M. Altini, *Normalized heart rate variability for population comparisons in hrv4training*, Jul. 2016. [Online]. Available: https://www.hrv4training.com/blog2/normalized-heart-rate-variability-for-population-comparisons-in-hrv4training.

[38] [Online]. Available: https://neuropsychology.github.io/NeuroKit/.

[39] J. J. Braithwaite, D. G. Watson, R. Jones and M. Rowe, 'A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments,' *Psychophysiology*, vol. 49, no. 1, pp. 1017–1034, 2013.

[40] H. F. Posada-Quintero and K. H. Chon, 'Innovations in electrodermal activity data collection and signal processing: A systematic review,' *Sensors*, vol. 20, no. 2, p. 479, Jan. 2020. DOI: 10.3390/s20020479.

[41] A. Schäfer and J. Vagedes, 'How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram,' *International Journal of Cardiology*, vol. 166, no. 1, pp. 15–29, 2013, ISSN: 0167-5273. DOI: https://doi.org/10.1016/j.ijcard.2012.03.119. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167527312003269.

[42] C.-A. Chun, R. H. Moos and R. C. Cronkite, 'Culture: A fundamental context for the stress and coping paradigm,' *Handbook of Multicultural Perspectives on Stress and Coping*, pp. 29–53, DOI: 10.1007/0-387-26238-5_2.