A Survey of Psychological Personality Classification Approaches

Mervat Ragab Bakry[1,a], Mona Mohamed Nasr[2,b], and Fahad Kamal Al-sheref [1,c]

[1]*Faculty of Computers and Artificial Intelligence, Beni-Suef University, Egypt*
[2]*Faculty of Computers and Artificial Intelligence, Helwan University, Egypt*

[a]*mervatragab@fcis.bsu.edu.eg* [b] *m.nasr@helwan.edu.eg* [c]*drfahad@fcis.bsu.edu.eg*

A B S T R A C T

Online social networks (OSNs) have become essential ways for users to socially share information and feelings, communicate, and thoughts with others through online social networks. Online social networks such as Twitter and Facebook are some of the most common OSNs among users. Users' behaviors on social networks aid researchers for detecting and understanding their online behaviors and personality traits. Personality detection is one of the new difficulties in social networks. Machine learning techniques are used to build models for understanding personality, detecting personality traits, and classifying users into different kinds through user generated content based on different features and measures of psychological models such as PEN (Psychoticism, Extraversion, and Neuroticism) model, DISC (Dominance, Influence, Steadiness, and Compliance) model, and the Big-five model (Openness, Extraversion, Consciousness, Agreeableness, and Neurotic) which is the most accepted model of personality. This survey discusses the existing works on psychological personality classification.

*Keywords:*Online social networks, PEN, DISC, Big-five

1. Introduction

People's daily life influenced by social network sites(SNS) with a residual speed in order to become a significant social point for social interactions (Alireza Souri et al., 2018). Social network permit users to express themselves; feelings, thoughts, opinions, statuses, and socially communicate with others. Several psychological models of personality have been suggested, such as the big-five model, the DISC model (Dominance, Influence, Steadiness, and Compliance), and PEN model (Psychoticism, Extraversion, and Neuroticism) (Yasmín Hernández et al., 2018). Regarding to Table 1, persons in the big-five model differ in terms of the OCEAN.

This paper is structured as follows: section two presents machine learning algorithms, section three presents discussion, section four presents the most psychological personality model used for classifying user personality types, and finally section five presents conclusion of the survey.

Table 1 - Outline of the big-five model (H.J.Eysenck, 1991).

Personality traits	Characteristics
Openness	Imaginative, Curious, Creative, Pursuits of Self-actualization, Open-minded, and Independent.
Conscientiousness	Efficient, Well-organized, Dependable, Self-sufficient, Planner, and Reliable.
Extroversion	Talkative, Sociability, Friendly, and Use Positive Emotions.
Agreeableness	Trustworthy, Kind, Affectionate, Cooperative, Helpful, Optimistic, and Nurturing.
Neuroticism	Emotional Instability, Sadness, Moodiness, Angry, Upset, Sensitive, and Neurotic.

1. Machine learning algorithms

Machine learning techniques permit automatic and scalable methods in which visions from huge, multi-dimensional data can be gathered. Broadly, machine learning is the capability of computers in order to automatically learn patterns and create implications from data. These algorithms can be classified along many varies interlocutor (Praveen Murthy et al., 2014).

2.1 Supervised learning

Includes all machine learning algorithms which map input data to a known class label(s) or target value (Praveen Murthy et al., 2014). Table 2 contains all papers which used supervised machine learning algorithms.

Table 2 - Supervised machine learning algorithms.

Reference	ML algorithms	Proposed ML model	Acc.	Psychological model	Dataset	Features
Alireza Souri et al., 2018	Support Vector Machine, Naïve Bayes, Neural Network, Decision Trees, Boosting SVM, Boosting NB, Boosting NN, and Boosting DT.	Boosting DT. With 70% training and 30% testing.	82.2 %	Big-five.	100 volunteers of Face-book users.	• Number of(likes, languages, sports, books, works, activities, games, groups, links, movies, friends, notes, educations, questions, music, TV shows, interests).
Yasmín Hernández et al., 2018	Naïve Bayes, Support Vector Machine, and KNN.	Support Vector Machine using Percentag	75%	DISC.	120 Facebook users.	• Participants answered questions about personal information (gender, age, schooling, marital status, preferred

Reference	ML algorithms	Proposed ML model	Acc.	Psychological model	Dataset	Features
		e split.				social network, and number of friends).
MICHAEL M. TADESSE et al., 2018	Support Vector Machine, Logistic Regression. and XGBoost.	XGBoost with 10-fold cross validation.	74.2%	Big-five.	myPersonality dataset (250 users with 9917 status updates).	• Status updates. • Text posts. • Language use habit.
Thin Nguyen et al., 2014	Logistic Regression.	Logistic Regression. With 10-fold cross validation.	89%		10,000 people in 24 mental health communities.	• Topics of interest • Emotions expressed in the post. • Mood tag . • Sentiment information. • Language style.
Dawn Beverley Branley, and Judith Covey, 2017	Regression.	Regression.		Willingness.	Online survey of 1102 users from 77 countries.	• Photos shared. • Location shared. • Shared videos.
Reece Akhtar et al., 2018	Regression.	Regression.		Dark triad.	51,712 Facebook users.	• Status updates.
Namrata Sonawane et al., 2018	Naïve Bayes	Naïve Bayes.	Low perf orma nce.		80 Facebook users.	• Posts.
Hsin-Chang Yang ,and Zi-Rui Huang, 2018	Naïve Bayes and Support Vector Machine.	Support Vector Machine.		Big-five.		
YilunWang, 2015	Support Vector Machine, Linear Regression, and Random Forest.	Support Vector Machine, Linear Regression, and Random Forest.	All give simil ar resul ts.	MBTI.	Twitter.	• Human language used in twitter.

Reference	ML algorithms	Proposed ML model	Acc.	Psychological model	Dataset	Features
Jong Hyun Park, 2017	Classification.	Classification.		MBTI.	41 users.	• Predefined user preference. • User profiles, relationship between resources.
Joel Philip et al., 2019	Support Vector Machine and Naïve Bayes.	Support Vector Machine.		Big-five.	1196 users.	• Twitter: Twitter API provides tweets, likes, and number of followers . • Facebook: Graph API can help us get all the publicly available information like liked pages, liked movies. • Quora: Quora API helps us get all the questions or answers provided by the users.

Table 2 – (continued).

Reference	ML algorithms	Proposed ML model	Acc.	Psychological model	Dataset	Features
Tommy Tandera et al., 2017	Naive Bayes, Support Vector Machine, Logistic Regression, Gradient Boosting, Linear Discriminant Analysis, and Deep learning.		Average accuracy: 74.17 %	Big-five.	myPersonality dataset and 150 Facebook users.	• Status updates.
Chad Crowe, 2018	Text-based Neural Networks , Image-based convolutional Neural Networks, Word2Vec, combined NN, CNN model, Decision Trees.	Combination of NN and CNN models.	The study also demonstrates increased model performance from combining NN and CNN models.		The research collects 350k Facebook posts.	• Facebook Posts, comments.(word vectors and processed images).
Xuan-Son Vu and Lili Jiang, 2018	Majority and Naive Bayes.	Majority and Naive Bayes.	Majority : 0.848 Naive Bayes: 0.97	Big-five.	- myPersonality.	• Status updates. • 7111 Lexical features. • 250 Topic features. • 300 Semantic features.
Golnoosh Farnadi et al., 2016	Regression.	Regression.		Big-five.	myPersonality, YouTube vlog dataset, and Twitter dataset.	• Facebook: age, gender, number of likes, status updates. • Twitter: tweets.

2.2 Deep learning

Deep learning is a machine learning technique that uses neural networks to learn. Although deep learning is similar to a traditional neural network, it has many more hidden layers.

The more complex the problem, the more hidden layers there are in the model (IBM Corp, 2019). Table 3 contains all papers which used deep learning algorithms.

Table 3 – Deep learning algorithms.

Reference	ML algorithms	Proposed ML model	Acc.	Psychological model	Dataset	Features
Di Xue et al., 2018	Utilize a hierarchical deep neural network.	Utilize a hierarchical deep neural network.		Big-five.	115,864 of Facebook users.	• Text posts(we extract the following 5 special linguistic statistical features from the tokenized text posts: (1) rate of emoticons; (2) rate of tokens which have no less than 3 tandem duplicated letters or symbols; (3) rate of capital letters; (4) rate of capitalized words; (5) total number of text posts of each user).

Table 3 – (continued).

Reference	ML algorithms	Proposed ML model	Acc.	Psychological model	Dataset	Features
Guozhen An and Rivka Levitan, 2018	Deep Learning.	Deep Learning.	Average accuracy of 67%.	Big-five.	myPersonality dataset and 346 individual speakers.	• Status updates.
Xiu-Shen Wei, 2018	Deep learning (Deep Bimodal Regression).	Deep learning (Deep Bimodal Regression).	0.9130 mean Accuracy.	Big-five.	10,000 videos.	• Images.

2. Discussion

(Alireza Souri et al., 2018) Predicting user personality via personality test in the social network and investigating five types of personality traits for their profile pictures using big-five model. The proposed machine learning model is boosting decision tree with accuracy 82.2%, this study also providing a recommendation system for friends' relationships in social media groups. The limitations for this paper are that it was insignificant to recognize the precise degree of persons on each trait and low number of dataset.

(Yasmín Hernández et al., 2018) Discovering personality through statistical analysis of language and the DISC model. The proposed machine learning model is support vector machine with accuracy 75%. The limitation of this paper is low number of dataset.

(MICHAEL M. TADESSE et al., 2018) Elucidate the correlation between users' personality traits and their interaction behaviors in social network sites using big-five model. The proposed machine learning model is XGBoost with accuracy 74.2%. This study also concludes that extraversion positively correlated with network size (number of friends, social groups, and Facebook usage). The limitations of this paper are low number of dataset and performance of the outcomes inclined to be rather confined.

(Thin Nguyen et al., 2014) Deliver a set of foretells of depression, comprising of the traits which discriminate mental health societies from others, illustration the power of social media as a new canal for depression, and displaying, observing, and supplying basis for online caution systems. The proposed machine learning of this study is logistic regression with accuracy 89%.

(Dawn Beverley et al., 2017) Forecasting risky behaviors (sexing, sharing location with public, sending perplexing photos, prevalence of pranks), discovering the relationship between personality kind and the risks that fall on, and conclude that extraverted and conscientiousness users are exposed to these risks more. The proposed machine learning of this study is regression. The limitation of this paper is that the existing study used new measure of prototype similarity, which can also account for some of the variances in the forecasting ability of this feature compared to earlier studies; while earlier studies have generally depended on self-reported impressions of similarity.

(Reece Akhtar et al., 2018) Discover the dark side of personality via language utilized in status updates by using dark triad. The proposed machine learning of this study is regression. The limitations of this paper are that the dark triad is incomplete model of dark personalities, the reviewed literature has based on twitter data in which persons may have created contents that are not popularization to wider population, and depending only on language in status updates.

(Namrata Sonawane et al., 2018) Classifying depression into different levels and providing doctor's location near to user's location. The proposed machine learning of this study is naïve bayes with low performance. The limitation of this study is that it is only relied on posts in their classification in which users do not permanently clarify their stressful state straightly in Facebook posts.

(Hsin-Chang Yang et al., 2018) Predicted two classes; conscientiousness and neuroticism using big-five model.

The proposed machine learning model of this study is support vector machine. This study predicted only two personality kinds.

(YilunWang et al., 2015) Understanding personality via social media using Myers-Briggs Type Indicator (MBTI). This study used linear regression, support vector machine, and random forest; all give similar results.

(Jong Hyun Park, 2017) Analyzing and classifying personal user priority through resource practice history relied on the MyersBriggs type indicator (MBTI), proposed a technique to admonish tailored resources for classifying user kinds, proposed a method to reduce the recommendation time and increased user satisfaction. This study used classification model. The limitation of this paper is low number of dataset.

(Joel Philip et al., 2019) The proposed study tries to simulate a statistical report making system using big-five model. The proposed machine learning model is support vector machine with average precision: 0.98, average recall: 0.97, and average F1-score: 0.97. The outputs from this study are only two classes.

(Tommy Tandera et al., 2017) Constructing a system that can forecast a personality kind relied on information of Facebook user using big-five model. This study used traditional machine learning; support vector machine, naive bayes, logistic regression, Gradient Boosting, and linear discriminant analysis, also used deep learning. The limitation of this paper is low number of dataset.

(Chad Crowe, 2018) Demonstration of increased model performance from combining using Image-based convolutional neural networks (CNN), Text-Based neural networks (NN), Word2Vec, combined NN, CNN model, and decision trees models. This research also explored constructing models to predict user interaction, contributes to understanding advertisement success on social media, and will aid in the selection and transformation of social media data.

The limitations of this research are that model training time was a limitation, there were fifteen models; each model requires many hours of training time, and model overtraining was also a result of less than perfect hyper parameters and model characteristics.

(Xuan-Son Vu and Lili Jiang, 2018) Proposing a self-adaptive approach for privacy concern recognition based on user personality using big-five model. This study used majority and naive bayes classifier. The experimental studies demonstrate the effectiveness to address a suitable personalized privacy protection for cold-start users. The limitation of this study is there are no gold standard values (or labels) from users regarding their privacy concern.

(Golnoosh Farnadi et al., 2016) Performing a comparative analysis of state-of-the-art computational personality recognition approaches on a different set of social media ground truth data from Facebook, Twitter and YouTube using regression algorithm with big-five model. This study based on three datasets; myPersonality dataset, Twitter dataset, and YouTube vlog dataset. Bachrach et al. employed myPersonality dataset in order to explore how users' action on Facebook depends to their personality. One of the outcomes was that neuroticism has a generally significant negative relationship with the number of friends.

(Di Xue et al., 2018) Proposing a deep learning based approach to detect personality through text posts of online social network users using big-five model. A hierarchical deep neural network is the proposed model for this study.

(Guozhen An and Rivka Levitan, 2018) Proposing two deep learning structures to detect personality by using acoustic-prosodic, psycholinguistic, and lexical features and presenting empirical outcomes of several experimental configurations, involving a cross-corpus condition to assess robustness using big-five model. The accuracy of the proposed model is 67%. The limitation of this study is that the model does not perform better on the five traits.

(Xiu-Shen Wei, 2018) Relied on the proposed framework, this study come up with a solution for the apparent personality analysis competition track in the ChaLearn looking people challenge in a association with ECCV 2016. Their DBR is the winner (first place) of this challenge with 86 registered participants. The proposed deep learning is deep bimodal regression with 0.9130 of mean accuracy. This study also adapted the traditional CNNs as descriptor aggregation networks (DANs) to increase the visual regression accuracy.

(Danny Azucar et al., 2017)Conducting a sequence of meta-analyses to define the forecasting power of digital footprints gathered from social media over big-five model. Also this study investigating the effect of diverse kinds of digital footprints on forecast accuracy. Outcomes specify that performance of forecasts is stable through big-five traits.

(Alixe Lay and Bruce Ferwerda, 2018) Report the necessity for discovering the way of automatic personality estimation by using 'liked' photos on Instagram. Suggesting a research project that targets to forecast users' personality relied on the photos that they like on Instagram using big-five model. This study only relied on images that users liked extracting these features; color-based, content-based and visual sentiment-based features from the gathered photographic data through an Instagram API.

(Hsin-Chang Yang and Zi-Rui Huang, 2018) An approach for the process of recommendation computer games to players related to their identified personality kinds was proposed using text mining and big-five model. This study used myPersonality dataset. The estimation results indicated that reviewers were pleased about the recommended games with average rating score 3.30 over all configurations. However, it is not easy to match this result to other correlated work as there is no gold standard for the similarity between games. It is hard to determine if such identical games reflected their true symmetry. They did not cover the time issue in their model which supposes the players having proportionate game preferences during time.

4. Psychological personality models

Personality affects the whole things in human life. Yet many people never actually expend much time thinking about their personality traits. Understanding personality can provide understanding into human weaknesses and strengths. It can also support person understanding how others see him. Numerous psychological models for personality have been suggested, such as the big-five model; which will discuss in this section, the DISC model, and PEN model. But most modern day psychologists agree there are five main personality kinds; called the "big-five model" which is considered the most acceptable model for personality identification (MICHAEL M. TADESSE et al., 2018).

4.1 Big-five model

4.1.1 Conscientiousness
People are well-organized, efficient, self-sufficient, and dependable. They desire to plan things in progress and target for high accomplishment.

4.1.2 Extroversion
People get power from social action. They're fluent and outgoing and they're relaxed in the interest. Others might view them as officious and attention-seeking.

4.1.3 Agreeableness
People are trustworthy, nice, and friendly to others. They're known for their pro-social actions and they're often devoted to helper work and altruistic events. Other people can view them as inexperienced and very ineffectual.

4.1.4 Openness to experience
People are recognized for their wide domain of interests and imaginative. They're inquisitive and they usually like better variety over inelastic practices. They're known for their detections of self-actualization through extreme, overjoyed experiences such as contemplative retreats or living away from Motherland. Others may view them as changeable and unclear.

4.1.5 Neuroticism
Neurotic people practice a high level of emotional changeability. They're more reactive and excitable and they report greater scores of sad emotions such as worry and petulance.

Conclusion

This survey discusses machine learning algorithms that widely used for psychological personality classification. Social media is a good screening tool that researchers can use for classifying personality. Understanding personality is an important step for dealing, targeting, and predicting user behavior. We conclude that Big-five model is the most widely used for classifying persons.

References
Alireza Souri, Shafgheh Hosseinpour, and Amir Masoud Rahmani. (2018). Personality classifcation based on profiles of social networks' users and the five factor model of personality, Springer.
Yasmín Hernández, Carlos Acevedo Peña, and Alicia Martínez. (2018). Model for Personality Detection Based on Text Analysis, Springer Nature.
MICHAEL M. TADESSE, HONGFEI LIN, BO XU, AND LIANG YANG. (2018). Personality Predictions Based on User Behavior on the Facebook Social Media Platform, IEEE.
H.J.Eysenck. (1991). Dimensions of personality: 16, 5 or 3?—Criteria for a taxonomic paradigm, Elsevier.
Praveen Murthy, Anurag Bharadwaj, P. A. Subrahmanyam, Arnab Roy, Sree Rajan. (2014). Big Data Taxonomy, Cloud Security Allia
Guozhen An and Rivka Levitan. (2018). " Lexical And Acoustic Deep Learning Model For Personality Recognition", Interspeech.
IBM Corp. (2019). Introduction to machine learning, IBM.
Xiu-Shen Wei, Chen-Lin Zhang, Hao Zhang et al. (2018). Deep Bimodal Regression of

Apparent Personality Traits from Short Video Sequences, IEEE.

Xuan-Son Vu and Lili Jiang. (2018). Self-adaptive Privacy Concern Detection for User-generated Content, arXiv.

Golnoosh Farnadi1, Geetha Sitaraman, Shanu Sushmita, Fabio Celli, Michal Kosinski, David Stillwell, Sergio Davalos, Marie-Francine Moens, and Martine De Cock. (2016). Computational personality recognition in social media, Springer.

Danny Azucar, Davide Marengo, and Michele Settanni. (2017). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis, Elsevier.

Alixe Lay and Bruce Ferwerda.(2018). Predicting Users' Personality Based on Their 'Liked' Images on Instagram, UMANIZE.

Hsin-Chang Yang and Zi-Rui Huang. (2018). Mining personality traits from social messages for game recommender systems, Elsevier.

Hsin-Chang Yang , and Zi-Rui Huang. (2018). Mining personality traits from social messages for game recommender systems, Elsevier.

YilunWang. (2015). Understanding Personality through Social Media, Stanford University.

Jong Hyun Park. (2017). Resource recommender system based on psychological user type indicator, Springer.

Joel Philip, Dhvani Shah, Shashank Nayak, Saumik Patel, and Yagnesh Devashrayee. (2019). Machine Learning for Personality Analysis Based on Big Five Model, Springer Nature.

Tommy Tandera, Hendro, Derwin Suhartono*, Rini Wongso, and Yen Lina Prasetio. (2017). Personality Prediction System from Facebook Users, Elsevier.