



# Internet addiction disorder detection of Chinese college students using several personality questionnaire data and support vector machine



Zonglin Di<sup>a</sup>, Xiaoliang Gong<sup>a,\*</sup>, Jingyu Shi<sup>b</sup>, Hosameldin O.A. Ahmed<sup>c</sup>, Asoke K. Nandi<sup>c,a</sup>

<sup>a</sup> School of Electronic and Information Engineering, Tongji University, Shanghai, China

<sup>b</sup> East Hospital, Tongji University School of Medicine, Shanghai 200120, China

<sup>c</sup> Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, Middlesex, UK

## ARTICLE INFO

### Keywords:

Internet addiction (IA)  
IA detection  
Personality questionnaire  
Feature selection  
Support vector machine

## ABSTRACT

With the unprecedented development of the Internet, it also brings the challenge of Internet Addiction (IA), which is hard to diagnose and cure according to the state-of-art research. In this study, we explored the feasibility of machine learning methods to detect IA. We acquired a dataset consisting of 2397 Chinese college students from the University (Age:  $19.17 \pm 0.70$ , Male: 64.17%) who completed Brief Self Control Scale (BSCS), the 11th version of Barratt Impulsiveness Scale (BIS-11), Chinese Big Five Personality Inventory (CBF-PI) and Chen Internet Addiction Scale (CIAS), where CBF-PI includes five sub-features (Openness, Extraversion, Conscientiousness, Agreeableness, and Neuroticism) and BSCS includes three sub-features (Attention, Motor and Non-planning). We applied Student's *t*-test on the dataset for feature selection and Support Vector Machines (SVMs) including C-SVM and  $\nu$ -SVM with grid search for the classification and parameters optimization. This work illustrates that SVM is a reliable method for the assessment of IA and questionnaire data analysis. The best detection performance of IA is 96.32% which was obtained by C-SVM in the 6-feature dataset without normalization. Finally, the BIS-11, BSCS, Motor, Neuroticism, Non-planning, and Conscientiousness are shown to be promising features for the detection of IA.

## 1. Introduction

### 1.1. Background

In recent years, the development of the Internet has brought a lot of benefits in our society. However, it also causes the Internet Addiction Disorder (IAD) problems, which is also named as Pathological Internet Use (PIU). Since IAD was first put forward by Ivan Goldberg in 1995 (Abbott, Cramer, & Sherrets, 1995; Young, 1998b), it has become a social-psychological problem and many researchers have been working on this topic (Dongyun, Ni'na, & Yao, 2018; Griffiths, 2018; Wiederhold, 2018). Although IAD was not officially added into Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition (DSM-V) in 2013, Internet Gaming Disorder (IGD) has been included in Section III, illustrating the importance of this area for further study (Petry & O'brien, 2013; Cho et al., 2014; Hahn, Reuter, Spinath, & Montag, 2017; Association, A.P., 2013; Spada, 2014).

IAD is a compulsive-impulsive spectrum disorder which includes five specific types or addiction: cyber-sexual addiction, cyber-relationship addiction, net compulsions, information overload and computer

addiction (Young, 1998a, 1998b; Nordegren, 2002). As a result, IAD can lead to marriage break-down, job losses, financial problems, academic failures and even death (Whang, Lee, & Chang, 2003; Young, 2004). Many studies indicate that IAD is a multi-dimension construct which has many dependencies like mental health, age, peer influence, social support, family relationship, parental mental health, emotion dysregulation, alexithymia personality and so on (Mo, Chan, Chan, & Lau, 2018; Xiuqin et al., 2010). Among these factors, certain personality traits like self-control, impulsivity, items in Big Five personality including Openness, Extraversion, Conscientiousness, Agreeableness, Neuroticism are regarded to have close association with IAD (Ismail & Zawahreh, 2017; Lam, Peng, Mai, & Jing, 2009; Musetti et al., 2016; Treuer, Fa'bi'an, & Fu'redi, 2001; Zhou, Li, Li, Wang, & Zhao, 2017).

Although IAD can be found in any age group and every occupation, the youths are more vulnerable to IAD. Once they are addicted to the Internet, they will have a deeper addiction level. Influenced by the digital age, the Internet has resulted in the improvement of proficiency in certain courses. However, more and more reports point out the addictive Internet usage problem. Globally, it is estimated that 4–12% of adolescents may demonstrate IAD although the definition of IAD varies

\* Corresponding author.

E-mail address: [gxlshsh@tongji.edu.cn](mailto:gxlshsh@tongji.edu.cn) (X. Gong).

<https://doi.org/10.1016/j.abrep.2019.100200>

Received 3 September 2018; Received in revised form 23 June 2019; Accepted 23 June 2019

Available online 11 July 2019

2352-8532/© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

a lot (Petry & O'Brien, 2013; Yau, Crowley, Mayes, & Potenza, 2012). 15.3% university freshmen in Taiwan and 20.3% adolescents in South Korea were reported to have IAD (Ha et al., 2006; Lin, Ko, & Wu, 2011). In China, the rates ranged from between 2.4%–5.5% in Hunan Province and to 6.4% in Shanxi Province (Mei, Yau, Chai, Guo, & Potenza, 2016). Also, college students are more likely trapped into the Internet among the adolescents because of academic pressure, unlimited Internet accesses and newly experienced freedom from parental control (Young, 2004).

### 1.2. The previous work of IAD detection

IAD has been put forward over twenty years and many researchers have been working on it about either the factors of this disorder or the understanding of this disorder. (Ko et al., 2006) found adolescents who have high novelty seeking, high harm avoidance and low reward dependence are more likely to be addicted to IAD using *t*-test and logistic regression. (Kayi, et al., 2016) investigated relationship between Big Five Personality Traits and Internet Addiction using meta methods which includes 12-study meta-analysis and calculates 13 effect sizes and found that openness to new experiences, conscientiousness, extraversion and agreeableness were negatively related with IAD whereas neuroticism was positively related with IAD. Resilience was found to be protective in IAD according to (Robertson, Yan, & Rapoza, 2018). (Fumero, Marrero, Voltes, & Pen˜ate, 2018) shows that personal factors have a greater impact on IAD.

Machine learning has been widely used in bioinformatics, brain-machine interfaces, medical diagnosis and other medical areas (Baldi & Brunak, 2001; Choi et al., 2018; Müller et al., 2008; Zhang et al., 2018; Kononenko, 2001; Giger, 2018; Guo & Nandi, 2006). However, there are few studies about how to detect IAD using machine learning methods. (Wang, ZHANG, & ZHANG, 2008) proposed a Fuzzy Neural Network (FNN) method to forecast pattern of network addiction, where the number of layers is 3 and the features used are online-hours, frequency, the reason for the Internet, determination, socialization, the Internet skills, the attitude to the Internet and whether surf on the Internet all night. The dataset used in the work is not mentioned and the number of subjects is 10. (Gong et al., 2016) used clustering methods including K-Means (Hartigan & Wong, 1979; Wagstaff, Cardie, Rogers, Schroˆdl, et al., 2001), Hierarchical Clustering (Johnson, 1967; Navarro, Frenk, & White, 1997) and Fuzzy C-Means Clustering (Bezdek, Ehrlich, & Full, 1984) and personality data to predict Internet Game Disorder (IGD) among Chinese college freshmen. The dataset used comes from 580 freshmen from the University and the features used include Self-Control, Attention, Motor, Non-planning, Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness. Based on the same dataset, (Di, Gong, Shi, Ahmed, & Nandi, 2017) used Support Vector Machines (SVMs) and personality questionnaire data to detect IAD among Chinese college students. The author compared the performance of C-SVM and  $\nu$ -SVM and did some work to find the influence of sex and age on IAD. Their work indicated that IAD can be detected by personality questionnaires and SVMs.

In IAD detection, Chen Internet Addiction Scale (CIAS) is one of the standard questionnaires used as criteria of IAD. It has been frequently used in many areas related, such as resting state fMRI study, task-related fMRI study, the correlation studies of IAD with other factors. The internal reliability is in the range from 0.79 to 0.93, which shows its effectiveness in statistic research (Ko et al., 2009; Chen et al., 2015; Dong, Lin, & Potenza, 2015; Liu et al., 2014; Chen, Weng, Su, Wu, & Yang, 2003; Mak et al., 2014; Yen, Ko, Yen, Wu, & Yang, 2007; Mo et al., 2018; Chern & Huang, 2018; Lei, Li, Chiu, & Lu, 2018; Lau, Wu, Gross, Cheng, & Lau, 2017; Chang, Chiu, Lee, Chen, & Miao, 2014).

So far, many works (Dieris-Hirche et al., 2017; Geng, Han, Gao, Jou, & Huang, 2018; Lam, 2015; Mahapatra & Sharma, 2018; Robertson et al., 2018; Romano, Truzoli, Osborne, & Reed, 2014) of IAD only investigate the effect of a single factor or feature, which brings some

limitation to study this multi-dimension disorder completely. One of the distinctive and important aspects of the current study is the use of several questionnaires and multiple factors simultaneously.

### 1.3. The current study

To test the reliability of machine learning methods in IA detection, this work used a larger dataset, compared the performance of multi-SVM methods and FNN, and uses grid search to optimize the parameters. Different from other works, the data of several related questionnaires was collected to find the most distinguished features in the following steps, which includes Brief Self Control Scale (BSCS), the 11th version of Barratt Impulsiveness Scale (BIS-11), Chinese Big Five Personality Inventory (CBF-PI) and CIAS. The details of these questionnaires are given in the next section. After data acquisition, Student's *t*-test was used for feature selection to obtain several datasets, each with the same set of students but different features. Then the performance from these datasets using Support Vector Machines (SVMs) (Cortes & Vapnik, 1995) with grid search for parameter selection were compared and 10-fold cross validation was used to avoid the overfitting problem. Furthermore, the runtime of the proposed method and others was compared. Finally, a model for IA detection using C-SVM with the accuracy of about 96.3% was found.

## 2. Materials and methods

### 2.1. Questionnaires

#### 2.1.1. The 11th version of Barratt Impulsiveness Scale (BIS-11)

BIS-11 (Barratt, 1959; Patton, Stanford, et al., 1995) is a 30-item questionnaire to evaluate one's impulsivity by summing sub-scale values which are Attention, Motor and Non-planning impulsivity, which is widely used around the world to measure one's impulsivity for fifty years. Each item is scored according to the *Frequency* scale (1 is for *never*; 4 is for *Almost Always/Always*). The internal consistency coefficients of BIS-11 total score is from 0.79 to 0.83 (Mayhew & Powell, 2014) and Cronbach's  $\alpha$  is 0.794 in Chinese children (Li, 2006) and its validity is also confirmed among Chinese (Cao, Su, Liu, & Gao, 2007; Yang, 2007; Yao et al., 2007).

#### 2.1.2. Chinese Big Five Personality Inventory (CBF-PI)

CBF-PI (McCrae & John, 1992; Poropat, 2009; Zhou, Niu, & Zou, 2000) is a restricted version of Big Five Inventory (BFI) to evaluate Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism in one's personality, which works well for Chinese college students (Costa & McCrae, 1992; Thompson, 2008). It consists of 44-item and the object rated each item on a 5-point scale where 1 stands for *strongly disagree* and 5 is for *strongly agree*. The higher score indicates higher level in the specific sub-dimension. The internal consistency coefficients are from 0.78 to 0.85 (Wang, Dai, & Yao, 2011; Wang, Jackson, Zhang, & Su, 2012) and Cronbach's  $\alpha$  is in the range from 0.721 to 0.777 among Chinese (Carciofo, Yang, Song, Du, & Zhang, 2016). It is reported that Big Five Inventory can help BIS-11 evaluate one's impulsivity and give a more specific evaluation result (Whiteside & Lynam, 2001).

#### 2.1.3. Brief Self Control Scale (BSCS)

BSCS (Tangney, Baumeister, & Boone, 2004) is a 13-item questionnaire to measure dispositional self-regulatory behaviors using 13 items rated on a 5-point scale, ranging from 1 (*Not at all like me*) to 5 (*Very much like me*). The internal consistency coefficient is from 0.73 to 0.84 (Mathews, Youman, Stuewig, & Tangney, 2007). Cronbach's  $\alpha$  of Self Control Scale (SCS) is 0.89 (Mei et al., 2016) and the SCS in Chinese fit better ( $\chi^2/df = 3.96$ , GFI = 0.94, TLI = 0.81, RMSEA = 0.06) (Qu & Zou, 2009). BSCS has a very good reputation in self-control assessment and BSCS has been widely used in many studies, especially in the

domains of school/work, eating/weight, interpersonal functioning, and wellbeing/adjustment (Malouf et al., 2014).

### 2.1.4. Chen Internet Addiction Scale (CIAS)

CIAS (Chen et al., 2003) is a 26-item self-report measure for the Internet Addiction Evaluation whose score is from 1 (*Does not match my experience at all*) to 4 (*Definitely matches my experience*). The total score is in the range from 26 to 104. Compared with The Internet Addiction Test, CIAS shows its priority for Chinese Students. Cronbach's  $\alpha$  of the CIAS for the sample is 0.94 (Chang et al., 2014). Higher CIAS score indicates that increased severity of addiction to Internet activity. As for the cut-off points of CIAS, the screening cut-off point is 57 while the diagnostic point is 63 (Ko, Yen, Chen, Chen, & Yen, 2005). To make the study accurate and general, we used the intersection of two versions. Those with score upper than 63 are considered very likely to suffer with IAD while whose score lower than 57 are not, which is also consistent with the suggested threshold score of 63/64 to provide good diagnostic accuracy with respect to Internet Addiction among adolescents (Chang et al., 2014; Ko et al., 2005)

### 2.2. Participants

3123 college undergraduate students participated in the study, where ethics approval was obtained from the Human Research Ethics Committee of the University for data acquisition. Among these 3123 participants, 2397 students gave valid questionnaire answers. The results from those participants who did not finish all questionnaires or give the straight same answers of the whole questionnaire test were considered as invalid data. The age of the participants ranges from 16.91 to 25 years old. The details about the participants and the valid ones are shown in Table 1, which can be found that characteristics of the valid dataset and the whole dataset are the same in age and gender generally. As for Sex in Table 1, the male was labelled as 1 while the female was -1. Sex in Table 1 and Table 2 is processed in the same way and has the same meaning to illustrate the sex ratio, which is for the convenience of feature selection and calculation in our experiments. Because all the values of Sex are integer, standard deviation (std.) has no meaning for Sex and we cancelled the std. in Table 1 and Table 2.

### 2.3. Methods

In this study, we tried to explore IA detection using machine learning methods. First, we selected the valid ones from the collected questionnaires. After that, we applied Student's *t*-test for feature selection and 2 linear normalizations so that we got several datasets. Then, we compared the performance of C-SVM and  $\nu$ -SVM on different datasets with or without grid search. FNN was chose as comparison method with SVM. Finally, we analyzed the relationship among grid search parameters, accuracy and computation time. The whole process is illustrated in Fig. 1.

#### 2.3.1. Data pre-processing

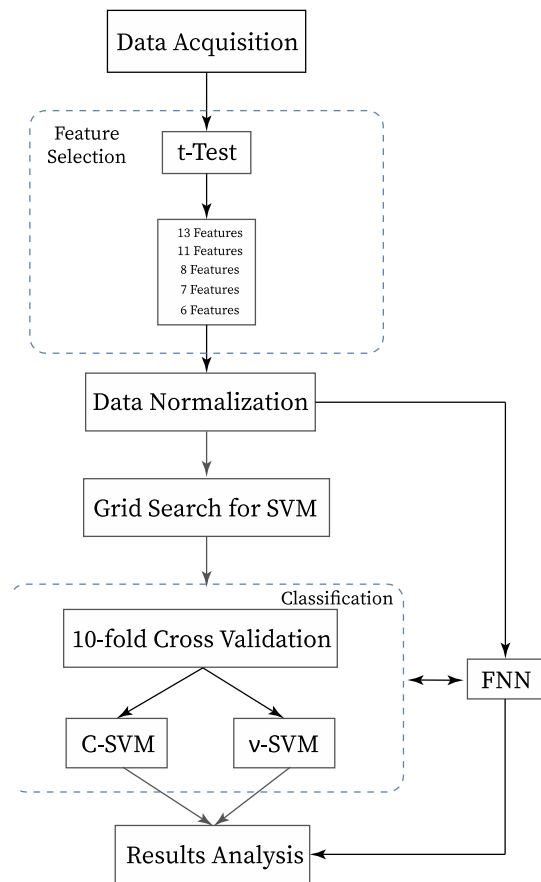
After data acquisition, we summed the score up individually based on 4 questionnaires. For BIS-11 and CBF-PI, we calculated the score of each subset feature, including Attention, Motor, Non-planning, Openness, Conscientiousness, Extroversion, Agreeableness,

**Table 1**  
The details about the participants and the valid ones.

Name	Participants	The valid
Male	2004 (64.17%)	1686 (60.98%)
Female	1119 (35.83%)	1079 (39.02%)
Total	3123 (100.00%)	2765 (100.00%)
Age	19.17 $\pm$ 0.70	19.17 $\pm$ 0.68
Gender	0.21	0.22

**Table 2**  
Mean value and standard deviation of every feature between two groups used in the experiments.

Name	Normal	IA
Gender	0.24	0.17
Age	19.19 $\pm$ 0.67	19.13 $\pm$ 0.73
Attention	37.98 $\pm$ 9.44	35.79 $\pm$ 4.69
Motor	20.16 $\pm$ 5.81	25.41 $\pm$ 5.94
Non-planning	22.39 $\pm$ 6.81	27.70 $\pm$ 5.55
Openness	51.66 $\pm$ 9.97	48.83 $\pm$ 9.19
Conscientiousness	50.92 $\pm$ 9.64	43.84 $\pm$ 9.44
Extroversion	50.71 $\pm$ 10.28	47.09 $\pm$ 9.69
Agreeableness	51.43 $\pm$ 10.15	47.49 $\pm$ 10.40
Neuroticism	48.82 $\pm$ 9.44	56.90 $\pm$ 10.58
BIS-11 total score	80.53 $\pm$ 18.07	88.89 $\pm$ 16.17
CBF-PI total score	253.55 $\pm$ 49.49	244.12 $\pm$ 49.30
Self-Control Scale	22.18 $\pm$ 6.81	27.70 $\pm$ 5.54
CIAS score	39.74 $\pm$ 10.44	76.42 $\pm$ 11.05



**Fig. 1.** The Framework of the present study.

Neuroticism, and the total score of 4 questionnaires. Then the dataset was divided into two groups according to the CIAS scores. Those whose score is less than 57 are regarded as the low IA (1598 participants) group while those with score over 63 belong to the high IA group (799 participants). The mean value and standard deviation value are shown in Table 2.

#### 2.3.2. Normalization

After the data pre-processing, two linear normalizations were applied to the dataset to avoid the impact of great attributes numeric value on those small ones and decrease the calculation cost, which mapped the range of every feature to  $[-1, 1]$  and  $[0, 1]$  respectively.

Thus, we had gotten three kinds of datasets with different value

ranges: original datasets, the datasets ranged in  $[0, 1]$  and the datasets ranged in  $[-1, 1]$ .

### 2.3.3. Student's *t*-test for feature selection

We applied Student's *t*-test for feature selection. The results of Student's *t*-test are  $h$ ,  $p$  and  $C_i$ . If  $h$  is 0, it indicates that the null hypothesis at significance level is rejected and vice versa.  $p$  is the probability of observing the validity of the null hypothesis. Small  $p$  shows doubt on the validity of the null hypothesis.  $C_i$  is the confidence interval containing the lower and upper boundaries of the 95% confidence interval.

### 2.3.4. Grid search

For SVM using radial basis function (RBF) function introduced in the following part, there are two parameters to choose ( $C$  and  $g$ ) and we don't know the best parameter in advance. Grid search is a method where various pairs of ( $C$ ,  $g$ ) in certain range are tried to find the best pair according to the best cross-validation accuracy.  $C$  and  $g$  both are in the range of  $[2^{-8}, 2^8]$ . The smaller the steps of  $C$  ( $C$ -step) and  $g$  ( $g$ -step) are, the more accurate results we can get. The original rule of  $C$  and  $g$  in the following range:

$$C = \{2^{C_{min}}, 2^{C_{min}+C-step}, 2^{C_{min}+2 \times C-step}, \dots, 2^{C_{max}}\},$$

$$g = \{2^{g_{min}}, 2^{g_{min}+g-step}, 2^{g_{min}+2 \times g-step}, \dots, 2^{g_{max}}\},$$

To make the result as accurate as possible, we set the updating rule as:

$$C = \{2^{C_{min}}, 2^{C_{min} + C - step}, 2^{C_{min} + 2 \times C - step}, \dots, 2^{C_{max}}\},$$

$$g = \{2^{g_{min}}, 2^{g_{min} + g - step}, 2^{g_{min} + 2 \times g - step}, \dots, 2^{g_{max}}\},$$

where  $C$ -step and  $g$ -step both are 1 in these two rules.

### 2.3.5. Support vector machine

Support Vector Machine (SVM) is widely used in binary classification because its original idea is to find a hyperplane to separate the two classes with the maximum margin (Mu & Nandi, 2007) (See Fig. 2). The main advantage of SVM is its classification ability to solve the non-linear problem using kernel function. The kernel function we used in

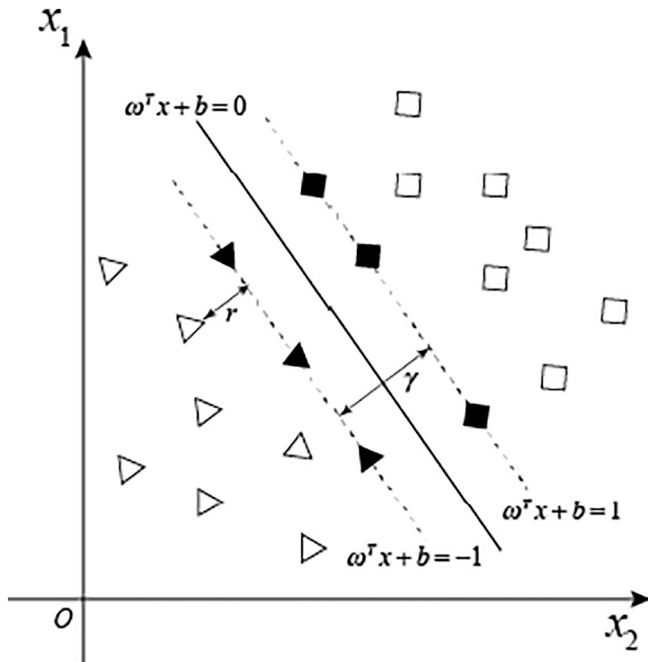


Fig. 2. The illustration of SVM in 2-feature dataset. The filled points on the dashed line indicate the support vectors.

the research is RBF function to nonlinearly map samples into a higher dimensional space, which also helps to decrease the hyper-parameter in calculation (Hsu, Chang, Lin, et al., 2003). The kernel function we used is:

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1)$$

where  $\sigma$  is the RBF function width adjusted by user.

In practice, most of the sample spaces cannot be separated completely. So, a parameter  $\xi$  is added to modify the classifier when there are some non-separable cases. To adjust the ability of modification, a new parameter  $C$  is defined as a matter of experience, often from range  $(0, +\infty)$  and its minimization of error function is:

$$\frac{1}{2}\omega^T \omega + C \sum_{i=1}^m \xi_i \quad (2)$$

which is what so-called C-SVM.

Like C-SVM,  $\nu$ -SVM can control the number of support vectors by adding two parameters  $\nu$  and  $\rho$ . The error function is:

$$\frac{1}{2}\omega^T \omega - \nu\rho + \frac{1}{m} \sum_{i=1}^m \xi_i \quad (3)$$

### 2.3.6. Cross validation

In every iteration, the dataset is split randomly, so it is reasonable to get different classifiers and results. However, we are more interested in the stable performance and its generalization on the whole dataset rather than the training set to avoid over-fitting. Among several methods in evaluation,  $k$ -fold cross validation is the most common method in statistics and machine learning (Kohavi et al., 1995; Seni & Elder, 2010).

In theory,  $k$  can be set randomly as long as  $k$  is bigger than 1 and smaller than the number of samples  $n$  in the dataset. However, it is meaningless to set  $k$  too big or too small. For example, if  $k$  is 2, the dataset is just split equally and the classifier is likely not to be trained well to learn the characters of the dataset to give correct prediction results. Meanwhile, if  $k$  is  $(n - 1)$ ,  $k$ -fold cross validation is exactly the leave-one-out cross-validation (Stone, 1974). In this experiment, we used 10-fold cross validation.

### 2.3.7. FNN

FNN combines the strengths of neural network and fuzzy logic while overcomes the weaknesses of their own like difficulties in explaining how they reached their decisions, acquiring the rules they use to make decisions automatically, which proves its effectiveness in many aspects for a long time, such as pattern recognition, regression and density estimation (Fuller, 1998; Kruse, 2008).

### 2.3.8. Experiment platform

In this study, the experimental platform is MATLAB R2017 on a PC with Intel (R) Xeon (R) CPU E5-2665 (2.4GHz) and the RAM is 64GB using the Microsoft Windows 10 Operating System.

## 3. Result

### 3.1. Feature selection

To find the most distinct features among 13 features, we applied Student's *t*-test on the dataset. In Table 3, Age and Sex had the largest  $p$ -value and their  $h$  are 0, which implied these two features are not so distinguishable. Except Age and Sex, an 11-feature dataset was obtained whose  $h$  is 1, the  $p$ -value of Extraversion, CBF and Openness were bigger than  $1e-20$ . Compared with the  $p$  value of each feature in the 11-feature dataset, we separately acquired 5 datasets with different features using Student's *t*-test according to  $h$  and  $p$ -value for detection



**Table 3**  
Student's *t*-test results for feature selection according to *p*-value degeneratively.

Feature	<i>h</i>	<i>p</i>	$C_i$
BIS	1	1.34e-130	[7.71, 8.99]
BSCS	1	8.74e-127	[4.16, 4.86]
Motor	1	1.02e-99	[4.78, 5.71]
Neuroticism	1	9.04e-84	[-8.76, -7.19]
Non-planning	1	2.66e-82	[4.71, 5.74]
Conscientiousness	1	8.32e-67	[-7.66, -6.13]
Attention	1	7.33e-23	[1.70, 2.53]
Agreeableness	1	4.59e-21	[-4.76, -3.13]
Extraversion	1	1.57e-17	[2.72, 4.33]
CBF-PI	1	1.81e-16	[7.00, 11.34]
Openness	1	3.08e-12	[-3.56, -2.00]
Sex	0	0.14	[-0.11, -0.00]
Age	0	0.50	[-0.01, 0.07]

and they were the following:

- 6 Features: BIS, BSCS, Motor, Neuroticism, Non-planning and Conscientiousness
- 7 Features: 6 Features and Attention
- 8 Features: 7 Features and Agreeableness
- 11 Features: 8 Features, Extraversion, CBF-PI and Openness
- 13 Features: 11 Features, Sex and Age

### 3.2. IA detection performance

In this study, we made a systematic comparison of IA detection performance among SVM methods and FNN methods as presented in (Wang et al., 2008). We also compared the performance of C-SVM and  $\nu$ -SVM with and without grid search. For SVMs, 10-fold cross validation was applied to avoid over-fitting. All the experiments were repeated for 1000 times to increase the reliability.

#### 3.2.1. IA detection performance of SVM (without grid search) and FNN

In Table 4, the detection performance of C-SVM,  $\nu$ -SVM, and FNN without grid search on different datasets are shown. The parameters of SVMs without grid search were set as the default values, where *C* and *g* both were 1. In 13-feature and 8-feature datasets, the best performances are from FNN with the accuracy of 69.73% and 75.58%, respectively. It can be observed that feature selection helped improve the performance of three methods. C-SVM had the largest increment (27.7%) and then

**Table 4**

Detection results of C-SVM,  $\nu$ -SVM and FNN without grid search on every dataset.

Dataset	C-SVM (%)	$\nu$ -SVM (%)	FNN (%)
13F <sub>0</sub>	51.00 ± 1.49	67.51 ± 1.53	69.73
13F <sub>1</sub>	51.46 ± 0.83	54.47 ± 1.32	57.41
13F <sub>2</sub>	51.95 ± 0.66	65.71 ± 1.91	57.39
11F <sub>0</sub>	52.66 ± 1.14	56.93 ± 0.73	54.98
11F <sub>1</sub>	51.30 ± 1.08	57.42 ± 1.30	57.35
11F <sub>2</sub>	51.59 ± 0.30	62.74 ± 2.81	57.73
8F <sub>0</sub>	72.51 ± 0.07	60.65 ± 0.79	<b>75.58</b>
8F <sub>1</sub>	77.52 ± 0.33	75.00 ± 0.72	74.91
8F <sub>2</sub>	77.23 ± 0.37	75.00 ± 0.71	74.67
7F <sub>0</sub>	74.89 ± 0.32	73.79 ± 0.68	74.45
7F <sub>1</sub>	77.60 ± 1.28	76.04 ± 0.28	73.89
7F <sub>2</sub>	<b>78.72 ± 1.46</b>	<b>77.43 ± 1.23</b>	72.47
6F <sub>0</sub>	76.05 ± 0.74	75.03 ± 0.73	71.74
6F <sub>1</sub>	75.15 ± 0.78	75.76 ± 0.71	72.47
6F <sub>2</sub>	75.08 ± 0.73	75.79 ± 0.72	72.39

F0 refers to the dataset without normalization. F1 refers to the dataset with the normalization in [-1, 1]. F2 refers to the dataset with the normalization in [0, 1].

The bold emphases the best detection accuracy by different classifiers.

**Table 5**

Detection results of C-SVM,  $\nu$ -SVM and FNN without grid search on every dataset.

Dataset	C-SVM (%)	$\nu$ -SVM (%)	The best in Table 3 (%)
13F <sub>0</sub> *	51.00 ± 1.48	67.77 ± 2.01	69.73 <sup>c</sup>
13F <sub>1</sub>	51.46 ± 0.83	56.91 ± 1.14	57.41 <sup>c</sup>
13F <sub>2</sub>	51.95 ± 0.66	65.71 ± 2.78	65.71 ± 1.91 <sup>b</sup>
11F <sub>0</sub>	51.16 ± 1.14	57.40 ± 1.65	56.93 ± 0.73 <sup>b</sup>
11F <sub>1</sub>	51.30 ± 1.10	57.42 ± 1.32	57.42 ± 1.30 <sup>b</sup>
11F <sub>2</sub>	51.59 ± 0.30	65.16 ± 2.81	62.74 ± 2.81 <sup>b</sup>
8F <sub>0</sub>	78.27 ± 0.41	73.70 ± 0.29	75.58 <sup>c</sup>
8F <sub>1</sub>	78.74 ± 0.35	77.34 ± 0.38	77.52 ± 0.33 <sup>a</sup>
8F <sub>2</sub>	78.90 ± 0.19	77.06 ± 0.53	77.23 ± 0.37 <sup>a</sup>
7F <sub>0</sub>	72.28 ± 0.34	75.08 ± 0.29	74.89 ± 0.32 <sup>a</sup>
7F <sub>1</sub>	84.78 ± 1.31	<b>80.94 ± 0.47</b>	77.60 ± 1.28 <sup>a</sup>
7F <sub>2</sub>	84.58 ± 1.47	78.33 ± 1.11	<b>78.72 ± 1.46<sup>a</sup></b>
6F <sub>0</sub>	<b>96.32 ± 0.18</b>	78.32 ± 0.36	76.05 ± 0.74 <sup>a</sup>
6F <sub>1</sub>	76.05 ± 0.74	78.54 ± 0.44	75.76 ± 0.71 <sup>b</sup>
6F <sub>2</sub>	76.02 ± 0.59	78.74 ± 0.43	75.79 ± 0.72 <sup>b</sup>

The bold emphases the best detection accuracy by different classifiers.

\* The feature is the same as that in Table 3.

<sup>a</sup> This indicates that the best results in Table 3 is C-SVM.

<sup>b</sup> This indicates that the best results in Table 3 is  $\nu$ -SVM.

<sup>c</sup> This indicates that the best results in Table 3 is FNN.

followed FNN (20.6%).  $\nu$ -SVM had the least, which was 20.5%.

Two types of normalization were also examined in this work. But among all the five datasets, these two ways of normalization did have much difference. Compared with the datasets without normalization, normalization made little contribution to the performance improvement.

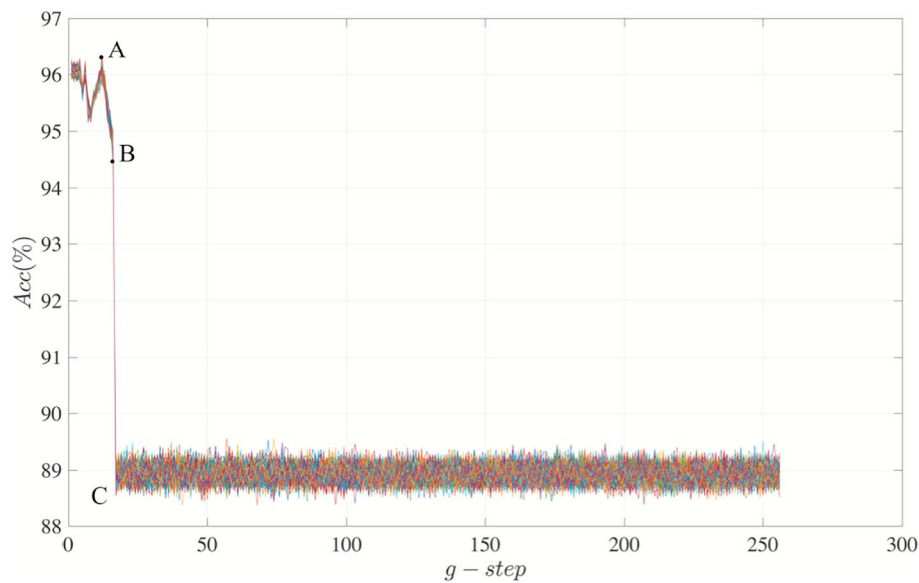
#### 3.2.2. IA detection performance of SVM with grid search

SVM algorithms were shown more flexible of our datasets. The next step of this study focused on the optimization of the parameters in SVM. Grid search is a method to optimize the parameters of RBF function in SVM to improve the detection performance. The comparison results are shown in Table 5. Based on 13-feature and 11-feature datasets, we found that the detection performance was not improved by the grid search. While in 8-feature datasets, C-SVM performed best and the best was from the dataset normalized in [0, 1] with the accuracy of 78.90%, which was increased about 3% by grid search. In 7-feature datasets, C-SVM also performed best and the best one was from the dataset normalized in [-1, 1], whose accuracy was 84.78%. In 6-feature datasets, C-SVM performed best in the dataset without normalization, which was also the best in this study and its accuracy and standard deviation were 96.32% and 0.18% respectively. It can be observed that grid search improved the performance in a large scale. Meanwhile, the normalization still made little contribution to the performance improvement and Student's *t*-test succeeded to select the features and improve the performance in a large scale. As for the increment, C-SVM (45.32%) was larger than  $\nu$ -SVM (24.03%).

#### 3.2.3. Accuracy, parameters and computation time in grid search of 6-feature dataset without normalization

Grid search is a time-consuming task to find the best parameters. Among these five datasets, 6-feature datasets were achieved with the best IA detection performance the least features. The following step of this study focused on extracting the best parameter value which can save the computation cost most. The smaller *C*-step and *g*-step are, the more time will be consumed. All the experiments were repeated 100 times to increase the dependency.

The relationship between (*C*-step, *g*-step) pair and accuracy in 6-feature datasets without normalization can be found in Figs. 3 and 4. There are 256 lines in the plots and each line with a different color stands for different *g*-step in Fig. 3 and different *C*-step in Fig. 4 which are both from 1 to 256.

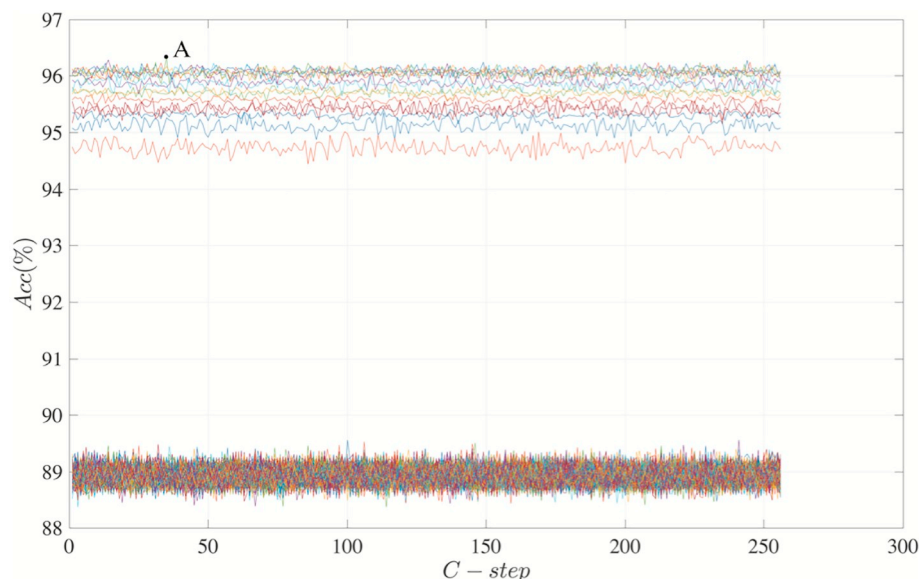


**Fig. 3.** The relationship between  $g$ -step and accuracy when  $g$ -step is fixed for C-SVM in 6-feature dataset without normalization. Each line with different color stands for different  $g$ -step from 1 to 256. The x-axis is the  $g$  from 1 to 256 while the y-axis is the accuracy. Point A has the best performance whose  $C$ -step is 35 and  $g$ -step is 12. Point B ( $g$ -step is 16) and C ( $g$ -step is 17) are the tuning points, which shows that  $g$ -step is the key parameter.

In Fig. 3, point A shows the overall best performance and the accuracy is 96.32%, where  $C$ -step is 35 and  $g$ -step is 12. It can be found that all lines are overlapped together from point B ( $C$ -step  $\in [1, 256]$ ,  $g$ -step = 16) to C ( $C$ -step  $\in [1, 256]$ ,  $g$ -step = 17).

In Fig. 4, point A corresponded to the overall best performance in Fig. 3. It was found that the accuracy did not vary a lot with the  $C$ -step and two part of lines were separated based on different  $g$ -step. Accordingly, compared with the  $C$ -step value,  $g$ -step was shown as the key parameter rather than  $C$ -step in this experiment.

According to the results we got above, the next step was to study the relationship among  $g$ -step, the calculating time and the accuracy when  $C$ -step was ignored to inspect the effect of key parameter  $g$ -step on calculating time and accuracy, which can be found in Fig. 5. Point E is the “compromised” point whose average accuracy is 96.06% and  $g$ -step is 37 while its average time is 33.94 s. Point D is the best point with the average accuracy 96.32% and  $g$ -step is 1 and took about 228.7 s, which is reasonable because the smaller step can lead to a larger chance to find the best result in grid search. Thus, an appropriate parameter pair was found ( $C$ -step  $\in [1, 256]$ ,  $g$ -step = 37) to balance the classification performance and calculating time of our IA detection task.



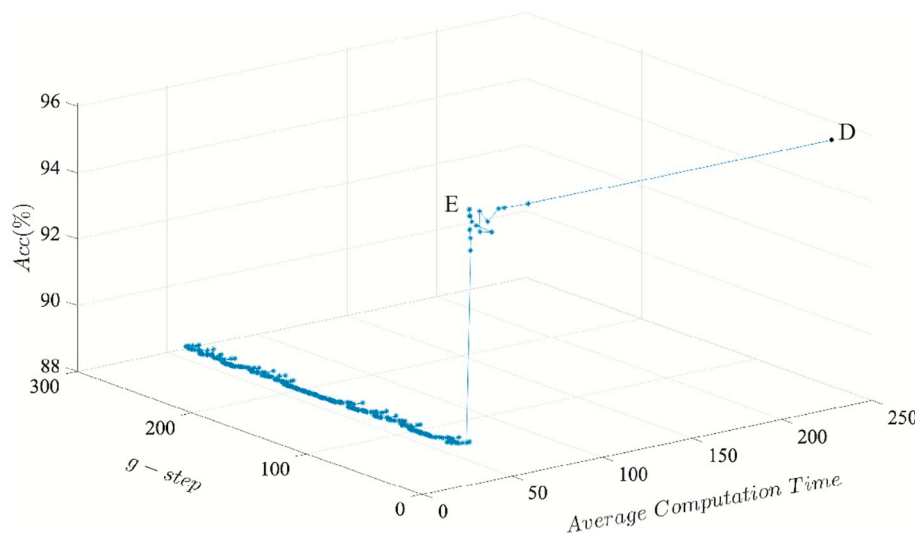
**Fig. 4.** The relationship between  $C$ -step and accuracy when  $C$ -step is fixed for C-SVM in 6-feature dataset without normalization. Each line with different color stands for different  $C$ -step from 1 to 256. The x-axis is the  $C$ -step from 1 to 256 while the y-axis is the accuracy. Point A has the best performance whose  $C$ -step is 35 and  $g$ -step is 12, which corresponds to Point A in Fig. 3.

#### 4. Discussion

In the previous works, the researchers used questionnaire and statistical methods to find the relationship between IAD and the features they used (Kuss, Griffiths, & Binder, 2013; Orsal, Orsal, Unsal, & Ozalp, 2013). Statistical methods can only give the degree of correlation, which cannot be used to predict whether the patients have IA or not directly. This work utilized machine learning using more questionnaires and a relative larger dataset (Di et al., 2017; Gong et al., 2016; Wang et al., 2008) and our results show its efficiency in this kind of task, which provided a new view for the researchers in this area.

In the works using multiple questionnaires (Kim et al., 2006; Tsai et al., 2009; Xin et al., 2018), some questions maybe are duplicated or some features maybe are correlated. It is necessary to find the relationship among these features. This work supports the previous works that the features which are found correlated with IA will lead to or prevent IA in some scale (Fumero et al., 2018; Robertson et al., 2018). Meanwhile, it also plays a role as supplement to find the relationship among features in the task using several questionnaires, which our results demonstrated.

As for the data collection and experiment design, it is important to



**Fig. 5.** The relationship among *g-step*, time and accuracy. Point E's accuracy is 96.06% whose *g-step* is 37 and calculating time is 33.94 s. Point D's accuracy is 96.08% whose *g-step* is 1 and calculating time is 228.7 s. The x-axis is the *g-step* value. The y-axis is the mean value of time for each *g-step*. The z-axis is the mean value of accuracy for each *g-step*.

do some data pre-processing including de-duplication and de-noising. It is better to use a larger dataset in a large age range. Cross validation is also necessary in the experiment, which will improve the generality and performance of the classifier especially when the dataset is relatively small (Kohavi et al., 1995; Seni & Elder, 2010). Parameter selection can provide a suitable parameter pair to balance the computation cost and performance, which provides possibility for engineering and usage in a really large scale.

Questionnaire is a kind of structured data which consists of numeric value according to the features it has. The results in this work demonstrated the efficiency of machine learning methods to deal with questionnaire data in IA.

For future study, researchers can refer to our work to build an IA prediction system. Although the dataset we used is relatively large, more participants in different ages and more questionnaires will make our work more general. SVM is a classical method and always plays a role as baseline of classification methods. However, the development of deep learning makes the performance of classifier much better than before, which is also a choice for IAD detection (Goodfellow, Bengio, & Courville, 2016).

As for practical meaning, this work provides a new choice to detect IA among teenagers in advance in a simple and quick way. People can use online or offline simple questionnaires mentioned above to detect IA more efficiently and precisely, allowing them to diagnose and cure IA among Chinese college students in time.

## 5. Conclusion

In this study, we carried out systematic experiments to detect IA using SVMs and FNN. The results proved the feasibility of using machine learning methods to detect IA. SVM methods were found more flexible for our questionnaire datasets. With grid search, a best parameter pair of C-SVM was achieved ( $C\text{-step} \in [1, 256]$ ,  $g\text{-step} = 37$ ,  $t = 33.94$  s, accuracy = 96.06% at the 6-feature dataset without normalization) and *g-step* is more important to the accuracy than *C-step* as the key parameter in this experiment. More interestingly, BIS-11, BSCS, Motor, Neuroticism, Non-planning and Conscientiousness are shown as a better detection feature combination of IA detection. This indicates the researchers may make more effort to study the relationship between these 6 features and IA. Based on these features to predict the IA risk, it may be a future research interest.

## Acknowledgment

This work was partly supported by the Fundamental Research Funds

for the Central Universities of Tongji University (22120170043, 22120180542) and the Natural Science Foundation of Shanghai grant number 16JC1401300.

## Accessibility requirements

None declared.

## Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, "Internet Addiction Disorder of Chinese College Students Using Several Personality Questionnaire Data and Support Vector Machine".

## References

- Abbott, D. A., Cramer, S. L., & Sherrets, S. D. (1995). Pathological gambling and the family: Practice implications. *Families in Society*, 76, 213.
- Association, A.P., et al. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®R)*. American Psychiatric Pub.
- Baldi, P., & Brunak, S. (2001). *Bioinformatics: The machine learning approach*. MIT press.
- Barratt, E. S. (1959). Anxiety and impulsiveness related to psychomotor efficiency. *Perceptual and Motor Skills*, 9, 191–198.
- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10, 191–203.
- Cao, F., Su, L., Liu, T., & Gao, X. (2007). The relationship between impulsivity and internet addiction in a sample of Chinese adolescents. *European Psychiatry*, 22, 466–471.
- Carciofo, R., Yang, J., Song, N., Du, F., & Zhang, K. (2016). Psychometric evaluation of Chinese-language 44-item and 10-item big five personality inventories, including correlations with chronotype, mindfulness and mind wandering. *PLoS One*, 11, 1–26.
- Chang, F. C., Chiu, C. H., Lee, C. M., Chen, P. H., & Miao, N. F. (2014). Predictors of the initiation and persistence of internet addiction among adolescents in Taiwan. *Addictive Behaviors*, 39, 1434–1440.
- Chen, C. Y., Huang, M. F., Yen, J. Y., Chen, C. S., Liu, G. C., Yen, C. F., & Ko, C. H. (2015). Brain correlates of response inhibition in internet gaming disorder. *Psychiatry and clinical neurosciences*, 69, 201–209.
- Chen, S., Weng, L., Su, Y., Wu, H., & Yang, P. (2003). Development of a Chinese internet addiction scale and its psychometric study. *Chinese Journal of Psychology*, 45, 279–294.
- Chern, K. C., & Huang, J. H. (2018). Internet addiction: Associated with lower health-related quality of life among college students in Taiwan, and in what aspects? *Computers in Human Behavior*, 84, 460–466.
- Cho, H., Kwon, M., Choi, J. H., Lee, S. K., Choi, J. S., Choi, S. W., & Kim, D. J. (2014). Development of the internet addiction scale based on the internet gaming disorder criteria suggested in dsm-5. *Addictive Behaviors*, 39, 1361–1366.
- Choi, Y., Liu, T. T., Pankratz, D. G., Colby, T. V., Barth, N. M., Lynch, D. A., ... Huang, J.



- (2018). Identification of usual interstitial pneumonia pattern using RNA-seq and machine learning: Challenges and solutions. *BMC Genomics*, 19, 101.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The neo personality inventory. *Psychological Assessment*, 4, 5–13.
- Di, Z., Gong, X., Shi, J., Ahmed, H. O., & Nandi, A. K. (2017). Detection of IAD based on personality questionnaires of Chinese college students and SVMs. *Image and signal processing, biomedical engineering and informatics (CISP-BMEI), 2017 10<sup>th</sup> international congress on* (pp. 1–6). IEEE.
- Dieris-Hirche, J., Bottel, L., Bielefeld, M., Steinbüchel, T., Kehyayan, A., Dieris, B., te Wildt, B., 2017. Media use and internet addiction in adult depression: A case-control study. *Computers in Human Behavior* 68, 96–103.
- Dong, G., Lin, X., & Potenza, M. N. (2015). Decreased functional connectivity in an executive control network is related to impaired executive function in internet gaming disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 57, 76–85.
- Dongyun, M., Ni'na, L., & Yao, G. (2018). Relationship between impulsive personality traits and cognitive function in college students with internet addiction disorder. *China Journal of Health Psychology*, 1, 029.
- Fuller, R. (1998). *Neural Fuzzy Systems*.
- Fumero, A., Marrero, R. J., Voltes, D., & Pen'ate, W. (2018). Personal and social factors involved in internet addiction among adolescents: A meta-analysis. *Computers in Human Behavior*, 86, 387–400.
- Geng, J., Han, L., Gao, F., Jou, M., & Huang, C. C. (2018). Internet addiction and procrastination among Chinese young adults: A moderated mediation model. *Computers in Human Behavior*, 84, 320–333.
- Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology*, 15, 512–520.
- Gong, X., Long, B., Fang, K., Di, Z., Hou, Y., & Cao, L. (2016). A prediction based on clustering and personality questionnaire data for IGD risk: A preliminary work. *Natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), 2016 12<sup>th</sup> international conference on* (pp. 1699–1703). IEEE.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Griffiths, M. D. (2018). Conceptual issues concerning internet addiction and internet gaming disorder: Further critique on Ryding and Kaye (2017). *International Journal of Mental Health and Addiction*, 16, 233–239.
- Guo, H., & Nandi, A. K. (2006). Breast cancer diagnosis using genetic programming generated feature. *Pattern Recognition*, 39, 980–987.
- Ha, J. H., Yoo, H. J., Cho, I. H., Chin, B., Shin, D., Kim, J. H., et al. (2006). Psychiatric comorbidity assessed in Korean children and adolescents who screen positive for internet addiction. *Journal of Clinical Psychiatry*, 67, 821–826.
- Hahn, E., Reuter, M., Spinath, F. M., & Montag, C. (2017). Internet addiction and its facets: The role of genetics and the relation to self-directedness. *Addictive Behaviors*, 65, 137–146.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28, 100–108.
- Hsu, C. W., Chang, C. C., Lin, C. J., et al. (2003). *A practical guide to support vector classification*.
- Ismail, A. B., & Zawahreh, N. (2017). Self-control and its relationship with the internet addiction among a sample of Najran university students. *Journal of Education and Human Development*, 6, 168–174.
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32, 241–254.
- Kayis, A. R., Satici, S. A., Yilmaz, M. F., Şimşek, D., Ceyhan, E., & Bakiog'lu, F. (2016). Big five-personality trait and internet addiction: A meta-analytic review. *Computers in Human Behavior*, 63, 35–40.
- Kim, K., Ryu, E., Chon, M. Y., Yeun, E. J., Choi, S. Y., Seo, J. S., & Nam, B. W. (2006). Internet addiction in Korean adolescents and its relation to depression and suicidal ideation: A questionnaire survey. *International Journal of Nursing Studies*, 43, 185–192.
- Ko, C. H., Liu, G. C., Hsiao, S., Yen, J. Y., Yang, M. J., Lin, W. C., ... Chen, C. S. (2009). Brain activities associated with gaming urge of online gaming addiction. *Journal of Psychiatric Research*, 43, 739–747.
- Ko, C. H., Yen, C. F., Yen, C. N., Yen, J. Y., Chen, C. C., & Chen, S. H. (2005). Screening for internet addiction: An empirical study on cut-off points for the Chen internet addiction scale. *The Kaohsiung Journal of Medical Sciences*, 21, 545–551.
- Ko, C. H., Yen, J. Y., Chen, C. C., Chen, S. H., Wu, K., & Yen, C. F. (2006). Tridimensional personality of adolescents with internet addiction and substance use experience. *The Canadian Journal of Psychiatry*, 51, 887–894.
- Ko, C. H., Yen, J. Y., Chen, C. C., Chen, S. H., & Yen, C. F. (2005). Gender differences and related factors affecting online gaming addiction among Taiwanese adolescents. *The Journal of Nervous and Mental Disease*, 193, 273–277.
- Kohavi, R., et al. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Montreal, Canada: Ijcai1137–1145.
- Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in Medicine*, 23, 89–109.
- Kruse, R. (2008). Fuzzy neural network. *Scholarpedia*, 3, 6043. <https://doi.org/10.4249/scholarpedia.6043> revision #91290.
- Kuss, D. J., Griffiths, M. D., & Binder, J. F. (2013). Internet addiction in students: Prevalence and risk factors. *Computers in Human Behavior*, 29, 959–966.
- Lam, L. T. (2015). Parental mental health and internet addiction in adolescents. *Addictive Behaviors*, 42, 20–23.
- Lam, L. T., Peng, Z. W., Mai, J. C., & Jing, J. (2009). Factors associated with internet addiction among adolescents. *Cyberpsychology & Behavior*, 12, 551–555.
- Lau, J. T., Wu, A. M., Gross, D. L., Cheng, K. M., & Lau, M. M. (2017). Is internet addiction transitory or persistent? Incidence and prospective predictors of remission of internet addiction among Chinese secondary school students. *Addictive Behaviors*, 74, 55–62.
- Lei, H., Li, S., Chiu, M. M., & Lu, M. (2018). Social support and internet addiction among mainland Chinese teenagers and young adults: A meta-analysis. *Computers in Human Behavior*, 85, 200–209.
- Li, F. (2006). Reliability and validity of the the Barratt impulsiveness scale in Chinese children. *Chinese Journal of Clinical Psychology*, 14, 115.
- Lin, M. P., Ko, H. C., & Wu, J. Y. W. (2011). Prevalence and psychosocial risk factors associated with internet addiction in a nationally representative sample of college students in Taiwan. *Cyberpsychology, Behavior and Social Networking*, 14, 741–746.
- Liu, G. C., Yen, J. Y., Chen, C. Y., Yen, C. F., Chen, C. S., Lin, W. C., & Ko, C. H. (2014). Brain activation for response inhibition under gaming cue distraction in internet gaming disorder. *The Kaohsiung Journal of Medical Sciences*, 30, 43–51.
- Mahapatra, A., & Sharma, P. (2018). Association of internet addiction and alexithymia—a scoping review. *Addictive Behaviors*, 81, 175–182.
- Mak, K. K., Lai, C. M., Ko, C. H., Chou, C., Kim, D. I., Watanabe, H., & Ho, R. C. (2014). Psychometric properties of the revised Chen internet addiction scale (CIAS-r) in Chinese adolescents. *Journal of Abnormal Child Psychology*, 42, 1237–1245.
- Malouf, E. T., Schaefer, K. E., Witt, E. A., Moore, K. E., Stuewig, J., & Tangney, J. P. (2014). The brief self-control scale predicts jail inmates' recidivism, substance dependence, and post-release adjustment. *Personality and Social Psychology Bulletin*, 40, 334–347.
- Mathews, J., Youman, K., Stuewig, J., & Tangney, J. (2007). Reliability and validity of the brief self-control scale in a sample of incarcerated offenders. *Annual meeting of the American Society of Criminology*. Atlanta, Georgia: Atlanta Marriott Marquis.
- Mayhew, M. J., & Powell, J. H. (2014). The development of a brief self-report questionnaire to measure 'recent' rash impulsivity: A preliminary investigation of its validity and association with recent alcohol consumption. *Addictive Behaviors*, 39(11), 1597–1605.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 175–215.
- Mei, S., Yau, Y. H., Chai, J., Guo, J., & Potenza, M. N. (2016). Problematic internet use, well-being, self-esteem and self-control: Data from a high-school survey in China. *Addictive Behaviors*, 61, 74–79.
- Mo, P. K., Chan, V. W., Chan, S. W., & Lau, J. T. (2018). The role of social support on emotion dysregulation and internet addiction among Chinese adolescents: A structural equation model. *Addictive Behaviors*, 82, 86–93.
- Müller, K. R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial eeg-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167, 82–90.
- Mu, T., & Nandi, A. K. (2007). Breast cancer detection from FNA using SVM with different parameter tuning systems and SOM-RBF classifier. *Journal of the Franklin Institute*, 344, 285–311.
- Musetti, A., Cattivelli, R., Giacobbi, M., Zuglian, P., Ceccarini, M., Capelli, F., ... Castelnovo, G. (2016). Challenges in internet addiction disorder: Is a diagnosis feasible or not? *Frontiers in Psychology*, 7, 842.
- Navarro, J. F., Frenk, C. S., & White, S. D. (1997). A universal density profile from hierarchical clustering. *The Astrophysical Journal*, 490, 493.
- Nordegren, T. (2002). *The AZ encyclopedia of alcohol and drug abuse*. Universal-Publishers.
- Orsal, O., Orsal, O., Unsal, A., & Ozalp, S. S. (2013). Evaluation of internet addiction and depression among university students. *Procedia-Social and Behavioral Sciences*, 82, 445–454.
- Patton, J. H., Stanford, M. S., et al. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, 51, 768–774.
- Petry, N. M., & O'Brien, C. P. (2013). Internet gaming disorder and the dsm-5. *Addiction*, 108, 1186–1187.
- Poropat, A. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135, 322–338. <https://doi.org/10.1037/a0014996>.
- Qu, Z., & Zou, H. (2009). Juvenile delinquency: The role of self-control, family environment and parental monitoring. *Psychological Science*, 32, 360–363.
- Robertson, T. W., Yan, Z., & Rapoza, K. A. (2018). Is resilience a protective factor of internet addiction? *Computers in Human Behavior*, 78, 255–260.
- Romano, M., Truzoli, R., Osborne, L. A., & Reed, P. (2014). The relationship between autism quotient, anxiety, and internet addiction. *Research in Autism Spectrum Disorders*, 8, 1521–1526.
- Seni, G., & Elder, J. F. (2010). Ensemble methods in data mining: Improving accuracy through combining predictions. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2, 1–126.
- Spada, M. M. (2014). An overview of problematic internet use. *Addictive Behaviors*, 39, 3–6.
- Stone, M. (1974). Cross-validated choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, 111–147.
- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72, 271–324.
- Thompson, E. R. (2008). Development and validation of an international English big-five mini-markers. *Personality and Individual Differences*, 45, 542–548.
- Treuer, T., Fa'bian, Z., & Fu'redi, J. (2001). Internet addiction associated with features of impulse control disorder: Is it a real psychiatric disorder. *Journal of Affective Disorders*, 66, 283.
- Tsai, H. F., Cheng, S. H., Yeh, T. L., Shih, C. C., Chen, K. C., Yang, Y. C., & Yang, Y. K. (2009). The risk factors of internet addiction—a survey of university freshmen. *Psychiatry Research*, 167, 294–299.
- Wagstaff, K., Cardie, C., Rogers, S., Schro'dl, S., et al. (2001). *Constrained k-means clustering with background knowledge*. ICML577–584.
- Wang, J. L., Jackson, L. A., Zhang, D. J., & Su, Z. Q. (2012). The relationships among the big five personality factors, self-esteem, narcissism, and sensation-seeking to Chinese



- university students' uses of social networking sites (SNSs). *Computers in Human Behavior*, 28, 2313–2319.
- Wang, M., Dai, X., & Yao, S. (2011). Development of the Chinese big five personality inventory (CBF-PI) iii: Psychometric properties of CBF-PI brief version. *Chin. J. Clin. Psychol*, 19, 454–457.
- Wang, Z.L., Zhang, W. D., & Zhang, J. J. (2008). Forecasting pattern of network addiction based on fuzzy neural network [J]. *Control Engineering of China*, 5, 556–559.
- Whang, L. S. M., Lee, S., & Chang, G. (2003). Internet overusers' psychological profiles: A behavior sampling analysis on internet addiction. *Cyberpsychology & Behavior*, 6, 143–150.
- Whiteside, S. P., & Lynam, D. R. (2001). The five-factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, 30, 669–689.
- Wiederhold, B. (2018). Stop scrolling, start living: The growing reality of internet addiction disorder. *Cyberpsychology, Behavior and Social Networking*, 21, 279–280.
- Xin, M., Xing, J., Pengfei, W., Houru, L., Mengcheng, W., & Hong, Z. (2018). Online activities, prevalence of internet addiction and risk factors related to family and school among adolescents in China. *Addictive Behaviors Reports*, 7, 14–18.
- Xiuqin, H., Huimin, Z., Mengchen, L., Jinan, W., Ying, Z., & Ran, T. (2010). Mental health, personality, and parental rearing styles of adolescents with internet addiction disorder. *Cyberpsychology, Behavior and Social Networking*, 13, 401–406.
- Yang, H. (2007). The Chinese version of the Barratt impulsiveness scale 11th version (bis-11) in college students: Its reliability and validity. *Chinese Mental Health Journal*, 21, 223.
- Yao, S., Yang, H., Zhu, X., Auerbach, R. P., Abela, J. R., Pulleyblank, R. W., & Tong, X. (2007). An examination of the psychometric properties of the Chinese version of the Barratt impulsiveness scale, 11th version in a sample of Chinese adolescents. *Perceptual and Motor Skills*, 104, 1169–1182.
- Yau, Y. H., Crowley, M. J., Mayes, L. C., & Potenza, M. N. (2012). Are internet use and video-game-playing addictive behaviors? Biological, clinical and public health implications for youths and adults. *Minerva Psichiatrica*, 53, 153.
- Yen, J. Y., Ko, C. H., Yen, C. F., Wu, H. Y., & Yang, M. J. (2007). The comorbid psychiatric symptoms of internet addiction: Attention deficit and hyperactivity disorder (ADHD), depression, social phobia, and hostility. *Journal of Adolescent Health*, 41, 93–98.
- Young, K. S. (1998a). *Caught in the net: How to recognize the signs of internet addiction—and a winning strategy for recovery*. John Wiley & Sons.
- Young, K. S. (1998b). Internet addiction: The emergence of a new clinical disorder. *Cyberpsychology & Behavior*, 1, 237–244.
- Young, K. S. (2004). Internet addiction: A new clinical phenomenon and its consequences. *American Behavioral Scientist*, 48, 402–415.
- Zhang, Y., Wang, Y., Zhou, G., Jin, J., Wang, B., Wang, X., & Cichocki, A. (2018). Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces. *Expert Systems with Applications*, 96, 302–310.
- Zhou, H., Niu, L., & Zou, H. (2000). Development of five-factor personality questionnaire for Chinese adolescents. *Psychological Development and Education*, 1, 48–54.
- Zhou, Y., Li, D., Li, X., Wang, Y., & Zhao, L. (2017). Big five personality and adolescent internet addiction: The mediating role of coping style. *Addictive Behaviors*, 64, 42–48.