Personality Evaluation:
Multi-modal Chinese Graphology Analysis

Written by Hang Yang[1*]
Zhihan Ruan[2]
College of Computer Science from Nankai University
No. 38, Tongyan Road, Haihe Education Park
Jinnan District, Tianjin 300350
innallyyang@hotmail.com

Abstract

Graphology evaluates personality through handwriting analysis, which is widely used in human resource management and psychology research. However, graphology analysis done manually by psychologists is time-consuming and the analytical subjectivity cannot be eliminated. In addition, researches into Chinese graphology is insufficient. To resolve such problems, we propose a new approach to evaluating people's personalities based on multi-modal Chinese graphology analysis, which leverages the tree-drawing test. Nevertheless, there is no dataset of Chinese graphology. Hence, we collect a dataset called Handwriting-Tree-Personality-184 (HTP-184) containing paired images — Chinese handwriting and tree-drawing images, with professional annotations on personalities. Experimental results demonstrate that our proposed model is a simple yet effective baseline for further study. To our knowledge, this paper is the first to study Chinese graphology and the first to explore this multi-modal task that adopts the tree-drawing test. We will release the dataset and code soon.

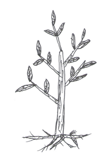
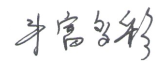
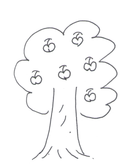
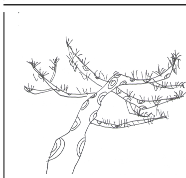
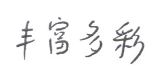
Tree-drawing Test	Handwriting	Personality Labels
		(1) ActionOriented (2) Extrovert (3) Rational (4) Compatible (5) Conservative (6) Realistic
		(1) ThinkingOriented (2) Introvert (3) Rational (4) Controlling (5) Conservative (6) Realistic
		(1) ThinkingOriented (2) Introvert (3) Emotional (4) Controlling (5) Ambitious (6) Idealistic

Table 1: Two modalities and annotation of three writers/-drawers. As the Table shows, different people have distinct styles of writing/drawing, along with variant personalities.

1 Introduction

Graphology is the analysis of the physical traits and patterns of handwriting, attempting to evaluate personality characteristics (Rollin 1984). It is a challenging but prospective task, which can be employed in human resource management, psychotherapy, psychology research, etc.

However, traditional manual graphology analysis is relatively controversial. Firstly, there are disagreements on the analytical criteria followed by graphologists. Most analytical methods are based on empirical experiences and without the support of quantitative analysis. Secondly, objectivity can be deviated due to negligence on details, the emotional status of the graphologists when analyzing, and the prior knowledge the graphologist has of the writer. Thirdly, Hiring a graphologist to do manual analysis is time-consuming and expensive. As these disadvantages manifested above, using machine learning (ML) methods to automate the procedure seems more feasible and promising.

Some previous works managed to automate the analysis process based on alphabetic handwriting (e.g. English handwriting). The pipeline of their works can be summarized as follows: (1) Select salient features according to the empirical graphology criteria. (2) Detect corresponding features from the handwriting image. For global features, digital image processing methods are employed. For example, lines segmentation is applied when calculating the left paper margin(Sheikholeslami, Srihari, and Govindaraju 1995). And for local features, ML approaches are adopted. For instance,(Gavrilescu and Vizireanu 2018) use Artificial Neural Networks (ANN) to classify different types of the alphabet "t". (3) Interpret the detected features and generate personality report according to graphology rules. Despite the efficient and tractable methodologies that previous works have proposed, none of them have verified the validity of the graphology criteria they follow. Therefore, some peo-

[*]With help from the Nankai Computer Vision Laboratary and Psychology Research Group from Tianjin Normal University.

ple remain skeptical about the effectiveness of the empirical graphology criteria(Ben-Shakhar et al. 1986; Dazzi and Pedrabissi 2009). Recently, some researchers have realized the problem related to the empirical criteria, and remodeled the analytical process. In other words, these researchers do not rely on existing criteria; rather, they utilize ML method to learn the criteria on its own. For example,(Chen and Lin 2017) applies Decision Tree (DT) to select significant features that have correlations with personality traits. After feature selection, classifiers such as Support Vector Machine (SVM)(Cortes and Vapnik 1995), Random Forest (RF), K-Nearest Neighbor (KNN) are employed to predict personality characteristics.

Previous works related to Artificial-Intelligence(AI)-aided graphology have only focused on alphabetic languages. Hence, there is no Chinese graphology dataset available. Under the guidance of a professional psychology research group we collaborate with, we collect the first Chinese graphology dataset called Handwriting-Tree-Personality-184 (HTP-184), containing Chinese handwriting, tree-drawing images, and personality annotations.

Traditional graphology only focuses on handwriting analysis. Recently, however, some studies have introduced drawing tests to facilitate manual graphology analysis. The tree-drawing test (Koch's Baum Test) is a projective psychological examination and simultaneously serves as an ancillary psychodiagnostic method(Kaneda et al. 2010). Compared to handwriting, the tree-drawing test allows writers to express themselves with relatively little resistance.

Inspired by the real-scenario psychological practice, we propose a multi-modal graphology model, introducing the tree-drawing test image for the first time to help improve the prediction performance, in addition to the handwriting modality. Figure 1 depicts the pipeline of our model. Firstly, we discard the controversial empirical graphology criteria. Instead, we design 59 tailored handcrafted features for Chinese handwriting images. After feature quantification, we find 23 features that are strongly associated with some personality traits (Table 3). Experimental result shows that the 23 features can well represent the pattern of handwriting modality. Secondly, we employ ShuffleNet-v2(Ma et al. 2018) as the backbone to extract deep features from tree-drawing modality. As a lightweight deep model, ShuffleNet-v2 can balance computational complexity and accuracy, making our model suitable for real-world applications. Then we fuse the two modalities by concatenating handcrafted features from handwriting and deep features from tree images. Finally, we carefully investigate ten multi-label classification approaches, finding that Relevance K-Nearest Neighbor (BRKNN) reaches an excellent performance on F-measure metrics, providing a simple but effective baseline for future studies.

Besides, based on the proposed model, we develop an easy-to-use application where people can evaluate their personalities by simply uploading both handwriting and tree-drawing images or one of them.

Our main contributions in this paper can be summarized as follows:

- We collect the first Chinese graphology dataset with professional annotations and for the first time introduce the tree-drawing test modality into AI-aided graphology analysis. The dataset will be released soon.
- We discard empirical graphology criteria, and design feature extraction algorithms to get handcrafted features tailored for Chinese handwriting. These algorithms can also be migrated to other ideographic languages that are similar to Chinese.
- We propose a multi-modal graphology anlysis model (Figure 1) based on our dataset HTP-184. Also, we develop an easy-to-use application for personality evaluation with the model.

No.	Handwriting Features	Description
1	ThrowNum(TLN)	Throw stroke number detected
2	ThrowAxis(TA)	Long to short axis ratio of throw
3	ThrowAxisStd(TAS)	Standard deviation(std) of TA
4	ThrowDirStd(TDS)	Std of throw direction
5	ThrowSolidStd(TSS)	Std of throw Solidity
6	VertOffset(VO)	Offset of Vertical to Horizontal
7	VertSolidStd(VSS)	Std if vertical Solidity
8	VertDir(VD)	Virtical direction
9	PressSolid(PS)	Solidity of press
10	PressSolidStd(PSS)	Std of PS
11	PressDirStd(PDS)	Std of Press direction
12	PressAreaStd(PAS)	Std of press area to writable region
13	PressVerti(PV)	Offset of press to vertical
14	HorizDirStd(HDS)	Std of horizontal direction
15	HorizSolid(HS)	Solidity of horizontal
16	HorizSolidStd(HSS)	Std of HS
17	HorizOffset(HO)	Offset of Horizontal to vertical
18	HorizNum(HLN)	Horizontal number detected
19	CentrOffset(CO)	Centroid offset of hull and skeleton
20	LineDist(LD)	Distance between lines
21	LineDistStd(LDS)	Std of LD
22	AreaRatio(AR)	Written to Writable area ratio
23	ConerRatio(CR)	Corner to writable area ratio

Table 2: The definitions of 23 handcrafted features. Using MLR method, We find these features strongly associate with several personality traits listed in Table 3.

2 Related Works

The study of handwriting, namely Graphology, is quite an old concept tracing back to the seventeenth century. However, researches investigating the factors associated with manual graphology have only focused on alphabetic languages until 1990, the world's first International Institute of Chinese Character Graphology was established in Brussels. This is the first time Ideograph (represented by Chinese characters) graphology has entered into the horizon of graphologists.

Handwriting Related Research As artificial intelligence technology emerges in recent years, handwriting study is a heated topic. There are some handwriting related subtopics such as handwriting recognition(Bhunia et al. 2019) , handwriting synthesis(Ghosh, Bhattacharya, and Chowdhury 2017; Tolosana et al. 2021), and handwriting style imitation(Gan and Wang 2021), handwriting writer identifica-

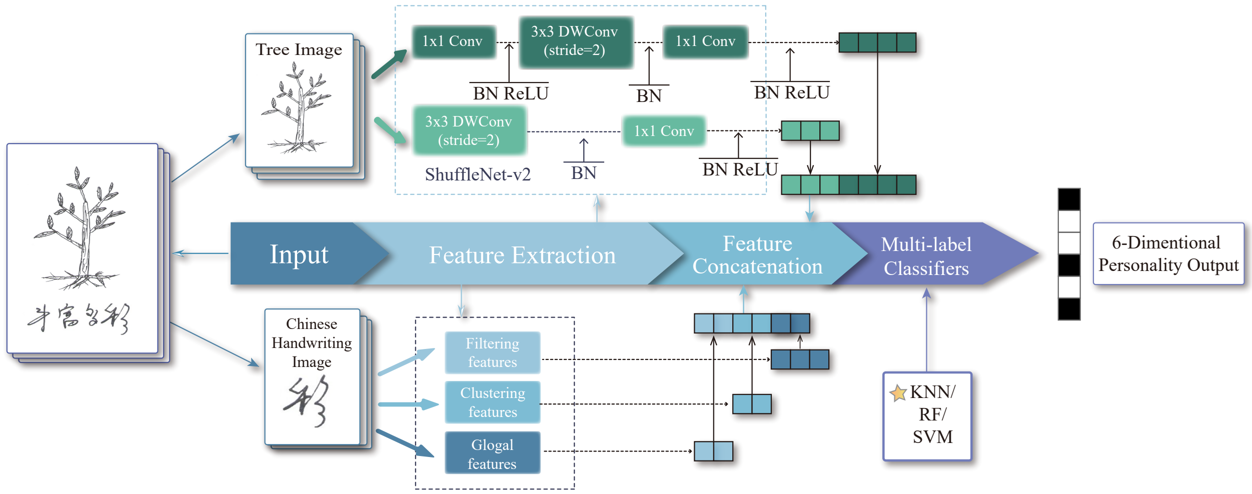


Figure 1: The pipeline of our multi-modal model. There are two modalities, tree-drawing and handwriting images. Our model adopts ShuffleNet-v2 to extract features from tree images, and extract handcrafted handwriting features using algorithms we design. Then we use ML classifiers to predict personalities. Experiments show that kNN reaches the best F1-Measure score.

tion(He and Schomaker 2019). Though subtopics mentioned above have been relatively well studied, personality evaluation through handwriting (a.k.a graphology) has not.

Empirical-Graphology-Based Model (Sheikholeslami, Srihari, and Govindaraju 1995) proposes a model called Computer Aided Graphology (CAG) that can generate a personality report of an input handwriting image by listing correlations between personality traits and specific handwriting features. For example, one of the criteria used in CAG is *Increasing left margin* $\Rightarrow$ *Tendency to fatigue as work progresses*. Other empirical-graphology-based models also follow the same procedure but they introduce new graphology criteria for their models. (Mutalib et al. 2007; Champa and AnandaKumar 2010) explores the “t” alphabet in their model, using ANN. (Prasad, Singh, and Sapre 2010) detects six handcrafted features and then uses SVM to classify them. (Djamal, Darmawati, and Ramdlan 2013) analyze autograph and handwriting simultaneously. (Djamal and Febriyanti 2015) combines global features (e.g., page margin) and local features (e.g., “t”), and introduces more alphabets such as a, d, i, and m. (Kedar et al. 2015) develops a graphology model to help identify negative emotions from the writer. And in the recent four years, some researchers start to employ convolutional neural networks (CNN) to tackle this task(Fatimah et al. 2019). Just like(Sheikholeslami, Srihari, and Govindaraju 1995) claims, they haven’t checked the validity of the graphology criteria they employ, so do all works mentioned above.

Graphology-Criteria-Learning Model Unlike most previous studies that only gave the association relations between personality traits and specific writing behaviors; instead, Zhi Chen et al. firstly proposed their notable works that provide a quantization of personality metrics(Chen and Lin 2017). In 2018, a Spanish graphology dataset called HWxPI(Ramirez De La Rosa, Villatoro tello, and Jimenez Salazar 2018) became available. HWxPI allows researchers to develop ML methods without the need to depend on existing graphology criteria. Several works have been done based on this dataset (Escalante 2019; Valdez-Rodríguez, Calvo, and Felipe-Riveron 2018; Gahmousse et al. 2020).

Besides alphabetic geography, there are also a few works research into Arabic handwriting(Mostafa, Al-Qurishi, and Mathkour 2019), Hebrew(Ben-Shakhar et al. 1986).

3 Dataset

3.1. Data Collection

We conduct Volunteer experiments to collect and annotate the data. Strictly following the psychology experimental standards, we try our best to guarantee the dataset quality.

There are two parts of the data we need to collect from each subject: (1) **The handwriting image collection**: this part requires subjects to write a designated paragraph in Chinese, and the writing area is limited to a rectangle region. 2) **The tree-drawing test image collection**: this part requests each subject to draw a tree without any limitation on the style or breed. Still, the drawing is required to be within the region we designate.

Eventually, we collect 117 samples from 117 subjects with annotations. The gender ratio of males to females is 35:64. The age distribution is from 18 to 28. The educational background of the subjects is all undergraduates or graduate students. Besides the 117 samples with annotations, 67 samples without annotations are provided by the psychology research group. The size of the dataset is relatively small because it’s economically costly and time-consuming to collect samples from each subject (it takes 20 to 60 minutes to collect one valid sample). To our knowledge, not only is our dataset the first Chinese handwriting dataset with personality annotation that will be publicly available, but also the first dataset with the tree-drawing test modality.

No.	Dimensions	Handwriting Features
1	Thinking-oriented	LDS
2	Action-oriented	PS, LD, HDS
3	Introvert	None
4	Extrovert	VD
5	Emotional	HSS, CO
6	Rational	TA, TSS, HO, TAS, PDS, VSS
7	Compatible	AR, VSS, CR, PV
8	Controlling	VSS, VO, PS, TDS, PAS
9	Conservative	TLN
10	Ambitious	TLN
11	Idealistic	AR, PSS, TDS
12	Realistic	VO
No.	**Paired Dimensions**	**Handwriting Features**
1	Thinking-Action	HDS
2	Intro-Extro	None
3	Emotion-Ration	HLN, CO
4	Compat-Control	AR,VSS, VD, HS
5	Conserv-Ambit	TLN
6	Idea-Real	AR, PSS

Table 3: Correlations found between handcrafted features and personality. We investigate the 12 personality traits individually and then in pairs. The definitions of handwriting features are listed in Table 2.

3.2. Data Annotation

Each sample is annotated by a rating scale. A rating scale is a kind of questionnaire that requires the subjects to assign a value to the rated object, as a measure of some rated attribute. The annotation is completed by the subjects filling up a scale provided by the psychology research group. And the scale assesses subject's personality, serving as the ground truth labels to describe the personality of each subject. It takes each subject about 30 minutes to fill up the scale. Finally, the personality label has 12 integer numbers referring to 12 different personality dimensions, including:

- $\{ThinkingOriented, ActionOriented\}$,
- $\{Introvert, Extrovert\}$,
- $\{Emotional, Rational\}$,
- $\{Compatible, Controlling\}$,
- $\{Conservative, Ambitious\}$,
- $\{Idealistic, Realistic\}$.

And each pair in the same brace is mutually exclusive.

3.3. Data Taxonomy

There are 6 pairs, in total 12 personality traits annotated on each subject. Meanwhile, the 2 personality characteristics in the same pair are exclusive. For example, Action-Oriented and Thinking-Oriented are a pair of personality traits. If one person is classified as Thinking-Oriented, then they cannot be Action-Oriented type at the same time. Consequently, according to this mutually exclusive attribute, the label space shrinks from 4096 (2^{12}) to 64 (2^6) kinds.

4 Chinese Graphology

As mentioned in related works, AI-aided graphology mostly focuses on alphabetic languages. Taking English as representative for alphabetic languages, in Section 4.1, we will talk about the differences between Chinese and English. In Section 4.2, we will discuss the challenges faced in Chinese graphology.

4.1. Differences between Chinese and English Graphology

The smallest unit in an English word is called the alphabet. There are 52 alphabets if we add up capital and lowercase letters. Unlike English, the smallest unit in a Chinese Character is called characters. Every character is different in shape. The number of most frequently used characters is around 2500 (General Chinese middle school requirement for students is to know 3500 characters). Such abundance in morphology aspect provides more potential handwriting features than English alphabets. This phenomenon also explains why previous works are more focused on global handwriting features. For example, page margins(Sheikholeslami, Srihari, and Govindaraju 1995). Despite the richness in shapes, Chinese characters are composed of some regular components called strokes (Figure 2), making it easier to extract local features. Several local features were also been taken into consideration in previous works, such as “t” bar position(Champa and AnandaKumar 2010), dot positions in “i”(Djamal and Febriyanti 2015). In the end, we successfully find 23 out of the 59 features have strong associations with personality traits (Figure 2).

4.2. Challenges Faced in Graphology

First of all, most of the previous works use empirical graphology criteria, causing doubts about the feasibility of the previous works. To avoid the same dilemma, we quantify 59 features and determine the correlation between features and personality traits. However, what makes a feature effective remains unknown. Therefore our strategy is to extract possible features, then select those who have correlations with personality traits. Secondly, The feature extraction algorithms are rule-based. Thus, it is challenging but significant to design stable, robust algorithms. Thirdly, considering practical reasons, the features should also be easy to detect and extract. In(Chen and Lin 2017), writing velocity and acceleration are the cardinal features. But detecting these features requires special measure devices (e.g. WACOMDTZ-1200W tablet), making it hardly possible for real-world application. And finally, the relatively small size of the dataset makes it a few-shot machine learning problem. Thus, overfitting is also an issue we should handle.

5 Handcrafted Feature

For handwriting images, we design handcrafted features. We haven't employed deep learning models that are more powerful and convenient. There are three reason: (1) The size of our dataset is small. Moreover, the textures in handwriting images are not as abundant as real world images or sketches. This might cause serious over-fitting problems in

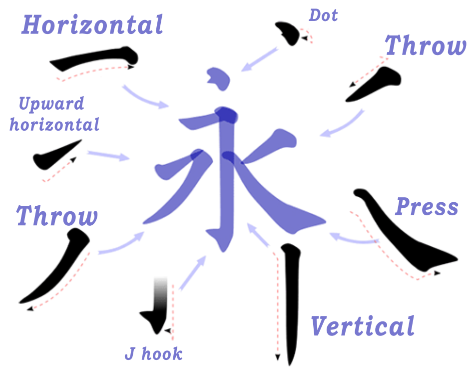


Figure 2: Stroke elements in Chinese. Among them the most basic four strokes are horizontal, vertical, press, and throw.

deep learning models. (2) Chinese handwriting follows regular writing rules (e.g. stoke orders), which makes it applicable to extract specific features. (3) Extracting specific features provide direct evidence of the existing association between handwriting and personality, which can be utilized in graphology psychology research. Here are three types of features we extract using rule-based methods: filtering-based feature, clustering-based feature, and global features.

5.1. Filtering-based Feature

Filtering-based methods erase useless information on the handwriting images. These features can be captured only after the filtering because they're relatively micro and local features.

Though there are approximately 3500 characters are frequently used, the most basic components of them are the same — **stokes** is the elements of characters. Among all the strokes presented in Figure 2, four of them are the most fundamental: horizontal stroke, vertical stroke, throw (left-falling) stroke, press (right-falling) stroke. Inspired by classical filter algorithms (e.g., Sober), we design several convolutional kernels to filter the handwriting images.

Let F denote a 2-dimensional convolutional kernel. The shape of the kernel is $k \times k$, and p, q denote the abscissa and ordinate index. Similarly, let H denote a handwriting image, and i, j as its abscissa and ordinate index. Finally, C represents the result.

The filtering operation is as follows:

$$C(p, q) = \sum_{i=0}^{k} \sum_{j=0}^{k} F(i, j) \cdot H(p - i, q - j)$$

After filtering, only the target stroke remains. For example, if we filter an image with the horizontal kernel (Table 4), only horizontal strokes will be left. Thus, by scanning the images with a specific kernel, we can get the corresponding stroke. In our works, we firstly extract the four stokes using the filtering-based method. Once we get all the strokes we need from a handwriting sample, we calculate attributes of the horizontal stroke such as: the average slant, the standard deviation of the slant, the ratio of the long axis to the short axis. In the end, we get 52 filtering-based local features.

0	**0**	**0**	**0**	**0**
1	**2**	**0**	**2**	**1**
3	**4**	**16**	**4**	**3**
1	**2**	**0**	**2**	**1**
0	**0**	**0**	**0**	**0**

Table 4: Horizontal kernel. Besides, kernels of vertical, throw, and press are horizontal rotated by specific angles.

5.2. Clustering-based Feature

Clustering-based features are relatively macro compared to those in the filtering-based features. These clustering-based features are line-level, which means the features are pertinent to lines of writing rather than local strokes inside characters. To get features from lines, we should first separate each line from the others. Using Density-Based Spatial Clustering of Applications with Noise (DBSCAN)(Ester et al. 1996) algorithms, we managed to get robust line clustering results. And finally, we've found two line-centered features: the slant of line, and the distance between lines.

5.3. Global Features

Five features depict the coarse-grained features of the handwriting image. They are area ratio of the written region to the whole writable area, the ratio of salient points to handwriting skeleton points, the ratio of handwriting convex hull and skeleton, the bias of convex hull center and skeleton center, and the coincidence rate of the focus of character convex hull and focus of character skeleton. They can describe a global scope of handwriting attributes.

5.4. Feature Selection

For handcrafted features, it is of great importance to select features that can better represent the data. Recently, some noteworthy works have proposed new methods to better select features, such as feature selection for multi-label classification(Wu et al. 2020), for meta learning(Schnapp and Sabato 2021). Multivariate Linear Regression (MLR) is a technique widely used in psychology to determine correlations between variables. Thus, we use MLR to select those features that have correlations with personality traits. Even though we have designed 59 handcrafted features, We will not adopt those features that fail to correlate with any personality traits. After feature selection, we find 23 features strongly associate with some personality traits (Table 3).

6 Experiment

There are two modalities in each sample: handwriting image and tree-drawing test image. We develop classifiers for them to see their prediction ability respectively. Then we fuse two modalities on the feature-level and test the combined features on the classifiers. In this section, we will conduct experiments to explore three aspects of our model: 6.2 is investigations into the two modalities; 6.3 explores the effectiveness of our handcrafted features; and in 6.4 we will perform ablation experiment to discuss the optimization process of our model.

6.1 Evaluation Metrics

The distribution of personality types in human society is unbalanced. For instance, in our dataset, the percentage of rational people is 17.9%. To evaluate the effectiveness of our multi-label classifiers, we utilize precision, recall ratio, F1-Measure score as the metrics. As multiple labels provide more than one metrics value, we choose the micro average method to get the mean score, When data distribution is not balanced, as a commonly used metric, accuracy can fail to evaluate the effectiveness of classifiers. The precision (denoted as Pre) refers to the proportion of true positive samples(TP) among all positive samples the classifier report, including true positive(TP) and false positive (FP). The recall ratio stands for the proportion of all positive samples (TP + FN) that are determined by the classifier to be real positive (TP). The F_β-Measure score is based on precision and recall ratio. It can be described as:

$$F_\beta = \frac{(1+\beta^2).Pre.Rec}{\beta^2 Pre + Rec}$$

The F_β-Measure score can be regarded as a weighted average of precision and recall, thus β is the coefficient to adjust the weight. In this task, we consider precision and recall are equally important, henceF1-Measure is employed. And for multi-label classification problems like our works, there is two mainstream of average method: micro average and macro average. Micro averages metrics scores on different labels respectively, yet macro averages overall scores and ignore the quantity variance among labels. Thus, the intrinsic unbalanced distribution of personality makes it better to use the micro average.

6.2. Investigations of Modalities

Handwriting Modality We extract 59 features from handwriting images via algorithms we design in section 6. However, after feature selection through multivariate linear regression, we find that only 23 out of the 59 features strongly associate with the 12 personality traits. Table 3 represents the correlations: Thus, we develop several classifiers by applying different multi-label classification methods such as Binary Relevance, Classifier Chain, LabelPowerset, Multi-label embedding.

Tree-drawing Modality The tree-drawing test provides free-hand sketch images, which contain more various textures than handwriting. Therefore, designing handcrafted features for such elusive patterns can be a challenging task than for handwriting images. As a result, handcrafted are employed for tree-drawing images. Finally, train ShuffleNet-v2 developed by Ningning Ma et al(Ma et al. 2018). ShuffleNet-v2 is a CNN architecture that trades off accuracy and efficiency. First of all, as a CNN model, it helps us to extract features from free-hand sketches. Secondly, unlike VGG, Resnet, etc., it reduces computational complexity, making it possible to deploy it on efficient applications. Finally, it consumes less time to re-train the model when new data is added to the dataset, which benefits the updating process for real-world application.

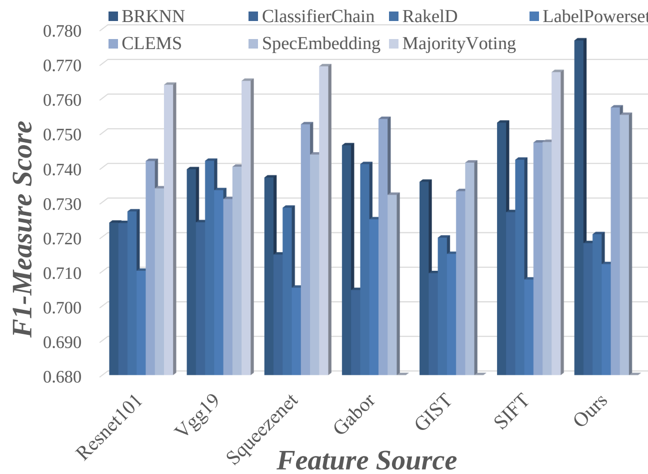


Figure 3: Average F1-Measure scores of different features on different classifiers. Abbreviation in the chart: RakelD for Distinct RAndom k-labELsets multi-label classifier; CLEMS for Cost-Sensitive Label Embedding with Multidimensional Scaling; spectral embedding for SpecEmbedding.

Multi-modal Fusion To bring out the best of the two different modalities and supplement one with another, we fuse the two modalities on the feature level. Now we have handcrafted features extracted from the handwriting images and deep features from the tree-drawing sketches. By concatenating them as the multi-modal representation, then use multi-label classifiers (e.g. BRKNN), we reach a better F1-score than the uni-modal circumstances. Figure 1 presents The multi-modal architecture.

6.3. Effectiveness of handcrafted Features

To demonstrate the effectiveness of our handcrafted features, we extract several other classical handcrafted features to make comparisons such as SIFT(Oliva and Torralba 2001), Gabor(Mehrotra, Namuduri, and Ranganathan 1992), Gist(Oliva and Torralba 2001). Besides these classical handcrafted features, we also extract deep features using mainstream deep models such as ResNet101(He et al. 2016), Vgg19(Simonyan and Zisserman 2014), and lightweight deep model SqueezeNet(Iandola et al. 2016) to make the comparison more comprehensive. We conduct experiments on different multi-label classifiers we build. Figure 3 shows that the handcrafted features we design reach the best F1-Measure score of 0.776 when using the BRKNN classifier (average of 100 rounds of training on random train-test split). In addition, Figure 3 also shows that deep features have a relatively steady performance on different classifiers. Besides, deep features from different models all have excellent F-Measure scores on the Majority Voting model. However, due to the model scale, deep models consume around 2.3 times more time than our features. In all the other handcrafted features, Gabor and Gist have weak and unstable per-

Classifiers	BRKNN			CLEMS			MAJORITYVOTING			BRRF		
#	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1	Pre	Rec	F1
1. Gabor Feature	0.694	0.787	0.736	0.694	0.789	0.738	0.653	0.502	0.565	0.690	0.734	0.710
2. Our 59 Features	0.703	0.844	0.766 $\uparrow_1$	0.710	0.842	**0.770** $\uparrow_1$	0.703	0.604	0.648 $\uparrow_1$	0.702	0.718	0.709
3. 59 Features+Tree	0.708	0.842	0.768 $\uparrow_1$	0.704	0.825	0.758	0.700	0.630	0.663 $\uparrow_1$	0.705	0.729	0.715
4. Selected 23 Features	0.708	0.841	0.768	0.705	0.833	0.763	0.694	0.672	0.682 $\uparrow_{2,3}$	0.705	0.710	0.706
5. **Selected+Tree**	0.714	0.850	**0.776** $\uparrow_4$	0.716	0.826	0.766	0.718	0.699	**0.707** $\uparrow_4$	0.709	0.758	**0.732** $\uparrow_4$

Table 5: The ablation experiments conducted in section 6.4. From top to bottom in this Table, the model gets optimized. The serial number in the first column indicates different stages of the optimization. Meanwhile, we use ↑ to denote a noteworthy improvement, and the subscript numbers indicate with which stage the improvements are compared. The bold numbers are the highest F1-Measure score in a column. We only note the best F1 score because it's the weighted average of Pre and Rec.

formance. SIFT feature reaches a decent F1 score, yet it cannot surpass our features. Finally, for the BRKNN method, we can see that only our features perform well and reach the highest F1-Score. On account of the innate characteristics of KNN that it works better when the features can better represent the data, it indirectly manifests that our features are more effective than others.

6.4. Ablation Experiment

In this paper, we optimize our model in the following 4 aspects.

- We extract 59 tailored handcrafted features for handwriting modality;
- We perform feature selection via MLR, gaining 23 effective features in the end;
- We add tree-drawing modality to utilize the supplementary information that handwriting image lacks.
- We try different multi-label machinle learning approaches to find out the best one is BRKNN.

Table 5 depicts the optimization process by comparing metrics score on different classification approaches. From top to bottom in Table 5, the model has been improved step by step. We use ↑ to denote notable improvements. First of all, we can see that no matter which classifier or which stage of the optimization, the precision only fluctuate a little around 0.7. Nevertheless, the **selected+Tree** is still has slightly higher precision than all other models that are not fully optimized. Secondly, the recall ratio is the main factor that has been improved, which directly influences the F1-measure score. Thirdly, From the MajorityVoting model, a relatively poor classifier, the improvement on it is more conspicuous than on those which are already strong. And even it's less obvious on those already effective classifiers, we can still see the ascending tendency of F1-Measure scores from stage 1 to stage 5. But there is an exception about CLEMS, it always outperforms the others when given 59 features and tree features together, which indicates it might have utilized information our feature selection method cannot capture. Finally, tree-drawing modality contributes relatively limited improvement. It is conceivable due to the relatively small size of dataset.

7 Application

Based on our model, we develop an application. This application can predict the writer's characteristics and will return a report with six-dimensional personality. This application can be employed not only in graphology psychology research, but also in many other practical scenarios, such as human resource management, adolescent development guidance, psychotherapy, interpersonal counseling.

Although our model is multi-modal, the application can work when given only one of the modalities as well. However, according to Table 5, it returns the most reliable evaluation when given two modalities simultaneously, renders a decent result when only given handwriting, and provides a relatively unsatisfactory result when only given tree image.

8 Conclusion

In this paper, we propose an ML-based multi-modal model for Chinese graphology, a challenging and understudied problem. And for the first time, we introduce a tree-drawing modality to improve the personality evaluation in AI-aided graphology. Also, we collect and publish the first Chinese graphology dataset, making our works reproducible and extensible. Meanwhile, we develop an application that helps people to make use of our research results. Moreover, the methodology can be used in not only Chinese graphology, but also transfer to other languages with corresponding adjustments. We hope our works will not only help advance psychology research but also facilitate machine learning application in social sciences.

9 Future Works

The high expense of data collecting makes it hard to obtain huge amounts of data with high-quality annotation. Thus, it seems like a promising choice to collect less fine-grain labeled data, which will make data collecting faster and more economical. Besides, autograph is also a common modality, and it is easier to collect. When the dataset size is large enough, we can take advantage of the deep models as it already show great ability with only small amount of data shown in Figure 3. Additionally, in our works, we only investigate four basic strokes of Chinese as local features. More strokes are waiting to be researched.

References

Ben-Shakhar, G.; Bar-Hillel, M.; Bilu, Y.; Ben-Abba, E.; and Flug, A. 1986. Can graphology predict occupational success? Two empirical studies and some methodological ruminations. *Journal of applied psychology*, 71(4): 645.

Bhunia, A. K.; Das, A.; Bhunia, A. K.; Kishore, P. S. R.; and Roy, P. P. 2019. Handwriting recognition in low-resource scripts using adversarial learning. In *CVPR*, 4767–4776.

Champa, H.; and AnandaKumar, K. 2010. Artificial neural network for human behavior prediction through handwriting analysis. *International Journal of Computer Applications*, 2(2): 36–41.

Chen, Z.; and Lin, T. 2017. Automatic personality identification using writing behaviours: an exploratory study. *Behaviour & Information Technology*, 36(8): 839–845.

Cortes, C.; and Vapnik, V. 1995. Support-vector networks. *Machine learning*, 20(3): 273–297.

Dazzi, C.; and Pedrabissi, L. 2009. Graphology and personality: An empirical study on validity of handwriting analysis. *Psychological Reports*, 105(3): 1255–1268.

Djamal, E. C.; Darmawati, R.; and Ramdlan, S. N. 2013. Application image processing to predict personality based on structure of handwriting and signature. *IC3INA 2013*, 163–168.

Djamal, E. C.; and Febriyanti. 2015. Identification of speed and unique letter of handwriting using wavelet and neural networks. *EECSI*, 2(August): 74–78.

Escalante, H. J. 2019. Recognition of Apparent Personality Traits from Text and Handwritten Images. In *ICPR 2018*, volume 11188, 146. Springer.

Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X.; et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, 226–231.

Fatimah, S. H.; Djamal, E. C.; Ilyas, R.; and Renaldi, F. 2019. Personality features identification from handwriting using convolutional neural networks. *ICITISEE 2019*, 6: 119–124.

Gahmousse, A.; Gattal, A.; Djeddi, C.; and Siddiqi, I. 2020. Handwriting based Personality Identification using Textural Features. *ICDABI 2020*.

Gan, J.; and Wang, W. 2021. HiGAN: Handwriting Imitation Conditioned on Arbitrary-Length Texts and Disentangled Styles. In *AAAI*, volume 35, 7484–7492.

Gavrilescu, M.; and Vizireanu, N. 2018. Predicting the Big Five personality traits from handwriting. *EURASIP Journal on Image and Video Processing*, 2018(1): 1–17.

Ghosh, A.; Bhattacharya, B.; and Chowdhury, S. B. R. 2017. Handwriting profiling using generative adversarial networks. In *AAAI*, volume 31.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *CVPR*, 770–778.

He, S.; and Schomaker, L. 2019. Deep adaptive learning for writer identification based on single handwritten word images. *Pattern Recognition*, 88: 64–74.

Iandola, F. N.; Han, S.; Moskewicz, M. W.; Ashraf, K.; Dally, W. J.; and Keutzer, K. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and¡ 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.

Kaneda, A.; Yasui-Furukori, N.; Saito, M.; Sugawara, N.; Nakagami, T.; Furukori, H.; and Kaneko, S. 2010. Characteristics of the tree-drawing test in chronic schizophrenia. *Psychiatry and Clinical Neurosciences*, 64(2): 141–148.

Kedar, S. V.; Bormane, D. S.; Dhadwal, A.; Alone, S.; and Agarwal, R. 2015. Automatic emotion recognition through handwriting analysis: A review. *ICCUBEA 2015*, 811–816.

Ma, N.; Zhang, X.; Zheng, H.-T.; and Sun, J. 2018. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *ECCV*, 116–131.

Mehrotra, R.; Namuduri, K. R.; and Ranganathan, N. 1992. Gabor filter-based edge detection. *Pattern recognition*, 25(12): 1479–1494.

Mostafa, M. A.; Al-Qurishi, M.; and Mathkour, H. I. 2019. Towards Personality Classification Through Arabic Handwriting Analysis. In *The International Research & Innovation Forum*, 557–565. Springer.

Mutalib, S.; Rahman, S. A.; Yusoff, M.; and Mohamed, A. 2007. Personality Analysis Based On Letter ' t ' Using Back Propagation Neural Network. *Proceedings of the International Conference on Electrical Engineering and Informatics Institut Teknologi Bandung*, (2): 496–499.

Oliva, A.; and Torralba, A. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3): 145–175.

Prasad, S.; Singh, V. K.; and Sapre, A. 2010. Handwriting Analysis based on Segmentation Method for Prediction of Human Personality using Support Vector Machine. *International Journal of Computer Applications*, 8(12): 25–29.

Ramirez De La Rosa, A. G.; Villatoro tello, E.; and Jimenez Salazar, H. 2018. HWxPI: A multimodal spanish corpus for personality identification. Latin American and Iberian Languages Open Corpora Forum.

Rollin, H. R. 1984. Longman Dictionary of Psychology and Psychiatry. Editor in Chief: Robert M. Goldenson. Harlow, Essex: Longman. 1984. Pp 815.£ 29.50. *The British Journal of Psychiatry*, 145(6): 682–683.

Schnapp, S.; and Sabato, S. 2021. Active Feature Selection for the Mutual Information Criterion. In *AAAI*, volume 35, 9497–9504.

Sheikholeslami, G.; Srihari, S. N.; and Govindaraju, V. 1995. COMPUTER AIDED GRAPHOLOGY 1 Introduction 2 Graphology 3 Computer Aided Graphology. *Master's thesis, State University of New York at Buffalo*, 1–7.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tolosana, R.; Delgado-Santos, P.; Perez-Uribe, A.; Vera-Rodriguez, R.; Fierrez, J.; and Morales, A. 2021. Deepwritesyn: On-line handwriting synthesis via deep short-term representations. In *AAAI*.

Valdez-Rodríguez, J. E.; Calvo, H.; and Felipe-Riveron, E. M. 2018. Handwritten Texts for Personality Identification Using Convolutional Neural Networks. In *ICPR*, 140–145. Springer.

Wu, X.; Jiang, B.; Yu, K.; Chen, H.; and Miao, C. 2020. Multi-label causal feature selection. In *AAAI*, volume 34, 6430–6437.