



# A new ontology-based multimodal classification system for social media images of personality traits

Kunal Biswas<sup>1</sup> · Palaiahnakote Shivakumara<sup>2</sup> · Umapada Pal<sup>1</sup> · Tong Lu<sup>3</sup>

Received: 20 October 2021 / Revised: 29 March 2022 / Accepted: 4 May 2022 / Published online: 1 June 2022  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

## Abstract

Number of users of social media is increasing exponentially. People are getting addicted to social media, and because of such addiction, it sometimes causes psychological and mental effects on the users. Understanding user interaction with social media is essential to study personality traits of the users. This paper focuses on classification of personality traits called Big Five factors, namely (i) agreeableness, (ii) conscientiousness, (iii) neuroticism, (iv) extraversion and (v) openness by combining image and textual features through ontology and fully connected neural network (FCNN). The intuition to classify the images of the five classes is that there is a strong correlation between images, profile picture, images of status, text in the images, description of the images uploaded on social media and the person's mind. To extract such observation, we explore an ontology-based approach, which constructs a weighted undirected graph (WUG) based on labels of the images, profile picture, banner image, text in the images and description of the images in a novel way for feature extraction. The proposed approach uses mini-batch gradient descent to convert WUG to feature vector form for each word. Furthermore, the feature vectors are fed to the FCNN for classification of personality traits-oriented images. The proposed classification method is evaluated by testing on our dataset of five classes containing 5000 images and two benchmark datasets, namely 5-class dataset which provides 33,556 images, and 10-class dataset which provides 2000 images. The results on different datasets show that the proposed approach is superior to the existing methods in terms of average classification rate.

**Keywords** Social Media · Ontology · Multimodal · Undirected graphs · Personality traits

## 1 Introduction

Social media platforms, such as Facebook, Twitter, Instagram and WhatsApp, are the integral parts of human activities for

sharing their happiness, feeling, emotions as well as for communication [1]. The impact and influence of social media on persons is huge, and therefore, it is causing mental and psychological problems for users, which include depression, hyper-aggressiveness, blissfulness, etc. [2, 3]. For example, the status of a user account indicates activities and actions of persons without communicating others. This motivated us to propose an approach for classifying social media images which represent five personality traits, such that our study can help systems to identify a person's mental status to avoid adverse effects. Personality traits classification using text is not new, and there are several methods developed in the past [4, 5]. However, most existing methods work well for text information because the approaches use natural language processing for identifying a person's behavior, traits, mental and psychological status.

The scope of the existing methods is limited to text information. In other words, the existing methods ignore other information, such as profile picture, images, banners and text in images for identifying personality traits. Similarly,

---

✉ Palaiahnakote Shivakumara  
shiva@um.edu.my

Kunal Biswas  
biswaskunal29@gmail.com

Umapada Pal  
umapada@isical.ac.in

Tong Lu  
lutong@nju.edu.cn

<sup>1</sup> Computer Vision and Pattern Recognition, Indian Statistical Institute, Kolkata, India

<sup>2</sup> Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia

<sup>3</sup> National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China



**Fig. 1** Examples of images of social media account information for all the five personality chosen from our dataset

there are approaches for emotions, expressions, action, gesture recognition, person behavior identification using facial, posture and video information in the literature [6–11]. However, these methods are inadequate to address the challenges of social media images of different personality traits. This is because the images uploaded on social media may not have face information all the time, and even if there is facial information, it may not be useful because of multiple emotions and expressions in a single image with a complex background.

We believe that images, profile pictures, banners, descriptions, texts and comments of users have a strong correlation syntactically and semantically, which reflects the personality traits of a person. Sample user images, profile pictures and banner images of different personality traits can be seen in Fig. 1a–e. For example, the person who is in depression may not upload images that convey happiness rather uploads

images with dark color background, which convey loneliness or sadness. In the same way, text may contain sad words and status (profile picture) may provide unusual photos. It is observed from Fig. 1a–e that images uploaded by the user, profile picture and banner images with text in (a) indicate the personality traits of Agreeableness. In the same way, in Fig. 1b–e, the images and texts indicate personality traits of conscientiousness, openness, extraversion and neuroticism, respectively. Since there is no constraint to post specific images, text and status for a particular personality trait as it is all depending on the individual mind, it is not easy to find coherent meaning among all the attributes mentioned above. Therefore, the classification of social media images which represent different personality traits is a complex problem.

To extract the above observations, inspired by intuition that although words and labels of images are different, the meanings of the words and labels are the same, and thus we propose to explore ontology that constructs a semantic graph by considering syntactic and spatial information, namely word-to-word co-occurrences, and syntactical information, namely dependency parsing within sentences. It is noted that the ontology has the ability to encode the semantic relationship between texts, and hence, it increases the discriminative power of classifiers [12]. Motivated by this, we explore ontology for classification of images of five personality traits in this work. Main contributions of the proposed work are as follows.

- (i) To the best of our knowledge, this is the first work to explore ontology for the classification of social media images of five personality traits.
- (ii) Integrating image features, image annotations, image text features and twitter text features for classification of images of five personality traits is different from the past methods.
- (iii) Exploring ontology to encode the semantic information extracted from different modalities, namely image features, image text features and twitter text features for classification of images of five personality traits with the help of mini-batch gradient descent approach with a fully connected neural network, is novel compared to the past methods.

The rest of the paper is organized as follows. The review of the past pieces of work is presented in Sect. 2, and it includes the methods of personality traits classification using handwriting analysis and images, emotions recognition using facial information, etc. Section 3 introduces the step for word formation, ontology concept for extracting syntactic and semantic information between the words and FCNN for classification. Section 4 presents discussion on experiments on different datasets to validate the proposed method. Section 5 concludes about the outcome of the proposed method.

## 2 Related work

In recent times, there have been several methods developed to study personality traits of people which can be classified as graphology, handwriting analysis, face and emotions recognition-based ones.

Graphology-based methods work well by studying characteristics of handwriting characters/strokes, and hypotheses are defined according to graphologists [3]. Patil et al. [13] used handwriting analysis for personality traits classification, which focuses on particular characters like “t” for extracting features. Chaudhari et al. [3] presented a survey on handwriting-based personality traits identification, where several methods were surveyed for identification. Bargshady et al. [14] proposed an algorithm to detect pain intensity from facial expression images using deep learning. Bozorgtabar et al. [15] proposed adversarial domain adaptation for capturing the visual appearance of a face using an image-to-image transfer model. Mungra et al. [16] proposed a CNN-based emotion recognition method using histogram equalization and data augmentation. Sharma et al. [7] proposed emotion recognition using facial expression by exploring key point descriptors and texture features. These constraints defined in the above-existing methods are not necessarily true for images considered in our work, where images can have a complex background, multiple emotions, different facial expressions and actions. With these limitations, we can conclude that the methods are not so capable of classifying personality traits oriented social media images.

Personality traits classification has been explored in different contexts in the literature. Palhano et al. [17] proposed a method for classifying personality traits such as agreeableness, conscientiousness, extraversion, neuroticism and openness. Lai et al. [18] proposed a method for studying the personality traits of students during e-learning. Liu et al. [6] proposed a method for analyzing personality traits through social media profile picture choice. Zhu et al. [1] proposed a method for studying personality traits from scene perception probability using CNN. Krishnani et al. [19] proposed a method for person behavior-oriented image classification using structural function and transform. Xue et al. [20] proposed a deep learning-based method for personality recognition using texts posted on online social networks. The model involves AttRCNN structure and inception structure to extract deep semantic features from the text information. The features are fed to traditional regression methods for classification of five personality traits. However, the scope of the method is limited to text information but not images of different personality traits. Xue et al. [21] proposed a method for personality traits recognition using text information and the method explores semantic-enhanced sequential networks. The method considers words and then it converts the words to vector form to extract word level semantics

based on context learning. A fully connected layer is used for recognition of personality traits. Since the scope of the method is limited to text information, it may not work for classification of images of different personality traits.

In summary, although the existing methods address challenges of personality traits-oriented image classification in the literature, none of them considers the combination of syntactic and spatial information of images uploaded by users, profile images, banner images and texts in images for classification. The existing methods overlook the strong relationship that exists between labels of image content, profile images and banners or texts in images. In addition, most methods use low-level features for classification but not high-level features which are used in our work. These limitations motivated us to propose a new approach for the classification of personality traits oriented social media images in this work.

## 3 Proposed approach

The objective of the proposed work is to classify social media images that represent Big Five factors (BFF), namely agreeableness, conscientiousness, neuroticism, extraversion and openness. As mentioned in the previous section, images uploaded by the user, profile picture, banner, text in images and description of images have strong correlation spatially and syntactically that represent each personality trait. This observation motivated us to explore an ontology-based approach for classification. For each input image, we use Google Cloud Vision API, which is available publicly for text detection, recognition and label detection in images, profile pictures and banner images. Further, description and other text information are used for extracting key information as reported in Table 1 for sample images of agreeableness and conscientiousness classes. Overall, the proposed method uses text recognition information, labels and descriptions to find the unique correlation among them for the classification of social media images of personality traits through constructing an ontology graph.

Inspired by ontology that is popular for representing concepts and relations, we explore the same for finding the unique spatial and semantic relations between the key words of different classes in the form of a weighted undirected graph (WUG). This graph describes the relation strength of every word to every other word in the vocabulary. To normalize the values in the ontology WUG, we propose a mini batch gradient descent approach, which results in feature vectors. The feature vectors are then fed to a fully connected neural network (FCNN) for classification of social media images that represent five personality traits. The block diagram of the proposed method can be seen in Fig. 2.

**Table 1** Extraction of the keywords from the input images of different classes

Classes	User image			Profile picture			Banner image		
	Recognition	Annotations	Description	Recognition	Annotations	Description	Recognition	Annotations	Annotations
Agreeableness	"FORGIVENESS IS ESSENTIAL TO OUR MARRIAGES ARE TO SURVIVE"	"Text font friendship"	"10 way to cultivate a more forgiving heart"	Not any	"People, smile"	"Life is wonderful! I am passionate about living out Christian principles in my marriage, family, workplace and ministry"	"The power of working together as a team brings"		"Text, product, font"
Conscientiousness	"Debugging is twice as hard as writing the code in the first place. Therefore, if you write the code as cleverly as possible, you are, by definition, not smart enough to debug it—Brian Kernighan"	"Event, presentation, text, font"	"Core data and marcudzarra at"	Not any	"Face hair facial hair beard white moustache chin head hairstyle forehead"	"Director of Technology at Lift Interactive. From business models to design to development to growth. A generalist with passion to be better at the breadth"	Not any		"Drawing"

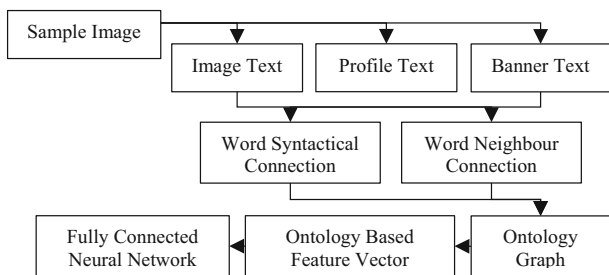


Fig. 2 Block diagram of the proposed method

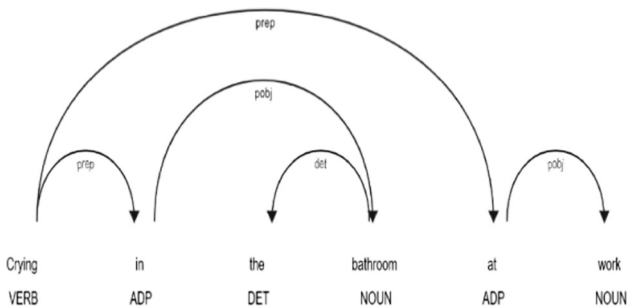


Fig. 3 Diagram of syntactical dependency tree

### 3.1 Forming word connections

This step considers two types of word connections, namely syntactical and spatial. Inspired by Tian et al. [22], word syntactical connections as well as word neighboring connections are considered for representation. Word syntactical connections are calculated by dependency parsing texts, where all the word dependencies are given equal weightage. The process is illustrated in Fig. 3 for the text recognition of the sample of the first image of the Neuroticism class shown in Fig. 1e. For the statement in the image, “Crying in the bathroom at work,” the syntactical dependency tree is constructed as shown in Fig. 3.

Here ADP, DET, Prep, Pobj stand for Adjectival Complement, Determiner, Prepositional Modifier and Object of a Preposition, respectively. Syntactic relation word pairs are (in, Crying), (the, bathroom), (bathroom, in), (at, Crying) and (work, at). The edges connecting these syntactical word pairs have one weight added to them. Initially, all syntactical connections (SyC) are set to zero as defined in Eq. (1).

$$SyC(w_i, w_j) = 0 \quad \forall w_i, w_j \in \text{Vocabulary} \quad (1)$$

When a new syntactical connection is discovered in a statement, it is updated incrementally as defined in Eq. (2).

$$SyC(w_i, w_j) = SyC(w_i, w_j) + 1 \quad (2)$$

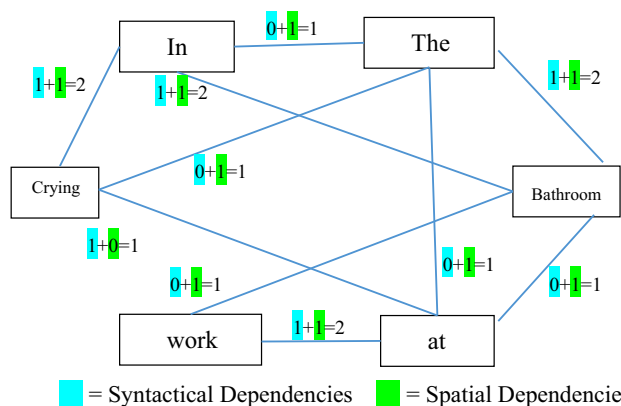


Fig. 4 Ontology weighted undirected graph

To obtain neighboring word connections, the proposed approach performs a sliding window operation, which outputs context words. All the neighboring words of the context word are considered for defining the relation between the words. Again, every word relation has equal weightage. The steps are illustrated for the same above statement of the images of Neuroticism class. Considering a context window of two, the spatial relation word pairs are (Crying, in), (Crying, the), (in, the), (in, bathroom), (the, bathroom), (the, at), (bathroom, at), (bathroom, work) and (at, work). The edges connecting these spatial word pairs have one weight added to them. Initially all spatial connections (SpC) are set to zero as defined in Eq. (3).

$$SpC(w_i, w_j) = 0 \quad \forall w_i, w_j \in \text{Vocabulary} \quad (3)$$

All the spatial connections that are presented in a statement are updated incrementally as defined in Eq. (4).

$$SpC(w_i, w_j) = SpC(w_i, w_j) + 1 \quad (4)$$

In this way, all the syntactical connections and spatial connections are discovered and recorded from text recognition results. The proposed method integrates all the spatial connections and syntactical connections and results in the weighted undirected graph ontology (WUG) connections as defined in Eq. (5). The final WUG can be seen in Fig. 4, where we can see spatial and syntactical relation word pairs.

$$\text{Ontology WUG}(w_i, w_j) = SyC(w_i, w_j) + SpC(w_i, w_j) \quad (5)$$

When a new word (concept) is encountered, it is considered as a new vertex and added to the current WUG. The connections are added as weights to edges to represent relations. As the process of discovering new words, forming word pairs (both syntactical and spatial), updating the ontology WUG by adding the edge weight continues, the

ontology WUG matures and better represents the unique relations among images of different classes. The ontology WUG is used for optimizing word vectors.

### 3.2 Classification of social media images of personality traits

In the previous section, ontology helps us to represent unique relations for images of different classes. However, in order to use the representation for classification, it is necessary to convert them to word vectors such that the vectors can be supplied to FCNN. In addition, to make the proposed method independent of the size of word vectors, inspired by the method [23] where min-batch gradient descent was proposed to optimize feature extraction, we explore the same and considered as an objective function as defined in Eq. (6) to optimize a set of vectors for representing words. This is done by initializing 25-dimensional vectors randomly for each word in the vocabulary and optimizing the vectors based on the ontology WUG obtained in the previous step. Thus, the essence of the ontology is extracted in each word vector.

$$\text{Obj}_{\text{Function}} = \sum_{i=0}^{\text{vocab size}} \sum_{j=0}^{\text{vocab size}} \text{Norm}(X_{ij}) \cdot ((\vec{w}_i \cdot \vec{w}_j) + b_i + b_j + X_{ij})^2 \quad (6)$$

where  $b_i$  and  $b_j$  are random initial biases of  $i$ th and  $j$ th word.  $X_{ij}$  is the weight of the edge between the  $i$ th and  $j$ th word in the ontology WUG.

$$\text{Norm}(X_{ij}) = \begin{cases} \frac{X_{ij}}{100}^{\frac{3}{4}}, & X_{ij} < 100 \\ 1, & \text{otherwise} \end{cases} \quad (7)$$

As the verbs tend to have too many syntactical connections, they are often the root of the dependency tree, and the stop words tend to have too many spatial connections. To overcome this problem, normalization is proposed as defined in Eq. (7). This process starts with random weight values and continues by updating the weights until it converges. When optimization is reached, weight values are recorded, and these are considered as ontology-based feature vectors. Since the step starts with 25D weights for all words in the vocabulary, it gives 25D vectors for each word in the vocabulary. The value of 25 is found according to our observation as it gives an optimal average classification rate in our proposed method.

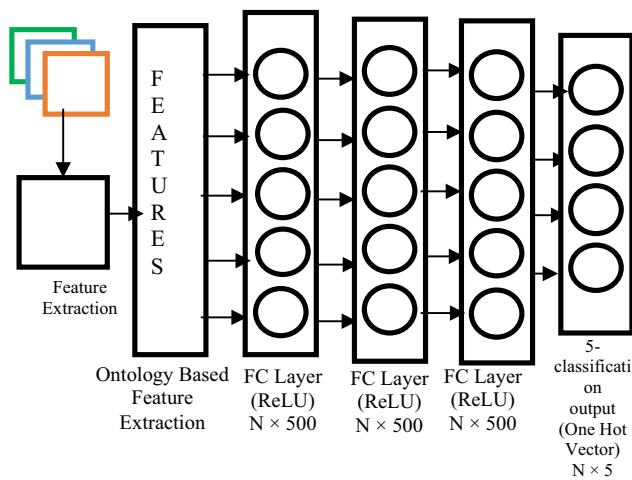
In this work, the number of unique keywords turned out to be  $\sim 20,000$  words which is the vocabulary size of the corpus. The number of words recognized in each image is variable. We need a constant input size for a neural network. Thus, we took the average of number of keywords

per image which was found to be  $\sim 400$  keywords, average number of keywords per profile and banner image, which was found to be around 35 keywords, average number of keywords of image-annotation, image-description, profile picture-annotation, profile picture-description and banner image-annotations was calculated to  $\sim 35$  words. Thus, we considered 400 keywords for image-text recognition and 35 keywords each for image-annotation, image-description, profile picture-text recognition, profile picture-annotation, profile picture-description, banner image-text recognition and banner image-annotations. Thus, in total, the number of keywords for each sample is  $400 * 1 + 35 * 7 = 645$ . Then for each word, the proposed method obtains 25D vectors. The total dimensionality of the semantic features is  $645 * 25 = 16,125$ .

Inspired by the strong discriminative power of FCNN, we propose to use it for classification considering the feature matrix as input. The details of the proposed FCNN are as follows. *Layer 1*: Takes in the final feature vector of 16,125 dimensions as input and outputs to the next hidden layer with 500 units. It has rectified linear activation units (ReLU). *Layer 2*: Takes in 500 inputs from Layer 1 and outputs to the next hidden layer with 500 units. It also has rectified linear activation units (ReLU). *Layer 3*: Takes in 500 inputs from Layer 2 and outputs to the next hidden layer with 500 units. It also has rectified linear activation units (ReLU). *Layer 4*: This is the output layer. It takes in 500 inputs from Layer 3 and outputs 5 values (as we have 5 classes). There is no activation function in this layer. To calculate the loss, we use mean squared error. The details of the hyperparameters that we have used in our proposed method are learning rate 1 = 0.001, learning rate 2 = 0.0001, batch size = 128, number of epochs = 200, train test split = 20%. We have used 200 epochs for training on the sample data. For the first 100 epochs we use learning rate 1 and for the subsequent 100 epochs we use learning rate 2. The architecture of FCNN can be seen in Fig. 5.

## 4 Experimental results

For validating the proposed method, we consider the dataset used in [24], where details of Twitter accounts of users, such as images uploaded by users, profile pictures, banner images and descriptions along with labels of different personality traits, are available. Since it is a huge dataset and many images do not have text information, we choose a subset of the dataset used in [24], which includes images containing text information. This dataset is considered as our dataset for experimentation. For extracting labels and other details, Tweepy API is used. For each class, we choose 1000 images. In total, our dataset provides 5000 images for five classes. The ground truth and labels for each image are available in

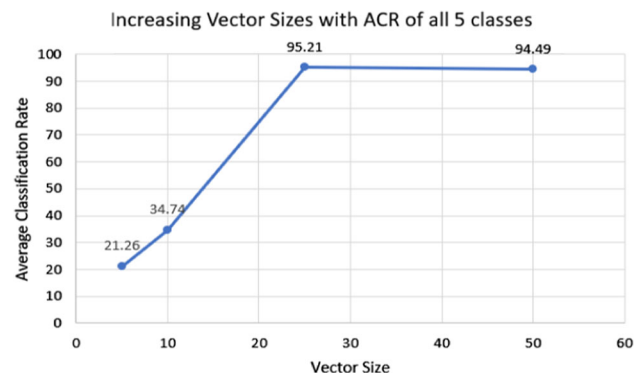


**Fig. 5** Architecture of fully connected neural network classifiers for classification

[24]. The same annotations are used for choosing the images for each class in this work.

To test the objectivity of the proposed method, we also consider two standard datasets. (i) Liu et al.'s [6] dataset (5 Class) which provides profile picture images extracted from Twitter account along with textual tweets, and it has 445 images belonging to the Agreeableness class, 5225 images of the Conscientiousness class, 1895 images of the Extraversion class, 12,722 images of the Neuroticism class and 13,269 images of the Openness class. This gives in total 33,556 images for experimentation. (ii) Krishnani et al. [19] dataset has 10 classes namely Activities (sports, action images), Extraversion (socializing, partying), Family (family photos), Fashion (fashion shows), Selfies (self-portraits), Bullying (harmful traits by groups), Depression (depressing), Neuroticism Sarcastic (sad feeling portrayed in irony), Psychopath (mentally disturbed) and Threatening (aggressive traits). Each class contains 200 images, and in total, the dataset provides 2000 images for experimentation.

To demonstrate that the proposed method is superior to the existing ones, we implemented the following existing methods for comparative study. Mungra et al. [16] proposed a method for emotion recognition using facial expressions. Krishnani et al. [19] proposed a method for classifying behavior-oriented social media images based on structural function transforms. The reason to consider the above existing methods for comparative study is that the objective of the methods is the same as that of the proposed method. In addition, to show that the method [16] developed for the classification of emotions using facial information may not work well for the classification of social media images that represent personality traits. Similarly, the method [19] developed for images of normal and abnormal behavior of the person with only image features may not be effective for the



**Fig. 6** Determining the feasible value for the vector size using our dataset

classification of personality oriented social media images. For measuring the performance of the proposed and existing methods, we calculated the Average Classification Rate (ACR), which is the mean of the diagonal elements in the confusion matrix. For all the experiments of the proposed and existing methods on our and two standard datasets, we use a 70:30 split for training and testing.

In the proposed methodology, the size of the feature vector is considered as 25 dimensions. This value is determined empirically by calculating the average classification rate for different sizes of feature vectors as shown in Fig. 6. It is also observed from Fig. 6 that the average classification rate is the highest for 25 dimensions of the feature vector, and the average classification rate slowly decreases as the size of the vector increases. Therefore, 25 is considered as an optimal value for all the experiments. For this experiment, we consider random samples across classes of our dataset.

#### 4.1 Ablation study

In the proposed methodology, to achieve the best classification results, the key features, namely features extracted from the images, profile picture, banner image, text in the images and description of the images are extracted. In addition, the above features are used for deriving syntactic and spatial features using ontology concepts for the classification of personality traits oriented social media images. To verify the effectiveness of the above features, we conduct the following experiments on our dataset and the Average Classification Rates (ACR) are reported in Table 2 for each experiment.

(i) ACR is calculated for the feature extracted from only images, which includes labels of the images, text in the images and description of the images. (ii) ACR is calculated for the feature extracted only from profile picture images (label, text and description). (iii) ACR is calculated for the extracted features only from banner images (label and text). (iv) ACR is calculated for the extracted feature from all the

**Table 2** ACR for the key features of the proposed approach in (%) on our dataset

#	Features	ACR	#	Features	ACR
(i)	Only Image	27.44	(v)	All Imagery	61.80
(ii)	Only Profile	44.12	(vi)	Only Spatial	72.16
(iii)	Only Banner	41.60	(vii)	Only Syntactical	66.75
(iv)	All Textual	66.62	(viii)	Proposed (vi)–(vii)	<b>86.44</b>

Bold indicates the best results of the method

textual data, which includes text in the images, profile picture, banner images and description of images. For this step, the proposed method combines the features of all textual information as a feature vector for classification. (v) ACR is calculated for the feature extracted from all the images, which includes labels of images, profile picture, banner images. (vi) ACR is calculated for the spatial features derived from WUG graphs for classification. (vii) ACR is calculated for syntactic features derived from WUG for the classification. (viii) ACR is calculated for combining spatial and syntactic features for the classification, which is the proposed method.

When we compare the results of (i)–(iii) reported in Table 2, profile picture contributes more for classification. This makes sense because usually users change profile pictures often according to their mind status compared to uploaded image and banner images. The results of (iv) and (v) show that textual information contributes more for classification compared to image information. This is true because text information provides exact meaning compared to images. Similarly, the results of (vi) and (vii) show that the spatial features contribute more compared to syntactic features. The reason is that due to the large size of the corpus, deriving correct syntactic using WUG is difficult, and hence, there are chances of missing the actual meaning of the text while spatial information does not.

The experiment of (viii), which is the proposed method, shows the combination of spatial and syntactical are effective in achieving high ACR for classification. Therefore, we can infer that both image and textual information are essential for addressing challenges of classification of personality traits oriented social media images. It is also noted from Table 2 that although individual features are effective, they are not capable of achieving the best results compared to the proposed method results.

## 4.2 Experimenting the proposed classification approach

Quantitative results of the proposed and existing methods [16, 19] on our dataset and two standard datasets [6, 19] are reported in Table 3, where it is noted that the proposed method is better than the two existing methods in terms of ACR for all the datasets. Since two standard datasets do not provide

**Table 3** Performance of the proposed and existing methods on different datasets and different sizes (in %)

Methods	Our 5 class dataset	Liu et al. [6] dataset	Krishnani et al. [19] dataset
	5 classes	5 classes	10 classes
Proposed	<b>86.44</b>	<b>85.88</b>	<b>61.17</b>
Krishnani et al. [19]	65.60	60.39	43.31
Mungra et al. [16]	21.40	22.73	47.52

Bold indicates the best results of the method

profile picture, banner and description information, the proposed method uses images and their labels for constructing ontology graphs for classification. From the results, it is also confirming that the proposed method is generic and independent of the number of classes and different datasets. The reason for the poor results of the existing methods compared to the proposed method is that the methods are designed for classification of emotions using faces and classification of normal and abnormal images but not social media images that represent personality traits, respectively.

On the other hand, the proposed approach explores ontology which gives high-level features and integrates image information with text information to achieve the best results. However, it is noted from Table 3 that for a dataset of 10 classes, all the methods including the method [19] report lower results compared to our and Liu et al.'s datasets. This is because of the dataset [19] which does not provide text information as the proposed method expects. In the same way, although the method [19] is developed for classifying the images of the same dataset [19], it reports lower results than the other two datasets. This is because the main objective of the method is to classify normal and abnormal images (two classes but not for 10 classes).

## 5 Conclusion and future work

In this work, we have proposed a novel idea of ontology for classification of personality traits oriented social media



images of Big Five factors (BFF), namely Agreeableness, Conscientiousness, Extraversion, Openness and Neuroticism. The ontology graph is constructed using image labels, profile picture labels, banner image labels, recognition results of text in images and finally description of image information. GOOGLE Cloud Vision API has been used for obtaining the labels and recognition results for the text in images. The proposed method combines images as well as text information as multimodal concept through ontology for the classification with the help of fully connected neural network. Experiments on our, two standard datasets and comparative studies with the existing methods show that the proposed method is effective and superior to the existing methods in terms of average classification rate. However, for images that are affected by distortion and large variations on content, the performance of the proposed method degrades, which is beyond the scope of the present work.

**Acknowledgements** This work receives partial support from FRGS Grant (FP104-2020), Ministry of Higher Education, Malaysia.

## References

- Zhu, H., Li, L., Zhao, S., Jiang, H.: Evaluating attributed personality traits from scene perception probability. *Pattern Recognit. Lett.* **116**, 121–126 (2018)
- Raj, C., Meel, P.: ConvNet frameworks for multi-modal fake news detection. *Appl. Intell.* **51**, 8132–8148 (2021)
- Chaudhari, K., Thakkar, A.: Survey on handwriting-based personality trait identification. *Expert Syst. Appl.* **124**, 282–308 (2019)
- Antonakaki, D., Fragopoulou, P., Ioannidis, S.: A survey of Twitter research: data model, graph structure, sentiment analysis and attacks. *Expert Syst. Appl.* **164**, 114006 (2021)
- Rahman, M.A., Al Faisal, A., Khanam, T., Amjad, M., Siddik, M.S.: Personality detection from text using convolutional neural network. In: *Proceedings of the ICASERT* (2019)
- Liu, L., Preotiuc-Pietro, D., Samani, Z.R., Moghaddam, M.E., Ungar, L.: Analyzing personality through social media profile picture choice. In: *Proceedings of the AAAI* (2016)
- Sharma, M., Jalal, A.S., Khan, A.: Emotion recognition using facial expression by fusing key points descriptor and texture features. *Multimed. Tools Appl.* **78**, 16195–16219 (2019)
- Ma, H., Celik, T., Li, H.C.: Lightweight attention convolutional neural network through network slimming for robust faical expression recontoin. *Signal Image Video Process.* **15**, 1507–1515 (2021)
- Garaawi, N.A., Wu, O., Morris, T.: BRIEF-based face descriptor: an application to automatic faical expression recontion (AEER). *Signal, Video Image Process.* **15**, 371–379 (2021)
- Altuwairqi, K., Jarraya, S.K., Allinjaw, A., Hammami, M.: Student behavior analysis to measure engagement levels online learning environments. *Signal Video Image Process.* **15**, 1387–1395 (2021)
- Amraee, S., Vafei, A., Jamshidi, K., Adibi, P.: Abnormal event detection in crowded scenes using one-class SVM. *Signal Video Image Process.* **12**, 115–123 (2018)
- Mabrouk, O., Hlaoua, L., Omri, M.N.: Exploiting ontology information in fuzzy SVM social media profile classification. *Appl. Intell.* **51**, 3757–3774 (2021)
- Patil, V., Mathur, H.: A survey: machine learning approach for personality analysis and writer identification through handwriting. In: *Proceedings of the ICICT*, pp. 1–5 (2020). <https://doi.org/10.1109/ICICT48043.2020.9112449>
- Bargshady, G., Zhou, X., Deo, R.C., Soar, J., Whittaker, F., Wang, H.: Enhanced deep learning algorithm development to detect pain intensity from facial expression images. *Expert Syst. Appl.* **149**, 113305 (2020)
- Bozorgtabar, B., Mahapatra, D., Thiran, J.P.: ExprADA: adversarial domain adaptation for facial expression analysis. *Pattern Recognit.* **100**, 107111 (2020)
- Mungra, D., Agrawal, A., Sharma, P., Tanwar, S., Obaidat, M.S.: PRATIT: a CNN-based emotion recognition system using histogram equalization and data augmentation. *Multimed. Tools Appl.* **79**, 2285–2307 (2020)
- Palhano, D.B., Dos Machado, L.S., De Almeida, A.A.F.: Paki Mirabolandia: a serious game to identify player personality. In: *Proceedings of the ICCTD* (2020)
- Lai, S., Sun, B., Wu, F., Xiao, R.: Automatic personality identification using students' online learning behavior. *IEEE Trans. Learn. Technol.* **13**, 26–37 (2020)
- Krishnani, D., Shivakumara, P., Lu, T., Pal, U., Ramachandra, R.: Structure function based transform features for behavior-oriented social media image classification. In: Palaiahnakote, S., di Baja, G.S., Wang, L., Yan, W.Q. (eds.) *Pattern Recognition*, pp. 594–608. Springer, Cham (2020)
- Xue, D., Wu, L., Hong, Z., Guo, S., Gao, L., Wu, Z., Zhong, X., Sun, J.: Deep learning-based personality recognition from text posts of online social network. *Appl. Intell.* **48**, 4232–4246 (2018)
- Xue, X., Feng, J., Sun, X.: Semantic-enhanced sequential modeling for personality trait recognition from texts. *Appl. Intell.* **51**, 7705–7717 (2021)
- Tian, C., Rong, W., Zhou, S., Zhang, J., Ouyang, Y., Xiong, Z.: Texts learning word representation by jointly using neighbor and syntactic contexts. *Neurocomputing* **456**, 136–146 (2021)
- Pennington, J., Socher, R., Manning, C.D.: GloVe: global vectors for word representation. In: *Proceedings of the EMNLP* (2014)
- Guntuku, S.C., Lin, W., Carpenter, J., Ng, W.K., Ungar, L.H., Preotiuc-Pietro, D.: Studying personality through the content of posted and liked images on Twitter. In: *Proceedings of the ACM on Web Science Conference*, pp. 223–227 (2017). <https://doi.org/10.1145/3091478.3091522>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.