Recent trends in deep learning based personality detection

Yash Mehta[1] · Navonil Majumder[2] · Alexander Gelbukh[2] · Erik Cambria[3]

Abstract

Recently, the automatic prediction of personality traits has received a lot of attention. Specifically, personality trait prediction from multimodal data has emerged as a hot topic within the field of affective computing. In this paper, we review significant machine learning models which have been employed for personality detection, with an emphasis on deep learning-based methods. This review paper provides an overview of the most popular approaches to automated personality detection, various computational datasets, its industrial applications, and state-of-the-art machine learning models for personality detection with specific focus on multimodal approaches. Personality detection is a very broad and diverse topic: this survey only focuses on computational approaches and leaves out psychological studies on personality detection.

Keywords Personality detection · Multimodal interaction · Deep learning

1 Introduction

Personality is the combination of an individual's behavior, emotion, motivation, and characteristics of their thought patterns. Our personality has a great impact on our lives, affecting our life choices, well-being, health along with our preferences and desires. Hence, the ability to automatically detect one's personality traits has many important practical applications. The Woodworth Psychoneurotic Inventory (Papurt 1930) is commonly cited as the first personality test. It was developed during World War I for the U.S. military to screen recruits for shell shock risks (post-traumatic stress disorders). In the present day, a personality model which has an active community and is used by many is the Process Communication Model (PCM). It was developed by Taibi Kahler with NASA funding and was initially used to assist with shuttle astronaut selection. This measure is now being used mainly for consultancy purposes in order to help individuals to become more effective in communication and even avoid or resolve situations where communication has gone off track.

✉ Erik Cambria
cambria@ntu.edu.sg

1   Gatsby Computational Neuroscience Unit, University College London, London, UK

2   Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico City, Mexico

3   School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

Instead of directly determining an individual's personality, one may also be interested in knowing how they are perceived by the people around them. Unlike the case of automatic personality recognition, the target of perceived personality analysis is not the true personality of individuals, but the personality which is attributed to them by the people who interact with them. The surrounding people fill a similar personality questionnaire on the individual, which then determines the perceived personality of that person. The perceived personality is evaluated using the same personality measures used for their 'true' personality (e.g., Big-Five). A point to be noted is that the perceived personality varies from person to person and is subject to their mental make or their own personality. Many studies have shown that people form a first impression about a person by their facial characteristics within the first 10 s of meeting an individual. Apart from the physical characteristics of the encountered face, the emotional expressions, the ambient information, the psychological formation of the viewer also affect these impressions. First impressions influence the behavior of people towards a newly encountered person or a human-like agent.

1.1 Personality measures

Modern trait theory (Pervin and John 1999) tries to model personality by setting of a number of classification dimensions (usually following a lexical approach) and constructing a questionnaire to measure them (Matthews et al. 2003). Researchers have used various schemes for personality modeling such as 16PF (Cattell and Mead 2008), EPQ-R (Miles and Hempel 2004) and three trait personality model PEN (Eysenck 2012) where there are superfactors Psychoticism, Extraversion, and Neuroticism (PEN) at the top of the hierarchy. The Myers-Briggs Type Indicator (MBTI) (Briggs-Myers 1993) is one of the most widely administered personality test in the world, given millions of times a year to employees in thousands of companies. The MBTI personality measure categorizes people into two categories in each four dimensions: introversion versus extraversion, sensing versus intuiting, thinking versus feeling, and judging versus perceiving.

The most popular measure used in the literature on automated personality detection is by far the Big-Five personality traits (Digman 1990), which are the following binary (yes/no) values:

- *Extraversion (EXT)* Is the person outgoing, talkative, and energetic versus reserved and solitary?
- *Neuroticism (NEU)* Is the person sensitive and nervous versus secure and confident?
- *Agreeableness (AGR)* Is the person trustworthy, straightforward, generous, and modest versus unreliable, complicated, meager and boastful?
- *Conscientiousness (CON)* Is the person efficient and organized versus sloppy and careless?
- *Openness (OPN)* Is the person inventive and curious versus dogmatic and cautious?

As majority of the research makes use of the Big-Five personality measure for classification, we will refer to this measure by default, unless stated otherwise.

1.2 Applications

There are various diverse industrial applications of automated personality recognition systems in the present day scenario. We surmise that a huge market will open up in the near future and if models are able to measure personality accurately and consistently, there will be a

huge demand for automated personality recognition software in the industry. As research progresses in this field, soon better personality prediction models with much higher accuracy and reliability will be discovered. Artificial personality can be integrated with almost all human computer interactions going forward.

Any computational devices can be equipped with some sort of personality which enables it to react differently to different people and situations. For example, a phone can have different modes with different configurable personalities. This will pave way for more interesting and personalized interactions.

Another possibility is to use personality traits as one of the inputs for achieving higher accuracy in other tasks such as sarcasm detection, lie detection or word polarity disambiguation systems.

- *Enhanced personal assistants* Present day automated voice assistants such as Siri, Google Assistant, Alexa, etc. can be made to automatically detect personality of the user and, hence, give customized responses. Also, the voice assistants can be programmed in such a way that they display different personalities based on the user personality for higher user satisfaction.
- *Recommendation systems* People that share a particular personality type may have similar interests and hobbies. The products and services that are recommended to a person should be those that have been positively evaluated by other users with a similar personality type. For example, Yin et al. (2018) propose to model the automobile purchase intentions of customers based on their hobbies and personality. Yang and Huang (2019) have developed a system for recommending games to players based on personality which is modelled from their chats with other players.
- *Word polarity detection* Personality detection is a subtask of sentiment analysis (Cambria et al. 2017) as it can be exploited for word polarity disambiguation in sentiment lexica (where the same concept can convey different emotions to different types of people) and for disambiguation between sarcastic and non-sarcastic content (Majumder et al. 2019).
- *Specialized health care and counseling* As of 2016, nearly one-third of Americans have sought professional counseling for mental health related issues. This is yet another area where huge practical applications of personality trait prediction exists. According to an individual's personality, appropriate automated counseling may be given or a psychiatrist may make use this information to give better counseling advice.
- *Forensics* If the police are aware of the personality traits of the people who were present at the crime scene, it may help in reducing the circle of suspects. Personality detection also helps in the field of automated deception detection and can help in building lie detectors with higher accuracy.
- *Job screening* In human resource management, personality traits affect one's suitability for certain jobs. For example, a company may need to recruit someone who will motivate and lead a particular team. They can narrow down their screening by eliminating candidates who are highly nervous and sensitive, i.e., those having high values of neuroticism trait. Liem et al. (2018) discusses the job candidate screening problem from an interdisciplinary viewpoint of psychologists and machine learning scientists.
- *Psychological studies* Automated personality trait detection will help find more complex and subtle relations between people's behaviorism and personality traits. This will aid in discovering new dynamics of the human psyche.
- *Political forecasting* Large scale automated personality detection is being used as a guideline for politicians to come up with more effective and targeted campaigns. If an analytical firm is able to procure large scale behavioral data about the voters, the firm can then create

their psychographical profiles. These profiles can give an insight to the kind of advertisement that would be most effective in persuading a particular person at a particular location for some political event.

1.3 Fairness and ethics

Over the past century, various techniques to access personality have been developed and refined by the practitioners and researchers in the field. Several ethical guidelines have been set to ensure that proper procedures are followed. However, often, personality detection involves ethical dilemmas regarding appropriate utilization and interpretations (Mukherjee and Kumar 2016). Concerns have been raised regarding the inappropriate use of these tests with respect to invasion of privacy, cultural bias and confidentiality.

Recently, political parties are trying to leverage large-scale machine learning based personality detection for political forecasting. A rather infamous example showing the use of personality detection in political forecasting is the Facebook's Cambridge Analytica data scandal,[1] which involved the collection of personally identifiable information of 87 million Facebook users. Studies have shown that the political choices of voters have a strong correlation with their social characteristics (Caprara et al. 2006). Users' social data were allegedly used in an attempt to influence voter opinion in favor of the 2016 Trump election campaign by identifying persuasive advertisement types effective for a particular region. It also involved automatically accessing personality traits of voters and using this to predict who would be manipulated easily by political propaganda.

There is also concern regarding the continued use of these tests despite them lacking valid proof of being accurate measures of personality. These tests may have inherent biases, which are then inadvertently learned by the machine learning algorithm. One step to reduce biases in machine learning algorithms is through Algorithmic Impact Assessments (AIAs), as proposed by New York University's AI Now Institute. AIAs extend from the idea that the 'black box' methodology leads to a vicious cycle, continuously moving further away from understanding these algorithms and diminishing the ability to address any issues that may arise from them. AI Now suggests the use of AIAs to handle the use of machine learning in the public realm, creating a set of standard requirements. Through AIAs, AI Now aims to provide clarity to the public by publicly listing and explaining algorithm systems used while allowing the public to dispute these systems, developing an audit and assessment processes, and increasing public agencies' internal capabilities to understand the systems they use. Similar to the use of AIAs to promote transparency in machine learning, the Defense Advanced Research Projects Agency (DARPA) suggests Explainable Artificial Intelligence (XAI) as a part of the solution. The goal is to produce more explainable models that users can understand and trust (Table 1).

1.4 Organization

We have thoroughly reviewed machine learning-based personality detection, with focus on deep learning-based approaches. Section 2 gives the reader an idea of other review papers on machine learning-based automated personality detection. We have divided our analysis based on the modality of the input, such as text, audio, video, or multimodal. Section 3 briefly discusses the popular techniques and methods used for personality detection in each

---

[1] The Great Hack, a recent documentary about the Cambridge Analytica data scandal

Table 1 The various popular tools used for feature extraction for each of the modalities

Modality	Tool	Features extracted
Text	LIWC	Psychological and linguistic features
	Receptiviti API	Built on top of LIWC. Visual representation of personality traits
	Freeling	POS tagging
	SenticNet	Sentiment polarity
Audio	Praat	Intensity, pitch, loudness, MFCC, jitter, shimmer, LSP
	OpenSMILE	
Visual	FACS	Facial expressions, AUs
	CERT	8 facial expressions (e.g., joy, sadness, surprise, etc.)
	Face++	Face detection, 106 facial landmarks, face attributes (e.g., gender, age)
	EmoVu	Face detection, emotion recognition (with intensity), face attributes

of these modalities. In Sect. 4, we analyze multiple interesting papers in-depth in each of these modalities. Results with analysis and conclusions follow in Sects. 5 and 6, respectively.

2 Related works

Automated personality detection is a new and upcoming field. There have not been many comprehensive literature surveys done in personality detection and our paper is the first one which gives the reader a bird's-eye view of the recent trends and developments in the field.

There is no recent work which gives the reader an overall perspective of the advances in machine learning based automated personality detection. Section 3 gives the popular machine learning models for each modality, which gives the reader a wide perspective in this field.

A review of the trends in personality detection from text using shallow learning techniques such as Naïve Bayes, kNN, mLR, Gaussian Process is given by Agarwal (2014). Post 2014, end-to-end deep neural network architectures and models picked up and started beating the state-of-the-art accuracy of these models. A review of the the various image processing techniques and facial feature analysis for personality detection is given by Ilmini and Fernando (2016), in which their focus lies on an overview of machine learning techniques along with various facial image preprocessing techniques.

Popular image preprocessing techniques in this domain include facial landmark identification, Facial Action Coding System (FACS), which gives the AUs in the face, and the Viola Jones face detection algorithm (Viola and Jones 2004). Also, instead of using a standard measure of personality (such as Big-Five or MBTI), they have used slightly different metrics such as unhappy, weird, intelligent, confident, etc. This survey only focuses on the visual aspect whereas Vinciarelli and Mohammadi (2014) gives a very systematic and clear survey of personality detection from text as well as audio. Instead of dividing personality detection by modality as it is done in this paper, they have chosen a broader topic which they call personality computing. This is further divided into 3 fundamental problems, namely Automatic Personality Perception (APP), Automatic Personality Synthesis (APS) and Automatic Personality Recognition (APR). This paper focuses on methods following APR and APP (similar to perceived personality). The paper cited above gives a very detailed insight of the different papers dealing with recognition and perception of personality from social media, mainly Twitter and Facebook. They present a thorough analysis of personality recognition

from text, non-verbal communication (e.g., interpersonal distances, speech and body movements), social media, mobiles, wearable devices and finally from computer games as well. A recent paper, Junior et al. (2018) provides a comprehensive survey on computer vision-based perceived personality trait analysis. Escalera et al. (2018) have made a compilation of the latest progress on automatic analysis of apparent personality in videos and images from multimedia information processing, pattern recognition and computer vision points of view. However, there are still several deep learning techniques, models and papers, especially for multimodal personality detection which have not been covered by any literature survey till date (Table 2).

3 Baseline methods

The following sections summarize the various popular models, architectures and techniques commonly used in deep learning-based personality detection. Table 3 provides a brief insight into the methods described by some of the pioneering papers in the field.

3.1 Text

Pertaining to the textual modality, data preprocessing is a very important step and choosing the correct technique can yield significantly better results. Usually, features from text are extracted such as Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al. 2001), Mairesse, Medical Research Council (MRC), etc. which are then fed into standard ML classifiers such as Sequential Minimum Optimizer, Support Vector Machine (SVM), Naïve Bayes, etc. Learning word embeddings and representing them as vectors (with GloVe or Word2Vec) is also a very commonly followed approach. These word vectors may also be created by feeding the word character-wise into a Long Short-Term Memory (LSTM) or a Gated Recurrent Unit (GRU). It was observed that combining text features (LIWC, MRC) with something else such as commonsense knowledge, convolutions, etc. results in better performance.

3.2 Audio

There are relatively few methods which focus on using only audio as the sole input for detecting personality. It is usually combined with visual modality for bimodal personality detection. Standard audio features such Mel-Frequency Cepstral Coefficients (MFCC), Zero Crossing Rate (ZCR), Logfbank, other cepstral and spectral ones serve as inputs into SVM and linear regressors.

3.3 Visual

As in the case of most visual deep learning tasks, Convolutional Neural Networks (CNN) are the most commonly used and yield state-of-the-art results in the field of personality detection as well. Most approaches analyze facial features and try to find a function which maps these features to personality traits. Many researchers have used pretrained deep CNNs (such as VGG-Face) and fine tuned it for the task of personality detection (transfer learning). They have experimented with different ways of extracting facial features such as EigenFace,

Table 2 Popular datasets, divided based on mode of input

Modality	Dataset	Personality measure	Description
Text	Essays I (Pennebaker and King 1999)[c]	Big-Five	2468 anonymous essays tagged with the author's personality traits. Stream-of-consciousness essays written by volunteers in a controlled environment and the authors of the essays were asked to label their own Big-Five personality traits
	Essays II (Tausczik and Pennebaker 2010	Big-Five	2400 essays were labelled manually with personality scores for five different personality traits. The data were then tuned and the regression scores were converted into class labels of five different traits
	MBTI Kaggle[a]	MBTI	Contains the MBTI personality labels for over 8600 people, along with 50 of their posts on the myPersonality cafe forum
	MyPersonality[b]	Big-Five	myPersonality was a Facebook App that allowed its users to participate in psychological research by filling in a personality questionnaire. However, from 2018 onwards they decided to stop sharing the data with other scholars
	Italian FriendsFeed (Celli et al. 2010)	Big-Five	Sample of 748 Italian FriendFeed users (1065 posts). The dataset has been collected from FriendFeed public URL, where new posts are publicly available
Audio	AMI Meeting Corpus (Carletta et al. 2005)	Big-Five	Video and audio recordings of monologues, dialogues and multi-party discussions with annotations to perceived personality which is found by the BFI-10 questionnaire
	Aurora2 corpus (Hirsch and Pearce 2000)	–	Connected digit corpus which contains 8440 sentences of clean and multi-condition training data and 70,070 sentences of clean and noisy test data
	CMU self-recorded database (Polzehl et al. 2010)	Big-Five	Experimental database with a professional speaker was recorded and generated the Big-Five factor scores for the recordings by conducting listening test using the NEO-FFI personality inventory
	Columbia deception corpus (Levitan et al. 2015)	–	Balanced corpus includes data from 126 (previously unacquainted) subject pairs, constituting 93.8 h of speech in English
Visual	PhychoFlikr dataset (Cristani et al. 2013)	Big-Five	Top 200 images favorited by 300 Flikr users (hence totally 60,000 images) labelled with their personality traits
	First Impressions V2 (CVPR'17)[c]	Big-Five	The first impressions data set, comprises 10,000 clips (average duration 15 s) extracted from more than 3000 different YouTube high-definition videos of people facing and speaking in English to a camera

[a] https://www.kaggle.com/datasnaek/mbti-type/
[b] https://sites.google.com/michalkosinski.com/mypersonality
[c] http://chalearnlap.cvc.uab.es/dataset/24/description/

Table 3 Brief description of the important personality-detection models

Modality	References	Architecture
Text	Majumder et al. (2017)	1D convolutions to extract n-grams combined with Mairesse features
	Poria et al. (2013)	Combined LIWC with MRC features extracted from text along with commonsense knowledge using sentic computing techniques
	Hernandez and Scott (2017)	GloVe embedding fed into an deep RNN (LSTM)
	Celli (2012)	Used a subset of Mairesse features along with unsupervised learning methods to measure correlation between features and personality traits
Audio	Valente et al. (2012)	Audio features such as prosodic features, speech activity features, word n-grams and dialogue act tags were extracted and fed into simple ML classifiers
	Polzehl et al. (2010)	Cepstral features of speech such as MFCC, Zero Crossing Rate (ZCR), intensity, pitch, loudness,formants were fed into a SVM regressor
	Levitan et al. (2016)	Combination of LIWC and prosodic features fed into ML classifiers such as AdaBoost and RandomForest
Visual	Rojas et al. (2011)	Histogram of Oriented Gradients (HOG) for modeling the face, EigenFace and specific points on the face were taken and fed into SVM, Gradient Boosting, bTree
	Biel (2012)	CERT used for extracting facial expressions followed by thresholding as well as Hidden Markov Models
	Gurpinar et al. (2016a)	Transfer Learning, pre-trained VGG-Face and VGG-19 to extract facial as well as background features followed by regularized regression with a kernel ELM classifier
	Al Moubayed et al. (2014)	Polynomial and Radial Basis Function (RBF) kernels for SVMs
Multimodal	Gucluturk et al. (2017)	Facial features extracted by VGG-Face, Scene features by VGG-VD19 and combined with audio features using Random Forest based score level fusion
	Kindiroglu et al. (2017)	Multi task learning of learning leadership and extraversion simultaneously

Histogram of Oriented Gradients (HOG) (Dalal and Triggs 2005), FACS (extracts AUs such as raised eyebrows, dimples, etc.) and Viola Jones Algorithm (face identification) to achieve higher accuracy.

3.4 Multimodal

Most of the multimodal approaches perform late fusion, that is, they take the average of individual results of predictions from the audio and visual modalities. Deep bimodal regression give state-of-the-art results by making use of slightly modified Deep Residual Networks (He 2016). The features extracted from each of the modalities may be used together to come up with the personality prediction. This technique is early fusion. Present research in the field takes the direction of finding more efficient ways of feature extraction along with multi-

modal feature combination. We see very few models which have dealt with trimodal fusion of features.

4 Detailed overview

In this section, we analyze individual papers in detail, discussing various techniques, approaches and methods. On average, we find that features extracted from the visual modality are most accurate in unimodal personality detection. Studies have found that combining inputs from more than one modality often results in a higher prediction accuracy, as one may expect.

4.1 Text

In the last decade, numerous studies have linked language use to a wide range of psychological correlates (Park et al. 2015; Ireland and Pennebaker 2010). Reliable correlations of writing style (e.g., frequency of word use) with personality were found by some of the earliest works (Pennebaker and King 1999). For example, individuals scoring higher on extraversion used more positive emotion words (e.g., great, amazing, happy) whereas those higher in neuroticism were found to use first-person singulars (e.g., I, mine, me) more frequently. Personality detection using these type of differences in the use of language is referred to as the *closed vocabulary* approach.

One of the most common closed-vocabulary methods for personality detection from text (especially social media) is LIWC. It categorizes the words into various psychologically relevant buckets like 'function words'(e.g., articles, conjunctions, pronouns), 'affective processes' (e.g., happy, nervous, cried) and 'social processes' (e.g., mate, talk, friend). It then counts the frequency of words in each of the buckets (over 60 psychologically relevant buckets) and then predicts the personality of the writer of the text. As this approach starts with predefined categories of words, it has been described as 'closed-vocabulary' and most of the features extracted by LIWC are language dependent.

Receptiviti API (Golbeck 2016) is a tool built on LIWC which used for personality analysis from text. A user's personality prediction is made using psycholinguistic features. The Receptiviti API allows users to submit a text sample which is analyzed and outputs a graphical representation of the predicted traits which can also be viewed on a web-based interface. However, this API does not perform well on social media derived text as closed vocabulary methods have lower accuracy while dealing with micro-text in general.

Personality traits may even be predicted from different sources of interpersonal communication (e.g., WhatsApp) on a mobile device. Saez et al. (2014) have built an application which is in charge of compiling and sending information about the user to a server application, which then runs the classification algorithms. One line of research this paper follows is checking whether personality is linked to happiness of a person, however it didn't result in anything conclusive. This is one of the few research papers which uses the PEN theory for classification instead of the Big-Five. Standard machine learning algorithms such as J48, Random Forest and SVM were tested.

In the recent years, personality detection from social media, especially Twitter sentiment analysis has gained a lot of popularity. This might be partially due to the fact that Twitter data collection is direct and easily accessible through the Twitter API. Kalghatgi et al. (2015) have used a combination of social behavior of the user (average number of links, average number
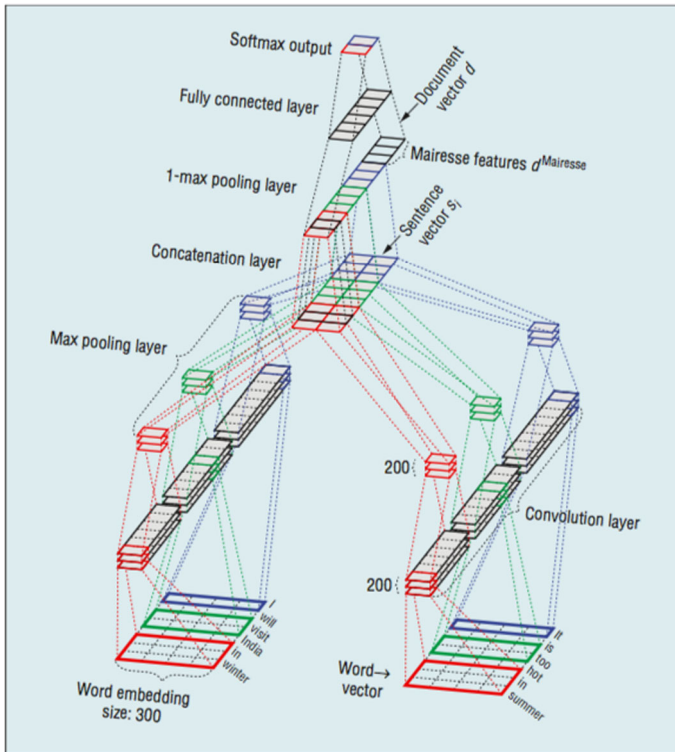
Fig. 1 CNN framework used by Majumder et al. (2017) makes use of Word2Vec embeddings for personality detection from text

of hash tags, average number of mentions, etc.) and grammatical information (average length of text, average number of positive and negative words, average number of special characters like comma, question mark, etc.) for user personality detection. Using these attributes, a feature vector is constructed which is then fed into a Multi-Layer Perceptron (MLP). The novelty of this approach lies in running the whole set up on the Hadoop framework (Hadoop Distributed File System and Map Reduce) which enables personality prediction of N users concurrently.

Apart from social media data, essays are also a popular mode of text and can be used for author profiling. A popular dataset is the James Pennebaker and Laura King's stream-of-consciousness essay dataset (Pennebaker and King 1999). It contains 2468 anonymous essays tagged with the authors' personality based on the Big-Five traits. Majumder et al. (2017) makes use of a deep CNN for document level personality detection tasks. The CNN helps in extracting the monogram, bigram and trigram features from the text and the architecture of this model is shown in Fig. 1. It is observed that the removal of neutral sentences gives a marked improvement in the prediction accuracy. Each word is represented in the input as a fixed-length feature vector using Word2Vec (Mikolov et al. 2013), and sentences are represented as a variable number of word vectors. In the end, document level Mairesse features (LIWC, MRC, etc.), totally 84 features, are concatenated with the feature vector extracted from the deep CNN. Lastly, this concatenated vector is then fed into a fully connected layer for the final personality trait predictions. Hernandez and Scott (2017) have tried modeling

temporal dependencies amongst sentences by feeding the input to Recurrent Neural Networks (RNNs). LSTMs were found to give better results compared to vanilla RNN, GRU, bi-LSTM. GloVe (Pennington et al. 2014) word embeddings of 50D were used for predicting MBTI personality of people based on 50 posts on a particular social personality discussion website called PersonalityCafe.

It was found that combining commonsense knowledge with psycho-linguistic features resulted in a remarkable improvement in the accuracy (Poria et al. 2013). In particular, authors use SenticNet (Cambria et al. 2018), a popular tool used for extracting commonsense knowledge along with associated sentiment polarity and affective labels from text. These features are used as inputs to five Sequential Minimal Optimization based supervised classifiers for the five personality traits, which are predicted by each of the classifiers independently.

While most researchers assume the five personality traits to be independent of each other, several studies (Shaver and Brennan 1992; Judge et al. 1999) claim that there exists certain correlations among the traits and it is inaccurate to build 5 completely independent classifiers. Zuo (2013) have modeled the inter-trait dependencies using a weighted ML-kNN algorithm which uses information entropy theory (Jaynes 1982) to assign weights to the extracted features. The linguistic and correspondingly dependent emotional features are extracted from text which can be then discretized on the basis of Kohonen's feature-maps algorithm (Kohonen 1990).

Instead of relying on prior word or category judgments, *open-vocabulary* methods rely on extracting a comprehensive collection of language features from text. These methods characterize a sample text by the relative use of non-word symbols (e.g., punctuation, emoticons), single uncategorized words, multi-word phrases and clusters of semantically related words identified through unsupervised methods (topics). Typically, Latent Dirichlet Allocation (LDA) (Blei et al. 2003) and its sentiment extensions (Poria et al. 2016) are used for forming clusters of semantically related words.

MyPersonality dataset is one of the famous textual social media datasets which comprises of status updates of over 66,000 Facebook users. Each of these volunteers also completed the Big-Five personality test. Park et al. (2015) use this dataset for automatic personality detection from the language used in social media. Firstly, features are extracted using standard techniques for text. Then, the dimensionality of these features is reduced followed by Ridge Regression (Hoerl and Kennard 1970). There is an individual regressor for prediction of each of the five personality traits. An end to end deep learning model was trained on a subset of this dataset (Facebook user status) where the authors Yu and Markov (2017) experimented with n-grams (extracted with CNN) and bi-directional RNNs (forward and backward GRU). Instead of using the pre-trained Word2Vec model, they trained the word embedding matrix using the skip-gram method (Mikolov et al. 2013) to incorporate internet slang, emoticons and acronyms. To improve accuracy, they perform late fusion of non-lexical features extracted from text and the penultimate layer of the neural network. This concatenated vector is then fed into a softmax output layer.

Liu et al. (2016) have used deep learning based models in combination with the atomic features of text, the characters. An individual's personality traits are predicted using hierarchical, vector word and sentence representations. This work is based on Ling et al. (2015), which provides a language independent model for personality detection. Instead of making use of standard word embeddings (e.g., GloVe or Word2Vec), the word vector representation is formed from a bi-RNN using GRU (Chung et al. 2014) as the recurrent unit. GRU is less computationally expensive than the LSTM, but results in similar performance. These words are fed to another bi-RNN which gives the representation of the sentence, followed by a feed forward neural network for the prediction of the values of the five traits. One of the best uses

of this deep learning based character-level word encoding is a corpus of Tweets. The data are noisy and consist of a large number of micro-texts, hence this model performs much better than the conventional LIWC/BoW models.

Recently, NLP tools have been developed for language-independent, unsupervised personality recognition from unlabeled text. The PR2 system (Celli and Poesio 2014) exploits language-independent features extracted from LIWC and MRC such as punctuation, question marks, quotes, exclamation marks, numbers, etc. Only feature values which are above average in the input text are mapped to personality traits. It was observed that text containing more punctuation fires negative correlations with extraversion and openness to experience. This approach is not as accurate as the state of the art in the field, but it much more computationally efficient along with being independent of language.

Instead of relying just on text as the only form of input, researchers at Nokia have developed a software (Chittaranjan et al. 2011) which make use of data from smart phones including information about which applications were used and how often, music preferences, anonymous call logs, SMS logs and Bluetooth scans for personality detection. The relationship between personality traits and usage features was systematically analyzed using standard correlation and multiple regression analyses. SVM and C4.5 classifiers were used to classify users' personality traits.

Social media is a huge source of people's opinions and thoughts and analyzing it can give important insights into people's personality. Gonzalez-Gallardo et al. (2015) have used stylistic features which can be divided in two groups: character N-grams and POS N-grams. The grammatical sequence of the writer is obtained from POS tagging by making use of the open source tool, Freeling. Character N-grams and POS N-grams are extracted from the POS and class files to create a feature vector which is then fed into a linerSVM for personality detection.

Sun et al. (2018) have introduced abstract feature combination based on closely connected sentences which they call the Latent Sentence Group. The authors use bidirectional LSTMs, concatenated with CNN to detect user's personality using the structures of text.

Sometimes, the input text may be a collection of chats between two individuals. Detecting personality from conversations is a harder task than with simple text as one needs to accurately model temporal dependencies. Interactions in dyadic conversations have various degrees of mutual influence caused by turn-taking dialogues between two individuals. Recurrent networks are good for taking care of language dependencies. Su et al. (2016) make use of vanilla RNNs for modeling short term temporal evolution in conversation. A 255-dimensional linguistic feature vector is coupled Hidden Markov Model HMM is employed for detecting the personalities of two speakers across speaker turns in each dialog by using long-term turn-taking temporal evolution and cross-speaker contextual information. The Mandarin Conversational Dialog Corpus (Tseng 2004) was used as the dyadic conversation corpus for training and evaluation of the model. In many studies, how the personality 'appears' is approximated rather than the 'true' personality (after taking the personality questionnaire) of an individual. Sometimes, trained judges label an individual's perceived personality depending on their behavior, giving us an insight into how they are perceived by those around them.

4.2 Audio

We observed that majority of the recent models for personality detection from audio modality work in two separate stages, feature extraction followed by feeding the features to a classifier.

The audio descriptors can be sub-divided into broadly 7 groups, intensity, pitch, loudness, formants, spectrals, MFCC and other features. Some researchers use one or more of these feature groups to assess personality which is derived at the utterance level (Polzehl et al. 2010). These audio features (e.g., intensity, pitch, MFCC) are extracted using Praat acoustic analysis program and fed to a SVM classifier. The authors of this paper Polzehl et al. (2010) found that the five personality traits are interdependent and the variation in one of the traits results in observable changes in the other four. The dataset is created by recording the utterances and labelling them by having a trained speaker mimic a particular type of personality. Hence, the accuracy of the training and evaluation sets is solely dependent on how good the trained speaker is at mimicking the voice of a certain personality.

Some researchers claim that non-linguistic features (prosody, overlaps, interruptions and speech activity) outperform linguistic features (dialog acts and word n-grams) for perceived personality detection. Valente et al. (2012) have tested their model on spoken conversations from the AMI corpus dataset (Carletta 2007), which is a collection of meetings captured in specially instrumented meeting rooms along with the audio and video of each of the participants. The novelty of their approach lies in incorporating dialog act tags (which capture the intention of the speaker in the discussion) as one of the input features fed into the classifier. Each speaker was been labeled with 14 Dialog Act tags and a correlation with their perceived personality traits was observed (especially extraversion). Speech activity features (the total and relative amount of speech time per speaker, average duration of pauses per speaker, etc.) and sentence features (total number of sentences, average duration, maximum duration, etc.) are also included in the algorithm using a Booster algorithm with multi-class Boosting (Schapire and Singer 2000).

Recently, end-to-end approaches which involve training deep neural networks have gained wide spread popularity. They yield the state-of-the-art accuracy along with automatically extracting the necessary acoustic features. Palaz et al. (2015) analyze the CNN to understand the speech information that is modeled between the first two convolution layers. The first convolutional layer acts as the filter bank and automatically learns to extract the relevant features from the unprocessed audio wave input. The key difference in this paper is that temporal raw speech is fed directly to the CNN instead of the conventional two step procedure of first extracting the feature vectors (e.g., MFCC) followed by feeding to a classifier.

It is often observed that knowing the personality traits of an individual may in turn help in other tasks, such as deception detection. Levitan et al. (2016) have used AdaBoost (Freund and Schapire 1996) and Random forest for personality detection, which is then used as input to their deception detection algorithm. They have made use of acoustic, prosodic and linguistic features (LIWC) for personality prediction. The contribution of each of the LIWC features is then analyzed. It was found that for NEU-score, 'money' and 'power' dimensions are the most useful, for EXT-score, 'focusfuture' and 'drives', for CON-score, 'time' and 'work' are the highest. Most of these are intuitive and show the power of using LIWC features for personality detection. The model performed better in deception detection after incorporating the personality traits of the subject and was evaluated on the Columbia X-Cultural Deception Corpus.

4.3 Visual

Physiognomy (Kamenskaya and Kukharev 2008) is the art of determining the character or personality traits of an individual from the features of the body, especially the face. Researchers have found that the face provides most of the descriptive information for perceived person-

ality trait inference (Bruce and Young 1986; Willis and Todorov 2006). A non deep learning approach which involves the analysis of facial features for personality prediction is given by Kamenskaya and Kukharev (2008). Their focus lies on extracting physical features from images like shape of nose, body shape (fat, muscular, thin), shape of eyebrows and finding their correlation with personality. Sometimes people of similar personality tend to have similar preferences. Cristani et al. (2013) have explored this idea and have studied if there is a relation between the aesthetic preferences of a person (e.g., pictures liked by them) and their personality. Features extracted from the images could be divided into two types, aesthetic (colorfulness, use of light, GLCM-features, etc.) and content (faces, objects, GIST descriptors), which are then fed into a lasso regressor.

Liu et al. (2016) have predicted personality traits of a person from the choice of their Twitter profile picture. The model was trained on data from more than 66,000 users whose personality type had been predicted from what they had tweeted previously, using the state-of-the-art techniques from text modality. A correlation was found between the aesthetic and facial features of the profile picture (extracted using face++ and EmoVu API) and users' personality traits. Agreeableness and Conscientiousness users display more positive emotions in their profile pictures, while users high in openness prefer more aesthetic photos. Self-presentation characteristics of the user are captured by the 3D face posture, which includes the pitch, roll and yaw angle of the face and eye openness (using EmoVu API). Appealing images tend to have increased contrast, sharpness, saturation and less blur, which is the case for people high in Openness.

As in many other computer vision problems, deep CNNs have performed remarkably well in the task of personality detection from images. However, despite of the success in terms of performance, CNN is widely regarded as the magic black box. It is not known what are the various internal representation which emerge in various hidden layers of the neural network. Ventura et al. (2017) presents an in-depth study on understanding why CNN models perform surprisingly well in this complex problem. They make use of class activation maps for visualization and it is observed that activation is always centered around the face as shown in Fig. 2. The neural network specifically focuses on the facial region, mainly areas corresponding to the eyes and mouth to discriminate among different personality traits and predict their values.
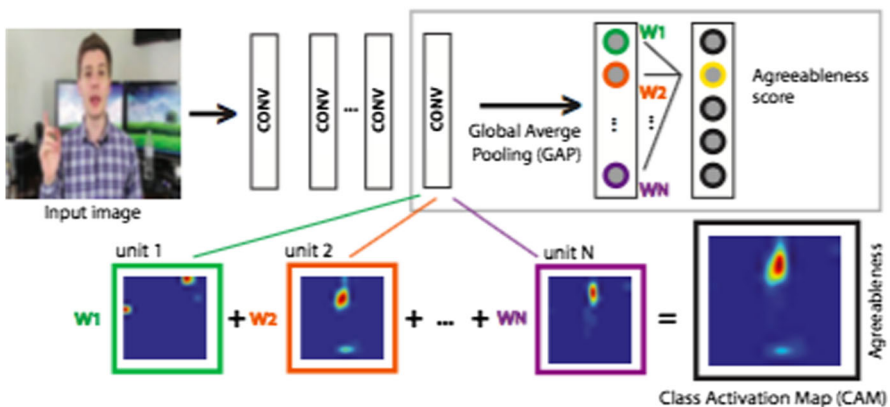

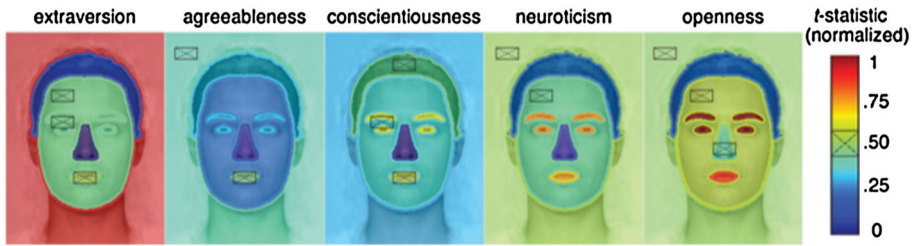**Fig. 2** Class activation map used by Ventura et al. (2017) for interpreting CNN models

Fig. 3 Segment-level occlusion analysis. Each image shows the changes in the prediction of the corresponding trait as a function of a predefined region (Gucluturk et al. 2017)

The deeper we go in the CNN, the more we find layers that are specialized in identifying high level features such as the eyes, nose, eyebrows, etc. Several Action Units (AU) such as raised eyebrows, dimple, etc. (subset of the FACS) are taken and a good correlation was observed between AU and perceived personality. A feature vector of the AUs was fed directly into a simple linear classifier and accuracy close to the state of the art was achieved.

The contribution of different parts of the image on personality trait prediction was studied by Gucluturk et al. (2017) using segment level occlusion analysis. Video frames were segmented into the following six regions with a deep neural network: background region, hair region, skin region, eye region, nose region and mouth region. It was observed that each region modulates at least one trait, and different traits were modulated by different regions. Background, skin and mouth regions modulated the fewest traits. Occlusion of background region increased extraversion trait but decreased conscientiousness trait. Occlusion of skin region decreased agreeableness and conscientiousness traits. Occlusion of mouth region increased neuroticism and openness traits. Finally, the eye region negatively modulates agreeableness trait, and positively modulates neuroticism and openness traits. Both the face region and the background region which includes clothing, posture and body parts have been shown to contain information regarding the perceived personality traits as can be seen in Fig. 3.

Psychological studies suggest that humans infer trait judgments of another person in just a few milliseconds after meeting them, that is from their first impression. First impressions and perceived personality are in-fact quite similar. One of the most important and widely used dataset in the field of multimodal perceived personality recognition is the ChaLearn First Impressions dataset. It comprises of 10,000 clips (average duration 15s) extracted from more than 3000 different YouTube high-definition videos of people facing and speaking into the camera. Chen et al. (2016) have proposed several methods for eliminating worker bias in labelling large datasets and have tested their method on the ChaLearn dataset. Gurpinar et al. (2016a) make use of the visual modality from this dataset for testing their model, which combines facial and ambient features extracted by a pre-trained deep CNN. These features are then fed into a kernel Extreme Learning Machine Regressor (Huang et al. 2012). The main contribution of this work is the combination of facial and ambient features along with making effective use of transfer learning.

A famous algorithm for detecting faces in images is the Viola Jones algorithm. Al Moubayed et al. (2014) have used this algorithm in the preprocessing stages after converting the image to grayscale, histogram equalizations (increases contrast) and rescaling. Then, EigenFace (Turk and Pentland 1991) features are extracted and inputted to a RBF SVM trained on Sequential Minimal Optimization approach. This approach was tested on a database created by making judges independently label images in the FERET corpus with the perceived personality traits. It is assumed that the judges have no ethnic or racial bias.

Rojas et al. (2011) have studied if there are specific structural relations among facial points that predict perception of facial traits. They have made a comparison between two approaches, first which uses the whole appearance and texture of the face (holistic approach) (Roberto and Poggio 1993) or just locations of specific fiducial facial points (structural approach). The holistic approach captures facial appearance information in two ways. Firstly, via the Eigen-Faces (Turk and Pentland 1991) method which analyzes the pixel information or secondly, making use of Histogram of Oriented Gradients (HOG), which captures facial appearance and shape. These extracted features are then fed into standard ML classifiers such as SVM or BTrees. Instead of using the Big-Five model, this paper predicts 9 different personality traits such as trustworthy, dominant, extroverted, threatening, competence, likability, mean, attractive and frightening. It was observed that HOG model found more correlation between the facial features and personality traits as compared to EigenFace.

A Vlog refers to a blog in which the postings are primarily in video form. A correlation between the facial expressions and personality of vloggers was analyzed by Biel (2012). In this paper, Computer Expression Recognition Toolbox (CERT) was used to determine the facial expressions of vloggers. CERT predicts 7 universal facial expressions (joy, disgust, surprise, neutral, sadness, contempt and fear). They found that facial expressions of emotions with positive valence such as 'joy' and 'smile' showed almost exclusively positive effects with personality impressions. For example, 'smile' is correlated with extraversion and agreeableness, among others.

Although the aforementioned approaches yield good results, they are very computationally expensive. Eddine Bekhouche et al. (2017) use a temporal face texture based approach for analyzing facial videos for job screening, which is much more computationally economical. After the preprocessing step, which involves face identification, 2D pose correction and cropping the region of interest (ROI), texture features of the image are extracted using Local Phase Quantization (LPQ) and Binarized Statistical Image features (BSIF). These features are fed to a Pyramid Multi Layer followed by five nonlinear Support Vector Regressors, one for each of the Big-Five personality traits. Finally, this perceived personality prediction is then input to a Gaussian Process Regression (GPR) which predicts a binary value corresponding to whether the candidate should be called for a job interview or not.

4.4 Bimodal

Most of the multimodal architectures for personality trait recognition are bimodular with the fusion of features from the audio and visual modalities. Kindiroglu et al. (2017) make use of multi task learning for recognizing extraversion and leadership traits from audio and visual modalities extracted from meeting videos of the Emergent Leader (ELEA) corpus. Predicting extraversion trait has attracted the attention of psychologists because it is able to explain a wide range of behaviors, predict performance and also assess the risk of an individual. The network architecture is initially trained on the larger VLOG corpus and afterwards on the ELEA corpus, hence effectively making use of transfer learning. Feature selection techniques such as Maximum Relevance Minimum Redundancy (MRMR) filtering feature selection results in a slight increase (about 2****) in performance.

Visualizing Deep Residual Networks (DRN) used for perceived personality analysis is done by Gucluturk et al. (2017) with the focus on explaining the workings of this end-to-end deep architecture. This DRN is a 2 stream network with one stream for the visual modality and the other one for the audio modality. What information do perceived personality trait recognition models rely on to make predictions? Gucluturk et al. (2017) concluded that

female faces with high levels of all traits seemed to be more colorful with higher contrast compared to those with low levels, who were more uniformly colored with lower contrast. Furthermore, looking at the unisex average faces, a bias for female faces for the high levels and male faces for the low levels was observed, more so in the case of average faces based on annotations. For instance, high levels of all traits were more bright and colorful, and with more positive expressions in line with the results observed by Walker and Vetter (2016).

The Deep Bimodal Regression (DBR) (Zhang et al. 2016) framework achieved the highest accuracy in the ChaLearn Challenge 2016 for perceived personality analysis. DBR consists of three main parts: the visual modality regression, audio modality regression and the last part is the ensemble process for fusing information of the two modalities. The network is trained end-to-end without the usage of any feature engine. The traditional CNN architecture is modified by discarding the fully connected layers. Also, the deep descriptors of the last convolution layer are both averaged and max pooled, followed by concatenation into a single 1024D vector (512 values from max pool and the other 512 values from average pooling). Finally, a regression (fc+sigmoid) layer is added for end-to-end training. This modified CNN model is called Descriptor Aggregation Network (DAN). For the audio modality, the log filter bank (logfbank) (Davis and Mermelstein 1990) features are extracted from the original audio of each video. Based on the logfbank features, a linear regressor is trained to obtain the Big-Five trait predictions. Finally, the two modalities are lately fused by averaging the five predicted scores of the visual and audio model.

Rai et al. (2016) propose a much more computationally efficient bimodal regression framework for the same, however with a slight decrease in accuracy. Instead of a single 15s audio clip, multiple audio segments (2–3 s length) are extracted from each video which increases the number of training samples resulting in considerably better results. FFMPEG is used for extracting audio clips from the original videos, followed by the OpenSMILE framework for extracting audio features. There are many studies (Shaver and Brennan 1992; Judge et al. 1999) which claim that the Big-Five personality traits are not independent of each other. The strength of a particular trait is quantified by the combination of two components, i.e., a global component and a trait specific component. In this model, each trait specific model is learned using a global component along with a trait specific component. Six models are created to predict the five personality traits where one model represents the global component and the other five represent the trait-specific components.

An (2018) have combined audio and text modalities for personality detection. The authors use acoustic-prosodic low level descriptor (LLD) features, word category features from LIWC and word scores for pleasantness, activation and imagery from the Dictionary of Affect in Language (DAL) (Whissell et al. 1986). A concatenation of these features are then fed into a MLP. They have compared this with a network consisting of word embedding with LSTM, and showed that the bimodal MLP network performs best on the myPersonality corpus. The MLP generalizes better when faced with out-of-vocabulary words. This points to the promise of acoustic-prosodic features, which are more robust with respect to language. They also find that models with early and late fusion of features achieve similar performance.

Subramaniam et al. (2016) makes use of temporally ordered deep audio and stochastic visual features for bimodal first impressions recognition. The novel idea of this paper is predicting perceived personality traits by making use of volumetric 3D convolution (Ji et al. 2013) based deep neural network. The temporal patterns in the audio and visual features are learned using a LSTM based deep RNN. Both the models concatenate the features extracted from audio and visual data in a later stage and finally feed into a fully connected layer.

Yang et al. (2017) follows a similar approach in trying to model the temporal dependencies by making use of LSTM. Audio features such as ZCR, MFCC and chroma vectors are
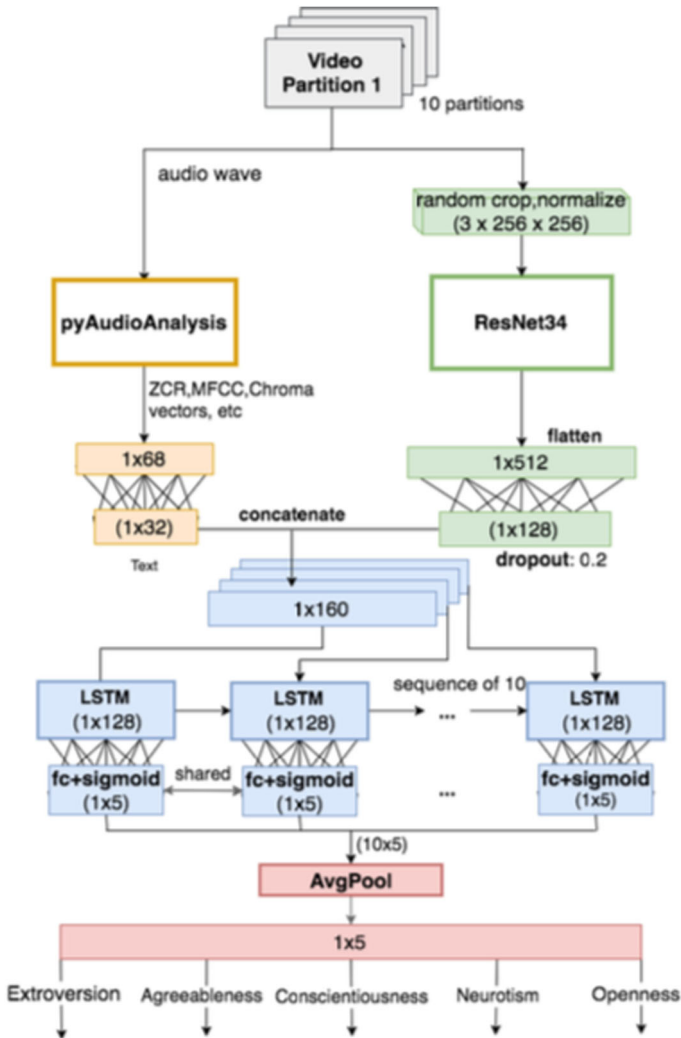
Fig. 4 Deep bimodal LSTM model architecture used by Yang et al. (2017)

extracted using PyAudio Analysis, whereas images are fed to a ResNet34 (He 2016) architecture after random cropping and normalization. This architecture is shown in Fig. 4 and achieves state-of-the-art performance for conscientiousness and openness traits. Gurpinar et al. (2016b) achieve state-of-the-art performance in this field by feeding the extracted feature vectors to a Extreme Learning Machine (ELM) and then fusing their predictions.

Acoustic features are extracted via the popularly used openSMILE tool while pre-trained Deep Convolutional Neural Network, VGG-VD-19 (Simonyan and Zisserman 2014) was used to extract facial emotion and ambient information from images. Local Gabor Binary Patterns from Three Orthogonal Planes (LGBP-TOP) (Zhang et al. 2005) along with VGGFace was used for facial feature extraction. The best results are achieved with multi-level fusion

of one acoustic and three visual sub-systems which are fed into an ELM and then combined with score level fusion.

Apart from perceived personality analysis of the Big-Five traits, an individual's attitudes such as amusement, impatience, friendliness, enthusiasm, frustration can also be estimated using similar architectures. Madzlan et al. (2014) automatically recognizes attitudes of Vloggers using prosodic and visual feature analysis. Prosodic features of audio such as pitch, intensity, quality, duration of utterance, etc. were extracted using TCL/TK script. Prosodic analysis of the attitudes shows 'impatience' having the highest pitch whereas 'frustration' the lowest. However, the analysis of visual modality is simply done by correlating the absolute movement of landmarks (eyebrows, mouth, etc.) of the face with respect to attitude.

Zhao et al. (2019) is one of the only papers to computationally study the influence of personality traits on emotion. The authors have jointly modeled physiological signals with personality. They have presented Vertex-weighted Multimodal Multitask Hypergraph Learning (VM2HL) learning model, where vertices are (subject, stimuli) pairs. Hyperedges formulate the relationship between personality and physiological signals. Hyperedges, modalities and vertices are given different levels of importance by introducing weights which are learned during training.

4.5 Trimodal

While most of the work focuses on fusion of inputs from audio and visual sources, (Gorbova et al. 2017) creates an additional stream that incorporates input from the text modality. These personality predictions are then used for automated screening of job candidates. They have achieved comparable performance to state-of-the-art while making use of a fully connected neural network with only 2 hidden layers. Facial features are extracted using OpenFace (Baltrusaitis 2016), an open source facial feature extractor toolkit. Audio features such as MFCC, ZCR, spectral energy distribution and speaking rate are extracted using OpenSMILE (Eyben et al. 2010) and SenticNet is used for polarity detection from text.

Poria et al. (2017) carries out multimodal sentiment analysis and make use of LSTMs to capture the context along with dependencies in between utterances. Initially, context independent features are extracted from each of the modalities. Features from the visual modality are extracted using 3D-CNN, OpenSMILE is used to capture low level descriptors from audio and Word2Vec embeddings followed by CNN for the text modality. To capture the flow of informational triggers across utterances, a LSTM-based RNN is used in a hierarchical scheme. Lastly, this is then fed to a dense layer for the final output. This architecture performs better than the state of the art on IEMOCAP (Busso et al. 2008), MOUD (Perez-Rosas et al. 2013) and MOSI (Zadeh et al. 2016) datasets and is shown in Fig. 5.

Deep networks being trained on multimodal data are generally very computationally intensive. Vo et al. (2018) have proposed a multimodal neural network architecture that combines different type of input data of various sizes with Discriminant Correlation Analysis (DCA) (Haghighat et al. 2016) Feature Fusion layer with low computational complexity. The fused features are used as inputs for layers in the Mixture Density Network to adjust for the information loss due to feature fusion. The final prediction is made using a state-of-art cascade network built on advanced gradient boosting algorithms.

Instead of having different trained neural networks for each of the 5 personality traits, (Kampman et al. 2018) train a single neural network with a multiclass approach with 3 channels (one for each of the modalities). Features are extracted from each of the 3 modalities using CNNs. The authors have experimented with using early fusion (concatenation of vectors
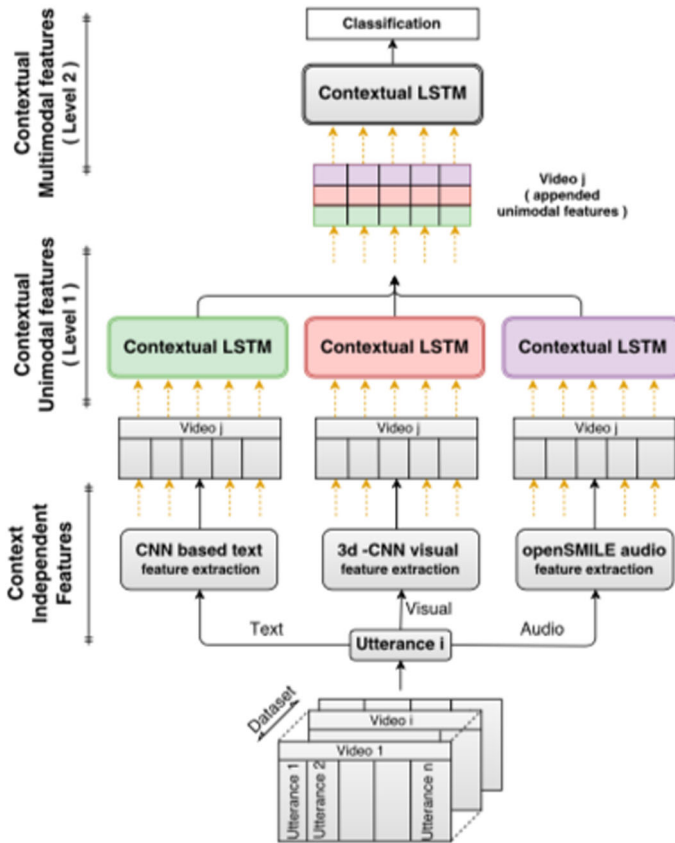
Fig. 5 Hierarchical architecture for extracting context- dependent multimodal utterance features (Poria et al. 2017)

to obtain shared representations of the input data) along with late fusion (each of the modalities makes an individual prediction and the prediction of the network is some weighted average of these predictions).

5 Results and discussions

Table 4 provides a systematic overview of the classification performance on popular personality-detection datasets. There has been a large increase in the number of works published in the field of personality detection in the last couple of years. This is due to the huge number of industrial applications of personality detection, as discussed in Sect. 1.2. Recently, systematically annotated datasets have been made publicly available and, hence, research can now be focused on creating better models and architectures rather than on data procurement and preprocessing. The CVPR First Impressions dataset (CVPR 2017) is the most popular multimodal dataset in personality detection. We see that there are many datasets for the Big-Five personality measure, however other personality measures (e.g., MBTI, PEN, 16PF) are lacking resources. The reliability of personality tests comes into question as it sometimes

Table 4 Performance of the state-of-the-art methods on popular personality-detection datasets

Modality	Dataset	Personality measure	Mean best accuracy	References
Text	Essays I (Pennebaker and King 1999)	Big 5	57.99	Majumder et al. (2017)
	Essays II (Tausczik and Pennebaker 2010)	Big 5	63.6	Poria et al. (2013)
	MBTI Kaggle	MBTI	67.77	Hernandez and Scott (2017)
	Italian FriendsFeed (Celli et al. 2010)	Big 5	63.1	Celli (2012)
Audio	The AMI Meeting Corpus (Carletta et al. 2005)	Big 5	64.84	Valente et al. (2012)
	CMU self recorded database (Polzehl et al. 2010)	Big 5	–	Polzehl et al. (2010)
	Columbia deception corpus (Levitan et al. 2015)	–	–	Levitan et al. (2016)
Visual	Face Evaluation (Oosterhof and Todorov 2008)	–	–	Rojas et al. (2011)
	YouTube Vlogs	Big 5	0.092 (R^2)	Biel (2012)
	ChaLearn First Impressions Dataset (Ponce-López et al. 2016)	Big 5	90.94	Gurpinar et al. (2016a)
	Color FERET database (Phillips et al. 1998)	Big 5	65	Al Moubayed et al. (2014)
Multimodal	ChaLearn First Impressions Dataset (Ponce-López et al. 2016)	Big 5	91.7	Gucluturk et al. (2017)
	ELEA (Emergent LEAder) corpus (Sanchez-Cortes et al. 2011)	Leadership/ extraversion	81.3	Kindiroglu et al. (2017)

happens that when a person takes the test it gives a certain personality, however when the same person takes the same test again, it gives a different type of personality. Escalante et al. (2018) and Escalante et al. (2017) study the interpretability and explainability of deep learning models in the context of computer vision for looking at people tasks. The MBTI personality measure is the most popular personality measure used across the world right now. Furnham (1990) have examined the relationship between subjects' actual test derived scores and their estimates of what those scores would be. Results showed significant positive correlations in majority of the dimensions. MBTI traits are harder and more complex to predict than the Big-Five traits (Furnham 1996), but we see current deep learning models are starting to achieve good accuracy in this area as well.

We see that the state of the art in personality detection has been achieved using deep learning techniques along with multimodal fusion of features. Techniques involving bimodal fusion are currently very popular. One can see that there is lot of scope for research and exploration for trimodal fusion for personality detection as presently, relatively few works have explored the fusion of all 3 modalities. If an individual's personality could be predicted with a little more reliability, there is scope for integrating automated personality detection in

almost all agents dealing with human-machine interaction such as voice assistants, robots, cars, etc. Research in this field is moving from detecting personality solely from textual data to visual and multimodal data.

We have seen the practical potential for automated personality detection is huge in the present day scenario. We expect that many of the deep learning models discussed in this paper will be implemented for industrial applications in the next few years. For training these models there will be a need for large publicly available datasets of various personality measures. In the last 5 years, we have seen a tremendous rise in the number of deep learning models that have been used for automated personality detection. The best performance in most modalities (especially visual) have been achieved by deep neural networks. We expect this trend to continue in the future with smarter and more robust deep learning models to follow.

6 Conclusion

As discussed in this paper, there are many diverse applications of automated personality detection which can be used in the industry, hence making it a very hot and upcoming field. However, machine learning models are as powerful as the data used to train them. For a large number of cases in this field, enough labelled data are not available to train huge neural networks. There is a dire need of larger, more accurate and more diverse datasets for personality detection. Almost all of the current datasets focus on the Big-Five personality model and very few for other personality measures such as the MBTI or PEN. Normally, personality is measured by answering multiple questions in a survey. Assuming that everyone taking the survey answers honestly, the credibility of this survey in correctly labelling an individual's personality is still in question. A more accurate and efficient way of labelling personality traits needs to be explored. Most of the current methods for creating personality detection datasets rely on manual annotation through crowd sourcing using Amazon Mechanical Turk.

Recent multimodal deep learning techniques have performed well and are starting to make reliable personality predictions. Deep learning offers a way to harness the large amount of data and computation power at our disposal with little engineering by hand. Various deep models have become the new state-of-the-art methods not only for personality detection, but in other fields as well. We expect this trend to continue with deeper models and new architectures which are able to map very complex functions. We expect to see more personality detection architectures that rely on efficient multimodal fusion.

References

Agarwal B (2014) Personality detection from text: a review. Int J Comput Syst 1:1–4

Al Moubayed N, Vazquez-Alvarez Y, McKay A, Vinciarelli A (2014) Face-based automatic personality perception. In: Proceedings of the 22nd ACM international conference on multimedia. ACM, pp 1153–1156

An G, Levitan R (2018) Lexical and acoustic deep learning model for personality recognition. In: Proceedings of interspeech, pp 1761–1765. https://doi.org/10.21437/Interspeech.2018-2263

Baltrusaitis T, Robinson P, Morency LP (2016) Openface: an open source facial behavior analysis toolkit. In: IEEE winter conference on applications of computer vision (WACV). IEEE, pp 1–10

Biel JI, Teijeiro-Mosquera L, Gatica-Perez D (2012) Facetube: predicting personality from facial expressions of emotion in online conversational video. In: Proceedings of the 14th ACM international conference on multimodal interaction. ACM, pp 53–56

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Lear Res 3:993–1022

Briggs-Myers I (1993) Introduction to type: a guide to understanding your results on the Myers-Briggs type indicator (revised by lk kirby & kd myers), Palo alto

Bruce V, Young A (1986) Understanding face recognition. Br J Psychol 77(3):305–327

Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS (2008) IEMOCAP: interactive emotional dyadic motion capture database. Lang Resour Eval 42(4):335

Cambria E, Poria S, Hazarika D, Kwok K (2018) SenticNet 5: discovering conceptual primitives for sentiment analysis by means of context embeddings. In: AAAI, pp 1795–1802

Cambria E, Poria S, Gelbukh A, Thelwall M (2017) Sentiment analysis is a big suitcase. IEEE Int Syst 32(6):74–80

Caprara GV, Schwartz S, Capanna C, Vecchione M, Barbaranelli C (2006) Personality and politics: values, traits, and political choice. Polit Psychol 27(1):1–28

Carletta J (2007) Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. Lang Resour Eval 41(2):181–190

Carletta J, Ashby S, Bourban S, Flynn M, Guillemot M, Hain T, Kadlec J, Karaiskos V, Kraaij W, Kronenthal M et al (2005) The ami meeting corpus: a pre-announcement. In: International workshop on machine learning for multimodal interaction. Springer, pp 28–39

Cattell HE, Mead AD (2008) The sixteen personality factor questionnaire (16pf). The SAGE handbook of personality theory and assessment 2:135–178

Celli F (2012) Unsupervised personality recognition for social network sites. In: Proceedings of sixth international conference on digital society

Celli F, Poesio M (2014) Pr2: a language independent unsupervised tool for personality recognition from text. arXiv preprint arXiv:1402.2796

Celli F, Di Lascio FML, Magnani M, Pacelli B, Rossi L (2010) Social network data and practices: the case of friendfeed. In: International conference on social computing, behavioral modeling, and prediction. Springer, pp 346–353

Chen B, Escalera S, Guyon I, Ponce-López V, Shah N, Simón MO (2016) Overcoming calibration problems in pattern labeling with pairwise ratings: application to personality traits. In: European conference on computer vision. Springer, pp 419–432

Chittaranjan G, Blom J, Gatica-Perez D (2011) Who's who with big-five: analyzing and classifying personality traits with smartphones. In: 15th Annual international symposium on wearable computers (ISWC). IEEE, pp 29–36

Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555

Cristani M, Vinciarelli A, Segalin C, Perina A (2013) Unveiling the multimedia unconscious: implicit cognitive processes and multimedia content analysis. In: Proceedings of the 21st ACM international conference on multimedia. ACM, pp 213–222

CVPR (2017) Chalearn dataset looking at people challenge. http://chalearnlap.cvc.uab.es/dataset/24/description/. Accessed 09 Oct 2019

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition, CVPR 2005, vol 1. IEEE, pp 886–893

Davis SB, Mermelstein P (1990) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. In: Readings in speech recognition. Elsevier, pp 65–74

Digman JM (1990) Personality structure: emergence of the five-factor model. Ann Rev Psychol 41(1):417–440

Eddine Bekhouche S, Dornaika F, Ouafi A, Taleb-Ahmed A (2017) Personality traits and job candidate screening via analyzing facial videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 10–13

Escalante HJ, Guyon I, Escalera S, Jacques J, Madadi M, Baró X, Ayache S, Viegas E, Güçlütürk Y, Güçlü U et al (2017) Design of an explainable machine learning challenge for video interviews. In: International joint conference on neural networks (IJCNN). IEEE, pp 3688–3695

Escalante HJ, Kaya H, Salah AA, Escalera S, Gucluturk Y, Guclu U, Baró X, Guyon I, Junior JJ, Madadi M et al (2018) Explaining first impressions: modeling, recognizing, and explaining apparent personality from videos. arXiv preprint arXiv:1802.00745

Escalera S, Baró X, Guyon I, Escalante HJ (2018) Guest editorial: apparent personality analysis. IEEE Trans Affect Comput 9(3):299–302

Eyben F, Wollmer M, Schuller B (2010) Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the 18th ACM international conference on multimedia. ACM, pp 1459–1462

Eysenck HJ (2012) A model for personality. Springer, Berlin

Freund Y, Schapire RE et al (1996) Experiments with a new boosting algorithm. ICML, Bari, Italy 96:148–156

Furnham A (1990) Can people accurately estimate their own personality test scores? Eur J Personal 4(4):319–327

Furnham A (1996) The big five versus the big four: the relationship between the Myers-Briggs type indicator (MBTI) and NEO-PI five factor model of personality. Personal Individ Differ 21(2):303–307

Golbeck J (2016) Predicting personality from social media text. AIS Trans Replication Res 2(1):2

Gonzalez-Gallardo CE, Montes A, Sierra G, Nunez-Juarez JA, Salinas-Lopez AJ, Ek J (2015) Tweets classification using corpus dependent tags, character and pos n-grams. In: CLEF (working notes)

Gorbova J, Lusi I, Litvin A, Anbarjafari G (2017) Automated screening of job candidate based on multimodal video processing. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 29–35

Gucluturk Y, Guclu U, Perez M, Balderas HJE, Baro X, Guyon I, Andujar C, Junior J, Madadi M, Escalera S et al (2017) Visualizing apparent personality analysis with deep residual networks. In: International conference on computer vision-ICCV 2017

Gurpinar F, Kaya H, Salah AA (2016) Combining deep facial and ambient features for first impression estimation. In: European conference on computer vision. Springer, pp 372–385

Gurpinar F, Kaya H, Salah AA (2016) Multimodal fusion of audio, scene, and face features for first impression estimation. In: 23rd International conference on pattern recognition (ICPR). IEEE, pp 43–48

Haghighat M, Abdel-Mottaleb M, Alhalabi W (2016) Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition. IEEE Trans Inf Forensics Secur 11(9):1984–1996

Hernandez RK, Scott I (2017) Predicting Myers-Briggs type indicator with text

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

Hirsch HG, Pearce D (2000) The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In: ASR2000-Automatic speech recognition: challenges for the new Millenium ISCA tutorial and research workshop (ITRW)

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12(1):55–67

Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern Part B (Cybern) 42(2):513–529

Ilmini K, Fernando TL (2016) Persons' personality traits recognition using machine learning algorithms and image processing techniques

Ireland ME, Pennebaker JW (2010) Language style matching in writing: Synchrony in essays, correspondence, and poetry. J Personal Soc Psychol 99(3):549

Jaynes ET (1982) On the rationale of maximum-entropy methods. Proc IEEE 70(9):939–952

Ji S, Xu W, Yang M, Yu K (2013) 3d convolutional neural networks for human action recognition. IEEE Trans Pattern Anal Mach Intell 35(1):221–231

Judge TA, Higgins CA, Thoresen CJ, Barrick MR (1999) The big five personality traits, general mental ability, and career success across the life span. Pers Psychol 52(3):621–652

Junior J, Jacques C, Güçlütürk Y, Pérez M, Güçlü U, Andujar C, Baró X, Escalante HJ, Guyon I, van Gerven MA et al (2018) First impressions: a survey on computer vision-based apparent personality trait analysis. arXiv preprint arXiv:1804.08046

Kalghatgi MP, Ramannavar M, Sidnal NS (2015) A neural network approach to personality prediction based on the big-five model. Int J Innov Res Adv Eng (IJIRAE) 2(8):56–63

Kamenskaya E, Kukharev G (2008) Recognition of psychological characteristics from face. Metody Inform Stosow 1(1):59–73

Kampman O, Barezi EJ, Bertero D, Fung P (2018) Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction. In: Proceedings of the 56th annual meeting of the association for computational linguistics (volume 2: short papers), vol 2, pp 606–611

Kindiroglu AA, Akarun L (2017) Aran O (2017) Multi-domain and multi-task prediction of extraversion and leadership from meeting videos. EURASIP J Image Video Proc 1:77

Kohonen T (1990) The self-organizing map. Proc IEEE 78(9):1464–1480

Levitan SI, An G, Wang M, Mendels G, Hirschberg J, Levine M, Rosenberg A (2015) Cross-cultural production and detection of deception from speech. In: Proceedings of the 2015 ACM on workshop on multimodal deception detection. ACM, pp 1–8

Levitan SI, Levitan Y, An G, Levine M, Levitan R, Rosenberg A, Hirschberg J (2016) Identifying individual differences in gender, ethnicity, and personality from dialogue for deception detection. In: Proceedings of the second workshop on computational approaches to deception detection, pp 40–44

Liem CC, Langer M, Demetriou A, Hiemstra AM, Wicaksana AS, Born MP, König CJ (2018) Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening. In: Explainable and interpretable models in computer vision and machine learning. Springer, pp 197–253

Ling W, Luis T, Marujo L, Astudillo RF, Amir S, Dyer C, Black AW, Trancoso I (2015) Finding function in form: Compositional character models for open vocabulary word representation. arXiv preprint arXiv:1508.02096

Liu L, Preotiuc-Pietro D, Samani ZR, Moghaddam ME, Ungar LH (2016) Analyzing personality through social media profile picture choice. In: ICWSM, pp 211–220

Madzlan NA, Han J, Bonin F, Campbell N (2014) Automatic recognition of attitudes in video blogs—prosodic and visual feature analysis. In: Fifteenth annual conference of the international speech communication association

Majumder N, Poria S, Gelbukh A, Cambria E (2017) Deep learning-based document modeling for personality detection from text. IEEE Intell Syst 32(2):74–79

Majumder N, Poria S, Peng H, Chhaya N, Cambria E, Gelbukh A (2019) Sentiment and sarcasm classification with multitask learning. IEEE Intell Syst 34(3):38–43

Matthews G, Deary IJ, Whiteman MC (2003) Personality traits. Cambridge University Press, Cambridge

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781

Miles J, Hempel S (2004) The Eysenck personality scales: the Eysenck personality questionnaire-revised (EPQ-R) and the Eysenck personality profiler (EPP). Comprehensive handbook of psychological assessment vol 2, pp 99–107

Mukherjee S, Kumar U (2016) Ethical issues in personality assessment. In: The Wiley handbook of personality assessment, pp 415–426

Oosterhof NN, Todorov A (2008) The functional basis of face evaluation. Proc Natl Acad Sci 105(32):11087–11092

Palaz D, Magimai-Doss M, Collobert R (2015) Analysis of cnn-based speech recognition system using raw speech as input. Technical report, Idiap

Papurt MJ (1930) A study of the woodworth psychoneurotic inventory with suggested revision. J Abnorm Soc Psychol 25(3):335

Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, Ungar LH, Seligman ME (2015) Automatic personality assessment through social media language. J Personal Soc Psychol 108(6):934

Pennebaker JW, King LA (1999) Linguistic styles: language use as an individual difference. J Personal Soc Psychol 77(6):1296

Pennebaker JW, Francis ME, Booth RJ (2001) Linguistic inquiry and word count: LIWC 2001, vol 71. Lawrence Erlbaum Associates, Mahway

Pennington J, Socher R, Manning C (2014) Glove: global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543

Perez-Rosas V, Mihalcea R, Morency LP (2013) Utterance-level multimodal sentiment analysis. In: Proceedings of the 51st annual meeting of the association for computational linguistics (volume 1: long papers), vol 1, pp 973–982

Pervin LA, John OP (1999) Handbook of personality: theory and research. Elsevier, Amsterdam

Phillips PJ, Wechsler H, Huang J, Rauss PJ (1998) The feret database and evaluation procedure for face-recognition algorithms. Image Vis Comput 16(5):295–306

Polzehl T, Moller S, Metze F (2010) Automatically assessing personality from speech. In: IEEE fourth international conference on semantic computing (ICSC). IEEE, pp 134–140

Ponce-López V, Chen B, Oliu M, Corneanu C, Clapés A, Guyon I, Baró X, Escalante HJ, Escalera S (2016) Chalearn lap 2016: first round challenge on first impressions-dataset and results. In: European conference on computer vision. Springer, pp 400–418

Poria S, Gelbukh A, Agarwal B, Cambria E, Howard N (2013) Common sense knowledge based personality recognition from text. In: Mexican international conference on artificial intelligence. Springer, pp 484–496

Poria S, Chaturvedi I, Cambria E, Bisio F (2016) Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis. In: International joint conference on neural networks (IJCNN), pp 4465–4473

Poria S, Cambria E, Hazarika D, Majumder N, Zadeh A, Morency LP (2017) Context-dependent sentiment analysis in user-generated videos. In: Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: long papers), vol 1, pp 873–883

Rai N (2016) Bi-modal regression for apparent personality trait recognition. In: 23rd International conference on pattern recognition (ICPR). IEEE, pp 55–60

Roberto B, Poggio T (1993) Face recognition: features versus templates. IEEE Tran Pattern Anal Mach Intell 15(10):1042–1052

Rojas M, Masip D, Todorov A, Vitria J (2011) Automatic prediction of facial trait judgments: appearance vs. structural models. PLoS ONE 6(8):e23323

Saez Y, Navarro C, Mochon A, Isasi P (2014) A system for personality and happiness detection. IJIMAI 2(5):7–15

Sanchez-Cortes D, Aran O, Gatica-Perez D (2011) An audio visual corpus for emergent leader analysis. In: ICMI-MLMI, Multimodal corpora for machine learning, pp 14–18

Schapire RE, Singer Y (2000) Boostexter: a boosting-based system for text categorization. Mach Learn 39(2–3):135–168

Shaver PR, Brennan KA (1992) Attachment styles and the "big five" personality traits: their connections with each other and with romantic relationship outcomes. Personal Soc Psychol Bull 18(5):536–545

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

Su MH, Wu CH, Zheng YT (2016) Exploiting turn-taking temporal evolution for personality trait perception in dyadic conversations. IEEE/ACM Trans Audio Speech Lang Process 24(4):733–744

Subramaniam A, Patel V, Mishra A, Balasubramanian P, Mittal A (2016) Bi-modal first impressions recognition using temporally ordered deep audio and stochastic visual features. In: European conference on computer vision. Springer, pp 337–348

Sun X, Liu B, Cao J, Luo J, Shen X (2018) Who am I? Personality detection based on deep learning for texts. In: IEEE international conference on communications (ICC). IEEE, pp 1–6

Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 29(1):24–54

Tseng SC (2004) Processing mandarin spoken corpora. Traitement Automatique des Langes, pp 89–108

Turk M, Pentland A (1991) Eigenfaces for recognition. J Cognit Neurosci 3(1):71–86

Valente F, Kim S, Motlicek P (2012) Annotation and recognition of personality traits in spoken conversations from the ami meetings corpus. In: Thirteenth annual conference of the international speech communication association

Ventura C, Masip D, Lapedriza A (2017) Interpreting cnn models for apparent personality trait regression. In: IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 1705–1713

Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57(2):137–154

Vo NN, Liu S, He X, Xu G (2018) Multimodal mixture density boosting network for personality mining. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, pp 644–655

Walker M, Vetter T (2016) Changing the personality of a face: Perceived big two and big five personality factors modeled in real photographs. J Personal Soc Psychol 110(4):609

Whissell C, Fournier M, Pelland R, Weir D, Makarec K (1986) A dictionary of affect in language: IV Reliability, validity, and applications. Percept Motor Skills 62(3):875–888

Willis J, Todorov A (2006) First impressions: making up your mind after a 100-ms exposure to a face. Psychol Sci 17(7):592–598

Yang Y, Mall S, Glaser N (2017) Prediction of personality first impressions with deep bimodal LSTM

Yang HC, Huang ZR (2019) Mining personality traits from social messages for game recommender systems. Knowl Based Syst 165:157–168

Vinciarelli A, Mohammadi G (2014) A survey of personality computing. IEEE Trans Affect Comput 5(3):273–291

Yin H, Wang Y, Li Q, Xu W, Yu Y, Zhang T (2018) A network-enhanced prediction method for automobile purchase classification using deep learning

Yu J, Markov K (2017) Deep learning based personality recognition from facebook status updates. In: IEEE 8th international conference on awareness science and technology (iCAST). IEEE, pp 383–387

Zadeh A, Zellers R, Pincus E, Morency LP (2016) Mosi: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. arXiv preprint arXiv:1606.06259

Zhang W, Shan S, Gao W, Chen X, Zhang H (2005) Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In: Tenth IEEE international conference on computer vision, ICCV 2005, vol 1. IEEE, pp 786–791

Zhang CL, Zhang H, Wei XS, Wu J (2016) Deep bimodal regression for apparent personality analysis. In: European conference on computer vision. Springer, pp 311–324

Zhao S, Gholaminejad A, Ding G, Gao Y, Han J, Keutzer K (2019) Personalized emotion recognition by personality-aware high-order learning of physiological signals. ACM Tran Multimed Comput Commun Appl (TOMM) 15(1s):14

Zuo X, Feng B, Yao Y, Zhang T, Zhang Q, Wang M, Zuo W (2013) A weighted ML-kNN model for predicting users' personality traits. In: Proceedings of international conference on information science and computer application (ISCA), pp 345–350

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